

# A unified approach to gradient type formulas for BSDEs and some applications

Xiliang Fan<sup>a),b)</sup>, Michael Röckner<sup>b),c)</sup>, Shao-Qin Zhang<sup>d)</sup>

<sup>a)</sup>School of Mathematics and Statistics, Anhui Normal University, Wuhu 241002, China

<sup>b)</sup>Fakultät für Mathematik, Universität Bielefeld, 33615 Bielefeld, Germany

<sup>c)</sup>Academy of Mathematics and Systems Science, CAS, Beijing 100190, China

<sup>d)</sup>School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, China  
fanxiliang0515@163.com(X. Fan), roeckner@math.uni-bielefeld.de(M. Röckner), zhangsq@cufe.edu.cn(S. Zhang)

March 10, 2021

## Abstract

In this paper we present a unified approach to establish gradient type formulas and Bismut type formulas for backward stochastic differential equations (BSDEs). This approach relies on a mix of derivative formulas with respect to the conditional probability of forward SDEs and the expression of the solution of BSDEs. Some concrete examples are given to illustrate the results. As applications, we provide representation formulas for the control solutions to McKean-Vlasov BSDEs and derive gradient estimates for related PDEs.

AMS Subject Classification (2010): 60H10; 60G22; 34F05

Keywords: Gradient type formula; Bismut type formula; BSDEs; McKean-Vlasov BSDEs; gradient estimate

## 1 Introduction

Let  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$  be a filtered probability space with  $(\mathcal{F}_t)_{t \in [0, T]}$  the natural completed and right continuous filtration generated by an  $m$ -dimensional Brownian motion  $(W_t)_{0 \leq t \leq T}$ . On  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$  we consider the following system of equations:

$$Y_t = g(X_T) + \int_t^T f(r, X_r, Y_r, Z_r) dr - \int_t^T Z_r dW_r, \quad t \in [0, T], \quad (1.1)$$

where  $X$  is the solution to a stochastic differential equation (SDE):

$$X_t = x + \int_0^t b(r, X_r) dr + \int_0^t \sigma(r, X_r) dW_r, \quad t \in [0, T]. \quad (1.2)$$

Here, the coefficients  $g : \mathbb{R}^d \rightarrow \mathbb{R}^l$ ,  $f : [0, T] \times \mathbb{R}^d \times \mathbb{R}^l \times \mathbb{R}^l \otimes \mathbb{R}^m \rightarrow \mathbb{R}^l$  and  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^m$  are deterministic functions. The system

(1.1)-(1.2) is called a (decoupled) forward-backward SDE (FBSDE), in which the processes  $X$  and  $Y$  are called the forward component and the backward component, respectively. The problem of existence and uniqueness for systems of this kind was first addressed by Pardoux and Peng [29, 30], and since then there are a large number of papers dedicated to the study of FBSDEs due to their increasing importance in stochastic control and mathematical finance (see, e.g., [19, 21, 22, 35, 37]).

In [30], Pardoux and Peng proved that the stochastic flow  $(X^{t,x}, Y^{t,x}, Z^{t,x}), t \in [0, T], x \in \mathbb{R}^d$ , which is the solution to the system (1.1)-(1.2) restricted to the interval  $[t, T]$  with  $X_t^{t,x} = x$ , provides a nonlinear Feynman-Kac formula for the solution to a quasi-linear parabolic partial differential equation (PDE) of the form:

$$\begin{cases} \partial_t u(t, x) + \mathbb{L}u(t, x) + f(t, x, u(t, x), (\nabla u \sigma)(t, x)) = 0, & (t, x) \in [0, T] \times \mathbb{R}^d, \\ u(T, x) = g(x), & x \in \mathbb{R}^d, \end{cases} \quad (1.3)$$

where  $\mathbb{L}u = (\mathcal{L}u_1, \dots, \mathcal{L}u_l)^*$  with  $\mathcal{L}u_i = \frac{1}{2} \text{Tr}(\sigma \sigma^* \nabla^2 u_i) + \langle b, \nabla u_i \rangle, i = 1, \dots, l$ , and the notation  $*$  denotes the transpose. More precisely, if the coefficients  $g, f$  and  $b, \sigma$  are sufficiently smooth, then

$$u(t, x) = Y_t^{t,x}, \quad (t, x) \in [0, T] \times \mathbb{R}^d \quad (1.4)$$

in the classical solution sense of PDE (1.3). Conversely, the solution  $u$  to PDE (1.3) admits the following representation formula for the solution to backward stochastic differential equation (BSDE) (1.1):

$$Y_s^{t,x} = u(s, X_s^{t,x}), \quad Z_s^{t,x} = \nabla u(s, X_s^{t,x}) \sigma(s, X_s^{t,x}), \quad s \in [t, T], \quad (1.5)$$

which is also valid for fully coupled FBSDEs, i.e.  $b$  and  $\sigma$  may depend on  $(Y, Z)$  (see [21] for further details). If the coefficients  $g, f$  and  $b, \sigma$  are Lipschitz continuous, the relations (1.4) and (1.5) between BSDE (1.1) and PDE (1.3) remain true in the viscosity solution sense of PDE (1.3) (see, e.g., [5, 28, 30]). We also point out that Crisan and Delarue [12], Zhu [38, 39] extended the above relations to the case of generalized solution of PDE (1.3).

When dealing with applications especially in the numerical analysis of BSDE, one needs to investigate regularity properties of the solution  $(Y, Z)$ . In [23], Ma and Zhang first studied the path regularity for the  $Z$  component when  $g, f$  are Lipschitz continuous and  $\sigma$  is uniformly non-degenerate. The key for their approach is to establish a probabilistic representation formula for  $\nabla u$  and then  $Z$  by using Malliavin's integration by parts formula (see Remark 2.2 and Remark 3.1 (i) below). That is, for any  $(t, x) \in [0, T] \times \mathbb{R}^d$ ,

$$\nabla u(t, x) = \mathbb{E} \left( g(X_T^{t,x}) N_T^t + \int_t^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) N_r^t dr \right), \quad (1.6)$$

where  $N_r^t = \frac{1}{r-t} \int_t^r \sigma^{-1}(s, X_s) \nabla X_s dW_s \cdot (\nabla X_t)^{-1}$ . Afterwards, Zhang [36] proved the path regularity of  $Z$  in the degenerate case ( $\sigma \neq 0$ ) where all processes concerned are one dimensional,  $f$  is linear on  $Z$  and  $g$  maybe discontinuous. The argument consists of using a representation formula for  $\nabla u$  similar to (1.6), in which  $N^t \nabla X_t$  is a bounded variation process rather than a martingale that makes the estimates of  $\nabla u$  more complicated. We remark that although the conditions imposed on  $\sigma$  in [36]

don't require the invertibility for  $\sigma\sigma^*$ , the invertibility of a linear functional of  $\sigma\sigma^*$  is required instead. As for another type of regularity, namely distributional regularity, we refer to, e.g., [1, 3, 15, 25, 26, 27] and the references therein.

The aim of this paper is to establish a gradient type formula for the solution  $Y$  of the system (1.1)-(1.2), which is stronger than (1.6). Our main result in that direction is Theorem 2.3 below, using a fundamentally different approach from [23, 36] and under more general assumptions. We first establish a gradient type formula in the sense of conditional expectation for Markov processes that is stronger than the Bismut type formula. To the best of our knowledge the result is new, and its proof relies heavily on the Markov property. Then, with the help of the relation (1.5) obtained under weaker conditions, we are able to provide a gradient type formula for the associated FBSDE in a general setting within which the formula can be applied to a larger class of examples. As a product of our main result, we establish a Bismut type formula for FBSDEs, which generalize and improve the corresponding ones in the existing literature (see Corollary 2.4 and Remark 2.2 below). In addition, we discuss a series of examples,

- FBSDEs with non-degenerate forward SDEs, for which two different types of gradient type formulas are given,
- FBSDEs with forward Gruschin type processes,
- FBSDEs with forward stochastic Hamiltonian systems.

These examples should illustrate the power and flexibility of our unified method. We believe that the method can also be used to handle other types of FBSDEs.

As an application of the main results, we wish to study McKean-Vlasov BSDEs and related nonlocal PDEs. McKean-Vlasov SDEs (whose coefficients depend on the law of the solution), initiated by Kac [18] and also known as mean-field equations, have been studied extensively in the past decade as they naturally arise in the context of statistical physics, and as they provide probabilistic representations for solutions of a class of nonlinear PDEs which may involve the Lions derivative introduced by Lions in his lectures [8]. Recently, Buckdahn, Djehiche, Li and Peng [6] first investigated nonlinear McKean-Vlasov BSDEs. Since then, existence and uniqueness results of McKean-Vlasov FBSDEs and the theory of the associated nonlocal PDEs of mean-field type have been studied in a variety of settings. For examples, Chassagneux, Crisan and Delarue [10] proved the existence and uniqueness of solutions to fully coupled McKean-Vlasov FBSDEs. Carmona and Delarue [9] investigated McKean-Vlasov FBSDEs through the stochastic maximum principle; Li [20] considered McKean-Vlasov FBSDEs driven by a Brownian motion and an independent Poisson random measure and showed the existence and uniqueness of solutions of this kind and provided a solution for the related nonlocal quasi-linear integral PDE of mean-field type via the value function. The second objective of the present paper is to deepen the investigation of McKean-Vlasov FBSDEs. With the main results above in hand, we shall establish representation formulas for the control solution  $Z_s^{t,\xi}$  of a McKean-Vlasov FBSDE which enable us to derive its path regularity, and obtain gradient estimates for the solution to the associated nonlocal PDE of mean-field type.

The remainder of this paper is organized as follows. In Section 2, we state and prove our main results concerning the gradient type formulas and Bismut type formulas of FBSDEs, which are then applied to concrete FBSDEs associated with various types of forward SDEs such as non-degenerate SDEs, Gruschin type processes and stochastic Hamiltonian systems. In Section 3, using these formulas, we establish

representation formulas for the control solutions of McKean-Vlasov FBSDEs, and provide gradient estimates for the corresponding PDEs. Section 4 will be devoted to the proofs of some auxiliary lemmas.

**Notation.** The following notations are used in the sequel.

- We use  $|\cdot|$  and  $\langle \cdot, \cdot \rangle$  for the Euclidean norm and the Euclidean inner product, respectively, and  $\|\cdot\|$  for either the operator norm or the Euclidean norm if this will not lead to confusion. For any  $p \in [1, \infty)$ , let  $\|\cdot\|_p$  denote the  $L^p(\mathbb{P}) := L^p(\Omega \rightarrow \mathbb{R}^d, \mathcal{F}, \mathbb{P})$  norm, and  $\mu$ -lim denotes the limit in measure  $\mu$ . For any  $0 \leq t \leq r \leq T$ ,  $\mathcal{F}_r^t$  denotes the completion of  $\sigma\{W_s - W_t : t \leq s \leq r\}$  with the  $\mathbb{P}$ -null sets of  $\mathcal{F}$ .

- $\mathcal{B}(\mathbb{R}^d; \mathbb{R}^l)$  denotes the space of all Borel measurable functions  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^l$  and  $\mathcal{B}_b(\mathbb{R}^d; \mathbb{R}^l)$  denotes the space of all those  $\varphi \in \mathcal{B}(\mathbb{R}^d; \mathbb{R}^l)$  that are bounded on  $\mathbb{R}^d$  with the norm  $\|\varphi\|_\infty := \sup_{x \in \mathbb{R}^d} |\varphi(x)|$ .  $C_b(\mathbb{R}^d; \mathbb{R}^l)$  is the set of all bounded continuous functions  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^l$ .  $C^1(\mathbb{R}^d; \mathbb{R}^l)$  denotes the collection of all continuously differentiable functions  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^l$  and  $C_b^1(\mathbb{R}^d; \mathbb{R}^l)$  denotes the collection of all those  $\varphi \in C^1(\mathbb{R}^d; \mathbb{R}^l)$  with bounded derivatives.  $C^{0,1}([0, T] \times \mathbb{R}^d; \mathbb{R}^l)$  stands for the class of all continuous functions  $\varphi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^l$  such that they are continuously differentiable on the space variable. For  $f \in C^1(\mathbb{R}^d; \mathbb{R}^l)$ , let  $\nabla f(x) = (\nabla f_1(x), \dots, \nabla f_l(x))^* \in \mathbb{R}^l \otimes \mathbb{R}^d$  denote the gradient matrix. When  $l = 1$ , we often suppress  $\mathbb{R}^l$  for simplicity.

- We let  $C, C_{k_1, k_2, T}, C_{T, q_1, q_2}$ , etc., denote generic constants, whose values may change from line to line.

## 2 Gradient type formulas for FBSDEs

The main objective of this section concerns the problem of gradient type formulas for FBSDEs. We first construct gradient type formulas for general Markov processes. In the second part of this section, we will show how to combine these results to establish gradient type formulas and Bismut type formulas for FBSDEs, and their proofs are addressed in the third part. Finally, we present some concrete examples to illustrate our main results.

### 2.1 The case of Markov processes

Let  $\{X_r^t\}_{0 \leq t \leq r \leq T}$  be a Markov process with respect to the filtration  $\{\mathcal{F}_r^t\}_{0 \leq t \leq r \leq T}$ . We also write  $X_r^{t,x}$  if  $X_t^t = x$ . Let  $P_{t,r}\psi(x) = \mathbb{E}\psi(X_r^{t,x})$ ,  $\psi \in \mathcal{B}(\mathbb{R}^d)$ , be the associated Markov semigroup. The current part is dedicated to gradient type formulas for  $X_r^{t,x}$  under the following assumptions on the  $\{X_r^{t,x}\}_{0 \leq t \leq r \leq T}$ :

(H1) There exists a  $\mathbb{R}^d \otimes \mathbb{R}^d$ -valued adapted process  $\{\nabla X_r^{t,x}\}_{0 \leq t < r \leq T}$  such that

$$\lim_{\epsilon \rightarrow 0^+} \mathbb{E} \left| \frac{X_r^{t,x+\epsilon v} - X_r^{t,x}}{\epsilon} - \nabla_v X_r^{t,x} \right|^2 = 0, \quad v \in \mathbb{R}^d, \quad (2.1)$$

where  $\nabla_v X_r^{t,x} = (\nabla X_r^{t,x})v$ .

(H2) For any  $0 \leq t < r \leq T$  and  $\psi \in C_b^1(\mathbb{R}^d)$ ,  $P_{t,r}\psi \in C^1(\mathbb{R}^d)$  and for any  $x \in \mathbb{R}^d$  there exists an integrable  $\mathbb{R}^d$ -valued,  $\mathcal{F}_r^t$ -measurable random variable  $M_r^t(x)$  such that

$$\nabla_v P_{t,r}\psi(x) = \mathbb{E}[\psi(X_r^{t,x}) \langle M_r^t(x), v \rangle], \quad \psi \in C_b^1(\mathbb{R}^d), \quad v \in \mathbb{R}^d. \quad (2.2)$$

(H3) For any  $0 \leq t \leq s < r \leq T$ ,

$$\lim_{y \rightarrow x} \mathbb{E} \left| M_r^s(X_s^{t,y}) - M_r^s(X_s^{t,x}) \right|^2 = 0, \quad (2.3)$$

$$\lim_{\delta \rightarrow 0^+} \mathbb{E} \left| M_r^s(X_s^{t,x}) - M_{r-\delta}^s(X_s^{t,x}) \right|^2 = 0. \quad (2.4)$$

With  $M_r^t(\cdot)$  in hand, we can provide a representation formula for  $\nabla_v \mathbb{E}[\phi(X_r^{t,x}) | \mathcal{F}_s^t]$  defined as follows:

$$\nabla_v \mathbb{E}[\phi(X_r^{t,x}) | \mathcal{F}_s^t] := \lim_{\epsilon \rightarrow 0^+} \frac{\mathbb{E}[\phi(X_r^{t,x+\epsilon v}) | \mathcal{F}_s^t] - \mathbb{E}[\phi(X_r^{t,x}) | \mathcal{F}_s^t]}{\epsilon},$$

if the limit exists in  $L^p(\mathbb{P})$  for some  $p \geq 1$  (see Proposition 2.1 below, which will play a crucial role in constructing gradient type formulas for FBSDEs in the next part).

**Proposition 2.1.** *Assume that (H1) and (H2) hold, and that for any  $0 \leq t \leq s < r \leq T$  there exist  $k_0 \geq 0$  and  $C(t, s, r) > 0$  such that*

$$K(t, s, r, x) := \|M_r^s(X_s^{t,x})\|_2 \leq C(t, s, r)(1 + |x|^{k_0}), \quad x \in \mathbb{R}^d, \quad (2.5)$$

and  $\mathbb{P}$ -a.s for any  $\psi \in \mathcal{B}_b(\mathbb{R}^d)$ ,

$$\mathbb{E} [\psi(X_r^{t,x}) G_r^s(X_s^{t,x}) | \mathcal{F}_s^t] = \mathbb{E} (\psi(X_r^{s,y}) G_r^s(y)) \Big|_{y=X_s^{t,x}}, \quad x, y \in \mathbb{R}^d, \quad (2.6)$$

where  $G = M, |M|$ . Assume moreover that for any  $k_1 \geq 1$ , there exist  $k_2 \geq 0$  and  $C_{k_1, k_2, T} > 0$  such that

$$\sup_{0 \leq s \leq r \leq T} \|X_r^{s,y}\|_{k_1} \leq C_{k_1, k_2, T}(1 + |y|^{k_2}), \quad y \in \mathbb{R}^d. \quad (2.7)$$

(1) For any  $\phi \in C(\mathbb{R}^d)$  with some  $q \geq 0$  such that

$$|\phi(x)| \leq C(1 + |x|^q), \quad x \in \mathbb{R}^d, \quad (2.8)$$

and for any  $0 \leq t \leq s < r \leq T$  and  $x, v \in \mathbb{R}^d$ ,  $\nabla_v \mathbb{E}[\phi(X_r^{t,x}) | \mathcal{F}_s^t]$  exists in  $L^p(\mathbb{P})$  for any  $p \in [1, 2)$  and moreover

$$\nabla_v \mathbb{E} [\phi(X_r^{t,x}) | \mathcal{F}_s^t] = \mathbb{E} [\phi(X_r^{t,x}) \langle M_r^s(X_s^{t,x}), \nabla_v X_s^{t,x} \rangle | \mathcal{F}_s^t]. \quad (2.9)$$

(2) If (H3) holds, and if furthermore  $C(t, s, \cdot)$  is bounded on any closed subinterval of  $(s, T]$  and

$$\lim_{y \rightarrow x} \mathbb{E} \|\nabla X_s^{t,y} - \nabla X_s^{t,x}\|^2 = 0, \quad (2.10)$$

then for any  $\phi \in \mathcal{B}(\mathbb{R}^d)$  satisfying (2.8), (2.9) also holds in  $L^p(\mathbb{P})$  for any  $p \in [1, 2)$ .

To prove this proposition, we first prove the following simple lemma.

**Lemma 2.2.** *Assume that  $\{X_r^t\}_{0 \leq t \leq r \leq T}$  satisfies (H2). Let  $\phi \in \mathcal{B}(\mathbb{R}^d)$  such that*

$$\mathbb{E} [|\phi(X_r^{t,x})| (|M_r^t(x)| + 1)] < \infty, \quad 0 \leq t < r \leq T, \quad x \in \mathbb{R}^d, \quad (2.11)$$

and suppose there exists a sequence  $\{\phi_n\}_{n=1}^\infty \subset C_b^1(\mathbb{R}^d)$  such that for any  $R > 0$

$$\lim_{n \rightarrow +\infty} \sup_{|x| \leq R} \mathbb{E} [ |(\phi - \phi_n)(X_r^{t,x})| (|M_r^t(x)| + 1) ] = 0. \quad (2.12)$$

Then  $P_{t,r}\phi \in C^1(\mathbb{R}^d)$  satisfying (2.2). If in addition, (2.5) and (2.7) hold and  $\phi$  satisfies (2.8), then there exist  $C > 0$  and  $q_0 > 0$  such that

$$|\nabla P_{t,r}\phi(y)| \leq C(1 + |y|^{q_0}), \quad y \in \mathbb{R}^d. \quad (2.13)$$

*Proof.* Since  $\phi_n \in C_b^1(\mathbb{R}^d)$ , (H2) yields that  $P_{t,r}\phi_n \in C^1(\mathbb{R}^d)$ . Moreover, for any  $R > 0$ , we have by (H2) and (2.12) that

$$\begin{aligned} & \lim_{n \rightarrow +\infty} \sup_{|x| \leq R} (|P_{t,r}\phi(x) - P_{t,r}\phi_n(x)| + |\mathbb{E}[\phi(X_r^{t,x})M_r^t(x)] - \nabla P_{t,r}\phi_n(x)|) \\ &= \lim_{n \rightarrow +\infty} \sup_{|x| \leq R} (|\mathbb{E}\phi(X_r^{t,x}) - \mathbb{E}\phi_n(X_r^{t,x})| + |\mathbb{E}[\phi(X_r^{t,x})M_r^t(x)] - \mathbb{E}[\phi_n(X_r^{t,x})M_r^t(x)]|) \\ &\leq \lim_{n \rightarrow +\infty} \sup_{|x| \leq R} \mathbb{E} [ |(\phi - \phi_n)(X_r^{t,x})| (1 + |M_r^t(x)|) ] \\ &= 0. \end{aligned} \quad (2.14)$$

Hence,  $P_{t,r}\phi \in C^1(\mathbb{R}^d)$  and it satisfies (2.2).

Furthermore, if (2.5) and (2.7) (with  $k_1 = 2q$ ) hold and  $\phi$  satisfies (2.8), then the Hölder inequality implies that

$$\begin{aligned} |\nabla_v P_{t,r}\phi(y)| &= |\mathbb{E}\langle \phi(X_r^{t,y})M_r^t(y), v \rangle| \\ &\leq |v| (\mathbb{E}|\phi(X_r^{t,y})|^2)^{\frac{1}{2}} (\mathbb{E}|M_r^t(y)|^2)^{\frac{1}{2}} \\ &\leq C|v| (\mathbb{E}(1 + |X_r^{t,y}|^q)^2)^{\frac{1}{2}} K(t, t, r, y) \\ &\leq C|v|(1 + |y|^{qk_2})(1 + |y|^{k_0}) \\ &\leq C(1 + |y|^{q_0})|v|, \end{aligned} \quad (2.15)$$

where  $q_0 = qk_2 + k_0$ .  $\square$

### Proof of Proposition 2.1

Let us first point out two facts for any  $\phi \in \mathcal{B}(\mathbb{R}^d)$  satisfying (2.8).

*Fact (i):* The relation (2.6) holds for all such  $\phi$   $\mathbb{P}$ -a.s. Indeed, similar to (2.15), by (2.5), (2.7) and (2.8) we first have

$$\mathbb{E}|\psi(X_r^{s,y})G_r^s(y)| \leq C(1 + |y|^{q_0}), \quad y \in \mathbb{R}^d \quad (2.16)$$

with  $q_0 = qk_2 + k_0$ , where  $G = M, |M|$ . Then by the monotone class theorem we derive the desired result.

*Fact (ii):* There exists a bounded measurable sequence  $\{\tilde{\phi}_n\}_{n=1}^\infty$  such that (2.12) holds. Indeed, setting  $\tilde{\phi}_n := (\phi \vee (-n)) \wedge n$ , we have

$$\begin{aligned} & \mathbb{E} \left[ |(\phi - \tilde{\phi}_n)(X_r^{s,y})| (|M_r^s(y)| + 1) \right] \\ &= \mathbb{E} \left[ |\phi(X_r^{s,y})| \mathbf{1}_{\{|\phi(X_r^{s,y})| \geq n\}} (|M_r^s(y)| + 1) \right] \\ &\leq \left[ \mathbb{E} (|\phi(X_r^{s,y})|^2 \mathbf{1}_{\{|\phi(X_r^{s,y})| \geq n\}}) \right]^{1/2} \left[ \mathbb{E} (|M_r^s(y)| + 1)^2 \right]^{\frac{1}{2}}. \end{aligned} \quad (2.17)$$

On one hand, for any  $\delta > 0$ , we deduce by using (2.7) (with  $k_1 = q(2 + \delta)$ ) and (2.8) that

$$\begin{aligned} \mathbb{E} \left( |\phi(X_r^{s,y})|^2 \mathbf{1}_{\|\phi(X_r^{s,y})\| \geq n} \right) &\leq \mathbb{E} \left[ |\phi(X_r^{s,y})|^2 \left( \frac{|\phi(X_r^{s,y})|}{n} \right)^\delta \mathbf{1}_{\|\phi(X_r^{s,y})\| \geq n} \right] \\ &\leq n^{-\delta} \mathbb{E} |\phi(X_r^{s,y})|^{2+\delta} \\ &\leq C n^{-\delta} \mathbb{E} \left( 1 + |X_r^{s,y}|^{q(2+\delta)} \right) \\ &\leq C n^{-\delta} \left( 1 + |y|^{q(2+\delta)k_2} \right). \end{aligned} \quad (2.18)$$

On the other hand, by the definition of the function  $K(t, s, r, x)$ , we get

$$\mathbb{E} |M_r^s(y)|^2 = \mathbb{E} |M_r^s(X_s^{s,y})|^2 = K^2(s, s, r, y), \quad (2.19)$$

which implies that  $\mathbb{E}(|M_r^s(y)| + 1)^2$  is locally bounded in  $y$ . Plugging (2.18) and (2.19) into (2.17), we obtain the second fact.

