Mean field stochastic control under sublinear expectation

Rainer Buckdahn$^{1,3}$, Bowen He$^{2,**}$, Juan Li$^{2,3,**}$

$^1$Laboratoire de Mathématiques de Bretagne Atlantique, Univ Brest, UMR CNRS 6205, 6 avenue Le Gorgeu, 29200 Brest, France.

$^2$School of Mathematics and Statistics, Shandong University, Weihai, Weihai 264209, P. R. China.

$^3$Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao 266237, P. R. China.

E-mails: rainer.buckdahn@univ-brest.fr, hbw@mail.sdu.edu.cn, juanli@sdu.edu.cn.

November 09, 2022

Abstract. Our work is devoted to the study of Pontryagin’s stochastic maximum principle for a mean-field optimal control problem under Peng’s $G$-expectation. The dynamics of the controlled state process is given by a stochastic differential equation driven by a $G$-Brownian motion, whose coefficients depend not only on the control, the controlled state process but also on its law under the $G$-expectation. Also the associated cost functional is of mean-field type. Under the assumption of a convex control state space we study the stochastic maximum principle, which gives a necessary optimality condition for control processes. Under additional convexity assumptions on the Hamiltonian it is shown that this necessary condition is also a sufficient one. The main difficulty which we have to overcome in our work consists in the differentiation of the $G$-expectation of parameterized random variables. As particularly delicate it turns out to handle with the $G$-expectation of a function of the controlled state process inside the running cost of the cost function. For this we have to study a measurable selection theorem for set-valued functions whose values are subsets of the representing set of probability measures for the $G$-expectation.

AMS subject classifications. 60H10, 60K35

Key words. $G$-expectation, stochastic control, Pontryagin’s stochastic maximum principle, mean-field SDE, differentiation with a sublinear expectation, time inconsistent control

*The work is supported by the NSF of P.R. China (NOS. 12031009, 11871037), National Key R and D Program of China (NO. 2018YFA0703900), and NSFC-RS (No. 11661130148; NA150344).

** Corresponding authors.
1 Introduction

Our work brings together two important subjects of actual intensive research, mean-field problems popularised by Lasry and Lions’ pioneering work [16] on mean-field games in 2007 on one side and stochastic control under Peng’s sublinear $G$-expectation (see, e.g., [20] and [21]) on the other side. More precisely, we study Pontryagin’s stochastic maximum principle (SMP) for a stochastic control problem over a $G$-expectation space, whose dynamics are given by a controlled $G$-stochastic differential equation ($G$-SDE) whose coefficients do not only depend on the control process and the associated controlled state process but also on its law under the $G$-expectation, which we consider as the $G$-expectation of a function of the controlled state space. Also in the associated cost functional both the terminal cost function and the running cost function depend not only on the controlled state process but also on its law with respect to (w.r.t.) the $G$-expectation.

Mean-field SDEs in form of McKean-Vlasov equations have been studied for a long time and have found a lot of applications in different domains. Recently, with their seminal paper [16] on mean-field games and their applications in economics, finance and game theory, Lasry and Lions have given new impulses to this research topic, opened the way to new applications and attracted a lot of researchers to this topic. One of these applications is the study of mean-field stochastic optimal control problems. Motivated by the rich literature on the stochastic maximum principle in the classical stochastic control, for example, Peng’s SMP [19], different authors studies the stochastic maximum principle in the context of mean-field control problems. Let us namely mention the work by Buckdahn, Djehiche and Li [3] in 2011, where the coefficients of the mean-field SDEs depend on the solution process, its expectation and the control. Li [18] studied Pontryagin’s SMP for mean-field SDEs, and obtained necessary and sufficient conditions for the optimality of a control process, while Buckdahn, Li, Ma [5,6] studied the optimal control problem for a class of general mean-field SDEs, in which the coefficients depend non linearly on both the state process as well as on its law. They extended the SMP of Buckdahn et al. [3] to this general case. Acciaio et al. [1] studied the stochastic maximum principle for an extended mean-field control problem.

However, for instance, in economy and in finance a vast field of applications requires to model Knightian uncertainty. Inspired by financial problems with uncertainty, Peng [20] introduced a fully non linear expectation, called $G$-expectation $\hat{E}[-]$, and he proved that it can well characterize the Knightian uncertainty. Under this $G$-expectation framework a new type of Brownian motion, the so-called $G$-Brownian motion, has been introduced and the stochastic calculus with respect to the $G$-Brownian motion has been developed. Recently, Hu et al. [12,13] developed the SDE and BSDE theory in this $G$-expectation framework. And they also studied the SMP for stochastic optimal control problems under $G$-expectation or uncertainty (see [11,15]).

For the reasons explained above we study the SMP for a mean-field stochastic control problem under $G$-expectation. We consider a stochastic control problem where the dynamic of the state process is given by a stochastic differential equation driven by a $G$-Brownian motion.
(G-SDE) of mean-field type. That is, the coefficients do not only depend on the control and the controlled state process but also on some functional of the law of the state process under the sublinear expectation. More precisely, we consider the dynamics

\[
\begin{align*}
\begin{cases}
\d X_t^u = \sigma(X_t^u, \mathbb{E}[\varphi_1(X_t^u)], u_t)dB_t + b(X_t^u, \mathbb{E}[\varphi_2(X_t^u)], u_t)dt + \beta(X_t^u, \mathbb{E}[\varphi_3(X_t^u)], u_t)d(B)_t, \\
X_0^u = x_0 \in \mathbb{R}^n, \quad t \in [0,T],
\end{cases}
\end{align*}
\]

for some functions \(b, \sigma, \beta, \varphi_i, \ i = 1, 2, 3\), and the G-Brownian motion \(B = (B_t)\). The admissible control process \(u = (u_t)\) takes its values in a convex state space \(U\). The objective is to minimize the associated cost functional of the form

\[
J(u) := \mathbb{E}[\Phi(X_T^u, \mathbb{E}[\varphi_3(X_T^u)])] + \int_0^T l(t, X_t^u, \mathbb{E}[\varphi_5(X_t^u)], u_t)dt,
\]

for given functions \(\Phi, l\) and \(\varphi_i, \ i = 4, 5\). Also this cost functional is of mean-field type, as the functions \(\Phi\) and \(l\), depend on the law under the sublinear expectation of the state process.

In this paper we derive necessary and sufficient conditions for optimality of this control problem in form of a stochastic maximum principle for a convex action space, using the convex perturbation. The stochastic maximum principle involves solving a family of adjoint equations, backward SDEs (BSDEs).

As concerns previous works related with the SMP under sublinear expectation, we have to mention mainly the recent works by Biagini, Meyer-Brandis and Øksendal [2], Sun [23] and Hu and Ji [11]. In [2] the authors study a stochastic control problem (without mean-field term), composed of a forward G-SDE and a cost functional under G-expectations (also without mean-field term). The sufficient but also the necessary optimality conditions for a control process \(\hat{u}\) they give need the assumption that in the adjoint equation, a G-BSDE (see Definition (2.11)), the non increasing G-martingale \(K\) is identically equal to zero. In [23] Sun studies a controlled system of G-forward and G-backward SDEs with solution \((X^u, Y^u, Z^u)\), and he associates the cost functional \(J(u) = \mathbb{E}[\psi(u)]\) with \(\psi(u) = \phi(X_T^u) + \int_0^T l(t, X_t^u, Y_t^u, Z_t^u, u_t)dt + \gamma(Y_T^u)\). In the deduction of the sufficient optimality condition for a control \(\hat{u}\) he uses convexity assumptions. Also in a non mean-field context, Hu and Ji [11] study a system of forward and backward G-SDEs, they consider as cost functional \(J(u) = Y_0^u\), and they investigate the SMP. For this they show namely that the cost functional for the perturbed optimal control \(\lambda \mapsto J(\hat{u} + \lambda(u - \hat{u}))\) is right-differentiable at \(\lambda = 0\), and they use the special form of this derivative and an application of Sion’s minimax theorem to derive a necessary optimality condition of the optimal control \(\hat{u}\). Their approach depends on the related G-BSDEs.

Inspired by above works we study the mean-field stochastic control problem under G-expectation. We investigate the SMP and give a necessary optimality condition for the optimal control and also a sufficient one for the optimality of a control. However, the fact that we have to do not only with the G-expectation of the definition of the cost functional \(J(u)\) but also with the G-expectations \(\mathbb{E}[\varphi_i(X_t^u)]\), \(1 \leq i \leq 5\), involves new difficulties. So, for instance, in the general case, Sion’s minimax theorem cannot be applied. As it plays a crucial role, a whole section (Section 4) is devoted to the study of the derivative of functions of laws under
$G$-expectation. For this we begin with the easy observation that, given two random variables $\xi, \eta$, the function $\lambda \mapsto F(\lambda) = \hat{E}[\xi + \lambda \eta]$ is convex, i.e., the right but also the left derivatives $F'_-(\lambda)$ and $F'_+(\lambda)$, respectively, exist. We determine them in a more direct approach than that in [11], without passing through the associated $G$-BSDE, and we also associate some essential result which has its own interest (see Proposition 4.2). The results are extended to the derivative of functions $F_f(\xi) := \sup_{P \in \mathcal{P}} f(P_\xi)$, where $P$ represents $\hat{E}[\cdot]$ (see Theorem 2.7). As the derivative of this latter function is not directly used for our SMP approach but has its own interest, it is shifted to Appendix 1. Section 5 is devoted to deduce the necessary optimality condition for an optimal control. Our main result is Theorem 5.5. The main difficulty here stems from the fact that our coefficients depend also on $\hat{E}[\phi_i(X_u^t)]$, $1 \leq i \leq 5$, and so all their derivatives have to be considered. The most delicate part comes from the dependence of the running cost on $\hat{E}[\phi_5(X_u^t)]$. To handle these difficulties we need a measurable selection theorem for a mapping $[0, T] \ni t \mapsto P_{\xi|\eta} \subset \mathcal{P}$ (see Theorem 6.8). Proving that this mapping is a weakly measurable set-valued function, we can use the Kuratowski and Ryll-Nardzewski measurable selection theorem to get Theorem 6.8, see Appendix 2. For the case that the running cost coefficient $l$ does not depend $\hat{E}[\phi_5(X_u^t)]$, Sion’s minimax theorem can be used to simplify the necessary optimality condition considerably; see Theorem 5.6. The second part of Section 5 is devoted to the study of a sufficient optimality condition for the general case and to an example.

Our paper is organized as follows: In Section 2, we recall some basic notions of $G$-expectation and results of $G$-SDEs and $G$-BSDEs. Section 3 introduces the formulation of the mean-field stochastic control problem, and Section 4 is devoted to the study of the derivative of the $G$-expectation of parameter depending random variables. In Section 5 we study the SMP and the Appendix is devoted to an extension of the studies made in Section 4 and to the proof of our measurable selection theorem.

2 Preliminaries

In this section, we review some notations and results in the $G$-expectation framework, which are mainly concerned with the $G$-Itô calculus and BSDEs driven by a $G$-Brownian motion. More relevant details can be found in [12,17,20,22].

2.1 $G$-expectation space

Let $\Omega$ be a given non empty set and $\mathcal{H}$ be a linear space of real-valued functions on $\Omega$ such that, for all $d \geq 1$, if $X_1, \ldots, X_d \in \mathcal{H}$, then also $\varphi(X_1, X_2, \ldots, X_d) \in \mathcal{H}$ for every $\varphi \in C_{b,Lip}(\mathbb{R}^d)$, where $C_{b,Lip}(\mathbb{R}^d)$ is the space of bounded Lipschitz functions on $\mathbb{R}^d$. The set $\mathcal{H}$ is considered as the space of random variables.

**Definition 2.1.** A sublinear expectation $\hat{E}$ on $\mathcal{H}$ is a functional $\hat{E} : \mathcal{H} \to \mathbb{R}$ having the following properties: For each $X, Y \in \mathcal{H}$,

(i) Monotonicity: $\hat{E}[X] \geq \hat{E}[Y]$, if $X \geq Y$;
(ii) Constant preserving: $\hat{E}[c] = c$, for $c \in \mathbb{R}$;
(iii) Sub-additivity: $\hat{E}[X + Y] \leq \hat{E}[X] + \hat{E}[Y]$;
(iv) Positive homogeneity: $\hat{E}[\lambda X] = \lambda \hat{E}[X]$, for all real $\lambda \geq 0$.

The triple $(\Omega, \mathcal{H}, \hat{E})$ is called a sublinear expectation space.

**Definition 2.2.** Two $d$-dimensional random vectors $X_1$ and $X_2$ defined, respectively, on sublinear expectation spaces $(\Omega_1, \mathcal{H}_1, \hat{E}_1)$ and $(\Omega_2, \mathcal{H}_2, \hat{E}_2)$ are called identically distributed, denoted by $X_1 \overset{d}{=} X_2$, if

$$\hat{E}_1[\varphi(X_1)] = \hat{E}_2[\varphi(X_2)], \quad \text{for every } \varphi \in C_{b,\text{Lip}}(\mathbb{R}^d).$$

**Definition 2.3.** On the sublinear expectation space $(\Omega, \mathcal{H}, \hat{E})$, an $n$-dimensional random vector $Y$ is said to be independent of a $d$-dimensional random vector $X$, denoted by $Y \perp X$, if

$$\hat{E}[\varphi(X, Y)] = \hat{E}[\hat{E}[\varphi(x, Y)|_{x=X}]], \quad \text{for every } \varphi \in C_{b,\text{Lip}}(\mathbb{R}^{d+n}).$$

A $d$-dimensional random vector $\tilde{X}$ is said to be an independent copy of $X$ if $\tilde{X} \overset{d}{=} X$ and $\tilde{X} \perp X$.

**Proposition 2.4.** Let $X, Y \in \mathcal{H}$ be such that $\hat{E}[Y] = -\hat{E}[-Y]$. Then we have

$$\hat{E}[X + Y] = \hat{E}[X] + \hat{E}[Y].$$

**Definition 2.5.** A $d$-dimensional random vector $X$ defined on $(\Omega, \mathcal{H}, \hat{E})$ is called $G$-normally distributed if for any $a, b \geq 0$,

$$aX + b\tilde{X} \overset{d}{=} \sqrt{a^2 + b^2} X,$$

where $\tilde{X}$ is an independent copy of $X$. Here the letter $G$ denotes the function $G(A) := \frac{1}{2}\hat{E}[(AX, X)]$, for $A \in \mathbb{S}(d)$, where $\mathbb{S}(d)$ is the space of all $d \times d$ symmetric matrices.

Throughout this paper, we denote by $\Omega := C([0, \infty); \mathbb{R}^d)$ the space of all $\mathbb{R}^d$-valued continuous paths $(\omega_t)_{t \geq 0}$, equipped with the distance

$$\rho_d(\omega^1, \omega^2) := \sum_{i=1}^{\infty} \frac{1}{2^i} (\|\omega^1 - \omega^2\|_{C([0, \infty]; \mathbb{R}^d)} \wedge 1),$$

where $\|\omega^1 - \omega^2\|_{C([0, T]; \mathbb{R}^d)} := \max_{t \in [0, T]} |\omega^1_t - \omega^2_t|$, for $T > 0$. Given any $T > 0$, we also define $\Omega_T := \{(\omega_{t \wedge T})_{t \geq 0} : \omega \in \Omega\}$.

Let $B_t(\omega) := \omega_t$, $\omega \in \Omega$, $t \geq 0$, be the coordinate process on $\Omega$. We introduce the space

$$L_{ip}(\Omega_T) := \{\varphi(B_{t_1}, B_{t_2}-B_{t_1}, \cdots, B_{t_n}-B_{t_{n-1}}) : n \in \mathbb{N}, \ 0 \leq t_1 < t_2 \cdots < t_n \leq T, \ \varphi \in C_{b,\text{Lip}}(\mathbb{R}^{d \times n})\},$$

as well as $L_{ip}(\Omega) := \bigcup_{m=1}^{\infty} L_{ip}(\Omega_m)$.

The $G$-expectation on $L_{ip}(\Omega)$ is defined by

$$\hat{E}[X] := \hat{E}[\varphi(\sqrt{t_1} \xi_1, \sqrt{t_2-t_1} \xi_2, \cdots, \sqrt{t_n-t_{n-1}} \xi_n)],$$
for all $X = \varphi(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}})$, $n \geq 1$, $0 \leq t_1 < \cdots < t_n < \infty$, where $\{\xi_i\}_{i=1}^n$ is a collection of $d$-dimensional identically distributed random variables on a sublinear expectation space $(\tilde{\Omega}, \tilde{\mathcal{H}}, \tilde{\mathbb{E}})$ such that, for all $1 \leq i \leq n$, $\xi_i$ is $G$-normally distributed and independent of $(\xi_1, \ldots, \xi_{i-1})$. Then under $\tilde{\mathbb{E}}$, the coordinate process $\tilde{B}_t = (\tilde{B}_t^1, \ldots, \tilde{B}_t^d)$ is a $d$-dimensional $G$-Brownian motion defined by the following properties:

(a) $B_0 = 0$;

(b) For every $t, s \geq 0$, the increment $\tilde{B}_{t+s} - \tilde{B}_t$ is independent of $(\tilde{B}_{t_1}, \ldots, \tilde{B}_{t_n})$, for all $n \in \mathbb{N}$ and $0 \leq t_1 \leq \cdots \leq t_n \leq t$;

(c) $\tilde{B}_{t+s} - \tilde{B}_t \overset{d}{=} \sqrt{s}\xi$, for $t, s \geq 0$, where $\xi$ is $G$-normally distributed.

**Remark 2.6.** (i) It is easy to check that the $G$-Brownian motion is symmetric, i.e., $(\tilde{B}_t)_{t \geq 0}$ is also a $G$-Brownian motion.

(ii) If, in particular, $G(A) = \frac{1}{2} \text{tr}(A)$, then the $G$-expectation is just a linear expectation with respect to the Wiener measure $P$, i.e., $\tilde{\mathbb{E}} = E_P$, and the $G$-Brownian motion is a classical Brownian motion over $(\Omega, \mathcal{B}(\Omega), P)$ ($\mathcal{B}(\Omega)$ denotes the Borel $\sigma$-field over $(\Omega, \rho_d)$).

The conditional $G$-expectation (knowing $\mathcal{B}(\Omega_t)$) for $X = \varphi(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}})$ at $t = t_j, 1 \leq j \leq n$, is defined by

$$\tilde{\mathbb{E}}_{t_j}[X] := \phi(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_j} - B_{t_{j-1}}),$$

where $\phi(x_1, \ldots, x_j) = \tilde{\mathbb{E}}[\varphi(x_1, \ldots, x_j, B_{t_{j+1}} - B_{t_j}, \ldots, B_{t_n} - B_{t_{n-1}})]$.

For every $p \geq 1$, we denote by $L^p_{G}(\Omega_t)$ ($L^p_{G}(\Omega)$, resp.) the completion of $L_{\text{lip}}(\Omega_t)$ ($L_{\text{lip}}(\Omega)$, resp.) under the norm $\|X\|_p := (\tilde{\mathbb{E}}[|X|^p])^{1/p}$. The conditional $G$-expectation $\tilde{E}_t[\cdot]$ ($t \geq 0$) can be extended continuously to $L^1_G(\Omega)$.