The rest of the proof is divided into three steps.

*Step 1: Claim: For  $\phi \in \mathcal{B}(\mathbb{R}^d)$  satisfying (2.8) and such that  $P_{s,r}\phi$  belongs to  $C^1(\mathbb{R}^d)$  and satisfies (2.2) and (2.13),  $\nabla_v \mathbb{E}[\phi(X_r^{t,x}) | \mathcal{F}_s^t]$  exists in  $L^p(\mathbb{P})$  for any  $p \in [1, 2)$  and moreover (2.9) holds. Since  $P_{s,r}\phi \in C^1(\mathbb{R}^d)$  and due to the Markov property of  $X$ , we first obtain for  $\epsilon > 0, v \in \mathbb{R}^d$  and  $0 \leq t \leq s < r \leq T$ ,*

$$\begin{aligned} &\frac{1}{\epsilon} \left( \mathbb{E}[\phi(X_r^{t,x+\epsilon v}) | \mathcal{F}_s^t] - \mathbb{E}[\phi(X_r^{t,x}) | \mathcal{F}_s^t] \right) - \nabla_{\nabla_v X_s^{t,x}} P_{s,r}\phi(X_s^{t,x}) \\ &= \frac{1}{\epsilon} \left( P_{s,r}\phi(X_s^{t,x+\epsilon v}) - P_{s,r}\phi(X_s^{t,x}) \right) - \nabla_{\nabla_v X_s^{t,x}} P_{s,r}\phi(X_s^{t,x}) \\ &= \int_0^1 \nabla_{\frac{X_s^{t,x+\epsilon v} - X_s^{t,x}}{\epsilon}} P_{s,r}\phi(X_s^{t,x} + \theta(X_s^{t,x+\epsilon v} - X_s^{t,x})) d\theta - \nabla_{\nabla_v X_s^{t,x}} P_{s,r}\phi(X_s^{t,x}) \\ &= \left\langle \int_0^1 \nabla P_{s,r}\phi(X_s^{t,x} + \theta(X_s^{t,x+\epsilon v} - X_s^{t,x})) d\theta - \nabla P_{s,r}\phi(X_s^{t,x}), \nabla_v X_s^{t,x} \right\rangle \\ &\quad + \left\langle \int_0^1 \nabla P_{s,r}\phi(X_s^{t,x} + \theta(X_s^{t,x+\epsilon v} - X_s^{t,x})) d\theta, \frac{X_s^{t,x+\epsilon v} - X_s^{t,x}}{\epsilon} - \nabla_v X_s^{t,x} \right\rangle. \end{aligned}$$

Then in order to prove that for any  $p \in [1, 2)$ ,

$$\lim_{\epsilon \rightarrow 0^+} \mathbb{E} \left| \frac{1}{\epsilon} \left( \mathbb{E}[\phi(X_r^{t,x+\epsilon v}) | \mathcal{F}_s^t] - \mathbb{E}[\phi(X_r^{t,x}) | \mathcal{F}_s^t] \right) - \nabla_{\nabla_v X_s^{t,x}} P_{s,r}\phi(X_s^{t,x}) \right|^p = 0, \quad (2.20)$$

according to the Hölder inequality, (H1), (2.13) and (2.7), it is enough to show that,

$$\lim_{\epsilon \rightarrow 0^+} \int_0^1 \nabla P_{s,r}\phi(X_s^{t,x} + \theta(X_s^{t,x+\epsilon v} - X_s^{t,x})) d\theta = \nabla P_{s,r}\phi(X_s^{t,x}) \quad (2.21)$$

in  $L^{\frac{2p}{2-p}}(\mathbb{P})$  for any  $p \in [1, 2)$ .

Noting that  $\nabla P_{s,r}\phi \in C(\mathbb{R}^d)$ , and that

$$\mathbb{P}\text{-}\lim_{\epsilon \rightarrow 0^+} X_s^{t,x+\epsilon v} = X_s^{t,x},$$

we deduce that

$$\mathbb{P}\text{-}\lim_{\epsilon \rightarrow 0^+} \nabla P_{s,r} \phi(X_s^{t,x} + \theta(X_s^{t,x+\epsilon v} - X_s^{t,x})) = \nabla P_{s,r} \phi(X_s^{t,x}), \quad \theta \in [0, 1]. \quad (2.22)$$

For any  $\tilde{p} \geq 1$ , it follows by (2.13) and (2.7) that

$$\begin{aligned} & \sup_{\epsilon \in (0,1], \theta \in [0,1]} \mathbb{E} |\nabla P_{s,r} \phi(X_s^{t,x} + \theta(X_s^{t,x+\epsilon v} - X_s^{t,x}))|^{\tilde{p}} \\ & \leq C \sup_{\epsilon \in (0,1]} \mathbb{E} (1 + (|X_s^{t,x}| + |X_s^{t,x+\epsilon v}|)^{q_0})^{\tilde{p}} \\ & < \infty. \end{aligned}$$

Then we conclude that

$$\left\{ |\nabla P_{s,r} \phi(X_s^{t,x} + \theta(X_s^{t,x+\epsilon v} - X_s^{t,x}))|^{\frac{2p}{2-p}} \right\}_{\epsilon \in (0,1]}$$

is uniformly integrable with respect to  $d\theta \otimes d\mathbb{P}$ . Combining this with (2.22), by the dominated convergence theorem we obtain

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0^+} \mathbb{E} \left| \int_0^1 (\nabla P_{s,r} \phi(X_s^{t,x} + \theta(X_s^{t,x+\epsilon v} - X_s^{t,x})) - \nabla P_{s,r} \phi(X_s^{t,x})) d\theta \right|^{\frac{2p}{2-p}} \\ & \leq \lim_{\epsilon \rightarrow 0^+} \int_0^1 \mathbb{E} |\nabla P_{s,r} \phi(X_s^{t,x} + \theta(X_s^{t,x+\epsilon v} - X_s^{t,x})) - \nabla P_{s,r} \phi(X_s^{t,x})|^{\frac{2p}{2-p}} d\theta \\ & = \int_0^1 \mathbb{E} \left( \lim_{\epsilon \rightarrow 0^+} |\nabla P_{s,r} \phi(X_s^{t,x} + \theta(X_s^{t,x+\epsilon v} - X_s^{t,x})) - \nabla P_{s,r} \phi(X_s^{t,x})|^{\frac{2p}{2-p}} \right) d\theta \\ & = 0, \end{aligned}$$

which means that (2.21) holds in  $L^{\frac{2p}{2-p}}(\mathbb{P})$ .

Hence, we obtain that  $\nabla_v \mathbb{E}[\phi(X_r^{t,x}) | \mathcal{F}_s^t]$  exists in  $L^p(\mathbb{P})$  for any  $p \in [1, 2)$ , and moreover by (2.20) and (2.2) the directional derivative satisfies

$$\begin{aligned} \nabla_v \mathbb{E} \left[ \phi(X_r^{t,x}) \middle| \mathcal{F}_s^t \right] &= \nabla_{\nabla_v X_s^{t,x}} P_{s,r} \phi(X_s^{t,x}) \\ &= \mathbb{E} [\phi(X_r^{s,y}) \langle M_r^s(y), w \rangle] \Big|_{y=X_s^{t,x}, w=\nabla_v X_s^{t,x}} \\ &= \left\langle \mathbb{E} \left[ \phi(X_r^{t,x}) M_r^s(X_s^{t,x}) \middle| \mathcal{F}_s^t \right], \nabla_v X_s^{t,x} \right\rangle, \end{aligned} \quad (2.23)$$

where we have used fact (i) in the last equality.

Note that again due to Fact (i), (2.16) and (2.7), we obtain by the Hölder inequality that

$$\begin{aligned} \mathbb{E} |\phi(X_r^{t,x}) \langle M_r^s(X_s^{t,x}), \nabla_v X_s^{t,x} \rangle| &\leq \mathbb{E} [|\phi(X_r^{t,x}) M_r^s(X_s^{t,x})| \cdot |\nabla_v X_s^{t,x}|] \\ &= \mathbb{E} \left\{ \mathbb{E} \left[ |\phi(X_r^{t,x}) M_r^s(X_s^{t,x})| \middle| \mathcal{F}_s^t \right] \cdot |\nabla_v X_s^{t,x}| \right\} \\ &= \mathbb{E} \left\{ \mathbb{E} [|\phi(X_r^{s,y}) M_r^s(y)|] \Big|_{y=X_s^{t,x}} \cdot |\nabla_v X_s^{t,x}| \right\} \\ &\leq C \left( \mathbb{E} (1 + |X_s^{t,x}|^{q_0})^2 \right)^{\frac{1}{2}} \left( \mathbb{E} |\nabla_v X_s^{t,x}|^2 \right)^{\frac{1}{2}} \end{aligned}$$



$< \infty$ .

Then we have

$$\left\langle \mathbb{E} \left[ \phi(X_r^{t,x}) M_r^s(X_s^{t,x}) \middle| \mathcal{F}_s^t \right], \nabla_v X_s^{t,x} \right\rangle = \mathbb{E} \left[ \phi(X_r^{t,x}) \left\langle M_r^s(X_s^{t,x}), \nabla_v X_s^{t,x} \right\rangle \middle| \mathcal{F}_s^t \right].$$

This, together with (2.23), yields (2.9).

*Step 2: Proof of assertion (1).* According to Step 1, it suffices to prove that for  $\phi \in C(\mathbb{R}^d)$  satisfying (2.8),  $P_{s,r}\phi \in C^1(\mathbb{R}^d)$  and that it satisfies (2.2) and (2.13). To this end, we will invoke Lemma 2.2. Since  $\phi \in C(\mathbb{R}^d)$ ,  $\tilde{\phi}_n$  defined as in Fact (ii) above belongs to  $C_b(\mathbb{R}^d)$ . Then for each  $n \geq 1$ , there exists  $\{\tilde{\phi}_{n,m}\}_{m=1}^\infty \subset C_b^1(\mathbb{R}^d)$  such that  $\|\tilde{\phi}_{n,m}\|_\infty \leq \|\tilde{\phi}_n\|_\infty$  and for any  $N > 0$

$$\lim_{m \rightarrow +\infty} \sup_{|x| \leq N} |\tilde{\phi}_{n,m}(x) - \tilde{\phi}_n(x)| = 0. \quad (2.24)$$

Thus, a simple application of the Hölder inequality yields that for any  $R > 0$ ,

$$\begin{aligned} & \sup_{|y| \leq R} \mathbb{E} \left[ |(\tilde{\phi}_{n,m} - \tilde{\phi}_n)(X_r^{s,y})| (|M_r^s(y)| + 1) \right] \\ & \leq 2 \|\tilde{\phi}_n\|_\infty \sup_{|y| \leq R} \mathbb{E} \left[ (|M_r^s(y)| + 1) \mathbb{1}_{\{|X_r^{s,y}| \geq N\}} \right] \\ & \quad + \left( \sup_{|x| \leq N} |\tilde{\phi}_{n,m}(x) - \tilde{\phi}_n(x)| \right) \sup_{|y| \leq R} \mathbb{E} (|M_r^s(y)| + 1) \\ & \leq \frac{2}{N} \|\tilde{\phi}_n\|_\infty \sup_{|y| \leq R} \left[ (1 + K(s, s, r, y)) (\mathbb{E}|X_r^{s,y}|^2)^{\frac{1}{2}} \right] \\ & \quad + \left( \sup_{|x| \leq N} |\tilde{\phi}_{n,m}(x) - \tilde{\phi}_n(x)| \right) \sup_{|y| \leq R} (1 + K(s, s, r, y)), \end{aligned}$$

which, along with (2.24), implies

$$\lim_{m \rightarrow +\infty} \sup_{|y| \leq R} \mathbb{E} \left[ |(\tilde{\phi}_{n,m} - \tilde{\phi}_n)(X_r^{s,y})| (|M_r^s(y)| + 1) \right] = 0.$$

Then, combining this with Fact (ii), we derive that there exists a sequence  $\{\phi_n\}_{n=1}^\infty \subset C_b^1(\mathbb{R}^d)$  such that (2.12) holds for  $\phi$ . Thanks to (2.16), (2.8) and (2.7), it is readily checked that (2.11) also holds for  $\phi$ . Therefore, applying Lemma 2.2, one sees that  $P_{s,r}\phi$  belongs to  $C^1(\mathbb{R}^d)$  and satisfies (2.2) and (2.13), from which the assertion (1) follows.

*Step 3: Proof of assertion (2).* Observe that assertion (1) holds for any  $\phi \in C_b(\mathbb{R}^d)$  thanks to  $C_b(\mathbb{R}^d) \subset C(\mathbb{R}^d)$  and (2.8) with  $q = 0$ . Then for any  $\phi \in C_b(\mathbb{R}^d)$ ,  $\epsilon > 0$ ,  $v \in \mathbb{R}^d$  and  $0 \leq t \leq s < r \leq T$ , we have

$$\begin{aligned} & \mathbb{E} \left[ \phi(X_r^{t,x+\epsilon v}) \middle| \mathcal{F}_s^t \right] - \mathbb{E} \left[ \phi(X_r^{t,x}) \middle| \mathcal{F}_s^t \right] \\ & = \int_0^\epsilon \nabla_v \mathbb{E} \left[ \phi(X_r^{t,x+\theta v}) \middle| \mathcal{F}_s^t \right] d\theta \\ & = \int_0^\epsilon \mathbb{E} \left[ \left\langle \phi(X_r^{t,x+\theta v}) M_r^s(X_s^{t,x+\theta v}), \nabla_v X_s^{t,x+\theta v} \right\rangle \middle| \mathcal{F}_s^t \right] d\theta. \quad (2.25) \end{aligned}$$

Let  $\mathbb{P}_{X_r^{t,x+\epsilon v}}$  and  $\mathbb{P}_{X_r^{t,x}}$  denote the law of  $X_r^{t,x+\epsilon v}$  and  $X_r^{t,x}$ , respectively, and let

$$\mu_{\epsilon,r}^{t,s}(A) = \int_0^\epsilon \mathbb{E} \left( \mathbb{1}_A(X_r^{t,x+\theta v}) |M_r^s(X_s^{t,x+\theta v})| \cdot |\nabla_v X_s^{t,x+\theta v}| \right) d\theta, \quad A \in \mathcal{B}(\mathbb{R}^d).$$

From (2.5) and (2.10), it is easy to see that  $\mu_{\epsilon,r}^{t,s}$  is a finite measure on  $\mathbb{R}^d$ . Then  $C_b(\mathbb{R}^d)$  is dense in  $L^1(\mathbb{P}_{X_r^{t,x+\epsilon v}} + \mathbb{P}_{X_r^{t,x}} + \mu_{\epsilon,r}^{t,s})$ . Hence, (2.25) holds for any  $\phi \in \mathcal{B}_b(\mathbb{R}^d)$ .

Next, we intend to prove that for any  $\phi \in \mathcal{B}_b(\mathbb{R}^d)$ ,

$$\mathbb{E} \left[ \left\langle \phi(X_r^{t,x}) M_r^s(X_s^{t,x}), \nabla_v X_s^{t,x} \right\rangle \middle| \mathcal{F}_s^t \right], \quad 0 \leq t \leq s < r \leq T,$$

is continuous with respect to  $x$  in  $L^p(\mathbb{P})$  for any  $p \in [1, 2)$ . Then it follows from (2.25) that (2.9) holds in  $L^p(\mathbb{P})$  for  $p \in [1, 2)$  for any  $\phi \in \mathcal{B}_b(\mathbb{R}^d)$ .

For any  $\phi \in \mathcal{B}_b(\mathbb{R}^d)$ ,  $x, y \in \mathbb{R}^d$  and  $0 < \delta < r - s$ , we obtain

$$\begin{aligned} & \mathbb{E} \left[ \left\langle \phi(X_r^{t,y}) M_r^s(X_s^{t,y}), \nabla_v X_s^{t,y} \right\rangle \middle| \mathcal{F}_s^t \right] - \mathbb{E} \left[ \left\langle \phi(X_r^{t,x}) M_r^s(X_s^{t,x}), \nabla_v X_s^{t,x} \right\rangle \middle| \mathcal{F}_s^t \right] \\ &= \mathbb{E} \left[ \left\langle \phi(X_r^{t,y}) M_r^s(X_s^{t,y}), \nabla_v X_s^{t,y} - \nabla_v X_s^{t,x} \right\rangle \middle| \mathcal{F}_s^t \right] \\ & \quad + \mathbb{E} \left[ \left\langle \phi(X_r^{t,y}) (M_r^s(X_s^{t,y}) - M_r^s(X_s^{t,x})), \nabla_v X_s^{t,x} \right\rangle \middle| \mathcal{F}_s^t \right] \\ & \quad + \mathbb{E} \left[ \left\langle (\phi(X_r^{t,y}) - \phi(X_r^{t,x})) (M_r^s(X_s^{t,x}) - M_{r-\delta}^s(X_s^{t,x})), \nabla_v X_s^{t,x} \right\rangle \middle| \mathcal{F}_s^t \right] \\ & \quad + \mathbb{E} \left[ \left\langle (\phi(X_r^{t,y}) - \phi(X_r^{t,x})) M_{r-\delta}^s(X_s^{t,x}), \nabla_v X_s^{t,x} \right\rangle \middle| \mathcal{F}_s^t \right] \\ &=: I_1 + I_2 + I_3 + I_4. \end{aligned}$$

For any  $p \in [1, 2)$ , by (2.6), (2.16) (with  $q = 0$ ), (2.7) and the Hölder inequality, we get

$$\begin{aligned} \mathbb{E} |I_1|^p &\leq \mathbb{E} \left[ \mathbb{E} \left( |\phi(X_r^{t,y})| \cdot |M_r^s(X_s^{t,y})| \middle| \mathcal{F}_s^t \right) |\nabla_v X_s^{t,y} - \nabla_v X_s^{t,x}|^p \right] \\ &= \mathbb{E} \left[ \mathbb{E} (|\phi(X_r^{s,z})| \cdot |M_r^s(z)|) \middle|_{z=X_s^{t,y}} |\nabla_v X_s^{t,y} - \nabla_v X_s^{t,x}|^p \right] \\ &\leq C \left( \mathbb{E} \left( 1 + |X_s^{t,y}|^{k_0} \right)^{\frac{2p}{2-p}} \right)^{\frac{2-p}{2}} (\mathbb{E} |\nabla_v X_s^{t,y} - \nabla_v X_s^{t,x}|^2)^{\frac{p}{2}} \\ &\leq C \left( 1 + |y|^{k_0 p k_2} \right) (\mathbb{E} |\nabla_v X_s^{t,y} - \nabla_v X_s^{t,x}|^2)^{\frac{p}{2}}. \end{aligned} \quad (2.26)$$

Then (2.10) yields that  $\lim_{y \rightarrow x} \mathbb{E} |I_1|^p = 0$ .

For the term  $I_2$ , from the Hölder inequality we have for any  $p \in [1, 2)$ ,

$$\mathbb{E} |I_2|^p \leq \mathbb{E} \left| \left\langle \phi(X_r^{t,y}) (M_r^s(X_s^{t,y}) - M_r^s(X_s^{t,x})), \nabla_v X_s^{t,x} \right\rangle \right|^p.$$

By (2.3) of (H3) we arrive at

$$\mathbb{P}\text{-}\lim_{y \rightarrow x} \left\langle \phi(X_r^{t,y}) (M_r^s(X_s^{t,y}) - M_r^s(X_s^{t,x})), \nabla_v X_s^{t,x} \right\rangle = 0.$$

Moreover, as in (2.26), for any  $p \in [1, 2)$  and  $R > 0$ , we have

$$\sup_{|y| \leq R} \mathbb{E} \left| \left\langle \phi(X_r^{t,y}) M_r^s(X_s^{t,y}), \nabla_v X_s^{t,x} \right\rangle \right|^p < \infty,$$

which yields that

$$\{|\langle \phi(X_r^{t,y})M_r^s(X_s^{t,y}), \nabla_v X_s^{t,x} \rangle|^p\}_{|y| \leq R}$$

is uniformly integrable. Consequently, we obtain that for  $p \in [1, 2)$  and  $R > 0$ ,

$$\{|\langle \phi(X_r^{t,y})(M_r^s(X_s^{t,y}) - M_r^s(X_s^{t,x})), \nabla_v X_s^{t,x} \rangle|^p\}_{|y| \leq R}$$

is also uniformly integrable. Then the dominated convergence theorem implies that  $\lim_{y \rightarrow x} \mathbb{E}|I_2|^p = 0$ .

For the term  $I_3$ , by the boundedness of  $\phi$  and the Hölder inequality we get for  $p \in [1, 2)$ ,

$$\begin{aligned} \mathbb{E}|I_3|^p &\leq 2^p \|\phi\|_\infty^p \mathbb{E} \left[ \mathbb{E} \left[ |M_r^s(X_s^{t,x}) - M_{r-\delta}^s(X_s^{t,x})| \middle| \mathcal{F}_s^t \right] |\nabla_v X_s^{t,x}| \right]^p \\ &\leq 2^p \|\phi\|_\infty^p \left( \mathbb{E} \left[ \mathbb{E} \left[ |M_r^s(X_s^{t,x}) - M_{r-\delta}^s(X_s^{t,x})| \middle| \mathcal{F}_s^t \right]^{\frac{2p}{2-p}} \right]^{\frac{2-p}{2}} (\mathbb{E}|\nabla_v X_s^{t,x}|^2)^{\frac{p}{2}} \right). \end{aligned}$$

Using (2.4) of (H3), we have

$$\mathbb{P}\text{-}\lim_{\delta \rightarrow 0^+} \mathbb{E} \left[ |M_r^s(X_s^{t,x}) - M_{r-\delta}^s(X_s^{t,x})| \middle| \mathcal{F}_s^t \right] = 0.$$

By (2.6) (with  $\psi \equiv 1$ ) and (2.5), we obtain that for any  $\tilde{p} \geq 1$ ,

$$\begin{aligned} &\mathbb{E} \left[ \mathbb{E} \left[ |M_r^s(X_s^{t,x}) - M_{r-\delta}^s(X_s^{t,x})| \middle| \mathcal{F}_s^t \right]^{\tilde{p}} \right] \\ &\leq \mathbb{E} \left[ \mathbb{E} \left[ |M_r^s(X_s^{t,x})| + |M_{r-\delta}^s(X_s^{t,x})| \middle| \mathcal{F}_s^t \right]^{\tilde{p}} \right] \\ &= \mathbb{E} \left[ \left( \mathbb{E}|M_r^s(z)| + \mathbb{E}|M_{r-\delta}^s(z)| \right) \middle|_{z=X_s^{t,x}} \right]^{\tilde{p}} \\ &\leq \mathbb{E} \left[ \left( K(s, s, r, z) + K(s, s, r - \delta, z) \right) \middle|_{z=X_s^{t,x}} \right]^{\tilde{p}} \\ &\leq (C(s, s, r) + C(s, s, r - \delta))^{\tilde{p}} \mathbb{E} \left( 1 + |X_s^{t,x}|^{k_0} \right)^{\tilde{p}}. \end{aligned}$$

Combining this with (2.7) and using the condition that  $C(s, s, \cdot)$  is bounded on any closed subinterval of  $(s, r]$ , we have

$$\sup_{\delta < \delta'} \mathbb{E} \left[ \mathbb{E} \left[ |M_r^s(X_s^{t,x}) - M_{r-\delta}^s(X_s^{t,x})| \middle| \mathcal{F}_s^t \right]^{\tilde{p}} \right] < \infty, \quad 0 < \delta' < r - s, \quad \tilde{p} \geq 1.$$

Then the dominated convergence theorem yields that  $\lim_{\delta \rightarrow 0^+} \overline{\lim}_{y \rightarrow x} \mathbb{E}|I_3|^p = 0$ .

For the term  $I_4$ , since  $s < r - \delta < r$  and  $M_{r-\delta}^s(X_s^{t,x})$  is measurable with respect to  $\mathcal{F}_{r-\delta}^t$ , we have

$$I_4 = \mathbb{E} \left[ \left\langle \mathbb{E} \left[ (\phi(X_r^{t,y}) - \phi(X_r^{t,x})) \middle| \mathcal{F}_{r-\delta}^t \right] M_{r-\delta}^s(X_s^{t,x}), \nabla_v X_s^{t,x} \right\rangle \middle| \mathcal{F}_s^t \right].$$

Putting  $\epsilon := |x - y|$  and  $v := \frac{y-x}{|y-x|}$ , by (2.25) we deduce that

$$\mathbb{E} \left[ (\phi(X_r^{t,y}) - \phi(X_r^{t,x})) \middle| \mathcal{F}_{r-\delta}^t \right]$$

$$= \int_0^{|x-y|} \mathbb{E} \left[ \left\langle \phi(X_r^{t,x+\theta v}) M_r^{r-\delta}(X_{r-\delta}^{t,x+\theta v}), \nabla_v X_{r-\delta}^{t,x+\theta v} \right\rangle \middle| \mathcal{F}_{r-\delta}^t \right] d\theta.$$

Then for any  $p \in [1, 2)$ , it follows from (2.6) (with  $\psi \equiv 1$ ), (2.5) and (2.7) that

$$\begin{aligned} & \mathbb{E} \left| \mathbb{E} \left[ (\phi(X_r^{t,y}) - \phi(X_r^{t,x})) \middle| \mathcal{F}_{r-\delta}^t \right] \right|^p \\ & \leq \|\phi\|_\infty |x-y|^{p-1} \int_0^{|x-y|} \mathbb{E} \left[ \mathbb{E} \left[ |M_r^{r-\delta}(X_{r-\delta}^{t,x+\theta v})| \middle| \mathcal{F}_{r-\delta}^t \right] |\nabla_v X_{r-\delta}^{t,x+\theta v}|^p \right] d\theta \\ & \leq C \|\phi\|_\infty |x-y|^{p-1} \int_0^{|x-y|} \mathbb{E} \left[ \left(1 + |X_{r-\delta}^{t,x+\theta v}|^{k_0}\right)^p |\nabla_v X_{r-\delta}^{t,x+\theta v}|^p \right] d\theta \\ & \leq C \|\phi\|_\infty |x-y|^{p-1} \int_0^{|x-y|} \left(1 + \|X_{r-\delta}^{t,x+\theta v}\|_{2k_0 p/(2-p)}^{\frac{2-p}{2}}\right) \|\nabla_v X_{r-\delta}^{t,x+\theta v}\|_2^p d\theta \\ & \leq C \|\phi\|_\infty |x-y|^{p-1} \int_0^{|x-y|} \left(1 + (1 + |x + \theta v|)^{\frac{(2-p)k_2}{2}}\right) \|\nabla_v X_{r-\delta}^{t,x+\theta v}\|_2^p d\theta. \end{aligned}$$

Observe that by (2.10),  $\|\nabla_v X_{r-\delta}^{t,x}\|_2^p$  is locally bounded with respect to  $x$ . Consequently, we have

$$\lim_{y \rightarrow x} \mathbb{E} \left| \mathbb{E} \left[ (\phi(X_r^{t,y}) - \phi(X_r^{t,x})) \middle| \mathcal{F}_{r-\delta}^t \right] \right|^p = 0,$$

which implies

$$\mathbb{P}\text{-}\lim_{y \rightarrow x} \mathbb{E} \left[ (\phi(X_r^{t,y}) - \phi(X_r^{t,x})) \middle| \mathcal{F}_{r-\delta}^t \right] = 0.$$

Note that, again as in (2.26), for any  $p \in [1, 2)$  we get,

$$\begin{aligned} \mathbb{E} \sup_{y \in \mathbb{R}^d} |I_4|^p & \leq 2^p \|\phi\|_\infty^p \mathbb{E} \left( \mathbb{E} \left[ |M_{r-\delta}^s(X_s^{t,x})| \middle| \mathcal{F}_s^t \right] |\nabla_v X_s^{t,x}| \right)^p \\ & \leq C \left(1 + |x|^{k_0 p k_2}\right) \left(\mathbb{E} |\nabla_v X_s^{t,x}|^2\right)^{\frac{p}{2}} < \infty. \end{aligned} \quad (2.27)$$

Then, by the dominated convergence theorem and (2.27) (with  $p = 1$ ) we obtain

$$\lim_{y \rightarrow x} \mathbb{E} |I_4| \leq \lim_{y \rightarrow x} \mathbb{E} \left\{ \left| \mathbb{E} \left[ (\phi(X_r^{t,y}) - \phi(X_r^{t,x})) \middle| \mathcal{F}_{r-\delta}^t \right] \right| \cdot |M_{r-\delta}^s(X_s^{t,x})| \cdot |\nabla_v X_s^{t,x}| \right\} = 0,$$

which leads to  $\mathbb{P}\text{-}\lim_{y \rightarrow x} |I_4| = 0$ . Therefore, by the dominated convergence theorem and (2.27) again, we derive that  $\lim_{y \rightarrow x} \mathbb{E} |I_4|^p = 0$  for any  $p \in [1, 2)$ .

Let us now prove that for  $\phi \in \mathcal{B}(\mathbb{R}^d)$  satisfying (2.8), (2.9) holds in  $L^p(\mathbb{P})$  for  $p \in [1, 2)$ . Notice that for such  $\phi$ , (2.11) holds true because of (2.8) and (2.16), and due to Fact (ii) there exists a sequence  $\{\tilde{\phi}_n\}_{n=1}^\infty \subset \mathcal{B}_b(\mathbb{R}^d)$  such that (2.12) holds. Since we have proved that (2.9) holds for each  $\tilde{\phi}_n$ , by letting  $s = t$  we get

$$(\nabla_v P_{t,r} \tilde{\phi}_n(x) =) \nabla_v \mathbb{E} \left[ \tilde{\phi}_n(X_r^{t,x}) \right] = \mathbb{E} \left[ \tilde{\phi}_n(X_r^{t,x}) \langle M_r^t(x), v \rangle \right], \quad v \in \mathbb{R}^d.$$

From the result proved in the above paragraphs, we know that  $\mathbb{E}[\tilde{\phi}_n(X_r^{t,x}) \langle M_r^t(x), v \rangle]$  is continuous with respect to  $x$ , that is,  $\nabla P_{s,r} \tilde{\phi}_n \in C(\mathbb{R}^d, \mathbb{R}^d)$ . Then along the same lines as in (2.14) and (2.15), we derive that  $P_{s,r} \phi \in C^1(\mathbb{R}^d)$  and it satisfies (2.2) and (2.13), which completes the proof of assertion (2), thanks to Step 1.  $\square$

## 2.2 Main results for FBSDEs

We now consider the following decoupled FBSDE for  $(t, x) \in [0, T] \times \mathbb{R}^d$ :

$$X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x}) dr + \int_t^s \sigma(r, X_r^{t,x}) dW_r, \quad s \in [t, T], \quad (2.28)$$

$$Y_s^{t,x} = g(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dr - \int_s^T Z_r^{t,x} dW_r, \quad s \in [t, T]. \quad (2.29)$$

Here  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^m$  and  $g : \mathbb{R}^d \rightarrow \mathbb{R}^l$ ,  $f : [0, T] \times \mathbb{R}^d \times \mathbb{R}^l \times \mathbb{R}^l \otimes \mathbb{R}^m \rightarrow \mathbb{R}^l$  are measurable mappings,  $W$  is an  $m$ -dimensional Brownian motion. Instead of concrete conditions imposed on  $b$  and  $\sigma$ , we assume that  $X_s^{t,x}$  has the following properties:

(C1) The system consisting of (2.28) and the equation

$$\nabla X_s^{t,x} = I_{d \times d} + \int_t^s \nabla b(r, X_r^{t,x}) \nabla X_r^{t,x} dr + \int_t^s \nabla \sigma(r, X_r^{t,x}) \nabla X_r^{t,x} dW_r, \quad s \in [t, T] \quad (2.30)$$

has a unique strong solution, where  $I_{d \times d}$  is the identity matrix in  $\mathbb{R}^d \otimes \mathbb{R}^d$ ,  $\nabla b$  and  $\nabla \sigma$  are the weak derivatives of  $b$  and  $\sigma$ , respectively. There exist  $k_2 \geq 1$ ,  $k_3 \geq 0$  such that for any  $k_1 \geq 1$ , (2.7) holds for some  $C_{k_1, k_2, T} > 0$  and

$$\sup_{0 \leq s \leq r \leq T} \|\nabla X_r^{s,y}\|_{k_1} \leq C_{k_1, k_3, T} (1 + |y|^{k_3}), \quad y \in \mathbb{R}^d, \quad (2.31)$$

for some  $C_{k_1, k_3, T} > 0$ . Moreover, for any  $t \leq s \leq T$  and  $x, v \in \mathbb{R}^d$ ,

$$\lim_{\epsilon \rightarrow 0^+} \mathbb{E} \sup_{s \in [t, T]} \left| \frac{X_s^{t, x + \epsilon v} - X_s^{t, x}}{\epsilon} - \nabla_v X_s^{t, x} \right|^2 = 0, \quad (2.32)$$

$$\mathbb{P}\text{-}\lim_{y \rightarrow x} \nabla_v X_s^{t, y} = \nabla_v X_s^{t, x}. \quad (2.33)$$

(C2) There exist continuously differentiable  $(b_n, \sigma_n)_{n \geq 1}$  with

$$\sup_{r \in [0, T], x \in \mathbb{R}^d} (\|\nabla b_n(r, x)\| + \|\nabla \sigma_n(r, x)\|) < \infty, \quad \forall n \in \mathbb{N}, \quad (2.34)$$

such that for any sequence  $x_n$  with  $\lim_{n \rightarrow +\infty} x_n = x$  and for the solution of (2.28) and (2.30) with  $(x, b, \sigma)$  replaced by  $(x_n, b_n, \sigma_n)$ , which we denote by  $(X_s^{n, t, x_n}, \nabla X_s^{n, t, x_n})$ , the following hold

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left( |X_T^{n, t, x_n} - X_T^{t, x}|^2 + \int_t^T |X_r^{n, t, x_n} - X_r^{t, x}|^2 dr \right) = 0, \quad (2.35)$$

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left( \|\nabla X_T^{n, t, x_n} - \nabla X_T^{t, x}\|^2 + \int_t^T \|\nabla X_r^{n, t, x_n} - \nabla X_r^{t, x}\|^2 dr \right) = 0, \quad (2.36)$$

$$\lim_{n \rightarrow +\infty} \|\sigma_n(s, x_n) - \sigma(s, x)\| = 0, \quad s \in [t, T]. \quad (2.37)$$

**Remark 2.1.** *It is clear that if  $b(r, \cdot)$  and  $\sigma(r, \cdot)$  are continuously differentiable and satisfy (2.76) below, then (C1) and (C2) hold. Besides, we may allow that  $\sigma$  has polynomial growth which also ensures that (C1) and (C2) are satisfied (see Example 2.9 below).*

Now, we introduce the hypotheses on the coefficients  $g$  and  $f$  of the BSDE (2.29), under which we will be able to establish gradient type and Bismut type formulas for the solution  $Y_s^{t,x}$ .

(A1)  $g$  is Lipschitz continuous, i.e. there exists a constant  $K_1 \geq 0$  such that

$$|g(x) - g(y)| \leq K_1|x - y|, \quad x, y \in \mathbb{R}^d.$$

(A1')  $g$  has  $q$ -th growth, i.e. there exist constants  $\tilde{K}_1 \geq 0$  and  $q \geq 0$  such that

$$|g(x)| \leq \tilde{K}_1(1 + |x|^q), \quad x \in \mathbb{R}^d.$$

(A2) There exist constants  $K_2 \geq 0$  and  $\tilde{K}_2 \geq 0$  such that for all  $(x_i, y_i, z_i) \in \mathbb{R}^d \times \mathbb{R}^l \times \mathbb{R}^l \otimes \mathbb{R}^m$ ,

$$\sup_{r \in [0, T]} |f(r, x_1, y_1, z_1) - f(r, x_2, y_2, z_2)| \leq K_2(|x_1 - x_2| + |y_1 - y_2| + \|z_1 - z_2\|),$$

$$\tilde{K}_2 := \sup_{r \in [0, T]} |f(r, 0, 0, 0)| < \infty.$$

To simplify the notations, for any  $t \leq s < r \leq T$  and  $x, v \in \mathbb{R}^d$ , we set  $\Theta_r^{t,x} := (X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})$  and  $M_r^{t,s}(x, v) := \langle M_r^s(X_s^{t,x}), \nabla_v X_s^{t,x} \rangle$ . In particular,  $M_r^{t,t}(x, v) = \langle M_r^t(x), v \rangle$ . Now we formulate our main result, the proof of which we present in the next subsection.

**Theorem 2.3.** *Let  $0 \leq t \leq s < T$  and  $x_0, v \in \mathbb{R}^d$ . Assume that (C1), (C2), (H2) and (A2) are satisfied and that (2.6) holds for any  $\psi \in \mathcal{B}_b(\mathbb{R}^d)$   $\mathbb{P}$ -a.s.*

(1) *Let (A1) hold, and assume furthermore that  $\sigma(r, \cdot)$  is continuous and there exist  $C > 0$  and  $k_4 \geq 0$  such that*

$$\sup_{r \in [0, T]} \|\sigma(r, x)\| \leq C(1 + |x|^{k_4}), \quad x \in \mathbb{R}^d, \quad (2.38)$$

and that  $K(t, s, r, \cdot)$  satisfies (2.5) and for some  $R > 0$

$$\int_s^T \sup_{|x - x_0| \leq R} K(t, s, r, x) dr < \infty. \quad (2.39)$$

If in addition, either  $g(\cdot)$  and  $f(r, \cdot, \cdot, \cdot)$  are continuously differentiable, or (2.3) holds, then the following gradient type formula in  $L^p(\mathbb{P})$  holds for any  $p \in [1, 2)$ :

$$\nabla_v Y_s^{t,x_0} = \mathbb{E} \left[ g(X_T^{t,x_0}) M_T^{t,s}(x_0, v) + \int_s^T f(r, \Theta_r^{t,x_0}) M_r^{t,s}(x_0, v) dr \middle| \mathcal{F}_s^t \right]. \quad (2.40)$$

(2) *Let (A1') hold, and assume furthermore that  $k_2 = 1$ , that  $\sigma$  is bounded on  $[0, T] \times \mathbb{R}^d$ , and that  $K(t, s, r, \cdot)$  satisfies (2.5) with  $k_0 = 0$  and there exists  $\beta < 1$  such that*

$$\sup_{0 \leq t \leq s < r \leq T} \left[ C(t, s, r)(r - s)^\beta \right] < \infty. \quad (2.41)$$

If in addition, either  $g \in C(\mathbb{R}^d; \mathbb{R}^l)$  and (2.3) holds, or (H3) is satisfied, then (2.40) also holds in  $L^p(\mathbb{P})$  for any  $p \in [1, 2)$ .

Concerning a Bismut type formula of  $Y_s^{t,x}$ , we have the following corollary. The proof of this corollary follows from the same kind of arguments as Theorem 2.3 with the conditional expectation  $\mathbb{E}[\cdot|\mathcal{F}_s^t]$  replaced by the expectation  $\mathbb{E}[\cdot]$ , and it is thus omitted here for the sake of conciseness.

**Corollary 2.4.** *Let  $0 \leq t \leq s < T$  and  $x_0, v \in \mathbb{R}^d$ . Assume that (C1), (C2), (H2) and (A2) are satisfied.*

(1) *Let (A1) hold, and assume furthermore that  $\sigma(r, \cdot)$  is continuous and there exist  $C > 0$  and  $k_4 \geq 0$  such that*

$$\sup_{r \in [0, T]} \|\sigma(r, x)\| \leq C(1 + |x|^{k_4}), \quad x \in \mathbb{R}^d,$$

and that  $K(t, t, r, \cdot)$  satisfies (2.5) and for some  $R > 0$

$$\int_s^T \sup_{|x-x_0| \leq R} K(t, t, r, x) dr < \infty. \quad (2.42)$$

If in addition, either  $g(\cdot)$  and  $f(r, \cdot, \cdot, \cdot)$  are continuously differentiable, or (2.3) (with  $s = t$ ) holds, then the following Bismut type formula in  $L^p(\mathbb{P})$  holds for any  $p \in [1, 2)$ :

$$\nabla_v \mathbb{E} Y_s^{t, x_0} = \mathbb{E} \left[ g(X_T^{t, x_0}) \langle M_T^t(x_0), v \rangle + \int_s^T f(r, \Theta_r^{t, x_0}) \langle M_r^t(x_0), v \rangle dr \right]. \quad (2.43)$$

(2) *Let (A1') hold, and assume furthermore that  $k_2 = 1$ , that  $\sigma$  is bounded on  $[0, T] \times \mathbb{R}^d$ , and that  $K(t, t, r, \cdot)$  satisfies (2.5) with  $k_0 = 0$  and for some  $\beta < 1$*

$$\sup_{0 \leq t < r \leq T} \left[ C(t, t, r)(r - t)^\beta \right] < \infty.$$

If in addition, either  $g \in C(\mathbb{R}^d; \mathbb{R}^1)$  and (2.3) (with  $s = t$ ) holds, or (H3) (with  $s = t$ ) is satisfied, then (2.43) also holds in  $L^p(\mathbb{P})$  for any  $p \in [1, 2)$ .

### 2.3 Proof of the main result

To prove Theorem 2.3, we need the following useful lemma. Its proof is elementary but lengthy, and we defer it to the Appendix.

**Lemma 2.5.** *Assume that (C1), (C2), (A1) and (A2) hold, and that  $g(\cdot)$  and  $f(r, \cdot, \cdot, \cdot)$  are continuously differentiable with any  $r \in [0, T]$ . For each  $(t, x) \in [0, T] \times \mathbb{R}^d$ , let  $(X^{t,x}, Y^{t,x}, Z^{t,x})$  be the solution of (2.28)-(2.29) and set  $u(t, x) := Y_t^{t,x}$ . Then for any  $t \in [0, T]$ ,  $u(t, \cdot)$  is continuously differentiable with*

$$\nabla u(t, x) = \mathbb{E} \left[ \nabla g(X_T^{t,x}) \nabla X_T^{t,x} + \int_t^T \nabla f(r, \Theta_r^{t,x}) \nabla \Theta_r^{t,x} dr \right], \quad x \in \mathbb{R}^d, \quad (2.44)$$

and there exist positive constants  $C_{K_1, K_2, k_3, T}$  and  $q_1$  such that

$$\sup_{t \in [0, T]} \|\nabla u(t, x)\| \leq C_{K_1, K_2, k_3, T} (1 + |x|^{q_1}), \quad x \in \mathbb{R}^d. \quad (2.45)$$

Moreover,  $Y_s^{t,x} = u(s, X_s^{t,x})$  and  $Z_s^{t,x} = \nabla u(s, X_s^{t,x}) \sigma(s, X_s^{t,x})$ ,  $ds \otimes d\mathbb{P}$ -a.e.

**Proof of Theorem 2.3**

First observe that, by (2.29) and the Fubini theorem we have

$$\begin{aligned} Y_s^{t,x} &= \mathbb{E} \left[ g(X_T^{t,x}) + \int_s^T f(r, \Theta_r^{t,x}) dr \middle| \mathcal{F}_s^t \right] \\ &= \mathbb{E} \left[ g(X_T^{t,x}) \middle| \mathcal{F}_s^t \right] + \int_s^T \mathbb{E} [f(r, \Theta_r^{t,x}) \middle| \mathcal{F}_s^t] dr. \end{aligned} \quad (2.46)$$

Below we shall apply Proposition 2.1 to  $f$  and  $g$  in two different cases (1) and (2), respectively. Without loss of generality, we assume that  $l = 1$ . Otherwise, we may use  $\langle Y_s^{t,x}, e \rangle, \langle g, e \rangle$  and  $\langle f, e \rangle$  to replace  $Y_s^{t,x}, g$  and  $f$  for a unit vector  $e \in \mathbb{R}^l$ .

We first prove assertion (1). If (A1) holds, then  $g \in C(\mathbb{R}^d)$  and it satisfies (2.8) with  $q = 1$ . So, with the help of Proposition 2.1 (1) we have that

$$\nabla_v \mathbb{E} \left[ g(X_T^{t,x_0}) \middle| \mathcal{F}_s^t \right] = \mathbb{E} \left[ g(X_T^{t,x_0}) M_T^{t,s}(x_0, v) \middle| \mathcal{F}_s^t \right] \quad (2.47)$$

in  $L^p(\mathbb{P})$  for any  $p \in [1, 2)$ .