We recall the following representation theorem.

**Theorem 2.7** ([14][17]). Let

$$\mathcal{P} = \{P \text{ probability on } (\Omega, \mathcal{B}(\Omega)) : E_P[X] \leq \tilde{\mathbb{E}}[X], \text{ for all } X \in L^1_G(\Omega)\}.$$

Then $\mathcal{P} \neq \emptyset$ is a convex, weakly compact subset of the space $\mathcal{P}(\mathbb{R}^d)$ of all probability measures over $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ endowed with the topology of weak convergence, and

$$\tilde{\mathbb{E}}[\xi] = \sup_{P \in \mathcal{P}} E_P[\xi], \text{ for all } \xi \in L^1_G(\Omega).$$

The set $\mathcal{P}$ is said to represent $\tilde{\mathbb{E}}$.

The following definition introduces the notion of distributions of random variables under $G$-expectation.

**Definition 2.8.** Let $X = (X_1, \cdots, X_n)$ be a given $n$-dimensional random vector on a $G$-expectation space $(\Omega, \mathcal{H}, \tilde{\mathbb{E}})$. We define the functional $\mathbb{F}_X$ on the space of Lipschitz functions $C_{\text{lip}}(\mathbb{R}^n)$ by putting

$$\mathbb{F}_X[\varphi] := \tilde{\mathbb{E}}[\varphi(X)], \varphi \in C_{\text{lip}}(\mathbb{R}^n).$$

The triple $(\mathbb{R}^n, C_{\text{lip}}(\mathbb{R}^n), \mathbb{F}_X)$ forms a nonlinear expectation space, and $\mathbb{F}_X$ is called the distribution of $X$ under $\tilde{\mathbb{E}}$. 


We also shall introduce the space
\[ M^0_G(0,T) = \{ \eta_\sigma(\omega) = \sum_{i=0}^{N-1} \xi_i(\omega)I_{(s_i,s_{i+1})}(s) : N \geq 1, s_0 < \cdots < s_N \} \]
partition of \([0,T], \xi_i \in L^p_G(\Omega_n), 0 \leq i \leq N - 1\).
By \( M^p_G(0,T) \) and \( H^p_G(0,T) \) we denote the completion of \( M^p_G(0,T) \) under the norm \( \| \cdot \|_{M^p_G} := (\mathbb{E}[\int_0^T | \cdot |^p ds])^\frac{1}{p} \) and \( \| \cdot \|_{H^p_G} := (\mathbb{E}[\int_0^T | \cdot |^2 ds])^\frac{1}{2} \), respectively.

Define \( S^0_G(0,T) := \{ \eta_s := h(s, B_{s_1 \wedge s}, \ldots, B_{s_n \wedge s}) : s_1, \ldots, s_n \in [0,T], h \in C_b, \text{Lip} (\mathbb{R}^{n+1}) \} \).
For \( p \geq 1 \), we denote by \( S^p_G(0,T) \) the completion of \( S^0_G(0,T) \) under the norm \( \| \eta \|_{S^p_G} := (\mathbb{E}[\sup_{s \in [0,T]} | \eta_s |^p])^\frac{1}{p} \), \( \eta \in S^p_G(0,T) \).

Let us now recall the stochastic integration under the G-expectation. We define
\[ \int_0^t \eta^n_s dB_s := \sum_{i=0}^{n-1} \xi^n_i(B_{t_{i+1}} - B_{t_i}), \quad \eta^n := \sum_{i=0}^{N-1} \xi^n_i I_{(t_i,t_{i+1})}(t) \in M^2_G(0,T), \text{ and for } \eta \in M^2_G(0,T) \text{ with } \| \eta^n - \eta \|_{M^2_G} \rightarrow 0 \text{ (} n \rightarrow \infty \text{), we define} \]
\[ \int_0^t \eta_s dB_s := \lim_{n \rightarrow \infty} \int_0^t \eta^n_s dB_s, \]
where \( L^2_G \) indicates the convergence in \( L^2_G(\Omega) : \mathbb{E}[\| \int_0^t \eta^n_s dB_s - \int_0^t \eta_s dB_s \|^2] \rightarrow 0, n \rightarrow \infty \).

Similarly, we define \( \int_0^t \xi_s dB_s \) and \( \int_0^t \xi_s ds \) for \( \xi \in M^1_G(0,T) \), where \( \langle B \rangle \) denotes the cross-variation process of \( B \).

Let’s not least we recall that, given a measurable space \((\mathcal{X}, \mathcal{F})\) and an \( \mathcal{X} \)-valued random variable \( \xi \) defined on \((\Omega, \mathcal{B}(\mathbb{R}^d), P)\), we denote by \( P_\xi := P \circ \xi^{-1} \) the law induced by \( \xi \) on \((\mathcal{X}, \mathcal{F})\).

### 2.2 SDEs and BSDEs driven by G-Brownian motion

For simplicity, we only consider the one-dimensional case \( d = 1 \), and so also the G-Brownian motion is supposed to be one-dimensional. Recall that in this one-dimensional case \( G(a) = \frac{1}{2} \mathbb{E}[a B_t^2] \), and for \( \sigma^2 := \mathbb{E}[B_t^2] \) and \( \mathbf{a}^2 := -\mathbb{E}[B_t^4] \), we have \( G(a) = \frac{1}{2} (\sigma^2 a^+ - \mathbf{a}^2 a^-) \). Let us suppose throughout what follows that \( \mathbf{a}^2 > 0 \), i.e., we have \( 0 < \mathbf{a}^2 \leq \sigma^2 < +\infty \). When \( \mathbf{a}^2 = \sigma^2 \), the G-expectation is just a linear expectation.

We consider the following G-SDE: For given \( 0 \leq t \leq T < \infty \),
\[
\begin{cases}
  dX^{t,x}_s = b(s, X^{t,x}_s) ds + h(s, X^{t,x}_s) dB_s + \sigma(s, X^{t,x}_s) dB_s, & s \in [t,T], \\
  X^{t,x}_t = x,
\end{cases}
\]
where \( x \in \mathbb{R} \), and \( b, h, \sigma : [0,T] \times \mathbb{R} \rightarrow \mathbb{R} \) are given functions satisfying the following assumptions:

(H1) For some \( p \geq 2 \) it holds \( b(\cdot, x), h(\cdot, x), \sigma(\cdot, x) \in M^p_G(0,T) \), for all \( x \in \mathbb{R} \);
(H2) There exists a constant \( L > 0 \) such that for all \( x, x' \in \mathbb{R}, t \in [0,T], \)
\[ |b(t,x) - b(t,x')| + |h(t,x) - h(t,x')| + |\sigma(t,x) - \sigma(t,x')| \leq L |x - x'|. \]
For simplicity, \( X_s^{0,x} \) will be denoted by \( X_s^x \), for \( s \in [0, T] \), \( x \in \mathbb{R} \). We have the following estimates for G-SDE \((2.1)\) which can be found in \cite{22}.

**Lemma 2.9.** Assume that the conditions (H1) and (H2) hold. Then G-SDE \((2.1)\) has a unique solution \((X_s^{t,x})_{s \in [t, T]} \) \( \in M^p(t, T) \). Moreover, there exists a constant \( C \in \mathbb{R} \) depending on \( p, T, L \) and \( G \) such that, for all \( x, y \in \mathbb{R}, \ t, t' \in [0, T], \) we have

\[
\begin{align*}
&i) \quad \mathbb{E} \left[ \sup_{s \in [0,t]} |X_s^x|^p \right] \leq C(1 + |x|^p); \quad ii) \quad \mathbb{E} [ |X_t^x - X_t^{y}|^p ] \leq C \left( |x - y|^p + (1 + |x|^p) |t - t'|^{p/2} \right).
\end{align*}
\]

We also consider the following BSDE driven by a G-Brownian motion:

\[
Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dB_s - (K_T - K_t), \ 0 \leq t \leq T, \quad (2.2)
\]

where the coefficient \( f(t, \omega, y, z) : [0, T] \times \Omega_T \times \mathbb{R} \times \mathbb{R} \to \mathbb{R} \) is supposed to satisfy the following conditions:

(H3) There exists some \( \beta > 1 \) such that, for all \( y, z, f(\cdot, \cdot, y, z) \in M^\beta_G(0, T); \)

(H4) \( |f(t, \omega, y, z) - f(t, \omega, y', z')| \leq L(|y - y'| + |z - z'|), \ (t, \omega) \in [0, T] \times \Omega, \ y, z, y', z' \in \mathbb{R}, \) for some constant \( L > 0. \)

For simplicity, we denote by \( \mathfrak{S}_G^p(0, T) \) the collection of all processes \((Y, Z, K)\) such that \( Y \in S^p_G(0, T), \ Z \in H^p_G(0, T), \) and \( K \) is a non-increasing G-martingale with \( K_0 = 0 \) and \( K_T \in L^p_G(\Omega_T). \)

**Definition 2.10 (12).** Let \( \xi \in L^\beta_G(\Omega_T) \) and \( f \) satisfy (H3) and (H4) for \( \beta > 1 \). A triplet of processes \((Y, Z, K)\) is called a solution of \((2.2)\), if for some \( 1 < p \leq \beta \) the following properties hold:

(a) \((Y, Z, K) \in \mathfrak{S}_G^p(0, T); \)

(b) \( Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dB_s - (K_T - K_t), \ t \in [0, T]. \)

**Theorem 2.11 (12).** Assume that \( \xi \in L^\beta_G(\Omega_T) \) and \( f \) satisfies (H3) and (H4) for \( \beta > 1. \) Then \((2.2)\) has a unique solution \((Y, Z, K).\)

### 3 Formulation of the Problem

We consider as control state space \( U \) a non-empty, closed and convex bounded subset of \( \mathbb{R}^d \). A process \( u : [0, T] \times \Omega \to U \) is said to be an admissible control on \([0, T]\), if \( u \in M^2_G(0, T; U) \). By \( \mathcal{U} = M^2_G(0, T; U) \) we denote the class of all admissible controls \( u \). For any \( u \in \mathcal{U}, \) we consider the following stochastic differential equation

\[
\begin{align*}
\left\{ \begin{array}{l}
\quad dX^u_t = \sigma(X^u_t, \tilde{\mathbb{E}}[\varphi_1(X^u_t)], u_t)dB_t + b(X^u_t, \tilde{\mathbb{E}}[\varphi_2(X^u_t)], u_t)dt \\
\quad + \beta(X^u_t, \tilde{\mathbb{E}}[\varphi_3(X^u_t)], u_t)(B)_t, \ t \in [0, T], \\
\quad X^u_0 = x \in \mathbb{R},
\end{array} \right.
\end{align*}
\]

where \( b, \beta : [0, T] \times \mathbb{R} \times \mathbb{R} \times U \to \mathbb{R}, \sigma : [0, T] \times \mathbb{R} \times \mathbb{R} \times U \to \mathbb{R}, \) and \( \varphi_1, \varphi_2, \varphi_3 : \mathbb{R} \to \mathbb{R}. \)
The associated cost functional is given by

\[ J(u) := \hat{E}[\Phi(X_T^u, \hat{E}[\varphi_4(X_T^u)])] + \int_0^T l(t, X_t^u, \hat{E}[\varphi_5(X_t^u)], u_t) dt, \tag{3.2} \]

where \( \Phi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \), \( l : [0, T] \times \mathbb{R} \times \mathbb{R} \times U \rightarrow \mathbb{R} \), and \( \varphi_4, \varphi_5 : \mathbb{R} \rightarrow \mathbb{R} \).

The following assumptions will be in force throughout this paper.

\[ \text{(A.1)} \]  The functions \( \varphi_i, i = 1, 2, 3, 4, 5 \), are continuously differentiable, \( \Phi \) and \( l \) are continuously differentiable w.r.t. \( (x, y) \), and \( b, \sigma, \beta \) are continuously differentiable w.r.t. \( (x, y, v) \).

\[ \text{(A.2)} \]  All the derivatives in (A.1) are Lipschitz continuous and bounded.

For given \( u(\cdot) \in U \), \( X^u \) is called a solution of the above mean-field \( G \)-SDE if \( X^u \in M_G^2(0, T; \mathbb{R}^n) \) satisfies (3.1). Under the above assumptions, due to Lemma 2.9, SDE (3.1) has a unique solution.

**Lemma 3.1** ([17]). *(Existence and uniqueness of the solution)* If (A.1) and (A.2) are satisfied, then (3.1) has a unique solution \( X^u \), for all \( u \in U \).

The optimal control problem consists in minimizing the functional \( J(\cdot) \) over \( U \). An admissible control that minimizes \( J \) is called optimal.

Our main objective is to characterise the optimal control with the help of Pontryagin’s stochastic maximum principle. For this the study of the derivative under the sublinear \( G \)-expectations is crucial. This is the subject of the following section.

## 4 Derivative of a function of a law under \( G \)-expectation

According to Section 2, \( (\Omega, \mathcal{H}, \hat{E}) \) is a sublinear expectation space, where we restrict now to \( \Omega = \Omega_T = C([0, T]; \mathbb{R}) \). Recall that, due to Theorem 2.7 \( \mathcal{P} = \{ P \text{ a probability on } (\Omega, \mathcal{B}(\Omega)) : E_P[X] \leq \hat{E}[X], \text{ for } X \in L^1_G(\Omega) \} \) is a non empty convex, weakly compact subset of \( \mathcal{P}(\mathbb{R}) \) endowed with the topology of weak convergence. Moreover,

\[ \hat{E}[\xi] = \sup_{P \in \mathcal{P}} E_P[\xi], \text{ for all } \xi \in L^1_G(\Omega), \]

where the supremum is in fact a maximum: For all \( \xi \in L^1_G(\Omega) \), there exists \( P \in \mathcal{P} \) such that \( \hat{E}[\xi] = E_P[\xi] \) (see [17]). Consequently, the set

\[ \mathcal{P}_[\xi] := \{ P \in \mathcal{P} : \hat{E}[\xi] = E_P[\xi] \} \]

is nonempty.

Let \( \xi, \eta \in L^1_G(\Omega) \) and put \( F(\lambda) := \hat{E}[\xi + \lambda \eta], \lambda \in \mathbb{R} \). Now we study the differentiability of \( F \). From the definition of the \( G \)-expectation \( \hat{E} \) we know that \( F \) is convex. Indeed, for all \( \lambda, \lambda' \in \mathbb{R} \) and \( \rho \in (0, 1) \),

\[ F(\rho \lambda + (1 - \rho) \lambda') = \hat{E}[\xi + (\rho \lambda + (1 - \rho) \lambda') \eta] = \hat{E}[\rho (\xi + \lambda \eta) + (1 - \rho)(\xi + \lambda' \eta)] \leq \rho \hat{E}[\xi + \lambda \eta] + (1 - \rho)\hat{E}[\xi + \lambda' \eta] = \rho F(\lambda) + (1 - \rho)F(\lambda'). \]
Consequently, for all $\lambda \in \mathbb{R}$, there exists the right-derivative of $F$ at $\lambda$

$$F'_+(\lambda) = \lim_{0<\varepsilon \to 0} \frac{F(\lambda + \varepsilon) - F(\lambda)}{\varepsilon}$$

and also the corresponding left-derivative

$$F'_-(\lambda) = \lim_{0>\varepsilon \to 0} \frac{F(\lambda + \varepsilon) - F(\lambda)}{\varepsilon},$$

and, for all $\lambda < \lambda'$, we have

$$F'_-(\lambda) \leq F'_+(\lambda) \leq \frac{F(\lambda') - F(\lambda)}{\lambda' - \lambda} \leq F'_-(\lambda').$$

Let us compute $F'_+(0)$ with avoiding the $G$-martingale representation (Recall the $G$-martingale representation from Theorem 2.11 obtained for $f = 0$. Let us also mention that the derivative with use of the $G$-martingale representation as essential tool was discussed in [11].) To this end, we first give the following lemma.

**Lemma 4.1.** Let $\xi, \eta \in L^1_{ip}(\Omega)$ and $0 < \varepsilon_l \downarrow 0$ ($l \to \infty$), and let $P_l \in \mathcal{P}_{\{\xi+\varepsilon_l\eta\}}$, $l \geq 1$. Then we have

i) There exists a subsequence of $(P_l)$, denoted by $(P_{l_k})$, and $P \in \mathcal{P}$, such that $P_{l_k} \rightharpoonup P$, as $l_k \to \infty$ (weak convergence of probability measures);

ii) If $P_l \rightharpoonup P$, as $\varepsilon_l \downarrow 0$ ($l \to \infty$), for some $P \in \mathcal{P}$, then $P \in \mathcal{P}_{\{\xi\}}$.

**Proof.** i) From the weak compactness of $\mathcal{P}$ we get i).

ii) Assume that $P_l \rightharpoonup P$, as $l \to \infty$, for some $P \in \mathcal{P}$. Note that the functions in $L_{ip}(\Omega)$ are bounded and uniformly continuous. Thus, for all $\theta \in L_{ip}(\Omega)$, $E_{P_l}[\theta] \to E_{P}[\theta]$, as $l \to \infty$. Given any $\delta > 0$, let $\theta \in L_{ip}(\Omega)$ be such that $\mathbb{E}[|\theta - \xi|] \leq \delta$. Then, as $E_{P_l}[\theta] \to E_{P}[\theta]$, $l \to \infty$, we have

$$|E_{P_l}[\xi] - E_{P}[\xi]| \leq |E_{P_l}[\theta] - E_{P}[\theta]| + |E_{P_l}[\xi - \theta]| + |E_{P}[\xi - \theta]| \leq 2\delta + |E_{P_l}[\theta] - E_{P}[\theta]| \to 2\delta,$$

as $l \to \infty$.

From the arbitrariness of $\delta > 0$, it follows that $E_{P_l}[\xi] \to E_{P}[\xi]$, as $l \to \infty$. But, as $P_l \in \mathcal{P}_{\{\xi+\varepsilon_l\eta\}}$, $l \geq 1$,

$$|\mathbb{E}[\xi + \varepsilon_l\eta] - E_{P_l}[\xi]| = |E_{P_l}[\xi + \varepsilon_l\eta] - E_{P_l}[\xi]| \leq \varepsilon_l E_{P_l}[|\eta|] \leq \varepsilon_l \mathbb{E}[|\eta|] \to 0,$$

and also

$$|\mathbb{E}[\xi + \varepsilon_l\eta] - \mathbb{E}[\xi]| \leq \varepsilon_l \mathbb{E}[|\eta|] \to 0,$$

it follows that $\mathbb{E}[\xi] = E_{P}[\xi]$, i.e., $P \in \mathcal{P}_{\{\xi\}}$. 