Now, we focus on dealing with the directional derivative of the second term of the right-hand side of (2.46). For the coefficients  $g(\cdot), f(r, \cdot, \cdot, \cdot)$  which fulfill (A1) and (A2) respectively, by standard approximation arguments there exist continuously differentiable sequences  $(g_n(\cdot))_{n \geq 1}, (f_n(r, \cdot, \cdot, \cdot))_{n \geq 1}$ , in which for each  $n \geq 1, g_n$  and  $f_n$  satisfy (A1) and (A2) with the same Lipschitz constants  $K_1, K_2$  and some positive constant  $\bar{K}_2$  (independent of  $n$ ), and moreover  $g_n$  converges to  $g$  uniformly on  $\mathbb{R}^d$  and  $f_n$  converges to  $f$  uniformly on  $[t, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^m$ . We now let  $(Y^{n,t,x}, Z^{n,t,x})$  be the solution of (2.29) with coefficients  $g, f$  replaced by  $g_n, f_n$ , respectively. By the Itô formula, one can show that

$$\lim_{n \rightarrow +\infty} \left( \sup_{r \in [t, T]} \mathbb{E} |Y_r^{n,t,x} - Y_r^{t,x}|^2 + \int_t^T \mathbb{E} |Z_r^{n,t,x} - Z_r^{t,x}|^2 dr \right) = 0. \quad (2.48)$$

The above step is partially borrowed from [23, Theorem 4.2, Page 1410-1411]. By Lemma 2.5, there exist  $u_n : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  continuously differentiable in the space variable, positive constants  $C_{K_1, K_2, \bar{K}_2, T}$  and  $q_1$  independent of  $n$  such that

$$\sup_{t \in [0, T]} (|u_n(t, x)| + |\nabla u_n(t, x)|) \leq C_{K_1, K_2, \bar{K}_2, T} (1 + |x|^{q_1}), \quad x \in \mathbb{R}^d \quad (2.49)$$

and

$$Y_s^{n,t,x} = u_n(s, X_s^{t,x}), \quad Z_s^{n,t,x} = \nabla u_n(s, X_s^{t,x}) \sigma(s, X_s^{t,x}), \quad ds \otimes d\mathbb{P}\text{-a.e.}$$

Let  $\Theta^{n,t,x} := (X^{t,x}, Y^{n,t,x}, Z^{n,t,x})$ . Then, we have

$$f(r, \Theta_r^{n,t,x}) = f(r, X_r^{t,x}, u_n(r, X_r^{t,x}), \nabla u_n(r, X_r^{t,x}) \sigma(r, X_r^{t,x})).$$

For any  $r \in [t, T]$ , let  $F_n(r, \cdot) := f(r, \cdot, u_n(r, \cdot), \nabla u_n(r, \cdot) \sigma(r, \cdot))$ . Then, by (A2), the continuity of  $\sigma(r, x)$  in  $x$ , (2.38) and (2.49), we see that for a.e.  $r \in [t, T]$ ,  $F_n(r, \cdot) \in C(\mathbb{R}^d)$  and

$$|F_n(r, x)| \leq \tilde{K}_2 + K_2 (|x| + |u_n(r, x)| + |\nabla u_n(r, x) \sigma(r, x)|)$$



$$\begin{aligned}
&\leq \tilde{K}_2 + K_2 \left[ |x| + C_{K_1, K_2, \bar{K}_2, T} (1 + |x|^{q_1}) + \right. \\
&\quad \left. + C_{K_1, K_2, \bar{K}_2, T} C (1 + |x|^{q_1}) (1 + |x|^{k_4}) \right] \\
&\leq C_{K_1, K_2, \bar{K}_2, \bar{K}_2, T} \left( 1 + |x|^{(q_1 + k_4) \vee 1} \right). \tag{2.50}
\end{aligned}$$

Consequently, using (C1), (H2) and (2.5)-(2.6), we then apply Proposition 2.1 (1) to conclude that for a.e.  $r \in [t, T]$ ,

$$\begin{aligned}
\nabla_v \mathbb{E} \left[ f(r, \Theta_r^{n,t,x}) \Big| \mathcal{F}_s^t \right] &= \nabla_v \mathbb{E} \left[ F_n(r, X_r^{t,x}) \Big| \mathcal{F}_s^t \right] \\
&= \mathbb{E} \left[ F_n(r, X_r^{t,x}) M_r^{t,s}(x, v) \Big| \mathcal{F}_s^t \right] \\
&= \mathbb{E} \left[ f(r, \Theta_r^{n,t,x}) M_r^{t,s}(x, v) \Big| \mathcal{F}_s^t \right], \quad v \in \mathbb{R}^d \tag{2.51}
\end{aligned}$$

in  $L^p(\mathbb{P})$  for any  $p \in [1, 2)$ . In view of (2.50), (2.7), (2.31) and the fact that  $C_{K_1, K_2, \bar{K}_2, \bar{K}_2, T}$  and  $q_1$  above are independent of  $n$ , we obtain that for a.e.  $r \in [t, T]$  and  $p \in [1, 2)$ ,

$$\begin{aligned}
&\left\| \mathbb{E} \left[ \sup_{n \geq 1} |f(r, \Theta_r^{n,t,x}) M_r^{t,s}(x, v)| \Big| \mathcal{F}_s^t \right] \right\|_p \\
&\leq C_{K_1, K_2, \bar{K}_2, \bar{K}_2, T} \left\| (1 + |X_r^{t,x}|^{(q_1 + k_4) \vee 1}) |M_r^s(X_s^{t,x})| \cdot |\nabla_v X_s^{t,x}| \right\|_p \\
&\leq C_{K_1, K_2, \bar{K}_2, \bar{K}_2, T} \left\| (1 + |X_r^{t,x}|^{(q_1 + k_4) \vee 1}) |\nabla_v X_s^{t,x}| \right\|_{\frac{2p}{2-p}} \|M_r^s(X_s^{t,x})\|_2 \\
&\leq C_{K_1, K_2, \bar{K}_2, \bar{K}_2, k_2, k_3, k_4, q_1, p, T} \left( 1 + |x|^{((q_1 + k_4) \vee 1) k_2} \right) (1 + |x|^{k_3}) K(t, s, r, x) |v|. \tag{2.52}
\end{aligned}$$

This, together with (2.39), leads to

$$\int_s^T \sup_{|x-x_0| \leq R} \left\| \mathbb{E} \left[ \sup_{n \geq 1} |f(r, \Theta_r^{n,t,x}) M_r^{t,s}(x, v)| \Big| \mathcal{F}_s^t \right] \right\|_p dr < \infty. \tag{2.53}$$

Then, using (2.51), (2.48) and the dominated convergence theorem, we deduce that the following relations hold in  $L^p(\mathbb{P})$  for  $p \in [1, 2)$  and  $x \in B_R(x_0) := \{z : |z - x_0| < R\}$ :

$$\begin{aligned}
\nabla_v \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{n,t,x}) \Big| \mathcal{F}_s^t \right] dr &= \int_s^T \nabla_v \mathbb{E} \left[ f(r, \Theta_r^{n,t,x}) \Big| \mathcal{F}_s^t \right] dr \\
&= \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{n,t,x}) M_r^{t,s}(x, v) \Big| \mathcal{F}_s^t \right] dr \tag{2.54}
\end{aligned}$$

and

$$\lim_{n \rightarrow +\infty} \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{n,t,x}) M_r^{t,s}(x, v) \Big| \mathcal{F}_s^t \right] dr = \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{t,x}) M_r^{t,s}(x, v) \Big| \mathcal{F}_s^t \right] dr. \tag{2.55}$$

Hence, (2.54)-(2.55) and the dominated convergence theorem imply that for any  $\epsilon > 0$  and  $v \in \mathbb{R}^d$  with  $x + \epsilon v \in B_R(x_0)$ ,

$$\int_s^T \mathbb{E} \left[ f(r, \Theta_r^{t,x+\epsilon v}) \Big| \mathcal{F}_s^t \right] dr - \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{t,x}) \Big| \mathcal{F}_s^t \right] dr$$

$$\begin{aligned}
&= \lim_{n \rightarrow +\infty} \left( \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{n,t,x+\epsilon v}) \middle| \mathcal{F}_s^t \right] dr - \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{n,t,x}) \middle| \mathcal{F}_s^t \right] dr \right) \\
&= \lim_{n \rightarrow +\infty} \int_0^\epsilon \left( \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{n,t,x+\theta v}) M_r^{t,s}(x+\theta v, v) \middle| \mathcal{F}_s^t \right] dr \right) d\theta \\
&= \int_0^\epsilon \left( \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{t,x+\theta v}) M_r^{t,s}(x+\theta v, v) \middle| \mathcal{F}_s^t \right] dr \right) d\theta, \tag{2.56}
\end{aligned}$$

where the first relation is due to (A2), (2.48) and (2.50).

Next, we will show that

$$\int_s^T \mathbb{E} \left[ f(r, \Theta_r^{t,x}) M_r^{t,s}(x, v) \middle| \mathcal{F}_s^t \right] dr$$

is continuous with respect to  $x$  in a neighbourhood of  $x_0$  in  $L^p(\mathbb{P})$  for  $p \in [1, 2)$ . Observe that this continuity, along with (2.56), easily yields that

$$\nabla_v \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{t,x_0}) \middle| \mathcal{F}_s^t \right] dr = \int_s^T \mathbb{E} \left[ f(r, \Theta_r^{t,x_0}) M_r^{t,s}(x_0, v) \middle| \mathcal{F}_s^t \right] dr \tag{2.57}$$

in  $L^p(\mathbb{P})$  for  $p \in [1, 2)$ , which due to our previous equality (2.47), leads to (2.40). We now invoke the condition (2.3) to get the continuity. Indeed, due to (A1), (A2) and (2.32), we can apply the Itô formula and the B-D-G inequality to get

$$\lim_{y \rightarrow x} \mathbb{E} \left[ \sup_{r \in [t, T]} |Y_r^{t,y} - Y_r^{t,x}|^2 + \int_t^T |Z_r^{t,y} - Z_r^{t,x}|^2 dr \right] = 0.$$

Combining this with (2.3) and (2.33) yields

$$dr \otimes d\mathbb{P} \text{-} \lim_{y \rightarrow x} f(r, \Theta_r^{t,y}) M_r^{t,s}(y, v) = f(r, \Theta_r^{t,x}) M_r^{t,s}(x, v). \tag{2.58}$$

Observe that by (2.50), we have

$$|f(r, \Theta_r^{n,t,x})| \leq C_{K_1, K_2, \tilde{K}_2, \bar{K}_2, T} \left( 1 + |X_r^{t,x}|^{(q_1+k_4) \vee 1} \right),$$

which, along with (2.48), implies

$$|f(r, \Theta_r^{t,x})| \leq C_{K_1, K_2, \tilde{K}_2, \bar{K}_2, T} \left( 1 + |X_r^{t,x}|^{(q_1+k_4) \vee 1} \right).$$

Then, as in (2.52) and (2.53), we derive that for any  $p \in [1, 2)$ ,

$$\left\{ |f(r, \Theta_r^{t,x}) M_r^{t,s}(x, v)|^p \right\}_{|x-x_0| \leq R} \tag{2.59}$$

is uniformly integrable and

$$\int_s^T \sup_{|x-x_0| \leq R} \left\| \mathbb{E} \left[ |f(r, \Theta_r^{t,x}) M_r^{t,s}(x, v)| \middle| \mathcal{F}_s^t \right] \right\|_p dr < \infty. \tag{2.60}$$

Thus, by (2.58) and (2.59) the dominated convergence theorem implies that

$$\lim_{y \rightarrow x} \mathbb{E} \left[ f(r, \Theta_r^{t,y}) M_r^{t,s}(y, v) \middle| \mathcal{F}_s^t \right] = \mathbb{E} \left[ f(r, \Theta_r^{t,x}) M_r^{t,s}(x, v) \middle| \mathcal{F}_s^t \right], \quad x \in B_R(x_0),$$

in  $L^p(\mathbb{P})$  for any  $p \in [1, 2)$  and a.e.  $r \in [t, T]$ . Consequently, again by the dominated convergence theorem and (2.60) we get the desired continuity result.

Note that if  $g(\cdot)$  and  $f(r, \cdot, \cdot, \cdot)$  are continuously differentiable, then we just set  $g_n \equiv g$  and  $f_n \equiv f$ , and the assertion (2.40) follows from (2.54) and (2.47).

We now prove assertion (2). Let  $g \in C(\mathbb{R}^d)$  satisfying (A1'), then we apply directly Proposition 2.1 (1) to get (2.47). While in the case of  $g$  only satisfying (A1'), we first observe that (2.10) holds due to (2.31) and (2.33), and then by (H3) and (2.41) it is easily checked in our context that the hypotheses of Proposition 2.1 (2) are all satisfied, which also yields (2.47).

Next, we investigate the directional derivative of the second term in the right-hand side of (2.46) by adopting a similar approximation argument as for assertion (1). For  $f$  satisfying (A2), let  $(f_n)_{n \geq 1}$  be as defined above. For  $g$  satisfying (A1'), we set  $\tilde{g}_n := (g \vee (-n)) \wedge n$  for each  $n \geq 1$ . Then it follows from the Lusin theorem (see, e.g., [11, Theorem 7.4.4]), that there exist  $\{\hat{g}_n\}_{n \geq 1} \subset C_b(\mathbb{R}^d)$  and compact sets  $\{\mathcal{K}_n\}_{n \geq 1}$  such that

$$\hat{g}_n \Big|_{\mathcal{K}_n} = \tilde{g}_n \Big|_{\mathcal{K}_n}, \quad \|\hat{g}_n\|_\infty \leq n, \quad \nu(\mathcal{K}_n^c) \leq \frac{1}{n^3}, \quad (2.61)$$

where  $\nu = \mathbb{P}_{X_T^{t,x}}$  is the law of  $X_T^{t,x}$ . By a standard approximation method, we may choose  $g_n \in C_b^1(\mathbb{R}^d)$  such that

$$\sup_{x \in \mathcal{K}_n} |g_n(x) - \hat{g}_n(x)| \leq \frac{1}{n}, \quad \|g_n\|_\infty \leq n. \quad (2.62)$$

Now, let  $(Y^{n,t,x}, Z^{n,t,x})$  be the solution of (2.29) with coefficients  $(g_n, f_n)$  replacing  $(g, f)$ . Applying the Itô formula and the B-D-G inequality, we obtain that

$$\begin{aligned} & \mathbb{E} \left( \sup_{r \in [t, T]} |Y_r^{n,t,x} - Y_r^{t,x}|^2 + \int_t^T |Z_r^{n,t,x} - Z_r^{t,x}|^2 dr \right) \\ & \leq C \left( \mathbb{E} |g_n(X_T^{t,x}) - g(X_T^{t,x})|^2 + \|f_n - f\|_\infty \right) \end{aligned} \quad (2.63)$$

and

$$\mathbb{E} \left( \sup_{r \in [t, T]} |Y_r^{n,t,x}|^2 + \int_t^T |Z_r^{n,t,x}|^2 dr \right) \leq C \left( \mathbb{E} |g_n(X_T^{t,x})|^2 + 1 \right) \quad (2.64)$$

with some constant  $C > 0$  independent of  $n$  and  $t$ . Note that from (2.61) and (2.62), we have

$$\begin{aligned} |(g_n - g)(x)| & \leq |(g_n - \hat{g}_n)(x)| + |(\hat{g}_n - \tilde{g}_n)(x)| + |(\tilde{g}_n - g)(x)| \\ & = |(g_n - \hat{g}_n)(x)| \mathbf{1}_{\mathcal{K}_n}(x) + |(g_n - \hat{g}_n)(x)| \mathbf{1}_{\mathcal{K}_n^c}(x) \\ & \quad + |(\hat{g}_n - \tilde{g}_n)(x)| \mathbf{1}_{\mathcal{K}_n^c}(x) + |(\tilde{g}_n - g)(x)| \mathbf{1}_{\{|g(x)| \geq n\}} \\ & \leq \frac{1}{n} + 4n \mathbf{1}_{\mathcal{K}_n^c}(x) + |g(x)| \mathbf{1}_{\{|g(x)| \geq n\}}, \end{aligned}$$

which yields

$$\mathbb{E} \left| g_n(X_T^{t,x}) - g(X_T^{t,x}) \right|^2 \leq \frac{3}{n^2} + 48n^2 \mathbb{P} \left( X_T^{t,x} \notin \mathcal{K}_n \right) + 3 \mathbb{E} \left[ |g(X_T^{t,x})|^2 \mathbf{1}_{\{|g(X_T^{t,x})| \geq n\}} \right].$$

Then, again by (2.61), (A1') and (2.7) we get

$$\lim_{n \rightarrow +\infty} \mathbb{E} |g_n(X_T^{t,x}) - g(X_T^{t,x})|^2 = 0.$$

Consequently, substituting this into (2.63) and using the uniform convergence of  $f_n$  to  $f$  we obtain

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left( \sup_{r \in [t, T]} |Y_r^{n,t,x} - Y_r^{t,x}|^2 + \int_t^T |Z_r^{n,t,x} - Z_r^{t,x}|^2 dr \right) = 0. \quad (2.65)$$

We now intend to find an upper bound for (2.64) independent of  $n$  and  $t$ . By (2.61), (2.62), (A1') and (2.7) with  $k_2 = 1$ , we have

$$\begin{aligned} \|g_n(X_T^{t,x})\|_2 &\leq \|\hat{g}_n(X_T^{t,x}) \mathbf{1}_{\mathcal{K}_n}(X_T^{t,x})\|_2 + \|\hat{g}_n(X_T^{t,x}) \mathbf{1}_{\mathcal{K}_n^c}(X_T^{t,x})\|_2 + \|g_n(X_T^{t,x}) - \hat{g}_n(X_T^{t,x})\|_2 \\ &\leq \|g(X_T^{t,x})\|_2 + \|\hat{g}_n(X_T^{t,x}) \mathbf{1}_{\mathcal{K}_n^c}(X_T^{t,x})\|_2 \\ &\quad + \frac{1}{n} + \|(g_n(X_T^{t,x}) - \hat{g}_n(X_T^{t,x})) \mathbf{1}_{\mathcal{K}_n^c}(X_T^{t,x})\|_2 \\ &\leq C(1 + |x|^q) + n\sqrt{\nu(\mathcal{K}_n^c)} + \frac{1}{n} + 2n\sqrt{\nu(\mathcal{K}_n^c)} \\ &\leq C(1 + |x|^q), \end{aligned} \quad (2.66)$$

where  $C$  is independent of  $n$  and  $t$ . Then, plugging this into (2.64) yields

$$\sup_{n \geq 1, t \in [0, T]} \mathbb{E} \left( \sup_{r \in [t, T]} |Y_r^{n,t,x}|^2 + \int_t^T |Z_r^{n,t,x}|^2 dr \right) \leq C(1 + |x|^q)^2. \quad (2.67)$$

On the other hand, since  $g_n, f_n$  are continuously differentiable, we can apply Lemma 2.5 to derive that there exists  $u_n : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $\nabla u_n(r, \cdot)$  is continuous for each  $r \in [0, T]$  and

$$Y_r^{n,t,x} = u_n(r, X_r^{t,x}), \quad Z_r^{n,t,x} = \nabla u_n(r, X_r^{t,x}) \sigma(r, X_r^{t,x}), \quad dr \otimes d\mathbb{P}\text{-a.e.} \quad (2.68)$$

By (2.67) and the fact that  $Y_t^{n,t,x}$  is deterministic, we have

$$\sup_{n \geq 1, x \in \mathbb{R}^d, t \in [0, T]} \frac{|u_n(t, x)|}{1 + |x|^q} < \infty.$$

Combining this with (2.65) and (2.68), we get

$$\sup_{n \geq 1} |Y_r^{n,t,x}| + |Y_r^{t,x}| \leq C(1 + |X_r^{t,x}|^q), \quad dr \otimes d\mathbb{P}\text{-a.e., } x \in \mathbb{R}^d. \quad (2.69)$$

Next, we want to establish an analogue of (2.69) for  $Z_r^{n,t,x}$  and  $Z_r^{t,x}$ . For this, define

$$A_s^{[n]} := (T - s)^\beta \sup_{x \in \mathbb{R}^d} \frac{|\nabla u_n(s, x)|}{1 + |x|^{q\vee 1}}, \quad s \in [t, T], \quad B_t^{[n]} := \sup_{s \in [t, T]} A_s^{[n]}.$$

Using an argument from step 1 in the proof of [36, Theorem 3.2], we shall prove that  $B_t^{[n]}$  has an upper bound independent of  $n$ . Note first that from assertion (1), there holds a gradient type formula in  $L^p(\mathbb{P})$  for  $Y_s^{n,t,x}$  for any  $s \in [t, T]$ :

$$\nabla_v Y_s^{n,t,x} = \mathbb{E} \left[ g_n(X_T^{t,x}) M_T^{t,s}(x, v) + \int_s^T f_n(r, \Theta_r^{n,t,x}) M_r^{t,s}(x, v) dr \middle| \mathcal{F}_s^t \right].$$

Then letting  $s = t$  and using (2.5) with  $k_0 = 0$ , (2.41) with  $s = t$ , (2.66), (2.68) and the Hölder inequality, we obtain for any  $\delta \in (0, T - t)$ ,

$$\begin{aligned}
\|\nabla Y_t^{n,t,x}\| &\leq \mathbb{E} \left[ |g_n(X_T^{t,x})| \cdot |M_T^t(x)| \right] + \int_t^T \mathbb{E} |f_n(r, \Theta_r^{n,t,x}) M_r^t(x)| dr \\
&\leq C(T-t)^{-\beta} \|g_n(X_T^{t,x})\|_2 \\
&\quad + C \int_t^T (r-t)^{-\beta} (1 + \|X_r^{t,x}\|_2 + \|Y_r^{n,t,x}\|_2 + \|Z_r^{n,t,x}\|_2) dr \\
&\leq \frac{C(1 + |x|^q)}{(T-t)^\beta} + C(T-t)^{1-\beta} \left( 1 + \sup_{r \in [t, T]} (\|X_r^{t,x}\|_2 + \|Y_r^{n,t,x}\|_2) \right) \\
&\quad + C \left( \int_t^{t+\delta} \frac{\|\nabla u_n(r, X_r^{t,x}) \sigma(r, X_r^{t,x})\|_2}{(r-t)^\beta} dr + \delta^{-\beta} \int_{t+\delta}^T \|Z_r^{n,t,x}\|_2 dr \right).
\end{aligned}$$

For the last two terms of the right-hand side of the above inequality, according to the boundedness of  $\sigma$ , (2.7) with  $k_2 = 1$  and (2.67), we deduce that

$$\begin{aligned}
\int_t^{t+\delta} \frac{\|\nabla u_n(r, X_r^{t,x}) \sigma(r, X_r^{t,x})\|_2}{(r-t)^\beta} dr &\leq \|\sigma\|_\infty \int_t^{t+\delta} \frac{A_r^{[n]} [\mathbb{E}(1 + |X_r^{t,x}|^{qV1})^2]^{\frac{1}{2}}}{(r-t)^\beta (T-r)^\beta} dr \\
&\leq \frac{C\delta^{1-\beta} B_t^{[n]} (1 + |x|^{qV1})}{(T-t-\delta)^\beta}
\end{aligned}$$

and

$$\int_{t+\delta}^T \|Z_r^{n,t,x}\|_2 dr \leq (T-t-\delta)^{\frac{1}{2}} \left( \int_t^T \mathbb{E} |Z_r^{n,t,x}|^2 dr \right)^{\frac{1}{2}} \leq CT^{\frac{1}{2}} (1 + |x|^q).$$

In addition, due to (2.7) with  $k_2 = 1$  and again (2.67), it is easily seen that

$$\sup_{r \in [t, T]} (\|X_r^{t,x}\|_2 + \|Y_r^{n,t,x}\|_2) \leq C(1 + |x|^{qV1}).$$

Thus, there exists a positive constant  $C$  independent of  $n$  and  $t$  such that

$$\|\nabla Y_t^{n,t,x}\| \leq C \left( \frac{1}{(T-t)^\beta} + 1 + \frac{\delta^{1-\beta} B_t^{[n]}}{(T-t-\delta)^\beta} + \delta^{-\beta} \right) (1 + |x|^{qV1}).$$

Consequently, we get

$$\begin{aligned}
A_t^{[n]} &= (T-t)^\beta \sup_{x \in \mathbb{R}^d} \frac{\|\nabla Y_t^{n,t,x}\|}{1 + |x|^{qV1}} \\
&\leq C \left( 1 + (T-t)^\beta + \frac{(T-t)^\beta \delta^{1-\beta} B_t^{[n]}}{(T-t-\delta)^\beta} + \delta^{-\beta} (T-t)^\beta \right) \\
&\leq C \left( 1 + \frac{(T-t)^\beta \delta^{1-\beta} B_t^{[n]}}{(T-t-\delta)^\beta} + \delta^{-\beta} (T-t)^\beta \right).
\end{aligned}$$

Letting  $\delta = \frac{T-t}{m}$  with some  $m \geq 2$  chosen later, we have

$$A_t^{[n]} \leq C \left( 1 + m^\beta + 2^\beta (T-t)^{1-\beta} m^{-(1-\beta)} B_t^{[n]} \right)$$

$$\leq \tilde{C}(1 + m^\beta + m^{-(1-\beta)} B_t^{[n]})$$

for some  $\tilde{C} > 0$  independent of  $n$  and  $t$ . Here we have used  $\frac{T-t}{T-t-\delta} \leq 2$ . Since  $B_t^{[n]}$  is non-increasing with respect to  $t$ , it follows that

$$B_t^{[n]} \leq \tilde{C}(1 + m^\beta + m^{-(1-\beta)} B_t^{[n]}).$$

Taking  $m$  such that  $\tilde{C}m^{-(1-\beta)} < 1$ , we arrive at

$$\sup_{n \geq 1} B_t^{[n]} \leq \frac{\tilde{C}(1 + m^\beta)}{1 - \tilde{C}m^{-(1-\beta)}} < \infty, \quad (2.70)$$

which means that  $B_t^{[n]}$  has an upper bound independent of  $n$ . Combining this with (2.68) and (2.70), we may find a constant  $C > 0$  independent of  $n$  such that for any  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} |Z_r^{n,t,x}| &\leq |\nabla u_n(r, X_r^{t,x})| \cdot \|\sigma(r, X_r^{t,x})\| \\ &\leq \frac{\|\sigma\|_\infty B_t^{[n]}(1 + |X_r^{t,x}|^{q\nu 1})}{(T-r)^\beta} \leq \frac{C(1 + |X_r^{t,x}|^{q\nu 1})}{(T-r)^\beta}, \quad dr \otimes d\mathbb{P}\text{-a.e.} \end{aligned} \quad (2.71)$$

Since (2.65) ensures that  $Z_r^{n,t,x}$  converges to  $Z_r^{t,x}$  in the measure  $dr \otimes d\mathbb{P}$ , we have

$$|Z_r^{t,x}| \leq \frac{C(1 + |X_r^{t,x}|^{q\nu 1})}{(T-r)^\beta}, \quad dr \otimes d\mathbb{P}\text{-a.e.}, \quad x \in \mathbb{R}^d, \quad (2.72)$$

which, along with (2.71), is the desired analogue of (2.69) for  $Z_r^{n,t,x}$  and  $Z_r^{t,x}$ .

By (2.41), (2.69) and (2.71), we obtain as in (2.52) that

$$\left\| \mathbb{E} \left[ \sup_{n \geq 1} |f(r, \Theta_r^{n,t,x}) M_r^{t,s}(x, v)| \middle| \mathcal{F}_s^t \right] \right\|_p \leq \frac{C}{(T-r)^\beta (r-s)^\beta} (1 + |x|^{q\nu 1})(1 + |x|^{k_3})|v|,$$

which implies that (2.53) holds in the present case due to  $\beta < 1$ . Consequently, (2.54), (2.55) and (2.56) are also true. Using (2.3), (2.69) and (2.72) and following the same argument as in assertion (1), one obtains that (2.57) holds in  $L^p(\mathbb{P})$  for  $p \in [1, 2)$ . This completes our proof.  $\square$

## 2.4 Examples

In this subsection, we present some examples to illustrate the above results. As preparation we first state a lemma for verifying (2.6) that holds for any  $\psi \in \mathcal{B}_b(\mathbb{R}^d)$   $\mathbb{P}$ -a.s. and (2.39) in the case of  $M_r^t(x)$  being given by a stochastic integral, whose proof is elementary and therefore postponed to the Appendix.

We set, for  $T \geq r \geq t \geq 0$ ,

$$\begin{aligned} \tilde{W}(r-t) &= W(r) - W(t), \\ W_{r,m}^{t,0} &= \{\gamma \in C([t, r], \mathbb{R}^m) \mid \gamma(t) = 0\}, \end{aligned}$$

and let  $\tilde{W}([t, r])$  denote the path of  $\tilde{W}(\cdot - t)$  on  $[t, r]$ . Let

$$H_{t,r} : [t, r] \times \mathbb{R}^d \times W_{r,m}^{t,0} \times \mathbb{R}^d \times \mathbb{R}^d \otimes \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^m$$

be measurable and let  $\{H_{t,r}(\theta, x, \tilde{W}([t, r]), X_\theta^{t,x}, \nabla X_\theta^{t,x})\}_{\theta \in [t, r]}$  be an adapted process. We suppose that  $\{M_r^t(x)\}_{0 \leq t < r \leq T}$  in (H2) has the following form:

$$M_r^t(x) = \int_t^r H_{t,r}(\theta, x, \tilde{W}([t, r]), X_\theta^{t,x}, \nabla X_\theta^{t,x}) dW_\theta. \quad (2.73)$$

**Lemma 2.6.** *Assume that the system (2.28) and (2.30) has a unique nonexplosive strong solution and satisfies the moment conditions (2.7) and (2.31). Assume moreover that there exist positive constants  $q_1, q_2$  and  $C_{q_1, q_2, T}$  such that for all  $0 \leq t \leq s \leq \theta < r \leq T$ ,*

$$\begin{aligned} & \left\| H_{s,r} \left( \theta, X_s^{t,x}, \tilde{W}([s, r]), X_\theta^{s, X_s^{t,x}}, \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right) \right\| \\ & \leq C_{q_1, q_2, T} \left( K_\theta(t, s, r, x) + |X_\theta^{t,x}|^{q_1} + \left\| \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right\|^{q_2} \right), \end{aligned} \quad (2.74)$$

where  $\{K_\theta(t, s, r, x)\}_{s \leq \theta < r}$  is an adapted process depending on  $t, s, r, x$  such that for any  $t \leq s < r$ ,  $\mathbb{E} \int_s^r K_\theta^2(t, s, r, \cdot) d\theta$  has polynomial growth. Then for all  $0 \leq t \leq s < r \leq T$ , (2.6) holds for any  $\psi \in \mathcal{B}_b(\mathbb{R}^d)$   $\mathbb{P}$ -a.s. and  $K(t, s, r, \cdot)$  has polynomial growth. If in addition,

$$\int_s^T \sup_{|x-x_0| \leq R} \left( \mathbb{E} \int_s^r K_\theta^2(t, s, r, x) d\theta \right)^{\frac{1}{2}} dr < \infty, \quad 0 \leq t \leq s \leq T, \quad (2.75)$$

then (2.39) holds.

The first example is an FBSDE with non-degenerate forward SDE.

**Example 2.7.** *Consider the FBSDE (2.28)-(2.29). Assume that for any  $r \in [0, T]$ ,  $b(r, \cdot)$  and  $\sigma(r, \cdot)$  are continuously differentiable with*

$$\sup_{r \in [0, T], x \in \mathbb{R}^d} (\|\nabla b(r, x)\| + \|\nabla \sigma(r, x)\|) < \infty, \quad (2.76)$$

and that  $\sigma$  is non-degenerate, i.e.  $\sigma\sigma^*$  is invertible and

$$\sup_{r \in [0, T], x \in \mathbb{R}^d} \|(\sigma^*(\sigma\sigma^*)^{-1})(r, x)\| < \infty.$$

Assume moreover that (A2) holds and one of the following assumptions is satisfied:

(i) (A1) holds.

(ii) (A1') holds and  $\sigma$  is bounded on  $[0, T] \times \mathbb{R}^d$ .

Then the gradient type formula (2.40) in Theorem 2.3 holds for

$$M_r^{t,s}(x_0, v) = \int_s^r \frac{1}{r-s} \left\langle (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{t,x_0}) \nabla_v X_\theta^{t,x_0}, dW_\theta \right\rangle.$$

*Proof.* By (2.76), we know that the system (2.28) and (2.30) has a unique strong solution and the moment conditions (2.7) and (2.31) hold with  $k_2 = 1$  and  $k_3 = 0$ , respectively. It is readily checked that (C1) and (C2) hold.

Owing to the condition that  $\sigma$  is non-degenerate and (2.76) again, it is well known that the derivative formula (2.2) holds with  $\langle M_r^t(x), v \rangle$  given as follows (see, e.g., [14, Theorem 2.1]):

$$\langle M_r^t(x), v \rangle = \int_t^r \frac{1}{r-t} \left\langle (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{t,x}) \nabla_v X_\theta^{t,x}, dW_\theta \right\rangle, \quad v, x \in \mathbb{R}^d, \quad 0 \leq t < r \leq T. \quad (2.77)$$

Moreover, since  $\nabla b(t, \cdot)$  and  $\nabla \sigma(t, \cdot)$  are bounded and continuous, we can derive from (2.2) that  $\nabla P_{t,r} \psi \in C(\mathbb{R}^d)$  for any  $\psi \in C_b^1(\mathbb{R}^d)$ . Hence (H2) holds.

Next, we shall apply Lemma 2.6 to verify (2.6) and (2.39). First note that for any  $x \in \mathbb{R}^d$  and  $0 \leq t < r \leq T$ , by (2.77) we can write  $M_r^t(x)$  as follows:

$$M_r^t(x) = \int_t^r H_{t,r}(\theta, x, \tilde{W}([t, r]), X_\theta^{t,x}, \nabla X_\theta^{t,x}) dW_\theta, \quad (2.78)$$

where

$$H_{t,r}(\theta, x, \tilde{W}([t, r]), X_\theta^{t,x}, \nabla X_\theta^{t,x}) = \frac{1}{r-t} \left( (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{t,x}) \nabla X_\theta^{t,x} \right)^*.$$

Then we deduce that for any  $0 \leq t \leq s \leq \theta < r \leq T$ ,

$$\begin{aligned} & \left\| H_{s,r} \left( \theta, X_s^{t,x}, \tilde{W}([s, r]), X_\theta^{s, X_s^{t,x}}, \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right) \right\| \\ &= \frac{1}{r-s} \left\| \left( (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{s, X_s^{t,x}}) \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right)^* \right\| \\ &\leq \frac{1}{r-s} \sup_{\theta \in [s, r], x \in \mathbb{R}^d} \left\| (\sigma^*(\sigma\sigma^*)^{-1})(\theta, x) \right\| \cdot \left\| \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right\| \\ &=: K_\theta(t, s, r, x). \end{aligned}$$

Combining this with (2.78) yields that for any  $x \in \mathbb{R}^d$  and  $0 \leq t \leq s < r \leq T$ ,

$$\begin{aligned} K^2(t, s, r, x) &= \mathbb{E} |M_r^s(X_s^{t,x})|^2 \\ &= \mathbb{E} \left| \int_s^r H_{s,r} \left( \theta, X_s^{t,x}, \tilde{W}([s, r]), X_\theta^{s, X_s^{t,x}}, \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right) dW_\theta \right|^2 \\ &= \mathbb{E} \int_s^r \left\| H_{s,r} \left( \theta, X_s^{t,x}, \tilde{W}([s, r]), X_\theta^{t,x}, \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right) \right\|^2 d\theta \\ &\leq \mathbb{E} \int_s^r K_\theta^2(t, s, r, x) d\theta \leq \frac{C_{\sigma, T}}{r-s} \sup_{\theta \in [s, r], y \in \mathbb{R}^d} \mathbb{E} \|\nabla X_\theta^{s,y}\|^2 \leq \frac{C_{b, \sigma, T}}{r-s}, \end{aligned} \quad (2.79)$$

where the last inequality is due to (2.31) with  $k_1 = 2$  and  $k_3 = 0$ . Hence, by Lemma 2.6 we get (2.6) and (2.39). In addition,  $K(t, s, r, \cdot)$  satisfies (2.5) with  $k_0 = 0$  and (2.41) holds with  $\beta = 1/2$ .

Note that by the continuity and the non-degeneracy of  $\sigma$  and the continuity of  $\nabla b(r, \cdot)$  and  $\nabla \sigma(r, \cdot)$ , it is routine to show that (H3) holds. Moreover, we have

$$\begin{aligned} M_r^{t,s}(x, v) &= \langle M_r^s(X_s^{t,x}), \nabla_v X_s^{t,x} \rangle \\ &= \left\langle \int_s^r \frac{1}{r-s} \left( (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{s, X_s^{t,x}}) \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right)^* dW_\theta, \nabla_v X_s^{t,x} \right\rangle \end{aligned}$$



$$\begin{aligned}
&= \int_s^r \frac{1}{r-s} \left\langle (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{s, X_s^{t,x}}) \nabla_{\nabla_v X_s^{t,x}} X_\theta^{s,y} \Big|_{y=X_s^{t,x}}, dW_\theta \right\rangle \\
&= \frac{1}{r-s} \int_s^r \left\langle (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{t,x}) \nabla_v X_\theta^{t,x}, dW_\theta \right\rangle. \tag{2.80}
\end{aligned}$$

Therefore, taking into account the assumptions (i) and (ii), the assertions follow from Theorem 2.3 (1) and (2), respectively.  $\square$

**Remark 2.2.** For the model studied in Example 2.7, two other works [23, 16] give some partial papers related to our gradient type formulas. In [23, Theorem 4.2] dealing with the case of  $l = 1$  and  $m = d$ , the authors aim to provide a representation formula for  $Z$ . Their proof, however, actually leads to the same gradient type formula as that of Example 2.7, when  $g \in C_b^1(\mathbb{R}^d)$  and  $f \in C_b^{0,1}([0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d)$ . In [16, Theorem 3.10], under some regularity conditions on the coefficients the authors proved the following ‘‘Bismut type’’ formula:

$$\mathbb{E} \nabla_v Y_s^{t,x_0} = \mathbb{E} \left[ g(X_T^{t,x_0}) \widetilde{M}_T^{t,s}(x_0, v) + \int_s^T f(r, \Theta_r^{t,x_0}) \widetilde{M}_r^{t,s}(x_0, v) dr \right],$$

where  $\widetilde{M}_r^{t,s}(x_0, v) = \frac{1}{r-t} \int_t^r \langle \sigma^{-1}(\theta, X_\theta^{t,x_0}) \nabla_v X_\theta^{t,x_0}, dW_\theta \rangle$ , which is slightly different from  $M_r^{t,s}(x_0, v)$  in Example 2.7. Both papers use the same arguments, which consist of utilizing a relationship between the Malliavin derivatives  $(DX, DY, DZ)$  and partial derivatives in the initial value  $x$  of  $(X, Y, Z)$  (see, e.g., [30, Lemma 2.4] and [23, Lemma 2.4]), together with the Malliavin integration by parts formula. In comparison with [23, 16], the gradient type formula in Example 2.7 is obtained under more general conditions, and moreover our method is more powerful and flexible, which is illustrated by Examples 2.8, 2.9 and 2.10 below. We also point out that under non-degenerate assumptions which additionally require  $\sigma$  to be independent of  $x$ , [24] extended the above ‘‘Bismut type’’ formula stated in [16, Theorem 3.10] to the case of  $f$  with quadratic growth with respect to  $Z$ .

For the FBSDE stated in Example 2.7, we can provide an alternative version of the gradient type formula.

**Example 2.8.** Let the same assumptions as for Example 2.7 hold. For each  $0 \leq t < s \leq T$ , define

$$\varrho_{t,s}(\theta) := c^{-1} \left( 1 - e^{-c(s-\theta)} \right), \quad \theta \in [t, s],$$

with some positive constant  $c$  depending only on  $b$  and  $\sigma$ , and let  $\{G_s^{t,x}(r)\}_{r \in [t,s]}$  satisfy the linear equation on  $\mathbb{R}^d \otimes \mathbb{R}^d$ :

$$G_s^{t,x}(r) = I_{d \times d} + \int_t^r \left( \nabla b(\theta, X_\theta^{t,x}) G_s^{t,x}(\theta) - \frac{G_s^{t,x}(\theta)}{\varrho_{t,s}(\theta)} \right) d\theta + \int_t^r \nabla \sigma(\theta, X_\theta^{t,x}) G_s^{t,x}(\theta) dW_\theta, \tag{2.81}$$

where  $(\int_t^r \nabla \sigma(\theta, X_\theta^{t,x}) G_s^{t,x}(\theta) dW_\theta) v := \int_t^r \nabla_{G_s^{t,x}(\theta)v} \sigma(\theta, X_\theta^{t,x}) dW_\theta$ ,  $v \in \mathbb{R}^d$ . Then the gradient type formula (2.40) in Theorem 2.3 holds for

$$M_r^{t,s}(x_0, v) = \int_s^r \frac{1}{\varrho_{s,r}(\theta)} \left\langle (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{t,x_0}) G_r^{s, X_s^{t,x_0}}(\theta) \nabla_v X_s^{t,x_0}, dW_\theta \right\rangle. \tag{2.82}$$

*Proof.* For any  $0 \leq t < s \leq T$ , we first introduce the following linear equation on  $\mathbb{R}^d$ :

$$\begin{aligned} \Gamma_s^{t,x}(r) &= v + \int_t^r \left( \nabla_{\Gamma_s^{t,x}(\theta)} b(\theta, X_\theta^{t,x}) - \frac{\Gamma_s^{t,x}(\theta)}{\varrho_{t,s}(\theta)} \right) d\theta \\ &\quad + \int_t^r \nabla_{\Gamma_s^{t,x}(\theta)} \sigma(\theta, X_\theta^{t,x}) dW_\theta, \quad r \in [t, s]. \end{aligned}$$

Obviously, our assumptions imply that the equation has a unique solution  $\{\Gamma_s^{t,x}(r)\}_{r \in [t,s]}$ , and moreover,  $\Gamma_s^{t,x}(r) = G_s^{t,x}(r)v$  for any  $r \in [t, s]$ . According to [32, Theorem 3.5.1 or Theorem 4.3.7], the derivative formula (2.2) holds with  $\langle M_r^t(x), v \rangle$  given as follows:

$$\langle M_r^t(x), v \rangle = \int_t^r \frac{1}{\varrho_{t,r}(\theta)} \left\langle (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{t,x}) \Gamma_r^{t,x}(\theta), dW_\theta \right\rangle, \quad v, x \in \mathbb{R}^d, \quad 0 \leq t < r \leq T.$$

This, along with  $\Gamma_r^{t,x}(\theta) = G_r^{t,x}(\theta)v$  for  $\theta \in [t, r]$ , implies that for any  $x \in \mathbb{R}^d$  and  $0 \leq t < r \leq T$ ,

$$M_r^t(x) = \int_t^r H_{t,r}(\theta, x, \tilde{W}([t, r]), X_\theta^{t,x}, \nabla X_\theta^{t,x}) dW_\theta,$$

where

$$H_{t,r}(\theta, x, \tilde{W}([t, r]), X_\theta^{t,x}, \nabla X_\theta^{t,x}) = \frac{1}{\varrho_{t,r}(\theta)} \left( (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{t,x}) G_r^{t,x}(\theta) \right)^*.$$

Then we get that for any  $0 \leq t \leq s \leq \theta < r \leq T$ ,

$$\begin{aligned} &\left\| H_{s,r} \left( \theta, X_s^{t,x}, \tilde{W}([s, r]), X_\theta^{s, X_s^{t,x}}, \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right) \right\| \\ &= \frac{1}{\varrho_{s,r}(\theta)} \left\| \left( (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{s, X_s^{t,x}}) G_r^{s, X_s^{t,x}}(\theta) \right)^* \right\| \\ &\leq \frac{1}{\varrho_{s,r}(\theta)} \sup_{\theta \in [s,r], x \in \mathbb{R}^d} \left\| (\sigma^*(\sigma\sigma^*)^{-1})(\theta, x) \right\| \cdot \left\| G_r^{s, X_s^{t,x}}(\theta) \right\| \\ &=: K_\theta(t, s, r, x). \end{aligned}$$

Similar to (2.79) in Example 2.7, we have

$$\begin{aligned} K^2(t, s, r, x) &= \mathbb{E} |M_r^s(X_s^{t,x})|^2 \leq \mathbb{E} \int_s^r K_\theta^2(t, s, r, x) d\theta \\ &\leq C_{\sigma, T} \mathbb{E} \int_s^r \frac{\|G_r^{s, X_s^{t,x}}(\theta)\|^2}{\varrho_{s,r}^2(\theta)} d\theta \\ &\leq C_{b, \sigma, T} \left( 1 + \frac{1}{1 - e^{-c(r-s)}} \right), \end{aligned} \quad (2.83)$$

where the last inequality is due to [32, Theorem 3.5.1 and (3.65), or Theorem 4.3.7 and (4.51)]. Then, (2.6) and (2.39) follow from Lemma 2.6 and  $K(t, s, r, \cdot)$  satisfies (2.5) with  $k_0 = 0$ . Besides, a direct and easy computation shows that (2.41) also holds with  $\beta = 1/2$  as in Example 2.7.

Finally, we are to verify that (H3) holds and  $M_r^{t,s}(x, v)$  is given by (2.82). As in (2.80) of Example 2.7, we obtain

$$\begin{aligned} M_r^{t,s}(x, v) &= \langle M_r^s(X_s^{t,x}), \nabla_v X_s^{t,x} \rangle \\ &= \left\langle \int_s^r \frac{1}{\varrho_{s,r}(\theta)} \left( (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{s, X_s^{t,x}}) G_r^{s, X_s^{t,x}}(\theta) \right)^* dW_\theta, \nabla_v X_s^{t,x} \right\rangle \\ &= \int_s^r \frac{1}{\varrho_{s,r}(\theta)} \left\langle (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{t,x}) G_r^{s, X_s^{t,x}}(\theta) \nabla_v X_s^{t,x}, dW_\theta \right\rangle, \end{aligned}$$

which is consistent with (2.82). As for (H3), it is standard to prove that its first part (2.3) holds due to the continuity and non-degeneracy of  $\sigma$  and the continuity of  $\nabla b(r, \cdot)$  and  $\nabla\sigma(r, \cdot)$ . Let us now deal with its second part (2.4). Following the same arguments as in [32, Lemma 4.3.8] and [32, Theorem 4.3.7 and (4.51)], we derive that for any  $p > 1$ ,

$$\mathbb{E} \sup_{\theta \in [s, r]} \|G_r^{s, X_s^{t,x}}(\theta)\|^p \leq C_{b, \sigma, T} \quad (2.84)$$

and

$$\sup_{0 < \delta \leq \frac{r-s}{2}} \int_s^r \frac{\mathbb{E} \|G_{r-\delta}^{s, X_s^{t,x}}(\theta)\|^p}{\varrho_{s, r-\delta}^p(\theta)} \mathbb{1}_{[s, r-\delta]}(\theta) d\theta \leq \sup_{0 < \delta \leq \frac{r-s}{2}} C_{b, \sigma, T} \left( 1 + \frac{1}{1 - e^{-c(r-\delta-s)}} \right) < \infty, \quad (2.85)$$

respectively. By (2.81), we have for  $\theta \in [s, r - \delta]$ ,

$$\begin{aligned} G_{r-\delta}^{s, X_s^{t,x}}(\theta) - G_r^{s, X_s^{t,x}}(\theta) &= \int_s^\theta \nabla b(\tilde{\theta}, X_{\tilde{\theta}}^{s, X_s^{t,x}}) \left( G_{r-\delta}^{s, X_s^{t,x}}(\tilde{\theta}) - G_r^{s, X_s^{t,x}}(\tilde{\theta}) \right) d\tilde{\theta} \\ &\quad - \int_s^\theta \left( \frac{G_{r-\delta}^{s, X_s^{t,x}}(\tilde{\theta})}{\varrho_{s, r-\delta}(\tilde{\theta})} - \frac{G_r^{s, X_s^{t,x}}(\tilde{\theta})}{\varrho_{s, r}(\tilde{\theta})} \right) d\tilde{\theta} \\ &\quad + \int_s^\theta \nabla\sigma(\tilde{\theta}, X_{\tilde{\theta}}^{s, X_s^{t,x}}) \left( G_{r-\delta}^{s, X_s^{t,x}}(\tilde{\theta}) - G_r^{s, X_s^{t,x}}(\tilde{\theta}) \right) dW_{\tilde{\theta}}. \end{aligned}$$

Applying the Itô formula and using the boundedness of  $\nabla b$  and  $\nabla\sigma$  lead to

$$\begin{aligned} &\mathbb{E} \|G_{r-\delta}^{s, X_s^{t,x}}(\theta) - G_r^{s, X_s^{t,x}}(\theta)\|^2 \\ &\leq C_{b, \sigma} \mathbb{E} \int_s^\theta \|G_{r-\delta}^{s, X_s^{t,x}}(\tilde{\theta}) - G_r^{s, X_s^{t,x}}(\tilde{\theta})\|^2 d\tilde{\theta} - 2 \mathbb{E} \int_s^\theta \frac{\|G_{r-\delta}^{s, X_s^{t,x}}(\tilde{\theta}) - G_r^{s, X_s^{t,x}}(\tilde{\theta})\|^2}{\varrho_{s, r-\delta}(\tilde{\theta})} d\tilde{\theta} \\ &\quad + 2 \mathbb{E} \int_s^\theta \left( \frac{1}{\varrho_{s, r}(\tilde{\theta})} - \frac{1}{\varrho_{s, r-\delta}(\tilde{\theta})} \right) \left\langle G_{r-\delta}^{s, X_s^{t,x}}(\tilde{\theta}) - G_r^{s, X_s^{t,x}}(\tilde{\theta}), G_r^{s, X_s^{t,x}}(\tilde{\theta}) \right\rangle d\tilde{\theta} \\ &\leq C_{b, \sigma} \int_s^\theta \mathbb{E} \|G_{r-\delta}^{s, X_s^{t,x}}(\tilde{\theta}) - G_r^{s, X_s^{t,x}}(\tilde{\theta})\|^2 d\tilde{\theta} \\ &\quad + \int_s^\theta \left( \frac{1}{\varrho_{s, r}(\tilde{\theta})} - \frac{1}{\varrho_{s, r-\delta}(\tilde{\theta})} \right)^2 \mathbb{E} \|G_r^{s, X_s^{t,x}}(\tilde{\theta})\|^2 d\tilde{\theta}, \quad \theta \in [s, r - \delta], \end{aligned}$$

where the second inequality is due to the positivity of  $\varrho_{s,r-\delta}(\cdot)$  and the Young inequality. The Gronwall inequality implies that for any  $\theta \in [s, r - \delta)$ ,

$$\mathbb{E}\|G_{r-\delta}^{s, X_s^{t,x}}(\theta) - G_r^{s, X_s^{t,x}}(\theta)\|^2 \leq C_{b,\sigma,T} \int_s^\theta \left( \frac{1}{\varrho_{s,r}(\tilde{\theta})} - \frac{1}{\varrho_{s,r-\delta}(\tilde{\theta})} \right)^2 \mathbb{E}\|G_r^{s, X_s^{t,x}}(\tilde{\theta})\|^2 d\tilde{\theta}.$$

Then using (2.84) and taking into account the fact that  $\lim_{\delta \rightarrow 0^+} \varrho_{s,r-\delta}(\tilde{\theta}) = \varrho_{s,r}(\tilde{\theta})$ , by the dominated convergence theorem we obtain that for every  $r' \in [s, r)$ ,

$$\lim_{\delta \rightarrow 0^+} \sup_{\delta < r-r'} \sup_{s \leq \theta \leq r'} \mathbb{E}\|G_{r-\delta}^{s, X_s^{t,x}}(\theta) - G_r^{s, X_s^{t,x}}(\theta)\|^2 = 0,$$

Consequently, we obtain that for every  $\theta \in [s, r)$ ,

$$\mathbb{P}\text{-}\lim_{\delta \rightarrow 0^+} \left( \frac{G_{r-\delta}^{s, X_s^{t,x}}(\theta)}{\varrho_{s,r-\delta}(\theta)} \mathbf{1}_{[s,r-\delta)}(\theta) - \frac{G_r^{s, X_s^{t,x}}(\theta)}{\varrho_{s,r}(\theta)} \right) = 0.$$

Note that the term involving  $\delta$  above is well defined as long as  $\delta$  is small enough. Thus, combining this with the boundedness of  $\sigma^*(\sigma\sigma^*)^{-1}$  and (2.85), we apply the dominated convergence theorem to get

$$\begin{aligned} & \lim_{\delta \rightarrow 0^+} \mathbb{E} \left| M_{r-\delta}^s(X_s^{t,x}) - M_r^s(X_s^{t,x}) \right|^2 \\ &= \lim_{\delta \rightarrow 0^+} \mathbb{E} \left| \int_s^r \frac{1}{\varrho_{s,r-\delta}(\theta)} \left( (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{s, X_s^{t,x}}) G_{r-\delta}^{s, X_s^{t,x}}(\theta) \right)^* \mathbf{1}_{[s,r-\delta)}(\theta) dW_\theta \right. \\ & \quad \left. - \int_s^r \frac{1}{\varrho_{s,r}(\theta)} \left( (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{s, X_s^{t,x}}) G_r^{s, X_s^{t,x}}(\theta) \right)^* dW_\theta \right|^2 \\ &= \lim_{\delta \rightarrow 0^+} \mathbb{E} \int_s^r \left\| (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{s, X_s^{t,x}}) \left( \frac{G_{r-\delta}^{s, X_s^{t,x}}(\theta)}{\varrho_{s,r-\delta}(\theta)} \mathbf{1}_{[s,r-\delta)}(\theta) - \frac{G_r^{s, X_s^{t,x}}(\theta)}{\varrho_{s,r}(\theta)} \right) \right\|^2 d\theta \\ &= 0, \end{aligned}$$

which means that (2.4) holds.