We recall that the set $\mathcal{P}$ endowed with the weak convergence of probability measures is a compact metrisable space. Let $d(\cdot, \cdot)$ be a metric on $\mathcal{P}$ which is compatible with the weak convergence, e.g., we can choose the Lévy-Prokhorov metric (see Theorem 11.3-3, [8]):

$$d(P, Q) := \sup \left\{ \int_{\Omega} f d(P - Q), |f|_{BL} \leq 1 \right\},$$

10
where $|f|_{BL} = \sup_{\omega \in \Omega} |f(\omega)| + \sup_{\omega \neq \omega'} |f(\omega) - f(\omega')|/|\omega - \omega'|^{\xi(0,t)}$.

Observe also that, as $(\mathcal{P}, d)$ is a compact metric space, it is, in particular, also separable.

For $A, B \subset \mathcal{P}$, we put

$$d(P, B) := \text{dist}_B(P) = \inf\{d(P, Q) : Q \in B\}, \text{ for } P \in \mathcal{P}, \text{ and } \Gamma(A, B) := \sup_{P \in A} d(P, B).$$

Note that $\Gamma(A, B)$ is the maximal distance from $B$ of the probabilities in $A$. In particular, $\Gamma(A, B) = 0$, if $A \subset B$. Of course, $\Gamma(\cdot, \cdot)$ is not symmetric, its symmetrisation is just the Hausdorff distance $d_H(A, B) = \max\{\Gamma(A, B), \Gamma(B, A)\}$, $A, B \subset \mathcal{P}$.

**Proposition 4.2.** We have $\Gamma(\mathcal{P}_{[\xi, \eta]}, \mathcal{P}_{[\xi]}) \to 0$, as $\varepsilon \downarrow 0$.

**Proof.** Let $0 < \varepsilon_l \downarrow 0$ and $P_l \in \mathcal{P}_{[\xi, \eta]}$, be such that

$$\Gamma(\mathcal{P}_{[\xi, \eta]}, \mathcal{P}_{[\xi]}) - \frac{1}{l} \leq d(P_l, \mathcal{P}_{[\xi]}), \ l \geq 1.$$

Due to Lemma 4.1, for all subsequence $(P_{k_l})_{k \geq 1} \subset (P_l)_{l \geq 1}$, there exists some sub-subsequence $(P_{k_{l_n}})_{n \geq 1} \subset (P_{k_l})_{k \geq 1}$ and some $P \in \mathcal{P}_{[\xi]}$ such that $P_{k_{l_n}} \rightarrow P$, as $n \rightarrow \infty$. Then,

$$\Gamma(\mathcal{P}_{[\xi, \eta]}, \mathcal{P}_{[\xi]}) \leq d(P_{k_{l_n}}, \mathcal{P}_{[\xi]}) + \frac{1}{l_{k_n}} \leq d(P_{k_{l_n}}, P) + \frac{1}{l_{k_n}} \rightarrow 0, \text{ as } n \rightarrow \infty.$$

This implies

$$\Gamma(\mathcal{P}_{[\xi, \eta]}, \mathcal{P}_{[\xi]}) \rightarrow 0 \quad (l \rightarrow \infty),$$

for any $0 < \varepsilon_l \downarrow 0$, and, consequently, $\Gamma(\mathcal{P}_{[\xi, \eta]}, \mathcal{P}_{[\xi]}) \rightarrow 0$, as $\varepsilon \downarrow 0$. \hfill \Box

**Remark 4.3.** Lemma 4.1 can also be regarded as a consequence of Proposition 4.2. Indeed, for any $0 < \varepsilon_l \downarrow 0$, let $P_l \in \mathcal{P}_{[\xi, \eta]}$, $l \geq 1$. Then,

$$d(P_l, \mathcal{P}_{[\xi]}) \leq \Gamma(\mathcal{P}_{[\xi, \eta]}, \mathcal{P}_{[\xi]}) \rightarrow 0, \text{ as } 0 < \varepsilon_l \downarrow 0.$$

Let $Q_l \in \mathcal{P}_{[\xi]}$ be such that $d(P_l, Q_l) \leq d(P_l, \mathcal{P}_{[\xi]}) + \frac{1}{l}, l \geq 1$. As $\mathcal{P}$ is weakly compact, there is a subsequence $(Q_{k_l})_{k \geq 1} \subset (Q_l)_{l \geq 1}$ and some $Q \in \mathcal{P}$ such that $Q_{k_l} \rightarrow Q$. Consequently, due to the corresponding argument in the proof of Lemma 4.1, $\hat{E}[\xi] = E_Q[\xi] \rightarrow E_Q[\xi]$, i.e., also $Q \in \mathcal{P}_{[\xi]}$. Finally, from $d(P_{k_l}, Q) \leq d(P_{k_l}, Q_{k_l}) + d(Q_{k_l}, Q) \leq \Gamma(\mathcal{P}_{[\xi, \eta]}, \mathcal{P}_{[\xi]}) + \frac{1}{l_{k_l}} + d(Q_{k_l}, Q) \rightarrow 0, k \rightarrow \infty$, we see that $P_{k_l} \rightarrow Q \in \mathcal{P}_{[\xi]}$, as $k \rightarrow \infty$.

Let us now come to the computation of the right-derivative $F_{\lambda}'(0)$ of $F(\lambda) = \hat{E}[\xi + \lambda \eta]$ $(\xi, \eta \in L^1_G(\Omega))$ at $\lambda = 0$. For this we let $0 < \varepsilon_l \downarrow 0$, $P_l \in \mathcal{P}_{[\xi, \eta]}$ and $P \in \mathcal{P}_{[\xi]}$ be such that $P_l \rightarrow P$ (Due to Lemma 4.1, this choice is possible). In analogy to the fact that $P_l \rightarrow P$ implies $E_{P_l}[\xi] \rightarrow E_P[\xi]$, we get that, for any $\zeta \in L^1_G(\Omega)$, $E_{P_l}[\zeta] \rightarrow E_P[\zeta]$, as $l \rightarrow \infty$, and so $E_{P_l}[\eta] \rightarrow E_P[\eta]$, as $l \rightarrow \infty$. Then, as $P_l \in \mathcal{P}_{[\xi, \eta]}$,

$$F_{\lambda}'(0) = \lim_{\varepsilon \downarrow 0} \frac{\hat{E}[\xi + \varepsilon \eta] - \hat{E}[\xi]}{\varepsilon} \leq \lim_{l \rightarrow \infty} \frac{E_{P_l}[\xi + \varepsilon \eta] - E_{P_l}[\xi]}{\varepsilon_l} = E_P[\xi].$$
i.e., $F'(0) \leq E_P[\eta]$. On the other hand, for all $Q \in \mathcal{P}(\xi)$,

$$F'_+(0) = \lim_{\varepsilon \downarrow 0} \frac{\hat{E}[\xi + \varepsilon \eta] - \hat{E}[\xi]}{\varepsilon} \geq \lim_{\varepsilon \downarrow 0} \frac{E_Q[\xi + \varepsilon \eta] - E_Q[\xi]}{\varepsilon} = E_Q[\eta].$$

Consequently, we get the following lemma.

**Lemma 4.4.** For $\xi, \eta \in L^1_\mathcal{G}(\Omega)$ and $F(\lambda) := \hat{E}[\xi + \lambda \eta]$, we have

$$F'_+(0) = \max_{P \in \mathcal{P}(\xi)} E_P[\eta] = \hat{E}_{\{\xi\}}[\eta],$$

where $\hat{E}_{\{\xi\}}[\eta] := \sup_{P \in \mathcal{P}(\xi)} E_P[\eta]$, $\eta \in L^1_\mathcal{G}(\Omega)$, is a new sublinear expectation, and $\hat{E}_{\{\xi\}}$ is dominated by $\hat{E}$, i.e., $\hat{E}_{\{\xi\}}[\cdot] \leq \hat{E}[\cdot]$.

**Remark 4.5.** From the above lemma it follows that

$$F'_-(0) = \lim_{0 < \varepsilon \downarrow 0} \frac{\hat{E}[\xi - \varepsilon \eta] - \hat{E}[\xi]}{-\varepsilon} = \lim_{0 < \varepsilon \downarrow 0} \frac{\hat{E}[\xi + \varepsilon(-\eta)] - \hat{E}[\xi]}{\varepsilon} = -\hat{E}_{\{\xi\}}[-\eta].$$

This shows in particular that $F(\lambda) = \hat{E}[\xi + \lambda \eta]$ is differentiable at $\lambda = 0$ if and only if $\hat{E}_{\{\xi\}}[\eta] = -\hat{E}_{\{\xi\}}[-\eta]$.

We also observe that, for all $\lambda \in \mathbb{R}$,

$$F'_+(\lambda) = \lim_{0 < \varepsilon \downarrow 0} \frac{\hat{E}[(\xi + \lambda \eta) - \varepsilon \eta] - \hat{E}[\xi + \lambda \eta]}{\varepsilon} = \hat{E}_{\{\xi+\lambda\eta\}}[\eta],$$

$$F'_-(\lambda) = \lim_{0 < \varepsilon \downarrow 0} \frac{\hat{E}[(\xi + \lambda \eta) - \varepsilon(-\eta)] - \hat{E}[\xi + \lambda \eta]}{\varepsilon} = -\hat{E}_{\{\xi+\lambda\eta\}}[-\eta],$$

and as $F$ is convex, for all $\lambda < \lambda'$,

$$-\hat{E}_{\{\xi+\lambda\eta\}}[-\eta] \leq \hat{E}_{\{\xi+\lambda\eta\}}[\eta] \leq \hat{E}_{\{\xi+\lambda\eta\}}[\eta] \leq -\hat{E}_{\{\xi+\lambda'\eta\}}[-\eta].$$

**Corollary 4.6.** Let $\varphi \in C^1(\mathbb{R})$ have a bounded Lipschitz derivative $\partial \varphi : \mathbb{R} \to \mathbb{R}$, and let $\xi, \eta \in L^1_\mathcal{G}(\Omega)$. Then, for $H(\lambda) := \hat{E}[\varphi(\xi + \lambda \eta)]$, $\lambda \in \mathbb{R}$, we have

i) $H'_+(0) = \hat{E}_{\{\varphi(\xi)\}}[\partial \varphi(\xi) \eta]$; and ii) $H'_-(0) = -\hat{E}_{\{\varphi(\xi)\}}[-\partial \varphi(\xi) \eta]$.

**Proof.** Let $\xi' := \varphi(\xi)$ and $\eta' := \partial \varphi(\xi) \eta$. Then, $\xi', \eta' \in L^1_\mathcal{G}(\Omega)$, and

$$\varphi(\xi + \varepsilon \eta) = \varphi(\xi) + \int_0^1 \partial \varphi(\xi + \lambda \varepsilon \eta) d\lambda = \varphi(\xi) + \int_0^1 \partial \varphi(\xi + \lambda \eta) d\lambda \cdot \varepsilon \eta$$

$$= \varphi(\xi) + \varepsilon (\partial \varphi(\xi) \eta) + R_\varepsilon,$$

where $R_\varepsilon := \int_0^1 (\partial \varphi(\xi + \lambda \eta) - \partial \varphi(\xi)) d\lambda \cdot \varepsilon \eta$, and from the Lipschitz property of $\partial \varphi$ we have $|R_\varepsilon| \leq C_\eta^2 |\varepsilon|^2$ and $\hat{E}[|R_\varepsilon|] \leq C_\eta^2 \varepsilon > 0$. Hence, as $\varphi(\xi + \eta) = \xi' + \varepsilon \eta' + R_\varepsilon$,

$$\left| \frac{\hat{E}[\varphi(\xi + \varepsilon \eta)] - \hat{E}[\varphi(\xi) \eta]}{\varepsilon} - \hat{E}_{\{\varphi(\xi)\}}[\partial \varphi(\xi) \eta] \right| \leq \left| \frac{\hat{E}[\varphi(\xi + \eta)] - \hat{E}[\xi'] \eta'}{\varepsilon} - \hat{E}_{\{\xi'\} \eta'} \right| + \hat{E}_{\{\xi'\}}[R_\varepsilon] \to 0,$$
as $\varepsilon \downarrow 0$. This proves i). For ii) we note that, from i) with $-\eta$ instead of $\eta$, it follows that
\begin{equation}
H'_-(0) = -\lim_{0<\varepsilon\downarrow 0} \frac{\mathbb{E}[(\varphi(\xi + \varepsilon(-\eta)) - \mathbb{E}[\varphi(\xi))] - \mathbb{E}[\varphi(\xi)]]}{\varepsilon} = -\mathbb{E}[(\varphi(\xi))] \frac{-\partial \varphi(\xi)\eta}{\varepsilon}.
\end{equation}

\[\square\]

**Remark 4.7.** Observe that Corollary 4.4 can be extended in a straight-forward way to $d$-dimensional random variables $\xi, \eta \in (L^1_{G_t}(\Omega))^d$. Indeed, in the proof of the corollary it suffices the derivative $\partial \varphi$ replaced by the gradient $\nabla \varphi$.

### 5 Pontryagin’s SMP for mean field stochastic control problems under $G$-expectation

#### 5.1 Necessary conditions for optimality

In the next both sections, to simplify the dynamics and the related computations, we put $\beta = 0$, and so SDE (3.1) becomes
\begin{equation}
\begin{cases}
da_t^u = b(x_t^u, \mathbb{E}[\varphi_1(x_t^u)], u_t)dt + \sigma(x_t^u, \mathbb{E}[\varphi_2(x_t^u)], u_t)dB_t, \ t \in [0, T],
x_0^u = x \in \mathbb{R}^n.
\end{cases}
\end{equation}

The cost functional is still given by (3.2),
\begin{equation}
J(u) := \mathbb{E}[\Phi(x_T^u, \mathbb{E}[\varphi_4(x_T^u)]) + \int_0^T l(t, x_t^u, \mathbb{E}[\varphi_5(x_t^u)], u_t)dt], \text{ for any } u \in \mathcal{U}.
\end{equation}

We suppose that there exists an optimal control $\hat{u} \in \mathcal{U}$, that is, $J(\hat{u}) \leq J(u)$, for all $u \in \mathcal{U}$. Let us denote $\hat{x}_t := x_t^u, t \in [0, T]$.

#### 5.1.1 Taylor expansions

Let $u$ be an arbitrary admissible control process in $\mathcal{U}$. By $x^u$ we denote the state process defined by SDE (5.1) with the control process $u^\theta$ defined as convex perturbation of $\hat{u}$:
\begin{equation}
u_t^\theta = \hat{u}_t + \theta(u_t - \hat{u}_t), \ t \in [0, T], \ \theta \in [0, 1].
\end{equation}

We put $v_t = u_t - \hat{u}_t, t \in [0, T]$, and introduce the following notations:
\begin{equation}
\begin{aligned}
\hat{b}(t) &= b(t, \hat{x}_t, \mathbb{E}[\varphi_1(t)], \hat{u}_t), & \hat{b}_x(t) &= b_x(t, \hat{x}_t, \mathbb{E}[\varphi_1(t)], \hat{u}_t), \\
\hat{\sigma}(t) &= \sigma(t, \hat{x}_t, \mathbb{E}[\varphi_2(t)], \hat{u}_t), & \hat{\sigma}_x(t) &= \sigma_x(t, \hat{x}_t, \mathbb{E}[\varphi_2(t)], \hat{u}_t), \\
\hat{\Phi}(t) &= \Phi(\hat{x}_t, \mathbb{E}[\varphi_4(t)]), & \hat{\Phi}_x(t) &= \Phi_x(\hat{x}_t, \mathbb{E}[\varphi_4(t)]), \\
\hat{l}(t) &= l(t, \hat{x}_t, \mathbb{E}[\varphi_5(t)], \hat{u}_t), & \hat{l}_x(t) &= l_x(t, \hat{x}_t, \mathbb{E}[\varphi_5(t)], \hat{u}_t),
\end{aligned}
\end{equation}

with $\varphi_2(t) = \varphi_1(t), \varphi_4(t) = \varphi_1(t), i = 1, 2, 4, 5$, and similarly are defined $\hat{b}_y(t), \hat{b}_v(t), \hat{\sigma}_y(t), \hat{\sigma}_v(t), \hat{l}_y(t)$ and $\hat{l}_v(t)$. Here, for notational convenience, we denote by $b_x, b_y, b_v$ the derivative of $b$ w.r.t. the state trajectory, the expected value and the control variable, respectively, and similarly for the other functions.

The objective of this section is to determine the directional derivative of the cost functional in terms of the first order Taylor expansion of the state process. We begin with identifying the Taylor expansion.
Lemma 5.1. Let \( z = (z_t) \in M^2(0, T; \mathbb{R}) \) be the unique solution of the following SDE

\[
\left\{ \begin{array}{l}
dz_t = \left( \hat{b}_x(t)z_t + \hat{b}_y(t)\tilde{E}_{\{\varphi_1(z_t)\}}[\varphi_1'(t)z_t] + \hat{b}_v(t)v_t \right) dt \\
\quad + \left( \hat{\sigma}_x(t)z_t + \hat{\sigma}_y(t)\tilde{E}_{\{\varphi_2(z_t)\}}[\varphi_2'(t)z_t] + \hat{\sigma}_v(t)v_t \right) dB_t, \quad t \in [0, T], \\
\quad z_0 = 0.
\end{array} \right.
\] (5.4)

Then, it holds that

\[
\lim_{\theta \to 0} \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \frac{x^\theta_t - \hat{x}_t}{\theta} - z_t \right|^2 \right] = 0.
\]

Proof. First we observe that, thanks to our assumptions on the coefficients, we have the existence and the uniqueness for SDE (5.4). Now, to simplify our computations, but without loss of generality for the method of the proof, let \( b = 0 \). So SDE (5.1) becomes

\[
\left\{ \begin{array}{l}
dx^\theta_t = \sigma(x^\theta_t, \tilde{E}_{\{\varphi_2(x^\theta_t)\}}, u_t) dB_t, \quad t \in [0, T], \\
x^\theta_0 = x,
\end{array} \right.
\] (5.5)

while SDE (5.4) writes

\[
\left\{ \begin{array}{l}
dz_t = \left( \hat{\sigma}_x(t)z_t + \hat{\sigma}_y(t)\tilde{E}_{\{\varphi_2(z_t)\}}[\varphi_2'(t)z_t] + \hat{\sigma}_v(t)v_t \right) dB_t, \quad t \in [0, T], \\
z_0 = 0.
\end{array} \right.
\] (5.6)

Putting \( \hat{\rho}_t := \tilde{E}_{\{\varphi_2(\hat{x}_t)\}}, \hat{\rho}^\theta_t := \tilde{E}_{\{\varphi_2(x^\theta_t)\}}, \hat{u}_t := \hat{u}_t + \theta v_t \) and \( \Theta_t^\lambda := (\hat{x}_t + \lambda(x^\theta_t - \hat{x}_t), \hat{\rho}_t + \lambda(\hat{\rho}^\theta_t - \hat{\rho}_t), \hat{u}_t + \lambda \theta v_t) \), we have

\[
\frac{1}{\theta}(x^\theta_t - \hat{x}_t) = \frac{1}{\theta} \int_0^t \left( \sigma(x^\theta_s, \rho^\theta_s, u^\theta_s) - \sigma(\hat{x}_s, \hat{\rho}_s, \hat{u}_s) \right) dB_s = \frac{1}{\theta} \int_0^t \int_0^1 \partial_\lambda \left[ \sigma(\Theta^\lambda_s) \right] d\lambda dB_s
\]

\[
= \frac{1}{\theta} \int_0^t \int_0^1 \left\{ \sigma_x(\Theta^\lambda_s)(x^\theta_s - \hat{x}_s) + \sigma_y(\Theta^\lambda_s)(\rho^\theta_s - \hat{\rho}_s) + \sigma_v(\Theta^\lambda_s) \theta v_s \right\} d\lambda dB_s
\]

\[
= \int_0^t \left\{ \hat{\sigma}_x(s) \frac{1}{\theta}(x^\theta_s - \hat{x}_s) + \hat{\sigma}_y(s) \frac{1}{\theta}(\rho^\theta_s - \hat{\rho}_s) + \hat{\sigma}_v(s) v_s \right\} dB_s + R^\theta_t,
\] (5.7)

where

\[
R^\theta_t := \int_0^t \int_0^1 \left\{ \sigma_x(\Theta^\lambda_s) - \hat{\sigma}_x(s) \right\} \frac{x^\theta_s - \hat{x}_s}{\theta} + \left( \sigma_y(\Theta^\lambda_s) - \hat{\sigma}_y(s) \right) \frac{\rho^\theta_s - \hat{\rho}_s}{\theta} + \left( \sigma_v(\Theta^\lambda_s) - \hat{\sigma}_v(s) \right) v_s \right\} d\lambda dB_s.
\] (5.8)

We put \( \nabla \sigma := (\sigma_x, \sigma_y, \sigma_v) \) and \( \nabla \hat{\sigma}(s) := (\hat{\sigma}_x(s), \hat{\sigma}_y(s), \hat{\sigma}_v(s)) \). Then, thanks to Assumption (A.2),

\[
|\nabla \sigma(\Theta^\lambda_s) - \nabla \hat{\sigma}(s)| \leq C \left( |x^\theta_s - \hat{x}_s| + |\rho^\theta_s - \hat{\rho}_s| + \theta |v_s| \right).
\]

Since \( x^\theta_t - \hat{x}_t = \int_0^t \sigma(x^\theta_s, \rho^\theta_s, u^\theta_s) - \sigma(\hat{x}_s, \hat{\rho}_s, \hat{u}_s) \right) dB_s \), for all \( p \geq 2 \), we have

\[
\mathbb{E} \left[ \sup_{0 \leq s \leq t} |x^\theta_s - \hat{x}_s|^p \right] \leq C_p \mathbb{E} \left[ \left( \int_0^t \left( |x^\theta_s - \hat{x}_s| + |\rho^\theta_s - \hat{\rho}_s| + \theta |v_s| \right)^2 ds \right)^{\frac{p}{2}} \right]
\]

\[
\leq C_p \mathbb{E} \left[ \int_0^t \left( |x^\theta_s - \hat{x}_s|^p + (\mathbb{E} |x^\theta_s - \hat{x}_s|^p)^p + \theta^p |v_s|^p \right) ds \right]
\]

\[
\leq C_p (\theta^p + \int_0^t \mathbb{E} \left[ |x^\theta_s - \hat{x}_s|^p \right] ds), \quad t \in [0, T]
\]

14
Furthermore, we get
\[ \mathbb{E} \left[ \sup_{0 \leq s \leq T} |x^\theta_s - \hat{x}_s|^p \right] \leq C_p \theta^p, \quad \theta > 0. \tag{5.9} \]
Hence, as \( \varphi_2 \) is Lipschitz, also for \( \rho^\theta - \hat{\rho} \) we have
\[ \sup_{0 \leq s \leq T} |\rho^\theta_s - \hat{\rho}_s|^p \leq C_p \theta^p, \quad \theta > 0. \tag{5.10} \]
On the other hand, by standard estimates we have \( \mathbb{E} \left[ \sup_{0 \leq s \leq T} |z_s|^p \right] \leq C_p, \ p \geq 2. \)
From (5.8), (5.9) and (5.10), for some \( p \geq 2, \)
\[ \mathbb{E} \left[ \sup_{t \in [0,T]} |R^\theta_t|^p \right] \leq C_p \mathbb{E} \left[ \left( \int_0^T \left( \left( |x^\theta_s - \hat{x}_s| + |\rho^\theta_s - \hat{\rho}_s| + \theta |v_s| \right)^{2\theta} \right)^{\frac{1}{2\theta}} ds \right]^2 \right] \leq C_p \frac{1}{\theta^p} \left( \mathbb{E} \left[ \sup_{t \in [0,T]} |x^\theta_s - \hat{x}_s|^{2p} \right] + \sup_{t \in [0,T]} |\rho^\theta_s - \hat{\rho}_s|^{2p} + \theta^{2p} \right) \leq C_p \theta^p, \quad \theta > 0. \tag{5.11} \]
We put \( y^\theta_t := \frac{x^\theta_t - \hat{x}_t}{\theta} - z_t, \) \( t \in [0,T], \) and we have
\[ \varphi_2(x^\theta_s) - \varphi_2(\hat{x}_s) = \left( \int_0^1 \varphi_2'(\hat{x}_s + \lambda(x^\theta_s - \hat{x}_s)) \, d\lambda \right) (x^\theta_s - \hat{x}_s) = \varphi_2'(\hat{x}_s)(x^\theta_s - \hat{x}_s) + \tilde{R}^\theta_s \]
\[ = \theta (\varphi_2'(\hat{x}_s)z_s) + \varphi_2'(\hat{x}_s)(x^\theta_s - \hat{x}_s - \theta z_s) + \tilde{R}^\theta_s = \theta (\varphi_2'(\hat{x}_s)z_s) + \theta \varphi_2'(\hat{x}_s)y^\theta_s + \tilde{R}^\theta_s, \]
where
\[ \tilde{R}^\theta_s := \left( \int_0^1 \varphi_2'(\hat{x}_s + \lambda(x^\theta_s - \hat{x}_s)) - \varphi_2'(\hat{x}_s) \, d\lambda \right) (x^\theta_s - \hat{x}_s), \]
and \( |\tilde{R}^\theta_s| \leq C|x^\theta_s - \hat{x}_s|^2, \) i.e.,
\[ \varphi_2(x^\theta_s) = \varphi_2(\hat{x}_s) + \theta (\varphi_2'(\hat{x}_s)z_s) + \theta \varphi_2'(\hat{x}_s)y^\theta_s + \tilde{R}^\theta_s. \tag{5.12} \]
Furthermore, we get
\[ \frac{\rho^\theta_s - \hat{\rho}_s}{\theta} = \frac{1}{\theta} \left( \mathbb{E} \left[ \varphi_2(x^\theta_s) \right] - \mathbb{E} \left[ \varphi_2(\hat{x}_s) \right] \right) \]
\[ = \frac{1}{\theta} \left( \mathbb{E} \left[ \varphi_2(\hat{x}_s) + \theta (\varphi_2'(\hat{x}_s)z_s) \right] - \mathbb{E} \left[ \varphi_2(\hat{x}_s) + \theta \varphi_2'(\hat{x}_s)y^\theta_s \right] + \tilde{R}^\theta_s(y^\theta_s) \right), \tag{5.13} \]
where, thanks to (5.12),
\[ \tilde{R}^\theta_s(\eta) := \frac{1}{\theta} \left( \mathbb{E} \left[ \varphi_2(\hat{x}_s) + \theta (\varphi_2'(\hat{x}_s)z_s) + \theta \varphi_2'(\hat{x}_s)\eta + \tilde{R}^\theta_s \right] - \mathbb{E} \left[ \varphi_2(\hat{x}_s) + \theta (\varphi_2'(\hat{x}_s)z_s) \right] \right), \eta \in L^2(\tilde{\omega}). \]
Notice that, due to (5.9),
\[ |\tilde{R}^\theta_s(\eta)| \leq C \mathbb{E}[|\eta|] + C \mathbb{E} \left[ \left| \frac{x^\theta_s - \hat{x}_s}{\theta} \right|^2 \right] \leq C (\theta + \mathbb{E}[|\eta|]), \quad \theta > 0. \tag{5.14} \]
Now, we define \( F_s(\theta) := \mathbb{E} \left[ \varphi_2(\hat{x}_s) + \theta (\varphi_2'(\hat{x}_s)z_s) \right], \) \( \theta \geq 0. \) Then, as \( F_s : \mathbb{R}_+ \rightarrow \mathbb{R} \) is convex, we obtain
\[ 0 \leq H_s(\theta) := \frac{F_s(\theta) - F_s(0)}{\theta} - F_{s,+}'(0) \leq \frac{F_{s,+}'(\theta) - F_{s,+}'(0)}{\theta} \downarrow 0, \quad \text{as } \theta \downarrow 0, \tag{5.15} \]
where, due to Lemma 4.3, the right-derivative of $F_s(\lambda)$ at $\lambda = 0$ satisfies

$$F_{s,+}'(0) = \mathbb{E}_{\{\varphi_2(\hat{x}_s)\}}[\varphi_2'(\hat{x}_s)z_s].$$

Observe that, thanks to (5.13) and the above definition of $F_s(\theta)$ and $H_s(\theta)$, we have

$$\frac{\theta^\alpha - \hat{\theta}}{\theta} = H_s(\theta) + F_{s,+}'(0) + \hat{R}_t(\theta), \quad s \in [0, T].$$

Then, recalling that $y_s^\theta := \frac{x_t^\theta - \hat{x}_t}{\theta} - z_t$, from (5.6) and (5.7), we get

$$dy_t^\theta = d\left(\frac{1}{\theta} (x_t^\theta - \hat{x}_t) - z_t\right) = \left(\hat{\sigma}_x(t) y_t^\theta + \hat{\sigma}_y(t) \left(\frac{1}{\theta} (\rho_t^\theta - \hat{\rho}_t) - \mathbb{E}_{\{\varphi_2(\hat{x}_t)\}}[\varphi_2'(t)z_t]\right)\right)dB_t + dR_t^\theta$$

$$\begin{cases} d\left(\frac{1}{\theta} (x_t^\theta - \hat{x}_t) - z_t\right) = \left(\hat{\sigma}_x(t) y_t^\theta + \hat{\sigma}_y(t) (H_t(\theta) + \hat{R}_t^\theta(y_t^\theta))\right)dB_t + dR_t^\theta, \\
y_0^\theta = 0.\end{cases}$$

Consequently, from (5.11) and (5.14), for $p \geq 2$,

$$\mathbb{E}[\sup_{s \in [0,t]} |y_s^\theta|^p] \leq C_p \int_0^t \left(\mathbb{E}[|y_s^\theta|^p] + |H_s(\theta)|^p + (C\theta)^p\right)ds + C_p \theta^p, \quad t \in [0, T], \quad \theta > 0,$$

and, thanks to Gronwall’s inequality, $\mathbb{E}[\sup_{t \in [0,T]} |y_t^\theta|^p] \leq C_p \left(\theta^p + \int_0^T |H_s(\theta)|^p ds\right), \quad \theta > 0$.

Note that, for $0 < \theta \leq 1$, as $F_s(\cdot)$ is convex, we have

$$0 \leq H_s(\theta) \leq F_{s,+}'(\theta) - F_{s,+}'(0) \leq \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\mathbb{E}[\varphi_2(x_s) + \varphi_2'(x_s)z_s + \varepsilon \varphi_2'(x_s)z_s] - \mathbb{E}_{\{\varphi_2(\hat{x}_s)\}}[\varphi_2'(\hat{x}_s)z_s]\right)$$

$$\leq 2\mathbb{E}[|\varphi_2'(x_s)z_s|] \leq C\mathbb{E}[|z_s|] \leq C\mathbb{E}[\sup_{s \in [0,T]} |z_s|] =: C^* < \infty.$$

Thus, since $0 \leq H_s(\theta) \downarrow 0$, as $\theta \downarrow 0$, $s \in [0, T]$, it follows from the bounded convergence theorem applied to $\int_0^T |H_s(\theta)|^p ds$ that

$$\mathbb{E}[\sup_{s \in [0,T]} |y_s^\theta|^p] \to 0, \quad \text{as} \quad \theta \downarrow 0, \quad p \geq 2. \quad (5.16)$$

□

Remark 5.2. From (5.13), (5.14), (5.15) and (5.16), we have

$$\begin{align*}
\frac{1}{\theta} \left(\mathbb{E}[\varphi_2(x_t^\theta)] - \mathbb{E}[\varphi_2(\hat{x}_t)]\right) &= H_t(\theta) + \hat{R}_t(\theta, y_t^\theta) \\
&\leq H_t(\theta) + C(\theta + \mathbb{E}[|y_t^\theta|]) \to 0, \quad \text{as} \quad \theta \downarrow 0, \quad \text{i.e.,}
\end{align*}$$

$$\lim_{\theta \downarrow 0} \frac{1}{\theta} \left(\mathbb{E}[\varphi_2(x_t^\theta)] - \mathbb{E}[\varphi_2(\hat{x}_t)]\right) = \mathbb{E}_{\{\varphi_2(\hat{x}_t)\}}[\varphi_2'(\hat{x}_t)z_t], \quad t \in [0, T]. \quad (5.17)$$
Lemma 5.3. The directional derivative of the cost functional $J$ is given by

$$
\lim_{\theta \downarrow 0} \frac{J(\dot{u} + \theta v) - J(\dot{u})}{\theta} = \hat{E}_{\{\psi(\hat{u})\}} \left[ \hat{\Phi}_x(T) z_T + \hat{\Phi}_y(T) \hat{E}_{\{\varphi_4(\hat{x}_T)\}} [\varphi'_4(\hat{x}_T) z_T] 
+ \int_0^T \left( \hat{l}_x(t) z_t + \hat{l}_y(t) \hat{E}_{\{\varphi_5(\hat{x}_t)\}} [\varphi'_5(\hat{x}_t) z_t] + \hat{l}_v(t) v_t \right) dt \right],
$$

where $\psi(\hat{u}) = \hat{\Phi}(T) + \int_0^T \hat{l}(t) dt = \Phi(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)]) + \int_0^T \hat{l}(t, \hat{x}_t, \hat{E}[\varphi_5(\hat{x}_t)], \hat{u}_t) dt$; for the other abbreviating notations, see (5.3).

Proof. For simplicity, but without restriction of the generality of the arguments, we suppose that $l = 0$: $J(u) = \hat{E}[\hat{\Phi}(x_T^\theta, \hat{E}[\varphi_4(x_T^\theta)])], \ u \in U$.

From Lemma 5.1 for $y_t^\theta = \frac{x_T^\theta - \dot{x}_t}{\theta} - z_t, \ t \in [0, T]$, we have

$$
\hat{E} \left[ \sup_{t \in [0, T]} |y_t^\theta|^2 \right] \to 0, \ \text{as} \ \theta \downarrow 0.
$$

Hence, for $H(\theta) := \hat{E}[\varphi_4(\hat{x}_T + \theta z_T)], \ \theta \geq 0,$

$$
|\hat{E}[\varphi_4(x_T^\theta)] - H(\theta)| = |\hat{E}[\varphi_4(\hat{x}_T + \theta z_T + \theta y_T^\theta)] - \hat{E}[\varphi_4(\hat{x}_T + \theta z_T)]| \leq C\theta \hat{E}[|y_T^\theta|],
$$

and, thus, similar to the proof of (5.17), we have

$$
\lim_{0 < \theta \downarrow 0} \frac{\hat{E}[\varphi_4(x_T^\theta)] - \hat{E}[\varphi_4(\hat{x}_T)]}{\theta} = H'_+(0),
$$

where, thanks to Corollary 4.6, $H'_+(0) = \hat{E}_{\{\varphi_4(\hat{x}_T)\}} [\varphi'_4(\hat{x}_T) z_T]$.

Putting $r_\theta := \frac{1}{\theta} \left( \hat{E}[\varphi_4(x_T^\theta)] - \hat{E}[\varphi_4(\hat{x}_T)] \right) - \hat{E}_{\{\varphi_4(\hat{x}_T)\}} [\varphi'_4(\hat{x}_T) z_T], \ \theta > 0$, and $\psi(\theta) := \hat{E} \left[ \hat{\Phi} \left( (\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)]), \theta(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)]), \theta \varphi'_4(\hat{x}_T) z_T) \right) \right], \ \theta \geq 0$, we have

$$
|\hat{E}[\hat{\Phi}(x_T^\theta, \hat{E}[\varphi_4(x_T^\theta)])] - \psi(\theta)| \leq C\theta (\hat{E}[|y_T^\theta|^2] + r_\theta^2) \frac{1}{\theta}, \ \theta > 0,
$$

with $\hat{E}[|y_T^\theta|^2] + r_\theta^2 \to 0$, as $\theta \downarrow 0$. Consequently,

$$
\lim_{0 < \theta \downarrow 0} \frac{\hat{E}[\hat{\Phi}(x_T^\theta, \hat{E}[\varphi_4(x_T^\theta)])] - \hat{E}[\hat{\Phi}(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)])]}{\theta} = \lim_{0 < \theta \downarrow 0} \left( \frac{\hat{E}[\hat{\Phi}(x_T^\theta, \hat{E}[\varphi_4(x_T^\theta)])] - \psi(\theta)}{\theta} + \frac{\psi(\theta) - \psi(0)}{\theta} \right) = \psi'_+(0),
$$

and from Remark 4.7,

$$
\psi'_+(0) = \hat{E}_{\{\psi(\hat{u})\}} \left[ (\partial_x \Phi)(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)]) z_T + (\partial_y \Phi)(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)]) \hat{E}_{\{\varphi_4(\hat{x}_T)\}} [\varphi'_4(\hat{x}_T) z_T] \right].
$$

The proof is complete. □
5.1.2 Duality

In this section, we consider the special case where $\sigma$ and $b$ are independent of $y$, and we still put $\beta = 0$. More general cases can be studied with the same approach as that we develop here, but, of course, this is related with more involved computations. In the case we study here (3.1) becomes

$$\begin{align*}
    dx^u_t = \sigma(x^u_t, u_t)dB_t + b(x^u_t, u_t)dt, \ t \in [0, T], \\
    x^u_0 = x \in \mathbb{R}^n.
\end{align*}$$

(5.18)

Concerning the cost functional, we make the following assumption

(A.3) $\tilde{\Phi}_y(T) \geq 0, \hat{i}_y(t) \geq 0, \ t \in [0, T], \text{ quasi-surely}.$

Of course, this assumption is, in particular, satisfied, if the partial derivates $\partial_y \Phi(.,.)$ and $\partial_y l(.,.,.,.)$ are everywhere non negative.