Therefore, similar to Example 2.7, the assertions follow from Theorem 2.3 (1) and (2), respectively.  $\square$

We may also consider the following FBSDE with forward Grushin type process.

**Example 2.9.** Let  $d = d_1 + d_2$ ,  $x = (x^{(1)}, x^{(2)})$  and  $W = (W^{(1)}, W^{(2)})$  be a Brownian motion on  $\mathbb{R}^{d_1+d_2}$ , and consider the FBSDE (2.28)-(2.29), where the forward SDE (2.28) is of the following form:

$$\begin{cases} X_s^{(1),t,x} = x^{(1)} + \int_t^s dW_r^{(1)}, & s \in [t, T], \\ X_s^{(2),t,x} = x^{(2)} + \int_t^s \sigma(X_r^{(1),t,x}) dW_r^{(2)}, & s \in [t, T], \end{cases} \quad (2.86)$$

with  $\sigma \in C^1(\mathbb{R}^{d_1}; \mathbb{R}^{d_2} \otimes \mathbb{R}^{d_2})$  which might be degenerate. Assume that there exist  $\alpha \in [1, \frac{d_1}{2} + 1)$  and constants  $a_1, a_2 > 0$  such that

$$\|\sigma(x)\| \geq a_1|x|^\alpha, \quad \|\sigma(x)\| + \|\nabla\sigma(x)\| \cdot |x| \leq a_2|x|^\alpha, \quad x \in \mathbb{R}^{d_1}, \quad (2.87)$$

and that (A1) and (A2) are satisfied. Then the gradient type formula (2.40) in Theorem 2.3 holds for

$$\begin{aligned}
M_r^{t,s}(x_0, v) &= \frac{\langle v^{(1)}, W_r^{(1)} - W_s^{(1)} \rangle}{r-s} \\
&- \text{Tr} \left( Q_{s,r}^{-1}(x_0^{(1)} + W_s^{(1)} - W_t^{(1)}) \int_s^r \frac{r-\theta}{r-s} ((\nabla_{v^{(1)}} \sigma) \sigma^*)(x_0^{(1)} + W_\theta^{(1)} - W_t^{(1)}) d\theta \right) \\
&+ \left\langle Q_{s,r}^{-1}(x_0^{(1)} + W_s^{(1)} - W_t^{(1)}) \left[ \nabla_v X_s^{(2),t,x_0} + \int_s^r \frac{r-\theta}{r-s} \nabla_{v^{(1)}} \sigma(x_0^{(1)} + W_\theta^{(1)} - W_t^{(1)}) dW_\theta^{(2)} \right], \right. \\
&\quad \left. \int_s^r \sigma(x_0^{(1)} + W_\theta^{(1)} - W_t^{(1)}) dW_\theta^{(2)} \right\rangle
\end{aligned}$$

provided that  $x_0^{(1)} \neq 0$  when  $s = t$ . Here  $v = (v^{(1)}, v^{(2)}) \in \mathbb{R}^{d_1+d_2}$  and

$$Q_{s,r}(y) := \int_s^r (\sigma \sigma^*)(y + W_\theta^{(1)} - W_s^{(1)}) d\theta, \quad y \in \mathbb{R}^{d_1}. \quad (2.88)$$

*Proof.* By (2.86) and (2.87), one can show that for any  $k \geq 1$ ,

$$\sup_{0 \leq t \leq s \leq T} \mathbb{E} |X_s^{t,x}|^k \leq C_{k,\alpha,a_2,T} \left( 1 + |x|^k + |x^{(1)}|^{k\alpha} \right)$$

and

$$\sup_{0 \leq t \leq s \leq T} \mathbb{E} |\nabla X_s^{t,x}|^k \leq C_{k,\alpha,a_2,T} \left( 1 + |x^{(1)}|^{k(\alpha-1)} \right),$$

which imply that the moment conditions (2.7) and (2.31) hold with  $k_2 = \alpha$  and  $k_3 = \alpha - 1$ , respectively. Moreover, (2.32) and (2.33) in (C1) follow easily from (2.86) and (2.87). Hence (C1) holds.

Now, we are to verify that (C2) holds. Let

$$\phi_n(x) := \left[ 1 + (|x| - n)^+ \right]^{-\frac{1}{2}}, \quad x \in \mathbb{R}^{d_1}, \quad n \geq 1,$$

and  $\sigma_n(x) := \sigma(\phi_n(x)x)$ . Then  $\phi_n(x) = 1$  for any  $|x| \leq n$ , and

$$\lim_{n \rightarrow +\infty} \sup_{|x| \leq R} \|\sigma(\phi_n(x)x) - \sigma(x)\| = 0, \quad R > 0. \quad (2.89)$$

Moreover, we have  $\nabla \sigma_n(\cdot) \in C(\mathbb{R}^{d_1}; \mathbb{R}^{d_2} \otimes \mathbb{R}^{d_2} \otimes \mathbb{R}^{d_1})$  with

$$\begin{aligned}
\nabla \sigma_n(x) &= \nabla \sigma(\phi_n(x)x) (\nabla \phi_n(x) \otimes x + \phi_n(x) I_{d_1 \times d_1}) \\
&= \nabla \sigma(\phi_n(x)x) \left[ -\frac{(|x| - n)^+ (x \otimes x)}{\left( 1 + (|x| - n)^+ \right)^{\frac{3}{2}} |x|} + \phi_n(x) I_{d_1 \times d_1} \right]. \quad (2.90)
\end{aligned}$$

Consequently, by (2.87) we get

$$\|\nabla \sigma_n\|_\infty \leq a_2 \sup_{x \in \mathbb{R}^{d_1}} \left[ \left( \frac{(|x| - n)^+ |x|}{\left( 1 + (|x| - n)^+ \right)^{\frac{3}{2}}} + 1 \right) |\phi_n(x)x|^{\alpha-1} \right] \leq a_2 (n+2)^\alpha,$$

i.e. (2.34) holds. For any  $(x_n)_{n \geq 1}$  with  $\lim_{n \rightarrow +\infty} x_n = x$ , let  $(X_s^{n,t,x_n})_{s \in [t,T]}$  be the solution of the system (2.86) with  $\sigma$  replaced by  $\sigma_n$ . Since  $X_s^{(1),n,t,x_n} - X_s^{(1),t,x} = x_n^{(1)} - x^{(1)}$  and  $\nabla X_s^{(1),n,t,x_n} = \nabla X_s^{(1),t,x}$ , we only need to check (2.35) and (2.36) for  $X_s^{(2),n,t,x}$ . By the Itô formula and the B-D-G inequality, we have

$$\begin{aligned} & \mathbb{E} \sup_{s \in [t,T]} \left| X_s^{(2),n,t,x_n} - X_s^{(2),t,x} \right|^2 \\ & \leq C \left( \left| x_n^{(2)} - x^{(2)} \right|^2 + \int_t^T \mathbb{E} \left\| \sigma_n(X_r^{(1),n,t,x_n}) - \sigma(X_r^{(1),t,x}) \right\|^2 dr \right). \end{aligned} \quad (2.91)$$

By (2.89) and the continuity of  $\sigma$ , we obtain that for each  $r \in [t, T]$ ,

$$\begin{aligned} & \lim_{n \rightarrow +\infty} \left\| \sigma_n(X_r^{(1),n,t,x_n}) - \sigma(X_r^{(1),t,x}) \right\| \\ & = \lim_{n \rightarrow +\infty} \left\| \sigma_n(x_n^{(1)} + W_r^{(1)} - W_t^{(1)}) - \sigma(x^{(1)} + W_r^{(1)} - W_t^{(1)}) \right\| \\ & \leq \lim_{n \rightarrow +\infty} \left[ \left\| \sigma_n(x_n^{(1)} + W_r^{(1)} - W_t^{(1)}) - \sigma(x_n^{(1)} + W_r^{(1)} - W_t^{(1)}) \right\| \right. \\ & \quad \left. + \left\| \sigma(x_n^{(1)} + W_r^{(1)} - W_t^{(1)}) - \sigma(x^{(1)} + W_r^{(1)} - W_t^{(1)}) \right\| \right] \\ & = 0. \end{aligned}$$

Combining this with (2.91) and applying the dominated convergence theorem, we conclude that

$$\lim_{n \rightarrow +\infty} \mathbb{E} \sup_{s \in [t,T]} \left| X_s^{(2),n,t,x_n} - X_s^{(2),t,x} \right|^2 = 0,$$

i.e. (2.35) holds. As for (2.36), owing to (2.90) and the continuity of  $\nabla \sigma$ , we derive that for each  $r \in [t, T]$ ,

$$\begin{aligned} & \lim_{n \rightarrow +\infty} \left\| \nabla \sigma_n(X_r^{(1),n,t,x_n}) - \nabla \sigma(X_r^{(1),t,x}) \right\| \\ & = \lim_{n \rightarrow +\infty} \left\| \nabla \sigma(\phi_n(y)y) (\nabla \phi_n(y) \otimes y + \phi_n(y) I_{d_1 \times d_1}) \Big|_{y=X_r^{(1),n,t,x_n}} - \nabla \sigma(X_r^{(1),t,x}) \right\| \\ & = 0. \end{aligned}$$

Then using the Itô formula and the B-D-G inequality and applying the dominated convergence theorem again, we deduce that

$$\begin{aligned} & \lim_{n \rightarrow +\infty} \mathbb{E} \sup_{s \in [t,T]} \left\| \nabla X_s^{(2),n,t,x_n} - \nabla X_s^{(2),t,x} \right\|^2 \\ & \leq \lim_{n \rightarrow +\infty} \int_t^T \mathbb{E} \left\| \nabla \sigma_n(X_r^{(1),n,t,x_n}) - \nabla \sigma(X_r^{(1),t,x}) \right\|^2 dr \\ & = 0, \end{aligned}$$

where the choice of the dominating function can be determined by (2.87) and (2.90). That is, (2.36) holds. Besides, it is clear that (2.37) holds. Hence (C2) holds.

Note that, according to [33, Lemma 3.1] and (2.87), we obtain that for any  $0 \leq t < r \leq T$ ,  $z \in \mathbb{R}^{d_1}$  and  $p > 0$ ,  $Q_{t,r}(z)$  defined in (2.88) is invertible and

$$\mathbb{E}\|Q_{t,r}^{-1}(z)\|^p \leq \frac{C}{(r-t)^p[|z|^2 + (r-t)]^{p\alpha}}. \quad (2.92)$$

By [33, Theorem 1.1 and Corollary 1.2], it follows that for any  $0 \leq t < r \leq T$  and  $x = (x^{(1)}, x^{(2)})$ ,  $v = (v^{(1)}, v^{(2)}) \in \mathbb{R}^{d_1+d_2}$ ,

$$\nabla_v P_{t,r}\psi(x^{(1)}, x^{(2)}) = \mathbb{E} \left[ \psi(X_r^{(1),t,x}, X_r^{(2),t,x}) \langle M_r^t(x), v \rangle \right], \quad \psi \in C_b^1(\mathbb{R}^{d_1+d_2})$$

with

$$\begin{aligned} \langle M_r^t(x), v \rangle &= \frac{\langle v^{(1)}, W_r^{(1)} - W_t^{(1)} \rangle}{r-t} \\ &- \text{Tr} \left( Q_{t,r}^{-1}(x^{(1)}) \int_t^r \frac{r-\theta}{r-t} ((\nabla_{v^{(1)}}\sigma)\sigma^*)(x^{(1)} + W_\theta^{(1)} - W_t^{(1)})d\theta \right) \\ &+ \left\langle Q_{t,r}^{-1}(x^{(1)}) \left[ v^{(2)} + \int_t^r \frac{r-\theta}{r-t} \nabla_{v^{(1)}}\sigma(x^{(1)} + W_\theta^{(1)} - W_t^{(1)})dW_\theta^{(2)} \right], \right. \\ &\quad \left. \int_t^r \sigma(x^{(1)} + W_\theta^{(1)} - W_t^{(1)})dW_\theta^{(2)} \right\rangle. \end{aligned}$$

Moreover, since

$$Q_{s,r}^{-1}(y) - Q_{s,r}^{-1}(x) = Q_{s,r}^{-1}(x) (Q_{s,r}(x) - Q_{s,r}(y)) Q_{s,r}^{-1}(y),$$

by the dominated convergence theorem and (2.92) we get

$$\lim_{y \rightarrow x} \mathbb{E}\|Q_{s,r}^{-1}(y) - Q_{s,r}^{-1}(x)\|^2 = 0. \quad (2.93)$$

Then we conclude that  $\nabla P_{t,r}\psi \in C(\mathbb{R}^d)$  for any  $\psi \in C_b^1(\mathbb{R}^d)$ . Hence (H2) holds.

Now, let  $\{e_i\}_{i=1}^{d_1+d_2}$  be the canonical ONB of  $\mathbb{R}^{d_1+d_2}$ . Since  $Q_{s,r}$  is independent of  $\mathcal{F}_s^t$  for any  $0 \leq t \leq s < r \leq T$ , it is obvious that for every  $e_i$  and  $y \in \mathbb{R}^{d_1+d_2}$ ,  $\langle M_r^s(y), e_i \rangle$  is also independent of  $\mathcal{F}_s^t$ . Consequently, it is readily checked that (2.6) holds due to the fact that  $M_r^s(y) = \sum_{i=1}^{d_1+d_2} \langle M_r^s(y), e_i \rangle e_i$ .

Next, we shall verify that  $K(t, s, r, \cdot)$  satisfies (2.5) and (2.39). First note that,

$$\begin{aligned} K^2(t, s, r, x) &= \mathbb{E}|M_r^s(X_s^{t,x})|^2 = \sum_{i=1}^{d_1+d_2} \mathbb{E} \left[ \left( \mathbb{E}|\langle M_r^s(X_s^{t,x}), e_i \rangle|^2 \middle| \mathcal{F}_s^t \right) \right] \\ &= \sum_{i=1}^{d_1+d_2} \mathbb{E} \left[ \left( \mathbb{E}|\langle M_r^s(y), e_i \rangle|^2 \middle|_{y=X_s^{t,x}} \right) \right]. \end{aligned} \quad (2.94)$$

Along the same lines as in the proof of [33, Corollary 1.2], we have

$$\mathbb{E}|\langle M_r^s(y), e_i \rangle|^2 \leq C \left( \frac{|e_i^{(1)}|^2}{r-s} + \frac{|e_i^{(2)}|^2}{(r-s)(|y^{(1)}|^2 + (r-s))^\alpha} \right).$$

Plugging this into (2.94), we obtain

$$K^2(t, s, r, x) \leq \frac{C}{r-s} \left( 1 + \mathbb{E} \frac{1}{\left( |x^{(1)} + W_s^{(1)} - W_t^{(1)}|^2 + (r-s) \right)^\alpha} \right). \quad (2.95)$$

Consequently, it is easy to see that  $K(t, s, r, \cdot)$  satisfies (2.5) with  $k_0 = 0$  and  $C(t, s, r) = \frac{C}{\sqrt{r-s}} \left( 1 + \frac{1}{(r-s)^{\alpha/2}} \right)$ . Since  $\alpha \in [1, \frac{d_1}{2} + 1)$ , we can take  $\beta \in (\alpha - 1, \frac{d_1}{2})$ , which implies that

$$d_1 - 1 - 2\beta > -1 \quad \text{and} \quad (1 + \alpha - \beta)/2 < 1.$$

Then by (2.95), we deduce that there exists  $R > 0$  such that

$$\begin{aligned} & \int_s^T \sup_{|x-x_0| \leq R} K(t, s, r, x) dr \\ & \leq C \int_s^T \frac{1}{\sqrt{r-s}} dr + C \int_s^T \sup_{|x-x_0| \leq R} \left( \mathbb{E} \frac{(r-s)^{-1}}{\left( |x^{(1)} + W_s^{(1)} - W_t^{(1)}|^2 + (r-s) \right)^\alpha} \right)^{\frac{1}{2}} dr \\ & \leq C + C \int_s^T \sup_{|x-x_0| \leq R} \left( \mathbb{E} \frac{(r-s)^{-1} |x^{(1)} + W_s^{(1)} - W_t^{(1)}|^{-2\beta}}{\left( |x^{(1)} + W_s^{(1)} - W_t^{(1)}|^2 + (r-s) \right)^{\alpha-\beta}} \right)^{\frac{1}{2}} dr \\ & \leq C + C \left( \sup_{|x-x_0| \leq R} \mathbb{E} |x^{(1)} + W_s^{(1)} - W_t^{(1)}|^{-2\beta} \right)^{\frac{1}{2}} \int_s^T (r-s)^{-(1+\alpha-\beta)/2} dr \\ & < \infty, \end{aligned}$$

where we have used the condition that  $x_0^{(1)} \neq 0$  when  $s = t$ . We emphasize that here we only consider  $\alpha \geq \beta$  in the second to last inequality since the integral in the second inequality is finite when  $\alpha < \beta$ . Hence (2.39) holds.

Finally, we are to verify that (2.3) holds. Owing to (2.87), (2.92), (2.93) and since  $\nabla \sigma \in C(\mathbb{R}^{d_1}, \mathbb{R}^{d_2} \otimes \mathbb{R}^{d_2} \otimes \mathbb{R}^{d_1})$ , we can apply the dominated convergence theorem to get

$$\lim_{y \rightarrow x} \mathbb{E} |M_r^s(y) - M_r^s(x)|^2 = 0. \quad (2.96)$$

Consequently, using the independence of  $\{M_r^s(x)\}_{x \in \mathbb{R}^{d_1+d_2}}$  and  $\mathcal{F}_s^t$  and applying the dominated convergence theorem again, we obtain

$$\begin{aligned} \lim_{y \rightarrow x} \mathbb{E} |M_r^s(X_s^{t,y}) - M_r^s(X_s^{t,x})|^2 &= \lim_{y \rightarrow x} \mathbb{E} \left[ \left( \mathbb{E} |M_r^s(z_1) - M_r^s(z_2)|^2 \right) \Big|_{z_1=X_s^{t,y}, z_2=X_s^{t,x}} \right] \\ &= 0. \end{aligned}$$

Indeed, for  $\varphi(z_1, z_2) := \mathbb{E} |M_r^s(z_1) - M_r^s(z_2)|^2$ , it suffices to prove that for every sequence  $(y_n)_{n \geq 1}$  with  $\lim_{n \rightarrow +\infty} y_n = x$ , there exists a subsequence  $(y_{n_k})_{k \geq 1}$  such that

$$\lim_{k \rightarrow +\infty} \mathbb{E} \varphi(X_s^{t, y_{n_k}}, X_s^{t, x}) = 0.$$

By (2.86)-(2.87) and the Itô formula, it is easy to see that  $\lim_{n \rightarrow +\infty} \mathbb{E} |X_s^{t, y_n} - X_s^{t, x}|^2 = 0$ . Then there exists a subsequence  $(y_{n_k})_{k \geq 1}$  such that  $\lim_{k \rightarrow +\infty} X_s^{t, y_{n_k}} = X_s^{t, x}$ ,  $\mathbb{P}$ -a.s.



Due to the definition of  $M_r^s(\cdot)$  and (2.92), one can show that for each  $p > 1$ ,

$$\sup_{k \geq 1} \mathbb{E} \varphi^p(X_s^{t, y^{n_k}}, X_s^{t, x}) < \infty$$

So, by the dominated convergence theorem and (2.96) we get the desired result. Hence (2.3) holds.

Therefore, the assertion follows from Theorem 2.3 (1).  $\square$

**Remark 2.3.** *It is not hard to extend the above result to the FBSDE with the following general Grushin type process:*

$$\begin{cases} X_s^{(1), t, x} = x^{(1)} + \int_t^s b_1(X_r^{(1), t, x}) dr + \int_t^s \sigma_1(X_r^{(1), t, x}) dW_r^{(1)}, & s \in [t, T], \\ X_s^{(2), t, x} = x^{(2)} + \int_t^s b_2(X_r^{(1), t, x}) dr + \int_t^s \sigma_2(X_r^{(1), t, x}) dW_r^{(2)}, & s \in [t, T], \end{cases}$$

where  $b_1 \in C_b^1(\mathbb{R}^{d_1}; \mathbb{R}^{d_1})$  and  $b_2 \in C_b^1(\mathbb{R}^{d_1}; \mathbb{R}^{d_2})$ ,  $\sigma_1 \in C_b^1(\mathbb{R}^{d_1}; \mathbb{R}^{d_1} \otimes \mathbb{R}^{d_1})$  is invertible and  $\sigma_1^{-1}$  is bounded,  $\sigma_2 \in C^1(\mathbb{R}^{d_1}; \mathbb{R}^{d_2} \otimes \mathbb{R}^{d_2})$  might be degenerate.