Recall from (3.2) that the cost functional is given by

$$J(u) := \tilde{E}[\Phi(x^u_T, \tilde{E}[\varphi_4(x^u_T)]) + \int_0^T l(t, x^u_t, \tilde{E}[\varphi_5(x^u_t)], u_t)dt].$$

(5.19)

Then from the optimality of $\hat{u}$, thanks to Lemma 5.3 with the notation $\psi(\hat{u}) = \Phi(\hat{x}_T, \tilde{E}[\varphi_4(\hat{x}_T)]) + \int_0^T l(t, \hat{x}_t, \tilde{E}[\varphi_5(\hat{x}_t)], \hat{u}_t)dt$ and those introduced in (5.3) we have

$$0 \leq \lim_{\theta \downarrow 0} \frac{J(\hat{u} + \theta v) - J(\hat{u})}{\theta}$$

$$= \tilde{E}_{\{\psi(\hat{u})\}}[\hat{\Phi}_x(T)z_T + \hat{\Phi}_y(T)\tilde{E}_{\{\varphi_4(\hat{x}_T)\}}[\varphi'_4(\hat{x}_T)z_T]$$

$$+ \int_0^T (\hat{i}_x(t)z_t + \hat{i}_y(t)\tilde{E}_{\{\varphi_5(\hat{x}_t)\}}[\varphi'_5(\hat{x}_t)z_t] + \hat{i}_v(t)v_t)dt]$$

$$= \sup_{P^1 \in \mathcal{P}_{\{\psi(\hat{u})\}}} E_{P^1}[\hat{\Phi}_x(T)z_T + \hat{\Phi}_y(T)\sup_{P^2 \in \mathcal{P}_{\{\varphi_4(\hat{x}_T)\}}} E_{P^2}[\varphi'_4(\hat{x}_T)z_T]$$

$$+ \int_0^T (\hat{i}_x(t)z_t + \hat{i}_y(t)\sup_{P^3 \in \mathcal{P}_{\{\varphi_5(\hat{x}_t)\}}} E_{P^3}[\varphi'_5(\hat{x}_t)z_t] + \hat{i}_v(t)v_t)dt]$$

$$= \sup_{P^1 \in \mathcal{P}_{\{\psi(\hat{u})\}}} \left\{ E_{P^1}[\hat{\Phi}_x(T)z_T + \int_0^T (\hat{i}_x(t)z_t + \hat{i}_v(t)v_t)dt] \right\}$$

$$+ E_{P^1}[\hat{\Phi}_y(T)] \sup_{P^2 \in \mathcal{P}_{\{\varphi_4(\hat{x}_T)\}}} E_{P^2}[\varphi'_4(\hat{x}_T)z_T] + \int_0^T E_{P^1}[\hat{i}_y(t)] \sup_{P^3 \in \mathcal{P}_{\{\varphi_5(\hat{x}_t)\}}} E_{P^3}[\varphi'_5(\hat{x}_t)z_t]dt \right\}.$$

Let us now define

$$\mathcal{R}_{\{\varphi_5(\hat{x}_t)\}} : = \{ R = (R_t) : [0, T] \to \mathcal{P} \text{ Borel measurable} : R_t \in \mathcal{P}_{\{\varphi_5(\hat{x}_t)\}}, \ t \in [0, T]\},$$

$$\mathcal{R}_{\{\varphi_5(\hat{x}_t)||\varphi'_5(\hat{x}_t)z_t\}} : = \{ R = (R_t) : [0, T] \to \mathcal{P} \text{ Borel measurable} : R_t \in \mathcal{P}_{\{\varphi_5(\hat{x}_t)||\varphi'_5(\hat{x}_t)z_t\}}, \ t \in [0, T]\},$$

where

$$\mathcal{P}_{\{\varphi_5(\hat{x}_t)||\varphi'_5(\hat{x}_t)z_t\}} := \{ R \in \mathcal{P}_{\{\varphi_5(\hat{x}_t)\}} : E_R[\varphi'_5(\hat{x}_t)z_t] = \tilde{E}_{\{\varphi_5(\hat{x}_t)\}}[\varphi'_5(\hat{x}_t)z_t] \} \subset \mathcal{P}$$

(cf. Definition (6.3) in Appendix 2). Here $[0, T]$ and $(\mathcal{P}, d)$ are endowed with their Borel $\sigma$-algebras. Recall that $d$ is the Lévy-Prokhorov metric on $\mathcal{P}$. From Theorem 6.10 (a measurable
selection theorem) we know that \( R_{\{\phi_5(\hat{x})|\phi_5'(\hat{x})z\}} \neq \emptyset \), and so \( R_{\{\phi_5(\hat{x})\}} \supset R_{\{\phi_5(\hat{x})|\phi_5'(\hat{x})z\}} \neq \emptyset \).

Moreover, we observe that, for all \( R = (R_t) \in R_{\{\phi_5(\hat{x})\}} \),

\[
\int_0^T E_{P_1}[\hat{y}(t)]E_{R_t}[\phi_5'(\hat{x}_t)z_t]dt \leq \int_0^T E_{P_1}[\hat{y}(t)] \sup_{P^1 \in P_{\{\phi_5'(\hat{x})\}}} E_{P^1}[\phi_5'(\hat{x}_t)z_t]dt, \tag{5.21}
\]

and, if \( R = (R_t) \in R_{\{\phi_5(\hat{x})|\phi_5'(\hat{x})z\}} \), we have equality in (5.21). Consequently,

\[
\sup_{R \in R_{\{\phi_5(\hat{x})\}}} \int_0^T E_{P_1}[\hat{y}(t)]E_{R_t}[\phi_5'(\hat{x}_t)z_t]dt = \int_0^T E_{P_1}[\hat{y}(t)] \sup_{P^1 \in P_{\{\phi_5'(\hat{x})\}}} E_{P^1}[\phi_5'(\hat{x}_t)z_t]dt,
\]

and since \( E_{P_1}[\hat{y}(T)] \geq 0 \) and \( E_{P_1}[\hat{y}(t)] \geq 0, \ t \in [0, T] \), using the notation

\[
\mathcal{P}\{\hat{u}\} := \mathcal{P}_{\{\psi(\hat{u})\}} \times \mathcal{P}_{\{\phi_4(\hat{x}_T)\}} \times R_{\{\phi_5(\hat{x})\}},
\]

(Observe that this set does not depend on the perturbing control \( u = (u_t) \)) we obtain from (5.20)

\[
0 \leq \lim_{\theta \downarrow 0} \frac{J(\hat{u} + \theta v) - J(\hat{u})}{\theta} = \sup_{(P,Q,R) \in \mathcal{P}(\{\hat{u}\})} \left\{ E_P[\hat{y}(T)z_T + \int_0^T \left( \hat{y}_x(t)z_t + \hat{y}_v(t)v_t \right) dt] + E_{P_1}[\hat{y}(T)]E_{Q}[\phi_4'(\hat{x}_T)z_T] + \int_0^T E_{P_1}[\hat{y}(t)]E_{R_t}[\phi_5'(\hat{x}_t)z_t]dt \right\}. \tag{5.22}
\]

As for the special case we consider here, (5.4) becomes

\[
\begin{cases}
\begin{align*}
dz_t &= \left( b_x(t)z_t + b_v(t)v_t \right) dt + \left( \sigma_x(t)z_t + \sigma_v(t)v_t \right) dB_t, \ t \in [0, T], \\
z_0 &= 0.
\end{align*}
\end{cases} \tag{5.23}
\]

Relation (5.22) brings us to introduce the following family of adjoint BSDEs (These BSDEs are classical ones, as they are considered under a linear expectation):

1) Under \( P \in \mathcal{P}_{\{\psi(\hat{u})\}},\)

\[
\begin{align*}
dp_s(P) &= -\left( \hat{b}_x(s)p_s(P) + \hat{t}_x(s) \right) ds - \hat{\sigma}_x(s)q_s(P)d\langle B \rangle_s + q_s(P)dB_s + dN_s(P), \\
p_T(P) &= \hat{\phi}_x(T), \quad s \in [0, T], \\
N(P) &\in \mathcal{M}^{2,1}_P(0, T) \text{ with } N_0(P) = 0;
\end{align*} \tag{5.24}
\]

2) Under \( Q \in \mathcal{P}_{\{\phi_4(\hat{x}_T)\}},\)

\[
\begin{align*}
d\tilde{p}_s(Q) &= -\tilde{b}_x(s)p_s(Q)ds - \tilde{\sigma}_x(s)\tilde{q}_s(Q)d\langle B \rangle_s + \tilde{q}_s(Q)dB_s + d\tilde{N}_s(Q), \\
\tilde{p}_T(Q) &= \phi_4'(\hat{x}_T), \quad s \in [0, T], \\
\tilde{N}(Q) &\in \mathcal{M}^{2,1}_Q(0, T) \text{ with } \tilde{N}_0(Q) = 0; \tag{5.25}
\end{align*}
\]

3) Under \( R_t, \ t \in [0, T], \) for \( R = (R_t) \in R_{\{\phi_5(\hat{x})\}},\)

\[
\begin{align*}
p_s(t, R_t) &= -\hat{b}_x(s)p_s(t, R_t)ds - \hat{\sigma}_x(s)q_s(t, R_t)d\langle B \rangle_s + q_s(t, R_t)dB_s + dN_s(t, R_t), \\
p_t(t, R_t) &= \phi_5'(\hat{x}_t), \quad s \in [0, t], \\
N(t, R_t) &\in \mathcal{M}^{2,1}_R(0, T) \text{ with } N_0(t, R_t) = 0. \tag{5.26}
\end{align*}
\]
Remark 5.4. 1) For the above BSDEs we consider the measurable space \((\Omega, \mathcal{B}(\Omega))\) endowed with the filtration \(\mathbb{F}^B = (\mathcal{F}_s)\) generated by the \(G\)-Brownian motion \(B\) (Recall that \(B\) has been introduced as coordinate process on \(\Omega\)). For a given probability measure \(P\) over \((\Omega, \mathcal{B}(\Omega))\) the associated filtration is the one augmented by all \(P\)-null sets: \(\mathbb{F}^P = \mathbb{F}^B \vee \mathcal{N}_P\).

2) Note that, under any \(P \in \mathcal{P}\), the \(G\)-Brownian motion \(B\) is only a continuous square integrable martingale, and so the martingale representation may not hold for \((B, \mathbb{F}^P)\). So it is necessary to introduce the second square integrable martingale \(\langle B, N(P) \rangle^P\) with \(N_0(P) = 0\) and joint quadratic variation \(\langle B, N(P) \rangle^P = (\langle B, N(P) \rangle^P)\) \(= 0\) (We write \(N(P) \in \mathcal{M}^2_{\perp}(0, T))\).

3) Recall that \(\langle B \rangle\) is the quadratic variation process of the \(G\)-Brownian motion \(B\) under \(\mathbb{E}\): For all \(\pi^N = \{0 = t_0^N < t_1^N < \cdots < t_N^N = t\},\ N \geq 1\), sequence of partitions of \([0, t]\) with mesh \(\max_{0 \leq j \leq N-1} (t_{j+1}^N - t_j^N) \to 0\ (N \to \infty)\),

\[
\mathbb{E}\left[ \sum_{j=0}^{N-1} (B_{t_{j+1}^N} - B_{t_j^N})^2 - \langle B \rangle_t \right] \to 0 \ (N \to \infty).
\]

And so, for all \(P \in \mathcal{P}\), \(\langle B \rangle\) coincides \(P\)-a.s. with the quadratic variation process \(\langle B \rangle^P\) of \(B\) as \(P\)-martingale, \(\langle B \rangle^P_t = \langle B \rangle_t, t \in [0, T]\, P\)-a.s. Also recall that, under the \(G\)-expectation the increments of \(\langle B \rangle\) are independent and stationary, and \(\sigma^2 ds \leq d\langle B \rangle_s \leq \bar{\sigma}^2 ds\), \(ds\)-a.e., quasi-surely.

Following El Karoui and Huang \[9\] and Buckdahn et al. \[4\], we see that, for all \(P \in \mathcal{P}\), there exists a unique triplet of processes \((p(P), q(P), N(P)) \in \mathcal{M}^2_{R}(0, T) \times \mathcal{M}^2_{R}(0, T) \times \mathcal{M}^2_{\perp}(0, T)\) which solves the adjoint equations (5.24) and (5.25) (equation (5.25) with \(Q\) instead of \(P\), respectively. The same we also have for the BSDE (5.26), only that here the BSDE is considered over the time interval \([0, t]\), so that the unique solution triplet \((p(t, R_t), q(t, R_t), N(t, R_t))\) belongs to \(\mathcal{M}^2_{R}(0, t) \times \mathcal{M}^2_{R}(0, t) \times \mathcal{M}^2_{\perp}(0, t)\), \(t \in [0, T]\). Moreover, standard BSDE estimates using that the coefficients \(\hat{b}_x, \hat{\sigma}_x, \hat{l}_x\) are bounded, show that, for all \(p \geq 1\), there is some constant \(C_p \in \mathbb{R}_+\) (independent of the underlying probability measure \(P \in \mathcal{P}\)) s.t.

\[
\mathbb{E}_P \left[ \sup_{s \in [0, T]} |p_s(P)|^p + \int_0^T |q_s(P)|^2 d\langle B \rangle_s + \langle N(P) \rangle_T^{1/2} \right] \leq C_p. \tag{5.27}
\]

Similar estimates we have for the solution \((p(t, R_t), q(t, R_t), N(t, R_t)) \in \mathcal{M}^2_{R}(0, t) \times \mathcal{M}^2_{R}(0, t) \times \mathcal{M}^2_{\perp}(0, t)\) of BSDE (5.26), for all \(t \in [0, T]\), only that unlike in (5.27), here \(T\) has to be replaced by \(t\). The constant \(C_p\) in the estimate of \((p(t, R_t), q(t, R_t), N(t, R_t))\) is again independent of \(R = (R_t) \in \mathcal{R}(\varphi_0(\bar{x}))\) but also independent of \(t \in [0, T]\).

Applying now Itô’s formula to \(p_s(P)z_s\), we have

\[
d(p_s(P)z_s) = (p_s(P)\hat{b}_x(s)v_s - \hat{l}_x(s)z_s)ds + \zeta_s(P)dB_s + q_s(P)\hat{\sigma}_x(s)v_sdB_s + z_sdN_s(P), \tag{5.28}
\]

where \(\zeta_s(P) := p_s(P)(\hat{\sigma}_x(s)z_s + \hat{\sigma}_x(s)v_s) + z_s q_s(P)\). As \(z_0 = 0\,

\[
pt(P)z_T = \int_0^T (p_s(P)\hat{b}_x(s)v_s - \hat{l}_x(s)z_s)ds + \int_0^T \zeta_s(P)dB_s + \int_0^T q_s(P)\hat{\sigma}_x(s)v_dsdB_s + \int_0^T z_sdN_s(P), \tag{5.29}
\]

\[
\text{20}
\]
where \( \int_0^T \zeta_s(P)dB_s \) and \( \int_0^T z_s dN_s(P) \) are \( P \)-martingales. Indeed, from our estimates it follows that
\[
E_P \left[ \left( \int_0^T |z_t|^2 d(N(P))_t \right)^{\frac{1}{2}} \right] \leq E_P \left[ \sup_{0 \leq t \leq T} |z_t| (N(P))^{\frac{1}{2}}_T \right]
\]
\[
\leq \left( E_P \left[ \sup_{0 \leq t \leq T} |z_t|^2 \right] \right)^{\frac{1}{2}} \left( E_P \left[ (N(P))_T \right] \right)^{\frac{1}{2}} < +\infty,
\]
and with similar arguments we also see that \( E_P \left[ \left( \int_0^T |\zeta_t(P)|^2 d(B)_t \right)^{\frac{1}{2}} \right] < +\infty. \)

Thus, recalling that \( p_T(P) = \Phi_x(T) \), we have
\[
E_P [\Phi_x(T)z_T + \int_0^T (\hat{\ell}_x(s)z_s + \hat{\ell}_v(s)v_s)ds] = E_P \left[ \int_0^T v_s \left( (p_s(P)\hat{b}_v(s) + \hat{\ell}_v(s))ds + q_s(P)\hat{\sigma}_v(s)d(B)_s \right) \right]. \tag{5.30}
\]

An analogous argument but with using now the solution of BSDE (5.25) yields, for \( Q \in \mathcal{P} , \)
\[
E_Q[\varphi'_4(\hat{x}_T)z_T] = E_Q[\hat{P}_T(Q)z_T] = E_Q \left[ \int_0^T v_s \left( p_s(Q)\hat{b}_v(s)ds + \hat{\ell}_v(s)\hat{\sigma}_v(s)d(B)_s \right) \right]. \tag{5.31}
\]

Finally, making use in the same way of the solution \( (p(t,R),q(t,R),N(t,R)) \in M_R^2(0,t) \times M_R^{2,\perp}(0,t) \) of BSDE (5.25), we obtain, for \( t \in [0,T], \)
\[
E_{R_t}[\varphi'_0(\hat{x}_T)z_t] = E_{R_t}[p_t(t,R_t)z_t] = E_{R_t} \left[ \int_0^t v_s \left( p_s(t,R_t)\hat{b}_v(s)ds + q_s(t,R_t)\hat{\sigma}_v(s)d(B)_s \right) \right]. \tag{5.32}
\]

Let us introduce now
\[
\Theta[P,Q,R](v) = E_P [\Phi_x(T)z_T + \int_0^T (\hat{\ell}_x(s)z_s + \hat{\ell}_v(s)v_s)ds] + E_P [\Phi_y(T)] E_Q[\varphi'_4(\hat{x}_T)z_T] + \int_0^T E_P [\hat{\ell}_y(t)] E_{R_t}[\varphi'_0(\hat{x}_T)z_t], \tag{5.33}
\]

and from the above computation we see that
\[
\Theta[P,Q,R](v) = E_P \left[ \int_0^T v_s \left( (p_s(P)\hat{b}_v(s) + \hat{\ell}_v(s))ds + q_s(P)\hat{\sigma}_v(s)d(B)_s \right) \right] + E_P [\Phi_y(T)] E_Q \left[ \int_0^T v_s \left( p_s(Q)\hat{b}_v(s)ds + \hat{\ell}_v(s)\hat{\sigma}_v(s)d(B)_s \right) \right] \tag{5.34}
\]
\[
+ \int_0^T E_P [\hat{\ell}_y(t)] E_{R_t} \left[ \int_0^t v_s \left( p_s(t,R_t)\hat{b}_v(s)ds + q_s(t,R_t)\hat{\sigma}_v(s)d(B)_s \right) \right] dt.
\]

In order to give to (5.33) another form, we make the convention that \( p_s(t,R_t) := 0, q_s(t,R_t) := 0, \) for \( t < s \leq T \), and we define the probability measure \( \tilde{R} := \int_0^T \frac{1}{T} dt \cdot (\delta_t \otimes R_t) \) over the probability space \( ([0,T] \times \Omega, \mathcal{B}([0,T]) \otimes \mathcal{F}) \). Here \( \delta_t \) denotes the Dirac measure over \([0,T]\) with mass at \( t \). Then, with \( (t,\omega) \mapsto (p_s(t,R_t)(\omega),q_s(t,R_t)(\omega)) \) and \( t \mapsto E_P[\hat{\ell}_y(t)] \) interpreted as random variables over \([0,T] \times \Omega, \), we have
\[
\int_0^T E_{\tilde{R}}[\hat{\ell}_y(t)] E_{R_t} \left[ \int_0^t v_s \left( p_s(t,R_t)\hat{b}_v(s)ds + q_s(t,R_t)\hat{\sigma}_v(s)d(B)_s \right) \right] dt = T E_{\tilde{R}} \left[ \int_0^T v_s \left( E_P[\hat{\ell}_y(t)] \left( p_s(\cdot,R_t)\hat{b}_v(s)ds + q_s(\cdot,R_t)\hat{\sigma}_v(s)d(B)_s \right) \right) \right]. \tag{5.35}
\]
Let us define $\Omega_{\{T\}} := [0, T] \times \Omega$ and embed the probabilities $P$ and $Q$ in the space of probabilities over $(\Omega_{\{T\}}, \mathcal{B}([0, T]) \otimes \mathcal{F})$ in a canonical way by making the identification $P := \delta_T \otimes P$ and $Q := \delta_T \otimes Q$. Then, thanks to (5.34) and (5.35),

$$
\Theta[P, Q, R|(v) = \int_{\Omega_{\{T\}}} \int_0^T v_s \left\{ \left( (p_s(P)\hat{b}_v(s) + \hat{\ell}_v(s)) ds + q_s(P)\hat{\sigma}_v(s)d(B)_s \right) dP + EP[\hat{\Phi}_y(T)]\left( \hat{\rho}_s(Q)\hat{b}_v(s) ds + \hat{q}_s(Q)\hat{\sigma}_v(s)d(B)_s \right) dQ + TEP[\hat{\ell}_y(\cdot)]\left( p_s(\cdot, R)\hat{b}_v(s) ds + q_s(\cdot, R, \hat{\sigma}_v(s)d(B)_s \right) dR \right\}
$$

where

$$dp_s(P, Q, R) = p_s(P)dp + EP[\hat{\Phi}_y(T)]\hat{\rho}_s(Q)dQ + TEP[\hat{\ell}_y(\cdot)]p_s(\cdot, R)dR,$$

$$dq_s(P, Q, R) = q_s(P)dP + EP[\hat{\Phi}_y(T)]\hat{q}_s(Q)dQ + TEP[\hat{\ell}_y(\cdot)]q_s(\cdot, R)dR.$$

We remark that $dsdp_s(P, Q, R)$ and $d(B)_s dq_s(P, Q, R)$ are signed measures on $\Omega_{\{T\}} \times [0, T]$ not depending on $v$ and so neither on the perturbing control $u$. Then, from (5.22), (5.34) and (5.36), and with the Hamiltonian measure

$$dH_v(s, P, Q, R) := \hat{b}_v(s)dsdp_s(P, Q, R) + \hat{\ell}_v(s)dsdP + \hat{\sigma}_v(s)d(B)_s dq_s(P, Q, R)$$

we have, for all $u \in U$ (Recalling that $v = u - \hat{u}$) that

$$0 \leq \sup_{(P, Q, R) \in \mathcal{P}(\hat{u})} \Theta[P, Q, R|(u - \hat{u})$$

and

$$= \sup_{(P, Q, R) \in \mathcal{P}(\hat{u})} \int_{\Omega_{\{T\}}} \int_0^T (u_s - \hat{u}_s)dH_v(s, P, Q, R).$$

(5.37)

Observe that (5.37) gives a necessary condition for the optimality of the control $\hat{u} \in U$. We resume our main result:

**Theorem 5.5.** Suppose (A.1)-(A.3) where $b, \sigma$ are independent of $y$, and let $\hat{u} \in U$ be an optimal control with state trajectory $\hat{x} = (\hat{x}_1)$. Then (5.37) gives a necessary optimality condition satisfied by all $u \in U$.

In the particular case when $\hat{\ell}_y(t) = 0$, quasi-surely, $dt$-a.s., and $\hat{\Phi}_y(T)$ is deterministic, by using an argument developed by Hu and Ji [11] based on Sion’s minimax theorem, we can simplify the necessary optimality condition (5.37). Indeed, let us suppose

(A.3') $l(t, x, y, u) = l(t, x, u)$, $\Phi(x, y) = \Phi_1(x) + \Phi_2(y)$, $(t, x, y, u) \in [0, T] \times \mathbb{R} \times \mathbb{R} \times U$.

We observe that under Assumption (A.3') $\hat{\ell}_y(t) = 0$ everywhere on $[0, T] \times \Omega$ and $\hat{\Phi}_y(T) = (\Phi_2)(\hat{\Phi}[\varphi_4(t)])$ is deterministic. Then (5.22) takes the simpler form

$$0 \leq \lim_{\theta \downarrow 0} \frac{J(\hat{u} + \theta v) - J(\hat{u})}{\theta}$$

$$= \sup_{(P, Q) \in \mathcal{P}(\hat{u}) \times \mathcal{P}(\varphi_4(\hat{x}_T))} \left\{ EP[\hat{\Phi}_x(T)z_T + \int_0^T \hat{\ell}_v(t)z_t + \hat{\ell}_v(t)v_t] dt \right\} + \hat{\Phi}_y(T)EQ[\varphi_4(\hat{x}_T)z_T].$$

(5.38)
We remark that the function \( F : (\mathcal{P}_{\psi}(\hat{u})) \times \mathcal{P}_{\phi^4(\hat{x}_T)}) \times \mathcal{U} \to \mathbb{R} \), defined by
\[
F((P,Q),u) = E_P\left[ \tilde{\Phi}_x(T)z_T + \int_0^T \left( \tilde{I}_x(t)z_t + \tilde{I}_v(t)(u_t - \hat{u}_t) \right) dt \right] + \tilde{\Phi}_y(T)E_Q[\phi^4(\hat{x}_T)z_T],
\]
\[
((P,Q),u) \in (\mathcal{P}_{\psi}(\hat{u})) \times \mathcal{P}_{\phi^4(\hat{x}_T)}) \times \mathcal{U},
\]
is affine in \((P,Q)\) over \(\mathcal{P}_{\psi(\hat{u})} \times \mathcal{P}_{\phi^4(\hat{x}_T)})\) and affine in \(u\) over \(\mathcal{U}\) (Recall SDE (5.23) for \(z^u = z\):
\[
F(\lambda(P,Q) + (1 - \lambda)(P', Q'),u) = \lambda F((P,Q),u) + (1 - \lambda)F((P',Q'),u),
\]
\[
F((P,Q),\lambda u + (1 - \lambda)u') = \lambda F((P,Q),u) + (1 - \lambda)F((P,Q),u'),
\]
\[
(P',Q'),(P,Q) \in \mathcal{P}_{\psi(\hat{u})} \times \mathcal{P}_{\phi^4(\hat{x}_T)}), u,u' \in \mathcal{U}, \lambda \in [0,1]. \tag{5.39}
\]
The fact that \(\mathcal{P}_{\psi(\hat{u})} \times \mathcal{P}_{\phi^4(\hat{x}_T)})\) is a non-void convex and weakly compact subset of a linear topological space (that of the pairs of bounded signed measures) and \(\mathcal{U}\) is a convex subset (Recall that the control state space \(U\) is convex) of a linear topological space, Sion’s minimax theorem applies,
\[
0 \leq \inf_{u \in \mathcal{U}} \sup_{(P,Q) \in \mathcal{P}_{\psi(\hat{u})} \times \mathcal{P}_{\phi^4(\hat{x}_T)})} \left\{ E_P\left[ \tilde{\Phi}_x(T)z_T + \int_0^T \left( \tilde{I}_x(t)z_t + \tilde{I}_v(t)(u_t - \hat{u}_t) \right) dt \right] + \tilde{\Phi}_y(T)E_Q[\phi^4(\hat{x}_T)z_T] \right\}
= \sup_{(P,Q) \in \mathcal{P}_{\psi(\hat{u})} \times \mathcal{P}_{\phi^4(\hat{x}_T)})} \inf_{u \in \mathcal{U}} \left\{ E_P\left[ \tilde{\Phi}_x(T)z_T + \int_0^T \left( \tilde{I}_x(t)z_t + \tilde{I}_v(t)(u_t - \hat{u}_t) \right) dt \right] + \tilde{\Phi}_y(T)E_Q[\phi^4(\hat{x}_T)z_T] \right\},
\]
where the supremum is taken over all \((P,Q) \in \mathcal{P}_{\psi(\hat{u})} \times \mathcal{P}_{\phi^4(\hat{x}_T)})\). By using the weak compactness of \(\mathcal{P}_{\psi(\hat{u})} \times \mathcal{P}_{\phi^4(\hat{x}_T)})\) a standard argument allows to show that there exists \((P^*, Q^*) \in \mathcal{P}_{\psi(\hat{u})} \times \mathcal{P}_{\phi^4(\hat{x}_T)})\) for which the latter supremum is attained (see also [11], proof of Theorem 4.6), i.e.,
\[
0 \leq \inf_{u \in \mathcal{U}} \left\{ E_{P^*}\left[ \tilde{\Phi}_x(T)z_T + \int_0^T \left( \tilde{I}_x(t)z_t + \tilde{I}_v(t)(u_t - \hat{u}_t) \right) dt \right] + \tilde{\Phi}_y(T)E_{Q^*}[\phi^4(\hat{x}_T)z_T] \right\}. \tag{5.40}
\]
This makes that we only have to use the adjoint BSDEs (5.24) and (5.25) under \(P^*\) and \(Q^*\), respectively, and the necessary optimality condition (5.37) takes the form
\[
0 \leq \int_\Omega \int_0^T (u_s - \hat{u}_s)\left\{ \tilde{b}_v(s)dsp_s(P^*,Q^*) + \tilde{I}_v(s)dsp^* + \tilde{\sigma}_v(s)d(B)_s dq_s(P^*,Q^*) \right\}, u \in \mathcal{U}, \tag{5.41}
\]
where
\[
dp_s(P^*,Q^*) = p_s(P^*)dP^* + \tilde{\Phi}_y(T)\tilde{p}_s(Q^*)dQ^*,
\]
\[
dq_s(P^*,Q^*) = q_s(P^*)dP^* + \tilde{\Phi}_y(T)\tilde{q}_s(Q^*)dQ^*.
\]
Finally, from the arbitrariness of \(u \in \mathcal{U}\) we obtain

**Theorem 5.6.** Suppose (A.1), (A.2) and (A.3’) where \(b\) are \(\sigma\) do not depend on \(y\), and let \(\hat{u}\) be an optimal control with the associated state trajectory \(\hat{x} = (\hat{x}_t)\). Then there exists \((P^*, Q^*) \in \mathcal{P}_{\psi(\hat{u})} \times \mathcal{P}_{\phi^4(\hat{x}_T)})\) such that, all \(u \in \mathcal{U}\),
\[
0 \leq (u_s - \hat{u}_s)\left\{ \tilde{b}_v(s)dsp_s(P^*,Q^*) + \tilde{I}_v(s)dsp^* + \tilde{\sigma}_v(s)d(B)_s dq_s(P^*,Q^*) \right\}. \tag{5.42}
\]
5.2 Sufficient conditions for optimality

In this section, we continue to consider the case discussed in Section 5.1.2. We define the Hamiltonian random field

\[ dH(t, x, u, p, q) := H_1(x, u, p)dt + H_2(x, u, q)d(B)_t, \]

with \( H_1(x, u, p) := b(x, u)p \) and \( H_2(x, u, q) := \sigma(x, u)q \), we make the following additional assumption:

(A.4) The function \( \Phi \) is convex in \((x, y)\); the running cost \( l(t, \ldots, \cdot) \) is convex, for all \( t \in [0, T] \); the functions \( \varphi_4 \) and \( \varphi_5 \) are convex; the Hamiltonian random field \( dH(t, x, u, p, q) \) is convex in \((x, u)\) (defined by the convexity of \( H_1(\cdot, \cdot, p) \) and that of \( H_2(\cdot, \cdot, q) \)).

**Theorem 5.7.** Assume the conditions (A.1)-(A.4) are satisfied and let \( \hat{u} \in U \) be a control process with associated state process \( \hat{x} = (\hat{x}_t) \), and let \( (p(P), q(P), N(P)) \), \( (\hat{p}(Q), \hat{q}(Q), \hat{N}(Q)) \) and \( (p(t, R_t), q(t, R_t), N(t, R_t)) \), \( t \in [0, T] \), \( (P, Q, R) \in \mathcal{P} \{\hat{u}\} \), be the solution of BSDE (5.24), (5.25) and (5.26), respectively. If (5.37) holds for all \( u \in U \), then \( \hat{u} \) is an optimal control.

**Proof.** Let \( u \in U \) be any admissible control. From (5.19), with

\[ \xi^u := \Phi(x_T^u, \hat{E}[\varphi_4(x_T^u)]) + \int_0^T l(t, x_t^u, \hat{E}[\varphi_5(x_t^u)], u_t)dt, \]
\[ \hat{\xi} := \Phi(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)]) + \int_0^T l(t, \hat{x}_t, \hat{E}[\varphi_5(\hat{x}_t)], \hat{u}_t)dt, \]

we have \( J(u) - J(\hat{u}) = \hat{E}[\xi] - \hat{E}[\hat{\xi}] \).

Since the function \( F(\lambda) := \hat{E}[\xi + \lambda(\xi^u - \hat{\xi})] \), \( \lambda \in [0, 1] \), is convex, \( F(1) - F(0) \geq F_+(0) \). Thus, from Lemma 4.4 we have

\[ \hat{E}[\xi^u] - \hat{E}[\hat{\xi}] \geq \lim_{\lambda \to 0} \frac{1}{\lambda} (\hat{E}[\xi + \lambda(\xi^u - \hat{\xi})] - \hat{E}[\hat{\xi}]) = \hat{E}_{\xi}(\xi^u - \hat{\xi}). \]  (5.43)

On the other hand, from the convexity of \( \Phi \), we get

\[ \Phi(x_T^u, \hat{E}[\varphi_4(x_T^u)]) - \Phi(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)]) \geq \Phi_x(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)])(x_T^u - \hat{x}_T) + \Phi_y(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)])(\hat{E}[\varphi_4(x_T^u)] - \hat{E}[\varphi_4(\hat{x}_T)]). \]

Using now the convexity of \( \varphi_4 \) as well as (5.43), but now with \( \xi^u = \varphi_4(x_T^u) \) and \( \hat{\xi} = \varphi_4(\hat{x}_T) \), we see that

\[ \hat{E}[\varphi_4(x_T^u)] - \hat{E}[\varphi_4(\hat{x}_T)] \geq \hat{E}_{\varphi_4}(\hat{x}_T)[\varphi_4(x_T^u) - \varphi_4(\hat{x}_T)] \]
\[ \geq \hat{E}_{\varphi_4}(\hat{x}_T)[\varphi_4'(\hat{x}_T)(x_T^u - \hat{x}_T)], \]

and from the non negativity of \( \Phi_y(T) \) we obtain

\[ \Phi(x_T^u, \hat{E}[\varphi_4(x_T^u)]) - \Phi(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)]) \geq \Phi_x(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)])(x_T^u - \hat{x}_T) + \Phi_y(\hat{x}_T, \hat{E}[\varphi_4(\hat{x}_T)])(\hat{E}[\varphi_4(x_T^u)] - \hat{E}[\varphi_4(\hat{x}_T)])[\varphi_4'(\hat{x}_T)(x_T^u - \hat{x}_T)] \]
\[ = \Phi_x(T)(x_T^u - \hat{x}_T) + \Phi_y(T)[\varphi_4'(\hat{x}_T)(x_T^u - \hat{x}_T)] \]
(Recall the notations introduced in (5.3)). Similarly, we see that, thanks to the convexity of \( l(t, x^u_t, \hat{E}[\varphi_5(x^u_t)], u_t) \) and \( \varphi_5 \) as well as the non negativity of \( \hat{I}_y(T) \),
\[
\begin{align*}
& l(t, x^u_t, \hat{E}[\varphi_5(x^u_t)], u_t) - l(t, \hat{x}_t, \hat{E}[\varphi_5(\hat{x}_t)], \hat{u}_t) \\
& \geq \hat{I}_x(t) (x^u_t - \hat{x}_t) + \hat{I}_y(t) \hat{E}[\varphi_5(\hat{x}_t)] \left[ \varphi'_5(\hat{x}_t) (x^u_t - \hat{x}_t) \right] + \hat{I}_v(t) (u_t - \hat{u}_t).
\end{align*}
\]
Hence, with the notation \( \mathcal{P}_{\{\hat{x}\}} = \mathcal{P}_{\{\varphi(\hat{u})\}} \) (\( \varphi(\hat{u}) \) has been introduced in Lemma 5.3), by summarising the above computations we obtain
\[
J(u) - J(\hat{u}) = \hat{E}[\xi] - \hat{E}[\xi] \geq \hat{E}[\xi] [\xi^u - \xi] \\
\geq \hat{E}[\varphi(\hat{u}) \left[ \hat{\Phi}_x(T) (x^u_T - \hat{x}_T) + \hat{\Phi}_y(T) \hat{E}[\varphi_4(\hat{x}_T)] \right] \left[ \varphi'_4(\hat{x}_T) (x^u_T - \hat{x}_T) \right] \\
+ \int_0^T (\hat{I}_x(t) (x^u_t - \hat{x}_t) + \hat{I}_y(t) \hat{E}[\varphi_5(\hat{x}_t)] \left[ \varphi'_5(\hat{x}_t) (x^u_t - \hat{x}_t) \right] + \hat{I}_v(t) (u_t - \hat{u}_t)) dt \right] \tag{5.44}
\]
Let us introduce now the following notations related with our Hamiltonian:
\[
\begin{align*}
& dH^{u,P}(s) := b^u(s)p_s(P) ds + \sigma^u(s)q_s(P) d\langle B \rangle_s, \\
& d\hat{H}^P(s) := \hat{b}(s)p_s(P) ds + \hat{\sigma}(s)q_s(P) d\langle B \rangle_s, \\
& d\hat{H}^P_x(s) := \hat{b}_x(s)p_s(P) ds + \hat{\sigma}_x(s)q_s(P) d\langle B \rangle_s, \\
& d\hat{H}^P_v(s) := \hat{b}_v(s)p_s(P) ds + \hat{\sigma}_v(s)q_s(P) d\langle B \rangle_s, \\
\end{align*}
\]
where \( (b^u, \sigma^u)(s) := (b, \sigma)(x^u_s, u_s) \) and \( l^u(s) := l(x^u_s, \hat{E}[\varphi_5(x^u_s)], u_s) \); for the other notations we refer to (5.3). Then, using BSDE (5.24) and applying the Itô formula to \( p_s(P)(x^u_s - \hat{x}_s) \), we obtain, for \( P \in \mathcal{P}_{\{\varphi(\hat{u})\}} \),
\[
\begin{align*}
& EP \left[ \hat{\Phi}_x(T) (x^u_T - \hat{x}_T) + \int_0^T \hat{I}_x(t) (x^u_t - \hat{x}_t) dt \right] \\
& = EP \left[ - \int_0^T (x^u_s - \hat{x}_s) (\hat{b}_x(s)p_s(P) ds + \hat{\sigma}_x(s)q_s(P) d\langle B \rangle_s) \\
& + \int_0^T \left\{ p_s(P)(b^u(s) - \hat{b}(s)) ds + q_s(P)(\sigma^u(s) - \hat{\sigma}(s)) d\langle B \rangle_s \right\} \right] \tag{5.46}
\end{align*}
\]
and from the convexity of \( H \) we conclude that
\[
\begin{align*}
& EP \left[ \hat{\Phi}_x(T) (x^u_T - \hat{x}_T) + \int_0^T \hat{I}_x(t) (x^u_t - \hat{x}_t) dt \right] \\
& = EP \left[ \int_0^T \left\{ d(H^{u,P}(s) - \hat{H}^P(s)) - (x^u_s - \hat{x}_s) d\hat{H}^P_x(s) \right\} \right] \\
& = EP \left[ \int_0^T \left\{ d(H^{u,P}(s) - \hat{H}^P(s)) - (x^u_s - \hat{x}_s) d\hat{H}^P_x(s) - (u_s - \hat{u}_s) d\hat{H}^P_v(s) \right\} \right] \\
& \geq EP \left[ \int_0^T (u_s - \hat{u}_s) d\hat{H}^P_v(s) \right]. \
\end{align*}
\]
Similarly we see that, for all \( Q \in \mathcal{P}_{\{\varphi(\hat{x}_T)\}} \),
\[
E_Q \left[ \varphi'_4(\hat{x}_T) (x^u_T - \hat{x}_T) \right] \geq E_Q \left[ \int_0^T (u_s - \hat{u}_s) d\hat{H}^Q_v(s) \right], \
\]
for $\hat{H}^Q_v$ defined like $\hat{H}^P_v$, but with $(\hat{p}_s(Q), \hat{q}_s(Q))$ instead of $(p_s(P), q_s(P))$. Similarly, for all $R = (R_t) \in \mathcal{R}_{(\varphi_s(\hat{x}))}$, we have, $dt$-a.e.,

$$E_{R_t}\left[\varphi'_s(\hat{x}_t)(x^n_t - \hat{x}_t)\right] \geq E_{R_t}\left[\int_0^t (u_s - \hat{u}_s)dh_{R_t}(s)\right],$$