Now we apply Corollary 2.4 to the FBSDE with forward stochastic Hamiltonian system. Let  $d = d_1 + d_2$  and  $W$  be a  $d_2$ -dimensional Brownian motion. Consider the FBSDE (2.28)-(2.29), where the forward SDE (2.28) is of the following form:

$$\begin{aligned} X_s^{t, x} &= x + \int_t^s b(r, X_r^{t, x}) dr + \int_t^s (0, \sigma(r)) dW_r \\ &= x + \int_t^s (B X_r^{(2), t, x}, \tilde{b}(r, X_r^{(1), t, x}, X_r^{(2), t, x})) dr + \int_t^s (0, \sigma(r)) dW_r, \quad s \in [t, T], \end{aligned}$$

where  $B$  is a  $d_1 \times d_2$ -matrix with rank  $d_1$ ,  $\tilde{b} : [0, T] \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \rightarrow \mathbb{R}^{d_2}$  and  $\sigma(r)$  is an invertible  $d_2 \times d_2$ -matrix which is continuous in  $r \in [0, T]$ . Since  $B$  has rank  $d_1$ , we know that  $d_2 \geq d_1$  and for every  $y^{(1)} \in \mathbb{R}^{d_1}$ ,

$$B^{-1}y^{(1)} := \left\{ y^{(2)} \in \mathbb{R}^{d_2} : B y^{(2)} = y^{(1)} \right\} \neq \emptyset.$$

We set

$$|B^{-1}y^{(1)}| := \inf \left\{ |y^{(2)}| : y^{(2)} \in B^{-1}y^{(1)} \right\}.$$

Then it is easy to show that

$$\|B^{-1}\| := \sup \left\{ |B^{-1}y^{(1)}| : y^{(1)} \in \mathbb{R}^{d_1} \text{ and } |y^{(1)}| \leq 1 \right\} < \infty.$$

In addition, let  $L^\infty([0, T], C_b^1(\mathbb{R}^{d_1} \otimes \mathbb{R}^{d_2}; \mathbb{R}^{d_2}))$  be the space of all a.e. bounded measurable functions  $\varphi : [0, T] \rightarrow C_b^1(\mathbb{R}^{d_1} \otimes \mathbb{R}^{d_2}; \mathbb{R}^{d_2})$  such that

$$\operatorname{ess\,sup}_{s \in [0, T], y \in \mathbb{R}^{d_1} \otimes \mathbb{R}^{d_2}} (|\varphi(s, y)| + \|\nabla \varphi(s, y)\|) < \infty.$$

**Example 2.10.** *Assume that  $\tilde{b} \in L^\infty([0, T], C_b^1(\mathbb{R}^{d_1} \otimes \mathbb{R}^{d_2}; \mathbb{R}^{d_2}))$  and that (A1) and (A2) are satisfied. Then for  $0 \leq t < s \leq T$ , the Bismut type formula (2.43) in Corollary 2.4 holds for*

$$\langle M_r^t(x_0), v \rangle = \int_t^r \left\langle \sigma^{-1}(\theta) \left( \chi_{t,r}''(\theta) \tilde{v}^{(2)} - \kappa_{t,r}''(\theta) v^{(2)} + \nabla_{\Xi_{t,r}(v, \tilde{v}^{(2)}, \theta)} \tilde{b}(\theta, X_\theta^{t, x_0}) \right), dW_\theta \right\rangle,$$

where  $v = (v^{(1)}, v^{(2)}) \in \mathbb{R}^{d_1+d_2}$ ,  $\tilde{v}^{(2)} \in B^{-1}v^{(1)}$  and

$$\chi_{t,r}(\theta) = \frac{(\theta-t)^2(3r-t-2\theta)}{(r-t)^3}, \quad \kappa_{t,r}(\theta) = \frac{(\theta-t)(r-\theta)^2}{(r-t)^2}, \quad \theta \in [t, r], \quad (2.97)$$

$$\Xi_{t,r}(v, \tilde{v}^{(2)}, \theta) = \left( (1 - \chi_{t,r}(\theta))v^{(1)} + \kappa_{t,r}(\theta)Bv^{(2)}, \kappa'_{t,r}(\theta)v^{(2)} - \chi'_{t,r}(\theta)\tilde{v}^{(2)} \right). \quad (2.98)$$

*Proof.* By our assumptions, one can show that the moment conditions (2.7) and (2.31) hold with  $k_2 = 1$  and  $k_3 = 0$ , respectively. Furthermore, it is easy to check that (C1) and (C2) hold.

According to [17, Theorem 2.2], the derivative formula (2.2) holds with  $\langle M_r^t(x), v \rangle$  given as follows: for any  $x, v \in \mathbb{R}^{d_1+d_2}$ ,  $0 \leq t < r \leq T$ ,

$$\begin{aligned} \langle M_r^t(x), v \rangle &= \int_t^r \left\langle \sigma^{-1}(\theta) \left( \chi''_{t,r}(\theta)\tilde{v}^{(2)} - \kappa''_{t,r}(\theta)v^{(2)} + \nabla_{\Xi_{t,r}(v, \tilde{v}^{(2)}, \theta)} \tilde{b}(\theta, X_\theta^{t,x}) \right), dW_\theta \right\rangle \\ &=: \int_t^r \left\langle \left( H_{t,r}(\theta, x, \tilde{W}([t, r]), X_\theta^{t,x}, \nabla X_\theta^{t,x}) \right)^* v, dW_\theta \right\rangle, \end{aligned}$$

where  $\chi_{t,r}(\theta)$ ,  $\kappa_{t,r}(\theta)$  and  $\Xi_{t,r}(v, \tilde{v}^{(2)}, \theta)$  are defined in (2.97) and (2.98). Moreover, similarly to Example 2.7, we obtain that  $\nabla P_{t,r}\psi \in C(\mathbb{R}^d)$  for any  $\psi \in C_b^1(\mathbb{R}^d)$ . Hence (H2) holds.

Notice that by a direct calculation, we have

$$\begin{aligned} & \left| \left( H_{t,r}(\theta, x, \tilde{W}([t, r]), X_\theta^{t,x}, \nabla X_\theta^{t,x}) \right)^* v \right|^2 \\ & \leq C_T \sup_{\theta \in [0, T]} \|\sigma^{-1}(\theta)\|^2 \left[ \frac{\|B^{-1}\|^2 |v^{(1)}|^2}{(r-t)^4} + \frac{|v^{(2)}|^2}{(r-t)^2} \right. \\ & \quad \left. + \|\nabla \tilde{b}(\theta, \cdot)\|_\infty^2 \left( |v^{(1)}|^2 + (r-t)^2 |Bv^{(2)}|^2 + |v^{(2)}|^2 + \frac{\|B^{-1}\|^2 |v^{(1)}|^2}{(r-t)^2} \right) \right] \\ & \leq C_{B, \tilde{b}, \sigma, T} \left( 1 + \frac{1}{(r-t)^4} \right) |v|^2. \end{aligned}$$

Consequently, we get for any  $x, v \in \mathbb{R}^{d_1+d_2}$ ,

$$\begin{aligned} \mathbb{E} |\langle M_r^t(x), v \rangle|^2 &= \mathbb{E} \int_t^r \left| \left( H_{t,r}(\theta, x, \tilde{W}([t, r]), X_\theta^{t,x}, \nabla X_\theta^{t,x}) \right)^* v \right|^2 d\theta \\ &\leq C_{B, \tilde{b}, \sigma, T} \left( r - t + \frac{1}{(r-t)^3} \right) |v|^2. \end{aligned} \quad (2.99)$$

Then we obtain that for any  $0 \leq t < s < r \leq T$  and  $R > 0$ ,

$$\begin{aligned} \int_s^T \sup_{|x-x_0| \leq R} K(t, t, r, x) dr &= \int_s^T \sup_{|x-x_0| \leq R} (\mathbb{E} |M_r^t(x)|^2)^{\frac{1}{2}} dr \\ &= \int_s^T \sup_{|x-x_0| \leq R} \left( \sum_{i=1}^{d_1+d_2} \mathbb{E} |\langle M_r^t(x), e_i \rangle|^2 \right)^{\frac{1}{2}} dr \\ &\leq C_{T, B, \sigma} \int_s^T \left( \sqrt{r-t} + \frac{1}{(r-t)^{\frac{3}{2}}} \right) dr \end{aligned}$$

$< \infty$ ,

where  $\{e_i\}_{i=1}^{d_1+d_2}$  is the canonical ONB of  $\mathbb{R}^{d_1+d_2}$ . Hence,  $K(t, t, r, \cdot)$  satisfies (2.5) with  $k_0 = 0$  and (2.42) holds.

Finally, since  $\chi_{t,r}$  and  $\kappa_{t,r}$  are deterministic and  $\nabla \tilde{b}$  is bounded, it is readily verified that (2.3) (with  $s = t$ ) holds.

Therefore, the assertion follows from Corollary 2.4 (1).  $\square$

**Remark 2.4.** (i) If  $d_1 = d_2$ ,  $B = \sigma(r) = I_{d_1 \times d_1}$  and

$$\tilde{b}(r, x) = -\nabla \varphi(x^{(1)}) - Cx^{(2)},$$

the forward SDE is known as “stochastic damping Hamiltonian system” in probability (see, e.g., [4] and [34]).

(ii) We mention that  $1/(r-t)^3$  in (2.99) seems to be optimal, meaning that the term  $\mathbb{E} \int_s^T f(r, \Theta_r^{t,x_0}) \langle M_r^t(x_0), v \rangle dr$  in (2.43) might be ill-defined when  $t = s$ . For example, let  $t = s$  and suppose that  $f$  is bounded, then by (2.99)

$$\left| \mathbb{E} \int_s^T f(r, \Theta_r^{t,x_0}) \langle M_r^t(x_0), v \rangle dr \right| \leq C_{B, \tilde{b}, \sigma, T, f} |v| \int_s^T \left( \sqrt{r-s} + \frac{1}{(r-s)^{\frac{3}{2}}} \right) dr = \infty.$$

So, the Bismut type formula is thus not available in the case of  $t = s$ .

### 3 Applications to McKean-Vlasov FBSDEs and related PDEs

In this section, we will apply the gradient type and Bismut type formulas to the study of McKean-Vlasov FBSDEs. Our goal is twofold. Firstly, we want to establish the representation formulae for the control solution  $Z_s^{t,\xi}$ , which enables us to derive its path regularity. Secondly, we are concerned with the gradient estimates for the solution of a nonlocal PDE of mean-field type, which is associated with a McKean-Vlasov FBSDE.

Let  $W$  be a  $d$ -dimensional Brownian motion,  $(\mathcal{F}_t)_{t \in [0, T]}$  and  $\{\mathcal{F}_r^t\}_{0 \leq t \leq r \leq T}$  the corresponding filtrations as before, and denote by  $\mathcal{G} \subset \mathcal{F}$  a sub- $\sigma$ -algebra independent of  $W$ . For an initial datum  $(t, \xi) \in [0, T] \times L^2(\Omega \rightarrow \mathbb{R}^d, \mathcal{G}, \mathbb{P})$  we consider the following McKean-Vlasov FBSDE: for  $s \in [t, T]$ ,

$$\begin{cases} X_s^{t,\xi} = \xi + \int_t^s b(r, X_r^{t,\xi}, \mathbb{P}_{X_r^{t,\xi}}) dr + \int_t^s \sigma(r, X_r^{t,\xi}, \mathbb{P}_{X_r^{t,\xi}}) dW_r, \\ Y_s^{t,\xi} = g(X_T^{t,\xi}, \mathbb{P}_{X_T^{t,\xi}}) + \int_s^T f(r, X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi}, \mathbb{P}_{(X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi})}) dr + \int_s^T Z_r^{t,\xi} dW_r, \end{cases} \quad (3.1)$$

where  $\mathbb{P}_{X_r^{t,\xi}}$  and  $\mathbb{P}_{(X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi})}$  denote the distributions of  $X_r^{t,\xi}$  and  $(X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi})$  under  $\mathbb{P}$ , respectively. Note that the coefficients  $b, \sigma$  and  $g, f$  depend on the law of the solution, and the above forward-backward system is decoupled, as the first equation does not contain the solution  $(Y^{t,\xi}, Z^{t,\xi})$  of the second one.

Let  $\mathcal{P}_2(\mathbb{R}^d)$  be the collection of all probability measures on  $\mathbb{R}^d$  with finite second moment. Define the 2-Wasserstein distance on  $\mathcal{P}_2(\mathbb{R}^d)$  by

$$\mathbb{W}_2(\mu, \nu) := \inf_{\pi \in \mathcal{C}(\mu, \nu)} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \pi(dx, dy) \right)^{\frac{1}{2}},$$

where  $\mathfrak{C}(\mu, \nu)$  is the set of all probability measures on  $\mathbb{R}^d \times \mathbb{R}^d$  with marginal laws  $\mu$  and  $\nu$ . Then  $(\mathcal{P}_2(\mathbb{R}^d), \mathbb{W}_2)$  is a Polish space. In addition, let  $\delta_0$  be the Dirac measure at 0 and denote by  $\mathcal{S}^2([t, T]; \mathbb{R}^k)$  (respectively,  $\mathcal{H}^2([t, T]; \mathbb{R}^k)$ ) the space of all  $\mathbb{R}^k$ -valued and  $(\mathcal{F}_t \vee \mathcal{G})_{0 \leq t \leq T}$ -adapted continuous processes (respectively, predictable processes)  $\phi = (\phi_s)_{s \in [t, T]}$  with  $\mathbb{E}(\sup_{s \in [t, T]} |\phi_s|^2) < \infty$  (respectively,  $\mathbb{E}(\int_t^T |\phi_s|^2 ds) < \infty$ ). Throughout this section, we will make the following assumptions on the coefficients  $b, \sigma$  and  $g, f$ .

( $\tilde{\text{H}}1$ )  $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$  and  $\sigma : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$  are continuous in time and there exists a constant  $L_1 > 0$  such that

$$\begin{aligned} & |b(r, x_1, \mu_1) - b(r, x_2, \mu_2)| + \|\sigma(r, x_1, \mu_1) - \sigma(r, x_2, \mu_2)\| \\ & \leq L_1(|x_1 - x_2| + \mathbb{W}_2(\mu_1, \mu_2)), \quad r \in [0, T], \quad x_1, x_2 \in \mathbb{R}^d, \quad \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d). \end{aligned}$$

Besides,  $b(r, 0, \delta_0)$  and  $\sigma(r, 0, \delta_0)$  are bounded functions of  $r \in [0, T]$ .

( $\tilde{\text{H}}2$ )  $g : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  and  $f : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d) \rightarrow \mathbb{R}$  are Lipschitz continuous, that is, there exists a constant  $L_2 > 0$  such that for all  $r \in [0, T], x_1, x_2 \in \mathbb{R}^d, y_1, y_2 \in \mathbb{R}, z_1, z_2 \in \mathbb{R}^d, \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d), \mu, \tilde{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$ ,

$$|g(x_1, \mu) - g(x_2, \tilde{\mu})| \leq L_2(|x_1 - x_2| + \mathbb{W}_2(\mu, \tilde{\mu}))$$

and

$$|f(r, x_1, y_1, z_1, \mu_1) - f(r, x_2, y_2, z_2, \mu_2)| \leq L_2(|x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2| + \mathbb{W}_2(\mu_1, \mu_2)).$$

Besides, assume that  $|g(0, \delta_0)| + \sup_{r \in [0, T]} |f(r, 0, 0, 0, \delta_0)| < \infty$ .

We observe that under ( $\tilde{\text{H}}1$ ) and ( $\tilde{\text{H}}2$ ), the system (3.1) has a unique solution  $(X^{t, \xi}, Y^{t, \xi}, Z^{t, \xi}) \in \mathcal{S}^2([t, T]; \mathbb{R}^d) \times \mathcal{S}^2([t, T]; \mathbb{R}) \times \mathcal{H}^2([t, T]; \mathbb{R}^d)$ , see, for instance [20, Theorem A.1], where  $b, \sigma$  and  $f$  are independent of the time variable and the driven noises are a Brownian motion and an independent Poisson random measure. Since the proof of [20, Theorem A.1] also applies to the present case, we omit the proof here. In order to solve a class of PDEs of mean-field type (see (3.9) below), the authors of [7] and [20] introduce SDEs accompanying the system (3.1) to deal with the homogeneous and non-homogeneous case, respectively, which then allows to provide probabilistic representations for the solution of (3.9) (see also [13] for the homogeneous case). In this part, we will adopt this idea to establish the representation formulas for  $Z^{t, \xi}$  and the gradient estimates for (3.9) via the gradient type and Bismut type formulas for BSDEs obtained in the last section. To formulate the accompanying SDEs, let  $(X^{t, \xi}, Y^{t, \xi}, Z^{t, \xi})$  be the solution of (3.1). Now, for any  $x \in \mathbb{R}^d$ , consider the following SDEs: for  $s \in [t, T]$ ,

$$X_s^{t, x, \mathbb{P}^\xi} = x + \int_t^s b(r, X_r^{t, x, \mathbb{P}^\xi}, \mathbb{P}_{X_r^{t, \xi}}) dr + \int_t^s \sigma(r, X_r^{t, x, \mathbb{P}^\xi}, \mathbb{P}_{X_r^{t, \xi}}) dW_r \quad (3.2)$$

and

$$\begin{aligned} Y_s^{t, x, \mathbb{P}^\xi} &= g(X_T^{t, x, \mathbb{P}^\xi}, \mathbb{P}_{X_T^{t, \xi}}) + \int_s^T f(r, X_r^{t, x, \mathbb{P}^\xi}, Y_r^{t, x, \mathbb{P}^\xi}, Z_r^{t, x, \mathbb{P}^\xi}, \mathbb{P}_{(X_r^{t, \xi}, Y_r^{t, \xi}, Z_r^{t, \xi})}) dr \\ &\quad + \int_s^T Z_r^{t, x, \mathbb{P}^\xi} dW_r. \end{aligned} \quad (3.3)$$

Since the distribution dependence in the coefficients of (3.2) is  $\mathbb{P}_{X_r^{t,\xi}}$ , rather than that of the solution to itself, it becomes a classical SDE. Then, it is easy to see that under  $(\tilde{H}1)$  there exists a unique solution  $X^{t,x,\mathbb{P}_\xi}$  to (3.2). Similarly, under  $(\tilde{H}1)$  and  $(\tilde{H}2)$  (3.3) also has a unique solution  $(Y^{t,x,\mathbb{P}_\xi}, Z^{t,x,\mathbb{P}_\xi})$ . Moreover,

$$X^{t,\xi} = X^{t,x,\mathbb{P}_\xi}|_{x=\xi}, \quad Y^{t,\xi} = Y^{t,x,\mathbb{P}_\xi}|_{x=\xi}, \quad Z^{t,\xi} = Z^{t,x,\mathbb{P}_\xi}|_{x=\xi}. \quad (3.4)$$

For further details, we, for instance, refer to [7, 13, 20].

Below is our first result which asserts that the control solution  $Z_s^{t,\xi}$  of the system (3.1) admits a representation formula.

**Theorem 3.1.** *Let assumptions  $(\tilde{H}1)$  and  $(\tilde{H}2)$  hold. Assume moreover that for each  $r \in [0, T]$  and  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $b(r, \cdot, \mu)$  and  $\sigma(r, \cdot, \mu)$  are continuously differentiable with*

$$\sup_{(r,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)} (\|\nabla b(r, x, \mu)\| + \|\nabla \sigma(r, x, \mu)\|) < \infty,$$

and that  $\sigma$  is non-degenerate. Then for each  $s \in [t, T]$ ,

$$\begin{aligned} Z_s^{t,\xi} = & \mathbb{E} \left[ g(X_T^{t,\xi}, \mathbb{P}_{X_T^{t,\xi}}) N_T^{t,s}(\xi) + \int_s^T f(r, X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi}, \mathbb{P}_{(X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi})}) N_r^{t,s}(\xi) dr \middle| \mathcal{F}_s^t \vee \mathcal{G} \right] \\ & \times (\nabla X_s^{t,\xi, \mathbb{P}_\xi})^{-1} \sigma(s, X_s^{t,\xi}, \mathbb{P}_{X_s^{t,\xi}}). \end{aligned} \quad (3.5)$$

In particular,

$$\begin{aligned} Z_t^{t,\xi} = & \mathbb{E} \left[ g(X_T^{t,\xi}, \mathbb{P}_{X_T^{t,\xi}}) N_T^{t,t}(\xi) + \int_t^T f(r, X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi}, \mathbb{P}_{(X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi})}) N_r^{t,t}(\xi) dr \middle| \mathcal{G} \right] \\ & \times \sigma(t, \xi, \mathbb{P}_\xi) \end{aligned} \quad (3.6)$$

Here for  $t \leq s < r \leq T$ ,  $\nabla X_s^{t,\xi, \mathbb{P}_\xi} = \nabla X_s^{t,x, \mathbb{P}_\xi}|_{x=\xi}$  and

$$N_r^{t,s}(\xi) = \int_s^r \frac{1}{r-s} \left( (\sigma^*(\sigma\sigma^*)^{-1})(\theta, X_\theta^{t,\xi}, \mathbb{P}_{X_\theta^{t,\xi}}) \nabla X_\theta^{t,\xi, \mathbb{P}_\xi} \right)^* dW_\theta.$$

*Proof.* We first assume that  $g(\cdot, \mu)$  and  $f(r, \cdot, \cdot, \cdot, \mu)$  are also continuously differentiable for any  $r \in [0, T]$  and  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ . For fixed  $(t, \xi)$ , considering the coefficients  $\tilde{b}(s, x) := b(s, x, \mathbb{P}_{X_s^{t,\xi}})$ ,  $\tilde{\sigma}(s, x) := \sigma(s, x, \mathbb{P}_{X_s^{t,\xi}})$  and  $\tilde{g}(x) := g(x, \mathbb{P}_{X_T^{t,\xi}})$ ,  $\tilde{f}(s, x, y, z) := f(s, x, y, z, \mathbb{P}_{(X_s^{t,\xi}, Y_s^{t,\xi}, Z_s^{t,\xi})})$  and then applying Lemma 2.5 or [23, Theorem 3.1] to the system (3.2)-(3.3), we see that there exists  $\tilde{u} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $\nabla \tilde{u}(r, \cdot)$  is continuous for each  $r \in [0, T]$  and

$$Y_s^{t,x, \mathbb{P}_\xi} = \tilde{u}(s, X_s^{t,x, \mathbb{P}_\xi}), \quad Z_s^{t,x, \mathbb{P}_\xi} = \nabla \tilde{u}(s, X_s^{t,x, \mathbb{P}_\xi}) \tilde{\sigma}(s, X_s^{t,x, \mathbb{P}_\xi}), \quad s \in [t, T], \quad \mathbb{P}\text{-a.s.}$$

Then in view of Example 2.7, we obtain that for any  $s \in [t, T]$ ,

$$\begin{aligned} Z_s^{t,x, \mathbb{P}_\xi} = & \nabla Y_s^{t,x, \mathbb{P}_\xi} (\nabla X_s^{t,x, \mathbb{P}_\xi})^{-1} \tilde{\sigma}(s, X_s^{t,x, \mathbb{P}_\xi}) \\ = & \mathbb{E} \left[ \tilde{g}(X_T^{t,x, \mathbb{P}_\xi}) N_T^{t,s}(x) + \int_s^T \tilde{f}(r, X_r^{t,x, \mathbb{P}_\xi}, Y_r^{t,x, \mathbb{P}_\xi}, Z_r^{t,x, \mathbb{P}_\xi}) N_r^{t,s}(x) dr \middle| \mathcal{F}_s^t \right] \\ & \times (\nabla X_s^{t,x, \mathbb{P}_\xi})^{-1} \tilde{\sigma}(s, X_s^{t,x, \mathbb{P}_\xi}), \end{aligned} \quad (3.7)$$

where for any  $t \leq s < r \leq T$ ,

$$N_r^{t,s}(x) = \int_s^r \frac{1}{r-s} \left( (\tilde{\sigma}^* (\tilde{\sigma} \tilde{\sigma}^*)^{-1}) (\theta, X_\theta^{t,x,\mathbb{P}_\xi}) \nabla X_\theta^{t,x,\mathbb{P}_\xi} \right)^* dW_\theta. \quad (3.8)$$

By the same approximation argument as in the proof of Theorem 2.3 (1) or [23, Theorem 4.2, Page 1410-1411], we may extend the formula (3.7) to all  $f$  and  $g$  satisfying only ( $\tilde{H}2$ ). Therefore, letting  $x = \xi$  in (3.7)-(3.8) and taking into account (3.4) and the definitions of  $\tilde{\sigma}$ ,  $\tilde{g}$  and  $\tilde{f}$ , we derive (3.5). We stress here that based on the regular conditional probability and the monotone class theorem, the fact that  $\mathcal{F}_s^t$  is independent of  $\mathcal{G}$  converts  $\mathbb{E}[\cdot | \mathcal{F}_s^t]$  in (3.7) into  $\mathbb{E}[\cdot | \mathcal{F}_s^t \vee \mathcal{G}]$  in (3.5). The proof is standard, and hence, omitted here.

Finally, it is easy to see that (3.6) follows from (3.5) since the process  $\nabla X_s^{t,\xi,\mathbb{P}_\xi} := \nabla X_s^{t,x,\mathbb{P}_\xi}|_{x=\xi}$ ,  $s \in [t, T]$  satisfies the following SDE:

$$\nabla X_s^{t,\xi,\mathbb{P}_\xi} = I_{d \times d} + \int_t^s \nabla b(r, X_r^{t,\xi}, \mathbb{P}_{X_r^{t,\xi}}) \nabla X_r^{t,\xi,\mathbb{P}_\xi} dr + \int_t^s \nabla \sigma(r, X_r^{t,\xi}, \mathbb{P}_{X_r^{t,\xi}}) \nabla X_r^{t,\xi,\mathbb{P}_\xi} dW_r,$$

which is a slight variation of [7, Remark 4.1].  $\square$

**Remark 3.1.** (i) If the coefficients  $b, \sigma$  and  $g, f$  of FBSDE (3.1) are assumed to be functions with no dependence on the measure and  $\xi = x$  (namely,  $\xi$  is deterministic), then it is readily checked that our representation formula (3.5) coincides with that of [23, Theorem 4.2]. So, our formula (3.5) is a generalized version of [23, Theorem 4.2].