(5.49)

where $\hat{H}^{R_t}(s), s \in [0, T]$, is defined by (5.45), but with the solution $(p_s(t, R_t), q_s(t, R_t))$ of BSDE (5.20) instead of that of BSDE (5.24). Consequently, from (5.44), (5.47), (5.48) and (5.49), for all $(P, Q, R) \in \mathcal{P}\{\hat{u}\}$,

$$J(u) - J(\hat{u}) \geq E_P\left[\Phi_x(T)(x^n_T - \hat{x}_T) + \int_0^T \left(\hat{l}_x(t)(x^n_t - \hat{x}_t) + \hat{l}_v(t)(u_t - \hat{u}_t)\right)dt\right]$$

$$+ E_P[\hat{\Phi}_y(T)]E_Q\left[\varphi'_s(\hat{x}_T)(x^n_T - \hat{x}_T)\right] + \int_0^T E_P[\hat{l}_y(t)]E_{R_t}\left[\varphi'_s(\hat{x}_t)(x^n_t - \hat{x}_t)\right]dt$$

$$\geq E_P\left[\int_0^T \{(u_s - \hat{u}_s)d\hat{H}^P_v(s) + \hat{l}_v(s)(u_s - \hat{u}_s)ds\}\right]$$

$$+ E_P[\hat{\Phi}_y(T)]E_Q\left[\int_0^T (u_s - \hat{u}_s)d\hat{H}^Q_v(s)\right] + \int_0^T E_P[\hat{l}_y(t)]E_{R_t}\left[\int_0^t (u_s - \hat{u}_s)d\hat{H}^{R_t}(s)\right]dt.$$  

(5.50)

Finally, recalling the notations introduced in Subsection 5.1.2, we see that the latter expression in (5.50) coincides with $\Theta[P, Q, R]$ (see (5.34)), i.e., because of the arbitrariness of $(P, Q, R) \in \mathcal{P}\{\hat{u}\}$ in (5.50) we conclude that

$$J(u) - J(\hat{u}) \geq \sup_{(P, Q, R) \in \mathcal{P}\{\hat{u}\}} \Theta[P, Q, R](u - \hat{u}) \geq 0,$$

where the latter inequality comes from the assumption of our statement. This proves the optimality of the control $\hat{u}$. \(\square\)

**Example 5.1.** We consider the following linear-quadratic control problem. The state equation is given by

$$\left\{\begin{array}{l}
  dx^n_t = (Ax^n_t + Bu_t)dt + (Cx^n_t + Du_t)dB_t, \\
  x(0) = x \in \mathbb{R},
\end{array}\right.$$

where $u \in \mathcal{U}$ and $A, B, C, D$ are constants. We associate the cost functional

$$J(u) = \frac{1}{2} E\left[\int_0^T ((x^n_t)^2 + u_t^2)dt + (x^n_T)^2 + E[|x^n_T|^2]\right], u \in \mathcal{U}.$$

The stochastic optimal control problem consists in minimizing the cost functional over $\mathcal{U}$.

We remark that the running cost and the terminal cost in the cost functional $J(u)$ do not satisfy (A.2), but one checks rather easily that our arguments apply also here, as $x^n \in L^p_G(0, T), x^n_T \in L^p_G(\Omega)$, for all $p \geq 1$. We see in particular that the adjoint BSDEs (5.24) and (5.25) take the form

$$dp_s(P) = (-Ap_s(P)s + \hat{x}_s)ds - Cq_s(P)d(B)_s + q_s(P)dB_s + dN_s(P), t \in [0, T], p_T(P) = \hat{x}_T,$$

$$d\tilde{p}_s(Q) = -A\tilde{p}_s(Q)sds - C\tilde{q}_s(Q)d(B)_s + \tilde{q}_s(Q)dB_s + d\tilde{N}_s(Q), t \in [0, T], \tilde{p}_T(Q) = \hat{x}_T,$$

(5.51)
respectively. We also remark that the solution \((p(t, R_t), q(t, R_t), N(t, R_t))\) is identically equal to zero, for all \(t \in [0, T]\), since the running cost \(l\) only depends on \((x^*, u)\). So, with the notation \(\psi(\hat{u}) := \frac{1}{2} \left( \int_0^T \left( (\hat{x}_t)^2 + \hat{u}_t^2 \right) dt + (\hat{x}_T)^2 + \bar{E}[(\hat{x}_T)^2] \right)\), Theorem \[5.6\] says that there exists \((P^*, Q^*) \in \mathcal{P}_{\psi(\hat{u})} \times \mathcal{P}_{(\hat{x}_T)^2}\) such that, for all \(u \in \mathcal{U}\),

\[
0 \leq (u_s - \hat{u}_s) \left( Bds dp_s(P^*, Q^*) + \hat{u}_s dsdp^* + Dd\langle B \rangle_s dq_s(P^*, Q^*) \right) \\
= (u_s - \hat{u}_s) \left( Bds(p_s(P^*)dP^* + \hat{p}_s(Q)dQ^*) + \hat{u}_s dsdP^* + Dd\langle B \rangle_s(q_s(P^*)dP^* + \hat{q}_s(Q^*)dQ^*) \right).
\]

(5.52)

On the other hand, we see that our example also satisfies the assumptions (A.3)-(A.4). Consequently, we have the following:

**Lemma 5.8.** For our linear-quadratic control problem of Example 5.1 the condition \[(5.52)\] is a necessary but also sufficient optimality condition for an admissible control \(\hat{u}\).

6 Appendix

6.1 Appendix 1. An extension of the result of Section 4

Let us consider a function \(f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}\) which is Lipschitz, i.e., there exist \(C > 0\) such that

\[
|f(\mu) - f(\mu')| \leq CW_2(\mu, \mu'), \quad \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d).
\]

(6.1)

We put \(F_f(\xi) := \sup_{P \in \mathcal{P}} f(P_\xi), \xi \in L^2_G(\Omega; \mathbb{R}^d)\).

**Remark 6.1.** From (6.1) one sees immediately that \(F_f : L^2_G(\Omega; \mathbb{R}^d) \to \mathbb{R}\) is Lipschitz.

Indeed, we have

\[
|F_f(\xi) - F_f(\eta)| \leq C \sup_{P \in \mathcal{P}} W_2(P_\xi, P_\eta) \leq C(\bar{E}[|\xi - \eta|^2])^{\frac{1}{2}}, \quad \xi, \eta \in L^2_G(\Omega; \mathbb{R}^d).
\]

**Lemma 6.2.** Let \(\xi \in L^2_G(\Omega; \mathbb{R}^d)\), with \(\bar{E}[|\xi|^2 I_{\{(|\xi| \geq N)\}}] \to 0\) \((N \to \infty)\). Then,

\[
\mathcal{P}^f_{\{\xi\}} := \{ P \in \mathcal{P} : f(P_\xi) = F_f(\xi) \} \neq \emptyset.
\]

**Proof.** As \(F_f(\xi) = \sup_{P \in \mathcal{P}} f(P_\xi) \leq f(\delta_0) + CW_2(\delta_0, P_\xi) \leq f(\delta_0) + C(\bar{E}[|\xi|^2])^{\frac{1}{2}} + \infty\), where \(\delta_0\) is the Dirac measure at \(0 \in \mathbb{R}^d\), there exists \((\mathcal{P}_l)_{l \geq 1} \subset \mathcal{P}\) such that \(f(P_\xi) \uparrow F_f(\xi)\), as \(l \to \infty\). But, since \(\mathcal{P}\) is weakly compact, we can extract a subsequence \((\mathcal{P}_l')_{l \geq 1} \subset (\mathcal{P}_l)_{l \geq 1}\), and find some \(P \in \mathcal{P}\) such that \(P_l' \to P\) \((\text{weak convergence})\), as \(l' \to \infty\). Then

\[
E_{P_l'}[\theta] \to E_P[\theta], \quad \text{as} \quad l' \to \infty, \quad \text{for all} \quad \theta \in L_{lip}(\Omega).
\]

(6.2)

Following the argument of the proof of Lemma \[4.1\] in Section 4, it follows that

\[
E_{P_l'}[\varphi(\xi)] \to E_P[\varphi(\xi)], \quad \text{for all} \quad \varphi \in C_0^1(\mathbb{R}^d).
\]

(6.3)
Indeed, \( \varphi(\xi) \in L_G^2(\Omega) \), and for any \( \delta > 0 \), there is \( \theta \in L_{ip}(\Omega) \) such that

\[
\sup_{Q \in \mathcal{P}} E_Q[|\varphi(\xi) - \theta|^2] \leq \delta,
\]

and so (6.3) follows from (6.2). But (6.3) means that \( P'^{\epsilon}_{\xi} \rightharpoonup P_{\xi} \), as \( l' \to \infty \). As, on the other hand,

\[
\sup_{l' \geq 1} E_{P'^{\epsilon}}[|\xi|^2 I_{\{\xi \geq N\}}] \leq \mathcal{E}[|\xi|^2 I_{\{\xi \geq N\}}] \to 0 \quad (N \to \infty),
\]

Hence, \( f(P_{\xi}) = \lim_{l' \to \infty} f(P'^{\epsilon}_{\xi}) = F_f(\xi) \), i.e., \( P \in \mathcal{P}^f_{\{\xi\}} \).

Similar to Section 4 we have

**Lemma 6.3.** Let \( \xi, \eta \in L^2_G(\Omega; \mathbb{R}^d) \), with \( \mathcal{E}[|\xi|^2 + |\eta|^2] I_{\{\xi + |\eta| \geq N\}} \to 0 \quad (N \to \infty) \). Then, for all \( 0 < \epsilon \downarrow 0 \), as \( l \to \infty \), and any \( P^l \in \mathcal{P}^f_{\{\xi + \epsilon \eta\}}, l \geq 1 \), we have:

i) There exists a subsequence \( (P'^{l'})_{l' \geq 1} \subseteq (P^l)_{l \geq 1} \) and \( P \in \mathcal{P} \) such that \( P'^{l'} \rightharpoonup P \), as \( l' \to \infty \).

ii) If, for some \( P \in \mathcal{P} \), \( P^l \rightharpoonup P \), as \( l \to \infty \), then \( P \in \mathcal{P}^f_{\{\xi\}} \).

**Proof.** For \( 0 < \epsilon \downarrow 0 \) (\( l \uparrow \infty \)), let \( P^l \in \mathcal{P}^f_{\{\xi + \epsilon \eta\}}, l \geq 1 \). From the weak compactness of \( \mathcal{P} \) it follows that there is a subsequence \( (P'^{l'})_{l' \geq 1} \subseteq (P^l)_{l \geq 1} \) and some \( P \in \mathcal{P} \) such that \( P'^{l'} \rightharpoonup P \), as \( l' \to \infty \), i.e., for all \( \theta \in L_{ip}(\Omega) \), \( E_{P'^{l'}}[\theta] \to E_P[\theta] \), as \( l' \to \infty \), and as \( (E_{P'^{l'}}[\cdot])_{l' \geq 1} \) is dominated by \( \mathcal{E}[\cdot] \), we have \( E_{P'^{l'}}[\xi] \to E_P[\xi] \), as \( l' \to \infty \), for all \( \xi \in L^2_G(\Omega) \).

Hence, for \( \zeta = f(\xi), \varphi \in C_b^1(\mathbb{R}^d) \),

\[
|E_{P'^{l'}}[\varphi(\xi + \epsilon l' \eta)] - E_P[\varphi(\xi)]| \leq \left| E_{P'^{l'}}[\varphi(\xi)] - E_P[\varphi(\xi)] \right| + C_{\zeta, \epsilon, l'} \mathcal{E}[|\eta|] \to 0, \quad l' \to \infty.
\]

This combined with

\[
\sup_{l' \geq 1} E_{P'^{l'}}[|\xi + \epsilon l' \eta|^2 I_{\{\xi + \epsilon l' \eta \geq N\}}] \to 0, \quad N \to \infty
\]

(Recall the assumption on \( \xi \) and on \( \eta \)) yields \( W_2(P'^{l'}_{\xi + \epsilon l' \eta}, P_{\xi}) \to 0 \) (\( l' \to \infty \)). Consequently,

\[
F_f(\xi + \epsilon l' \eta) = f(P'^{l'}_{\xi + \epsilon l' \eta}) \to f(P_{\xi}), \quad \text{as} \quad l' \to \infty,
\]

since \( P'^{l'} \in \mathcal{P}^f_{\{\xi + \epsilon l' \eta\}} \); while, on the other hand,

\[
|F_f(\xi + \epsilon l' \eta) - F_f(\xi)| \leq C(\mathcal{E}[|\epsilon l' \eta|^2])^{\frac{1}{2}} \to 0, \quad l' \to \infty.
\]

It follows that \( f(P_{\xi}) = F_f(\xi) \), i.e., \( P'^{l'} \rightharpoonup P \in \mathcal{P}^f_{\{\xi\}} \).

From Lemma 6.3 we get

**Proposition 6.4.** \( \Gamma(\mathcal{P}^f_{\{\xi + \epsilon \eta\}}, \mathcal{P}^f_{\{\xi\}}) \to 0, \quad \text{as} \quad 0 < \epsilon \downarrow 0 \).

The proof is analogous to that of Proposition 4.2, and so we omit it here.

Our objective is to study the (right- and left-) differentiability of \( \lambda \to F_f(\xi + \lambda \eta) = \sup_{P \in \mathcal{P}} f(P_{\xi + \lambda \eta}) \), for \( \xi, \eta \in L^2(\Omega; \mathbb{R}^d) \), with \( \mathcal{E}[|\xi|^2 + |\eta|^2 I_{\{\xi \geq N\}}] \to 0, \quad N \to \infty \).
For this we suppose that \( f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \) is differentiable in Lion’s sense with Lipschitz continuous derivative \( \partial_\mu f : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R} \). Recall (see [7]) that \( f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \) is differentiable, if there exists a continuous function \( \partial_m f : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R} \) with \( \partial_m f(\mu, \cdot) : \mathbb{R}^d \to \mathbb{R} \) differentiable, for all \( \mu \in \mathcal{P}_2(\mathbb{R}^d) \), such that

\[
\lim_{0 < \lambda \downarrow 0} \frac{f((1 - \lambda)\mu + \lambda\mu') - f(\mu)}{\lambda} = \int_{\mathbb{R}^d} \partial_m f(\mu, y)(\mu' - \mu)(dy), \quad \text{for all } \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d),
\]

and the derivative of \( f \) w.r.t. the measure \( \mu \) is defined by \( \partial_\mu f(\mu, y) := \partial_y(\partial_m f)(\mu, y) \), for all \( \mu \in \mathcal{P}_2(\mathbb{R}^d) \), \( y \in \mathbb{R}^d \).

Note, for all \( P \in \mathcal{P} \),

\[
f(P_{\xi + \varepsilon \eta}) = f(P_\xi) + \int_0^1 \partial_\lambda [f(P_{\xi + \lambda \varepsilon \eta})] d\lambda = f(P_\xi) + \varepsilon E_P[(\partial_\mu f)(P_\xi, \xi)\eta] + \varepsilon R_\varepsilon^P, \quad \varepsilon \geq 0, \tag{6.4}
\]

where

\[
R_\varepsilon^P = \int_0^1 E_P[((\partial_\mu f)(P_{\xi + \lambda \varepsilon \eta}, \xi + \lambda \varepsilon \eta) - (\partial_\mu f)(P_\xi, \xi))\eta] d\lambda,
\]

and

\[
|R_\varepsilon^P| \leq C\varepsilon \hat{E}[|\eta|^2], \quad \varepsilon \geq 0. \tag{6.5}
\]

Let us put

\[
G(\lambda) := \sup_{P \in \mathcal{P}} \{ f(P_\xi) + \lambda E_P[(\partial_\mu f)(P_\xi, \xi)\eta] \}, \quad \lambda \in \mathbb{R}.
\]

Then, \( G : \mathbb{R} \to \mathbb{R} \) is convex, and, so, in particular, there exists its right-derivative \( G_+(0) \) at \( \lambda = 0 \). On the other hand, from our above estimates it follows that

\[
|F_\varepsilon(\xi + \varepsilon \eta) - G(\varepsilon)| \leq C\varepsilon^2 \hat{E}[|\eta|^2], \quad \varepsilon \geq 0.
\]

Hence,

\[
\left| \frac{F_\varepsilon(\xi + \varepsilon \eta) - F_\varepsilon(\xi)}{\varepsilon} - G_+(0) \right| \leq \left| \frac{G(\varepsilon) - G(0)}{\varepsilon} - G_+(0) \right| + C\varepsilon \hat{E}[|\eta|^2] \to 0, \quad \text{as } 0 < \varepsilon \downarrow 0,
\]

i.e., the right-derivative of \( \varepsilon \to F_\varepsilon(\xi + \varepsilon \eta) \) at \( \varepsilon = 0 \) exists and

\[
\lim_{0 < \varepsilon \downarrow 0} \frac{F_\varepsilon(\xi + \varepsilon \eta) - F_\varepsilon(\xi)}{\varepsilon} = G_+(0).
\]

**Proposition 6.5.** Let \( f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \) be differentiable, with Lipschitz derivative \( \partial_\mu f : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R} \), and let \( \xi, \eta \in L^2_{\mathcal{E}}(\Omega; \mathbb{R}^d) \), with \( \hat{E}[\|\xi\|^2 + |\eta|^2]I_{\{|\xi| + |\eta| \geq N\}}] \to 0 \) \( (N \to \infty) \). Then,

i) \( \lim_{0 < \varepsilon \downarrow 0} \frac{F_\varepsilon(\xi + \varepsilon \eta) - F_\varepsilon(\xi)}{\varepsilon} = \sup_{P \in \mathcal{P}_2(\xi)} E_P[(\partial_\mu f)(P_\xi, \xi)\eta]; \)

ii) \( \lim_{0 > \varepsilon \uparrow 0} \frac{F_\varepsilon(\xi + \varepsilon \eta) - F_\varepsilon(\xi)}{\varepsilon} = - \sup_{P \in \mathcal{P}_2(\xi)} \left( - E_P[(\partial_\mu f)(P_\xi, \xi)\eta] \right). \)
Proof. We remark that ii) follows from i) by replacing in i) $\eta$ by $(-\eta)$. Let us prove i).

For this, using Lemma 5.3 let $P_{t}^{l} \in \mathcal{P}_{\{x+E_l^\xi\}}^l$, $l \geq 1$, and $P \in \mathcal{P}$, such that, for $0 < \varepsilon_l \downarrow 0$ ($l \to \infty$), $W_2(P_{t}^{l}+\varepsilon_{l}P_{t}^{l}, P_{t}) \to 0$. Then $P \in \mathcal{P}_{\{x\}^\xi}$. Thanks to (6.4) and (6.5)

$$\frac{F_{f}(\xi + \varepsilon \eta) - F_{f}(\xi)}{\varepsilon_{l}} \leq \frac{f(P_{t}^{l}+\varepsilon_{l}P_{t}^{l}) - f(P_{t}^{l})}{\varepsilon_{l}} = E_{P_{t}^{l}}[(\partial_{\mu} f)(P_{t}^{l}, \xi)\eta] + R_{P_{t}^{l}}^{\varepsilon_{l}}$$

Moreover, as $W_2(P_{t}^{l}+\varepsilon_{l}P_{t}^{l}, P_{t}^{l}) \leq E_{\varepsilon}((\hat{E}|\eta|^2)^{\frac{1}{2}} \to 0$ ($l \to \infty$), also $W_2(P_{t}^{l}, P_{t}) \to 0$, and so $E_{P_{t}^{l}}[(\partial_{\mu} f)(P_{t}^{l}, \xi)\eta] \to E_{P}[(\partial_{\mu} f)(P_{t}, \xi)\eta]$, $l \to \infty$. This shows that

$$\lim_{0 < \varepsilon_{l} \downarrow 0} \frac{F_{f}(\xi + \varepsilon \eta) - F_{f}(\xi)}{\varepsilon_{l}} \leq \sup_{P \in \mathcal{P}_{\{x\}^\xi}} E_{P}[(\partial_{\mu} f)(P_{t}, \xi)\eta].$$

On the other hand, for all $Q \in \mathcal{P}_{\{x\}^\xi}$,

$$\frac{F_{f}(\xi + \varepsilon \eta) - F_{f}(\xi)}{\varepsilon_{l}} \geq \frac{f(Q_{t}^{l}+\varepsilon_{l}Q_{t}^{l}) - f(Q_{t}^{l})}{\varepsilon_{l}} = E_{Q_{t}^{l}}[(\partial_{\mu} f)(Q_{t}^{l}, \xi)\eta], \quad l \to \infty.$$  

This proves i),

$$\lim_{0 < \varepsilon_{l} \downarrow 0} \frac{F_{f}(\xi + \varepsilon \eta) - F_{f}(\xi)}{\varepsilon_{l}} = \sup_{Q \in \mathcal{P}_{\{x\}^\xi}} E_{Q}[(\partial_{\mu} f)(Q_{t}, \xi)\eta].$$ (6.6) □

Remark 6.6. Let $\xi$, $\eta \in L^{2}_{G}(\Omega; \mathbb{R}^{d})$, such that $\hat{E}((|\xi|^{2} + |\eta|^{2})I_{(|\xi|+|\eta|\geq N}) \to 0$ ($N \to \infty$), $\varphi \in C_{b}^{1}(\mathbb{R}^{d})$, $f(P_{t}) := E_{P}[(\varphi(\theta))$, and $F_{f}(\varphi) := \sup_{P \in \mathcal{P}} E_{P}[(\varphi(\theta))]$.

Then, as $\partial_{\mu} f(\mu, y) = \nabla \varphi(y)$, $(\mu, y) \in \mathcal{P}_{2}(\mathbb{R}^{d}) \times \mathbb{R}^{d}$, (6.6) gives the result of Section 4, but only for $\xi$, $\eta \in L^{2}_{G}(\Omega; \mathbb{R}^{d})$ with $\hat{E}((|\xi|^{2} + |\eta|^{2})I_{(|\xi|+|\eta|\geq N}) \to 0$ ($N \to \infty$), while in Section 4 we have considered $\xi$, $\eta \in L^{2}_{G}(\Omega; \mathbb{R}^{d})$.

6.2 Appendix 2. A measurable selection theorem

Let $\xi = (\xi_{t})$ and $\eta = (\eta_{t})$ be in $M_{G}^{2}(0, T; \mathbb{R})$ such that the following assumptions are satisfied:

(B.1) $\hat{E}[|\xi_{t} - \xi_{s}|^{2} + |\eta_{t} - \eta_{s}|^{2}] \leq C|t - s|$, $t, s \in [0, T]$, for some constant $C \geq 0$.

Remark 6.7. Recall from Lemma 2.9 that, for $\phi, \psi : \mathbb{R} \to \mathbb{R}$ Lipschitz functions, the processes $\xi = (\xi_{t} = \phi(\hat{x}_{t}))$ and $\eta = (\eta_{t} = \psi(\hat{x}_{t}))$ satisfy assumption (B.1), where $\hat{x}$ is the solution of SDE (5.1) with $u = \hat{u}$ optimal control.

We also observe that, for all $\xi = (\xi_{t})$, $\eta = (\eta_{t}) \in M_{G}^{2}(0, T; \mathbb{R})$ satisfying (B.1), the function $t \mapsto \hat{E}(\xi_{t})|\eta_{t}|$ is Borel measurable. Indeed, from (B.1) it follows that, for all $\varepsilon > 0$, the function
theorem 6.8. Assume that $\xi = (\xi_t)$, $\eta = (\eta_t) \in M_G^2(0,T;\mathbb{R})$ satisfy (B.1). Then the mapping
\[ [0,T] \ni t \mapsto \mathcal{P}_{\{\xi_t|\eta_t\}} := \{ R \in \mathcal{P}_{\{\xi_t\}} : E_R[\eta_t] = \hat{E}_{\{\xi_t\}}[\eta_t] \} \subset \mathcal{P} \tag{6.8} \]
is a weakly measurable set-valued function with non empty values which are compact subsets of $(\mathcal{P},d)$ (Recall that $d$ is the Lévy-Prokhorov metric on $\mathcal{P}$).

Remark 6.9. Recall that, if $(X,\mathcal{G})$ is a measurable space and $Y$ a topological space, a set-valued function $G : X \ni x \mapsto G(x) \subset Y$ for which the values $G(x)$ are non empty, closed subsets of $Y$, is called weakly measurable if, for all open subset $O$ of $Y$, it holds \( \{ x \in X : G(x) \cap O \neq \emptyset \} \in \mathcal{G}. \)

**Theorem 6.10.** Assume that $\xi = (\xi_t)$, $\eta = (\eta_t) \in M_G^2(0,T;\mathbb{R})$ satisfy (B.1). Then the mapping $[0,T] \ni t \mapsto \mathcal{P}_{\{\xi_t|\eta_t\}} \subset \mathcal{P}$ admits a $\mathcal{B}([0,T]) - \mathcal{B}(\mathcal{P})$- measurable selection ($\mathcal{B}([0,T])$ and $\mathcal{B}(\mathcal{P})$ are the Borel $\sigma$-field over $[0,T]$ and $(\mathcal{P},d)$, respectively), i.e., there is a selection $R_t \in \mathcal{P}_{\{\xi_t|\eta_t\}}$, $t \in [0,T]$, such that the mapping $t \mapsto R_t$ is $\mathcal{B}([0,T]) - \mathcal{B}(\mathcal{P})$- measurable.

The proof of this theorem is an immediate consequence of Theorem 6.8 and the Kuratowski and Ryll-Nardzewski measurable selection theorem (cf. [10]). For the proof of Theorem 6.8 we need the following well-known auxiliary result:

**Lemma 6.11.** For a given measurable space $(X,\mathcal{G})$ and a separable metric space $(Y,d)$ a set-valued function $G : X \ni x \mapsto G(x) \subset Y$ with non empty, closed values is weakly measurable if and only if, for every $y$ from a dense subset of $Y$, the function $X \ni x \mapsto d(G(x),y)$ is $\mathcal{G} - \mathcal{B}(Y)$- measurable ($d(G(x),y)$ is the distance of $y$ to $G(x)$ in $(Y,d)$ and $\mathcal{B}(Y)$ is the Borel-$\sigma$-field on $Y$).

We are now able to give the proof of Theorem 6.8.

**Proof.** Let us begin with observing that $\mathcal{P}_{\{\xi_t|\eta_t\}}$ is non empty, for every $t \in [0,T]$. Indeed, $\mathcal{P}_{\{\xi_t\}} \neq \emptyset$ and $\hat{E}_{\{\xi_t\}}[\eta_t] = \max_{R \in \mathcal{P}_{\{\xi_t\}}} E_R[\eta_t]$ (see Lemma 4.4). On the other hand, by writing
\[ \mathcal{P}_{\{\xi_t|\eta_t\}} = \{ R \in \mathcal{P} : E_R[\xi_t] = \hat{E}_{\{\xi_t\}}[\xi_t], E_R[\eta_t] = \hat{E}_{\{\xi_t\}}[\eta_t] \} \]
we see easily that $\mathcal{P}_{\{\xi_t|\eta_t\}}$ is closed and, hence, also compact, as $\mathcal{P}$ is.

Now, for any sequence $0 < \varepsilon_k \searrow 0$ ($k \nearrow +\infty$), we consider
\[ \mathcal{P}_{\{\xi_t|\eta_t\}}^k := \{ R \in \mathcal{P} : E_R[\xi_t] > \hat{E}_{\{\xi_t\}}[\xi_t] - \varepsilon_k, E_R[\eta_t] > \hat{E}_{\{\xi_t\}}[\eta_t] - \varepsilon_k \} \subset \mathcal{P}_{\{\xi_t|\eta_t\}}, k \geq 1. \]
Because of the compactness of $(\mathcal{P},d)$ also the closure $\bar{\mathcal{P}}_{\{\xi_t|\eta_t\}}^k \subset (\mathcal{P},d)$ is compact, and
\[ \mathcal{P}_{\{\xi_t|\eta_t\}} = \bigcap_{k \geq 1} \bar{\mathcal{P}}_{\{\xi_t|\eta_t\}}^k. \]
As $(\mathcal{P},d)$ is compact, this space is in particular separable, i.e., we can choose a dense countable subset $\mathcal{D} \subset \mathcal{P}$. Let us put $\mathcal{D}_{\{\xi_t|\eta_t\}}^k = \mathcal{D} \cap \mathcal{P}_{\{\xi_t|\eta_t\}}^k$, $t \in [0,T]$, $k \geq 1$. As for all $R \in \mathcal{P}_{\{\xi_t|\eta_t\}}^k$
there exists $(R_\ell)_{\ell \geq 1} \subset \mathcal{D}$ s.t. $R_\ell \to R$ and, thus, also $E_{R_\ell}[\xi_\ell] \to E_R[\xi]$ and $E_{R_\ell}[\eta] \to E_R[\eta]$ ($\ell \to +\infty$), it follows that $\mathcal{D}^k_{\{\xi, | \eta\}} \subset \mathcal{P}^k_{\{\xi, | \eta\}}$ is dense.

For $k \geq 1$, put
\[
F_k(t, Q) := \text{dist}_{\mathcal{P}^k_{\{\xi, | \eta\}}}(Q) = \inf \{ d(Q, R) : R \in \mathcal{D}^k_{\{\xi, | \eta\}} \}, \quad (t, Q) \in [0, T] \times \mathcal{P}.
\]
Then,
\[
F_k(t, Q) = \inf \{ d(R, Q) : R \in \mathcal{D}^k_{\{\xi, | \eta\}} \} = \inf_{R \in \mathcal{D}} \chi(d(R, Q), \alpha_k(t, R), \beta_k(t, R)),
\]
where
\[
\begin{align*}
\alpha_k(t, R) & : = (E_R[\xi] - (\mathbb{E}[\xi] - \varepsilon_k))^-, \\
\beta_k(t, R) & : = (E_R[\eta] - (\mathbb{E}[\eta] - \varepsilon_k))^-, \\
\chi(\rho, \alpha, \beta) & : = \begin{cases} 
\rho, & (\alpha, \beta) = (0, 0) \\
+\infty, & (\alpha, \beta) \neq (0, 0).
\end{cases}
\end{align*}
\]
Observe that, thanks to assumption (B.1), the functions $t \mapsto \alpha_k(t, R), \beta_k(t, R)$ are continuous and, hence, Borel measurable, for all $R \in \mathcal{D}$, and so $[0, T] \times \mathcal{P} \ni (t, Q) \mapsto \chi(d(Q, R), \alpha_k(t, R), \beta_k(t, R))$ is Borel measurable (more precisely, $\mathcal{B}([0, T]) \otimes \mathcal{B}(\mathbb{P}) - \mathcal{B}(\mathbb{R})$-measurable), for all $R \in \mathcal{D}$. But as $\mathcal{D}$ is countable, also the infimum w.r.t. $R \in \mathcal{D}$ over these Borel functions is Borel measurable. Consequently, $F_k : [0, T] \times \mathcal{P} \ni (t, Q) \mapsto F_k(t, Q)$ is a Borel function, for all $k \geq 1$.

On the other hand, since $\mathcal{P}^{\{\xi, | \eta\}} = \bigcap_{k \geq 1} \downarrow \mathcal{P}^k_{\{\xi, | \eta\}}$, we have
\[
F(t, Q) := \text{dist}_{\mathcal{P}^{\{\xi, | \eta\}}}(Q) = \lim_{k \to +\infty} \text{dist}_{\mathcal{P}^k_{\{\xi, | \eta\}}}(Q) = \lim_{k \to +\infty} F_k(t, Q), \quad (t, Q) \in [0, T] \times \mathcal{P},
\]
and, hence, $F : [0, T] \times \mathcal{P} \to \mathbb{R}$ is Borel measurable. From Lemma 6.1 we get now the weak measurability of the set-valued function $t \mapsto \mathcal{P}^{\{\xi, | \eta\}}$. \hfill \Box

References

[1] B. Acciaio, J. Backhoff-Veraguas, R. Carmona. Extended mean field control problems: stochastic maximum principle and transport perspective, SIAM J. of Control and Optimization. 57 (6), 3666-3693, 2019.

[2] F. Bagiani, T. Meyeer-Brandis, B. Oksendal. Optimal control with delayed information flow of systems driven by G-Brownian motion. PUQR, 3(1), 2014.

[3] R. Buckdahn, B. Djehiche, J. Li. A general stochastic maximum principle for SDEs of mean-field type. Applied Mathematics and Optimization, 64, 197-216, 2011.

[4] R. Buckdahn, B. Labed, C. Rainer, L. Tamer. Existence of an optimal control for stochastic control systems with nonlinear cost functional, Stochastics. An International Journal of Probability and Stochastics Processes. 82, 241-256, 2010.

[5] R. Buckdahn, J. Li, J. Ma. A stochastic maximum principle for general mean-field systems. Applied Mathematics and Optimization. 74, 507-534, 2016.
[6] R. Buckdahn, J. Li, J. Ma. A mean-field stochastic control problem with partial observations. The Annals of Applied Probability. 27 (5), 3201-3245, 2017.

[7] P. Cardaliaguet, F. Delarue, J.M. Lasry, P.L. Lions. The Master Equation and the Convergence Problem in Mean Field Games. Princeton University Press, 2019.

[8] R.M. Dudley. Real Analysis and Probability. Cambridge Studies in Advanced Mathematics, Series Number 74, 2002.

[9] N. EL Karoui, S. Huang. A general result of existence and uniqueness of backward stochastic differential equations, El Karoui, N. and Mazliak, L. eds., Paris, 1995-1996, Pitman Research Notes in Mathematics Series. 364, 27-36, 1997.

[10] K. Kuratowski, C. Ryll-Nardzewski. A general theorem on selectors. Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys. 13, 397-403, 1965.

[11] M. Hu, S. Ji. Stochastic maximum principle for stochastic recursive optimal control problem under volatility ambiguity. SIAM J. of Control and Optimization. 54 (2), 918-945, 2016.

[12] M. Hu, S. Ji, S. Peng, Y. Song. Backward stochastic differential equations driven by $G$-Brownian motion. Stochastic Processes and their Applications. 124, 759-784, 2014.

[13] M. Hu, S. Ji, S. Peng, Y. Song. Comparison theorem, Feynman-Kac formula and girsanov transformation for BSDEs driven by $G$-Brownian motion. Stochastic Processes and their Applications. 124, 1170-1195, 2014.

[14] M. Hu, S. Peng. On representation theorem of G-expectations and paths of G-Brownian motion, Acta Mathematicae Applicatae Sinica-English Series. 25 (3), 539-546, 2009.

[15] M. Hu, F. Wang. Maximum principle for stochastic recursive optimal control problem under model uncertainty. SIAM J. of Control and Optimization. 58 (3), 1341-1370, 2020.

[16] J.M. Lasry, P.L. Lions. Mean field games. Japanese J. Math, 2 (1), 229-260, 2007.

[17] S. Peng. Nonlinear expectations and stochastic calculus under uncertainty, Probability Theory and Stochastic Modelling. Volume 95, Springer, Berlin, 2019.

[18] J. Li. Stochastic maximum principle in the mean-field controls. Automatica. 48, 366-373, 2012.

[19] S. Peng. A general stochastic maximum principle for optimal control problems. SIAM J. of Control and Optimization. 28, 966-979, 1990.

[20] S. Peng. $G$-expectation, $G$-Brownian motion and related stochastic calculus of Itô's type . ArXiv: 0601035, 2006.

[21] S. Peng. Multi-dimensional $G$-Brownian motion and related stochastic calculus under $G$-expectation. Stochastic Processes and their Applications. 118, 2223-2253, 2008
[22] S. Peng. Nonlinear expectations and stochastic calculus under uncertainty. ArXiv: 1002.4546, 2010.

[23] Z. Sun. Maximum principle for forward-backward stochastic control system under $G$-Brownian motion and relation to dynamic programming. J. of Comput. and Appl. Mathematics, 296, 753-775, 2016.