(ii) Combining the above proof and Example 2.8, we can alternatively derive that

$$N_r^{t,s}(\xi) = \int_s^r \frac{1}{\varrho_{s,r}(\theta)} \left( (\sigma^* (\sigma \sigma^*)^{-1}) (\theta, X_\theta^{t,\xi}, \mathbb{P}_{X_\theta^{t,\xi}}) G_r^{s,y}(\theta) \Big|_{y=X_s^{t,\xi}} \nabla X_s^{t,\xi,\mathbb{P}_\xi} \right)^* dW_\theta,$$

where  $\{G_r^{s,y}(\theta)\}_{\theta \in [s,r]}$  solves (2.81) for  $\nabla \tilde{b}$  and  $\nabla \tilde{\sigma}$  replacing  $\nabla b$  and  $\nabla \sigma$  there, respectively. Besides, under conditions on  $\sigma$  similar to Example 2.9 and the assumptions ( $\tilde{H}1$ ) and ( $\tilde{H}2$ ), the representation formula for  $Z^{t,\xi}$  can be extended to McKean-Vlasov FBSDE with forward Grushin type process.

Now let  $V(t, x, \mathbb{P}_\xi) := Y_t^{t,x,\mathbb{P}_\xi}$ . Notice that  $V(t, x, \mathbb{P}_\xi)$  is deterministic since  $Y_t^{t,x,\mathbb{P}_\xi}$  is measurable both with respect to  $\mathcal{F}_t$  and  $\mathcal{F}_T^t$ . By the Itô formula for distribution dependent functionals (see, e.g., [7, Theorem 7.1] and [20, Theorem 2.1]), it is shown in [20, Theorem 9.2] that under appropriate regularity conditions on the coefficients of  $b, \sigma, g$  and  $f$ ,  $V$  is the solution to the following nonlocal PDE of mean-field type:

$$\begin{cases} (\partial_t + \mathbb{L}_{b,\sigma})V(t, x, \mathbb{P}_\xi) = f \left( x, V(t, x, \mathbb{P}_\xi), (\nabla V \sigma)(t, x, \mathbb{P}_\xi), \mathbb{P}_{(\xi, V(t,\xi,\mathbb{P}_\xi), (\nabla V \sigma)(t,\xi,\mathbb{P}_\xi))} \right), \\ \quad \quad \quad (t, x, \xi) \in [0, T] \times \mathbb{R}^d \times L^2(\Omega \rightarrow \mathbb{R}^d, \mathcal{F}_t, \mathbb{P}), \\ V(T, x, \mathbb{P}_\xi) = g(x, \mathbb{P}_\xi), \quad (x, \mathbb{P}_\xi) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d), \end{cases} \quad (3.9)$$

where the operator  $\mathbb{L}_{b,\sigma}$ , introduced in [7], is defined as follows: for any  $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ ,

$$\mathbb{L}_{b,\sigma} V(t, x, \mu) = \langle b, \nabla V \rangle(t, x, \mu) + \frac{1}{2} \text{Tr} (\sigma \sigma^* \nabla^2 V)(t, x, \mu)$$

$$\begin{aligned}
& + \int_{\mathbb{R}^d} \langle b(t, y, \mu), \partial_\mu V(t, x, \mu)(y) \rangle \mu(dy) \\
& + \frac{1}{2} \int_{\mathbb{R}^d} \text{Tr} \left( (\sigma \sigma^*)(t, y, \mu) \nabla (\partial_\mu V(t, x, \mu)(\cdot))(y) \right) \mu(dy).
\end{aligned}$$

Here  $\nabla V, \nabla^2 V$  and  $\partial_\mu V$  denote the gradient, the Hessian and the Lions derivative of  $V$ , respectively. We remark that, in contrast to the classical case, the last two terms in the description of  $\mathbb{L}_{b,\sigma} V(t, x, \mu)$  involve the derivatives with respect to the measure variable. Recall that the notion of this type of derivative was introduced by P.-L. Lions in his lectures [8] at the Collège de France, see also [2, 7, 31]. Below we will provide the uniform gradient estimates for PDE (3.9).

**Theorem 3.2.** *Let the assumptions in Theorem 3.1 hold and  $V(t, x, \mathbb{P}_\xi) = Y_t^{t,x,\mathbb{P}_\xi}$  be the solution to the PDE (3.9) (which is assumed to exist and to be unique). Then there exists a constant  $C_{L_1, L_2, T} > 0$  such that*

$$|\nabla V(t, x, \mathbb{P}_\xi)| \leq C_{L_1, L_2, T} \left( \frac{1}{\sqrt{T-t}} + \sqrt{T-t} \right) (1 + |x| + \|\xi\|_2). \quad (3.10)$$

*Proof.* We consider the functions  $\tilde{b}, \tilde{\sigma}$  and  $\tilde{g}, \tilde{f}$  introduced in the proof of Theorem 3.1. With these notations in hand, we can apply Corollary 2.4 and (2.78) in Example 2.7 to the system (3.2)-(3.3), and we obtain that for any  $0 \leq t \leq s < T$  and  $v \in \mathbb{R}^d$ ,

$$\nabla_v \mathbb{E} Y_s^{t,x,\mathbb{P}_\xi} = \mathbb{E} \left[ \tilde{g}(X_T^{t,x,\mathbb{P}_\xi}) \langle M_T^t(x), v \rangle + \int_s^T \tilde{f}(r, X_r^{t,x,\mathbb{P}_\xi}, Y_r^{t,x,\mathbb{P}_\xi}, Z_r^{t,x,\mathbb{P}_\xi}) \langle M_r^t(x), v \rangle dr \right], \quad (3.11)$$

where for each  $t < r \leq T$ ,

$$\begin{aligned}
M_r^t(x) &= \int_t^r \frac{1}{r-t} \left( (\tilde{\sigma}^* (\tilde{\sigma} \tilde{\sigma}^*)^{-1}) (\theta, X_\theta^{t,x,\mathbb{P}_\xi}) \nabla X_\theta^{t,x,\mathbb{P}_\xi} \right)^* dW_\theta \\
&= \int_t^r \frac{1}{r-t} \left( (\sigma^* (\sigma \sigma^*)^{-1}) (\theta, X_\theta^{t,x,\mathbb{P}_\xi}, \mathbb{P}_{X_\theta^{t,\xi}}) \nabla X_\theta^{t,x,\mathbb{P}_\xi} \right)^* dW_\theta.
\end{aligned}$$

Since  $Y_t^{t,x,\mathbb{P}_\xi}$  is deterministic, setting  $s = t$  in (3.11) implies

$$\begin{aligned}
& \nabla V(t, x, \mathbb{P}_\xi) = \nabla Y_t^{t,x,\mathbb{P}_\xi} \\
&= \mathbb{E} \left[ g(X_T^{t,x,\mathbb{P}_\xi}, \mathbb{P}_{X_T^{t,\xi}}) M_T^t(x) + \int_t^T f(r, X_r^{t,x,\mathbb{P}_\xi}, Y_r^{t,x,\mathbb{P}_\xi}, Z_r^{t,x,\mathbb{P}_\xi}, \mathbb{P}_{(X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi})}) M_r^t(x) dr \right].
\end{aligned} \quad (3.12)$$

In order to obtain an upper bound for  $\nabla V(t, x, \mathbb{P}_\xi)$ , we start with the following moment estimates of the solutions to the equations (3.1)-(3.3):

$$\mathbb{E} \left[ \sup_{s \in [t, T]} |X_s^{t,x,\mathbb{P}_\xi}|^2 + \sup_{s \in [t, T]} |Y_s^{t,x,\mathbb{P}_\xi}|^2 + \text{ess sup}_{s \in [t, T]} |Z_s^{t,x,\mathbb{P}_\xi}|^2 \right] \leq C_{L_1, L_2, T} (1 + |x|^2)$$

and

$$\mathbb{E} \left[ \sup_{s \in [t, T]} |X_s^{t,\xi}|^2 + \sup_{s \in [t, T]} |Y_s^{t,\xi}|^2 + \text{ess sup}_{s \in [t, T]} |Z_s^{t,\xi}|^2 \right] \leq C_{L_1, L_2, T} (1 + \|\xi\|_2^2),$$

which can be proved in the spirit of the proof of [23, Theorem 3.3]. Then by the Hölder inequality and the Lipschitz continuity of  $g$  and  $f$ , we have

$$\begin{aligned}
\left| \mathbb{E} \left[ g(X_T^{t,x,\mathbb{P}_\xi}, \mathbb{P}_{X_T^{t,\xi}}) M_T^t(x) \right] \right| &\leq \left[ \mathbb{E} \left| g(X_T^{t,x,\mathbb{P}_\xi}, \mathbb{P}_{X_T^{t,\xi}}) \right|^2 \right]^{\frac{1}{2}} \|M_T^t(x)\|_2 \\
&\leq C_{L_1, L_2, T} \left[ \mathbb{E} \left( 1 + |X_T^{t,x,\mathbb{P}_\xi}| + \|X_T^{t,\xi}\|_2 \right)^2 \right]^{\frac{1}{2}} \|M_T^t(x)\|_2 \\
&\leq C_{L_1, L_2, T} (1 + |x| + \|\xi\|_2) \|M_T^t(x)\|_2
\end{aligned} \tag{3.13}$$

and

$$\begin{aligned}
&\left| \mathbb{E} \int_t^T f(r, X_r^{t,x,\mathbb{P}_\xi}, Y_r^{t,x,\mathbb{P}_\xi}, Z_r^{t,x,\mathbb{P}_\xi}, \mathbb{P}_{(X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi})}) M_r^t(x) dr \right| \\
&\leq C_{L_1, L_2, T} \mathbb{E} \int_t^T \left[ \left( 1 + |X_r^{t,x,\mathbb{P}_\xi}| + |Y_r^{t,x,\mathbb{P}_\xi}| + |Z_r^{t,x,\mathbb{P}_\xi}| + \|X_r^{t,\xi}\|_2 + \|Y_r^{t,\xi}\|_2 + \|Z_r^{t,\xi}\|_2 \right) \right. \\
&\quad \left. \times |M_r^t(x)| \right] dr \\
&\leq C_{L_1, L_2, T} (1 + |x| + \|\xi\|_2) \int_t^T \|M_r^t(x)\|_2 dr.
\end{aligned} \tag{3.14}$$

Note that due to (2.79) (with  $s = t$ ) in Example 2.7, we can get for  $t < r \leq T$ ,

$$\|M_r^t(x)\|_2 \leq \frac{C_{L_1, T}}{\sqrt{r-t}}.$$

Therefore, substituting this into (3.13)-(3.14) and going back to relation (3.12), we get the desired assertion.  $\square$

**Remark 3.2.** (i) In [13, Remark 5.6 and Theorem 5.8], by developing integration by parts formulas for McKean-Vlasov SDEs with uniformly elliptic coefficients, the authors proved the existence of a classical solution to the homogeneous PDE (3.9) (namely,  $f = 0$  in (3.9)) with a class of non-differentiable terminal conditions  $g$  (including, for instance,  $g(x, \mu) = \int \varphi(x, y) \mu(dy)$ , where  $\varphi : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  is continuous with polynomial growth, more interesting examples can be found in [13, Example 5.4]), and obtained a gradient bound of the form  $C(T-t)^{-1/2}(1+|x|+\|\xi\|_2)^q$  with positive constants  $C$  and  $q$ , which essentially coincides with our estimate (3.10). Therefore, the result stated in Theorem 3.2 can be regarded as a generalization of that in [13].

(ii) One can provide an alternative version of the previous estimate, whose proof follows the same lines as the one of Theorem 3.2. Owing to (2.83) (with  $s = t$ ) in Example 2.8, we have for  $t < r \leq T$ ,

$$\|M_r^t(x)\|_2 \leq C_{L_1, T} \left( 1 + \frac{1}{\sqrt{1 - e^{-c(r-t)}}} \right).$$

Then, we obtain

$$|\nabla V(t, x, \mathbb{P}_\xi)| \leq C_{L_1, L_2, T} \left( 1 + \frac{1}{\sqrt{1 - e^{-c(T-t)}}} + \int_t^T \frac{1}{\sqrt{1 - e^{-c(r-t)}}} dr \right) (1 + |x| + \|\xi\|_2).$$



In particular, as  $T$  tends to  $t$ , we have

$$1 + \frac{1}{\sqrt{1 - e^{-c(T-t)}}} + \int_t^T \frac{1}{\sqrt{1 - e^{-c(r-t)}}} dr = O\left(\frac{1}{\sqrt{T-t}}\right),$$

which means that this type of gradient estimate has the same order as the right-hand side of (3.10) when  $T$  goes to  $t$ .

(iii) For PDEs (3.9) associated with a McKean-Vlasov FBSDE with forward Grushin type process, it follows from (2.95) (with  $s = t$ ) in Example 2.9 that

$$\|M_r^t(x)\|_2 \leq \frac{C_{L_1, T}}{\sqrt{r-t}} \left(1 + \frac{1}{(|x^{(1)}|^2 + (r-t))^{\frac{\alpha}{2}}}\right).$$

Then, we have

$$|\nabla V(t, x, \mathbb{P}_\xi)| \leq C_{L_1, L_2, T} \left[ \frac{1}{\sqrt{T-t}} \left(1 + \frac{1}{(|x^{(1)}|^2 + (T-t))^{\frac{\alpha}{2}}}\right) + \int_t^T \frac{1}{\sqrt{r-t}} \left(1 + \frac{1}{(|x^{(1)}|^2 + (r-t))^{\frac{\alpha}{2}}}\right) dr \right] (1 + |x| + \|\xi\|_2),$$

provided that  $x^{(1)} \neq 0$ .

(iv) For PDEs (3.9) associated with a McKean-Vlasov FBSDE with forward stochastic Hamiltonian system, in light of (2.99) in Example 2.10 there will appear

$$\int_t^T \|M_r^t(x)\|_2 dr \leq C_{B, \tilde{b}, \sigma, T} \int_t^T \left( \sqrt{r-t} + \frac{1}{(r-t)^{\frac{3}{2}}} \right) dr = \infty.$$

So, we cannot be sure that the left-hand side is finite and thus cannot directly adopt the above arguments to obtain an efficient estimate in this setting.

## 4 Appendix: proofs of auxiliary lemmas

### 4.1 Proof of Lemma 2.5

*Proof.* We split the proof into three steps.

*Step 1: Claim:*  $u(t, \cdot)$  is continuously differentiable for every  $t \in [0, T]$  and satisfies (2.44). Let  $(t, x) \in [0, T] \times \mathbb{R}^d$  be fixed. Owing to (A1), (A2) and the moment conditions (2.7) and (2.31) in (C1), the following equations have a unique solution  $(Y_s^{t,x}, Z_s^{t,x})_{s \in [t, T]}$  and  $(\nabla Y_s^{t,x}, \nabla Z_s^{t,x})_{s \in [t, T]}$ :

$$Y_s^{t,x} = g(X_T^{t,x}) + \int_s^T f(r, \Theta_r^{t,x}) dr - \int_s^T Z_r^{t,x} dW_r, \quad (4.1)$$

$$\nabla Y_s^{t,x} = \nabla g(X_T^{t,x}) \nabla X_T^{t,x} + \int_s^T \nabla f(r, \Theta_r^{t,x}) \nabla \Theta_r^{t,x} dr - \int_s^T \nabla Z_r^{t,x} dW_r. \quad (4.2)$$

Moreover, as in the proof of [23, Theorem 3.1, Page 1398-1399], one can show that for any  $v \in \mathbb{R}^d$ ,

$$\lim_{\epsilon \rightarrow 0^+} \mathbb{E} \left( \sup_{s \in [t, T]} \left| \frac{Y_s^{t,x+\epsilon v} - Y_s^{t,x}}{\epsilon} - \nabla_v Y_s^{t,x} \right|^2 + \int_t^T \left\| \frac{Z_r^{t,x+\epsilon v} - Z_r^{t,x}}{\epsilon} - \nabla_v Z_r^{t,x} \right\|^2 dr \right) = 0.$$

That is,  $u(t, \cdot)$  is differentiable along the direction  $v$  and  $\nabla_v u(t, x) = \nabla_v Y_t^{t,x}$ .

Next, we focus on proving that  $\nabla Y_t^{t,\cdot}$  is continuous. To the end, applying the Itô formula to (4.1) and using (A1)-(A2), we first obtain

$$\begin{aligned}
& |Y_s^{t,y} - Y_s^{t,x}|^2 + \int_s^T \|Z_r^{t,y} - Z_r^{t,x}\|^2 dr \\
&= |g(X_T^{t,y}) - g(X_T^{t,x})|^2 + 2 \int_s^T \langle f(r, \Theta_r^{t,y}) - f(r, \Theta_r^{t,x}), Y_r^{t,y} - Y_r^{t,x} \rangle dr \\
&\quad - 2 \int_s^T \langle Y_r^{t,y} - Y_r^{t,x}, (Z_r^{t,y} - Z_r^{t,x}) dW_r \rangle \\
&\leq K_1^2 |X_T^{t,y} - X_T^{t,x}|^2 + 2K_2 \int_s^T |\Theta_r^{t,y} - \Theta_r^{t,x}| \cdot |Y_r^{t,y} - Y_r^{t,x}| dr \\
&\quad + 2 \left| \int_s^T \langle Y_r^{t,y} - Y_r^{t,x}, (Z_r^{t,y} - Z_r^{t,x}) dW_r \rangle \right|.
\end{aligned}$$

By the Hölder inequality, the Gronwall inequality and the B-D-G inequality, we deduce that there exists a constant  $C > 0$  such that

$$\begin{aligned}
& \mathbb{E} \left( \sup_{s \in [t, T]} |Y_s^{t,y} - Y_s^{t,x}|^2 + \int_t^T \|Z_r^{t,y} - Z_r^{t,x}\|^2 dr \right) \\
& \leq C \left( \mathbb{E} |X_T^{t,y} - X_T^{t,x}|^2 + \mathbb{E} \int_t^T |X_r^{t,y} - X_r^{t,x}|^2 dr \right).
\end{aligned}$$

Observe that from (2.31) (with  $k_1 = 2$ ) and (2.32) of (C1), it follows that for each  $r \in [t, T]$ ,

$$\begin{aligned}
\mathbb{E} |X_r^{t,y} - X_r^{t,x}|^2 &= \mathbb{E} \left| \int_0^{|y-x|} \nabla_v X_r^{t,x+\theta v} d\theta \right|^2 \leq |y-x| \int_0^{|y-x|} \mathbb{E} |\nabla_v X_r^{t,x+\theta v}|^2 d\theta \\
&\leq C \left[ 1 + (|x| + |y-x|)^{k_3} \right]^2 |y-x|^2,
\end{aligned} \tag{4.3}$$

where  $v := \frac{y-x}{|y-x|}$ . Then we have

$$\lim_{y \rightarrow x} \mathbb{E} \left( \sup_{s \in [t, T]} |Y_s^{t,y} - Y_s^{t,x}|^2 + \int_t^T \|Z_r^{t,y} - Z_r^{t,x}\|^2 dr \right) = 0. \tag{4.4}$$

Similarly, applying the Itô formula to (4.2) and using (A1)-(A2) again, we get

$$\begin{aligned}
& \|\nabla Y_s^{t,y} - \nabla Y_s^{t,x}\|^2 + \int_s^T \|\nabla Z_r^{t,y} - \nabla Z_r^{t,x}\|^2 dr \\
&= \|\nabla g(X_T^{t,y}) \nabla X_T^{t,y} - \nabla g(X_T^{t,x}) \nabla X_T^{t,x}\|^2 \\
&\quad + 2 \int_s^T \langle \nabla f(r, \Theta_r^{t,y}) \nabla \Theta_r^{t,y} - \nabla f(r, \Theta_r^{t,x}) \nabla \Theta_r^{t,x}, \nabla Y_r^{t,y} - \nabla Y_r^{t,x} \rangle dr
\end{aligned}$$

$$\begin{aligned}
& - 2 \int_s^T \langle \nabla Y_r^{t,y} - \nabla Y_r^{t,x}, (\nabla Z_r^{t,y} - \nabla Z_r^{t,x}) dW_r \rangle \\
\leq & 2K_1^2 \|\nabla X_T^{t,y} - \nabla X_T^{t,x}\|^2 + 2\|\nabla g(X_T^{t,y}) - \nabla g(X_T^{t,x})\|^2 \cdot \|\nabla X_T^{t,x}\|^2 \\
& + 2K_2 \int_s^T \|\nabla \Theta_r^{t,y} - \nabla \Theta_r^{t,x}\| \cdot \|\nabla Y_r^{t,y} - \nabla Y_r^{t,x}\| dr \\
& + 2 \int_s^T \|\nabla f(r, \Theta_r^{t,y}) - \nabla f(r, \Theta_r^{t,x})\| \cdot \|\nabla \Theta_r^{t,x}\| \cdot \|\nabla Y_r^{t,y} - \nabla Y_r^{t,x}\| dr \\
& + 2 \left| \int_s^T \langle \nabla Y_r^{t,y} - \nabla Y_r^{t,x}, (\nabla Z_r^{t,y} - \nabla Z_r^{t,x}) dW_r \rangle \right|.
\end{aligned}$$

Then, with the help of the Hölder, Gronwall and B-D-G inequalities again, we derive that there exists a constant  $C > 0$  such that

$$\begin{aligned}
& \mathbb{E} \left( \sup_{s \in [t, T]} \|\nabla Y_s^{t,y} - \nabla Y_s^{t,x}\|^2 + \int_t^T \|\nabla Z_r^{t,y} - \nabla Z_r^{t,x}\|^2 dr \right) \\
\leq & C \left[ \mathbb{E} \|\nabla X_T^{t,y} - \nabla X_T^{t,x}\|^2 + \mathbb{E} \left( \|\nabla g(X_T^{t,y}) - \nabla g(X_T^{t,x})\|^2 \cdot \|\nabla X_T^{t,x}\|^2 \right) \right. \\
& + \mathbb{E} \int_t^T \|\nabla X_r^{t,y} - \nabla X_r^{t,x}\|^2 dr \\
& \left. + \mathbb{E} \int_t^T \|\nabla f(r, \Theta_r^{t,y}) - \nabla f(r, \Theta_r^{t,x})\|^2 \cdot \|\nabla \Theta_r^{t,x}\|^2 dr \right]. \tag{4.5}
\end{aligned}$$

By (2.31) and (2.33) of (C1), the dominated convergence theorem implies

$$\lim_{y \rightarrow x} \left[ \mathbb{E} \|\nabla X_T^{t,y} - \nabla X_T^{t,x}\|^2 + \mathbb{E} \int_t^T \|\nabla X_r^{t,y} - \nabla X_r^{t,x}\|^2 dr \right] = 0. \tag{4.6}$$

In view of (4.3) and (4.4), we obtain that as  $y$  goes to  $x$ ,  $X_T^{t,y}$  and  $\Theta_r^{t,y}$  converge to  $X_T^{t,x}$  and  $\Theta_r^{t,x}$  in measures  $d\mathbb{P}$  and  $dr \otimes d\mathbb{P}$ , respectively. Then, since  $\nabla g(\cdot)$  and  $\nabla f(r, \cdot, \cdot, \cdot)$  are continuous, we have

$$\mathbb{P}\text{-}\lim_{y \rightarrow x} \|\nabla g(X_T^{t,y}) - \nabla g(X_T^{t,x})\|^2 = 0$$

and

$$dr \otimes d\mathbb{P}\text{-}\lim_{y \rightarrow x} \|\nabla f(r, \Theta_r^{t,y}) - \nabla f(r, \Theta_r^{t,x})\|^2 = 0.$$

Consequently, due to the fact that  $\nabla g$  and  $\nabla f$  are bounded, we can apply the dominated convergence theorem to get

$$\begin{aligned}
& \lim_{y \rightarrow x} \left[ \mathbb{E} \left( \|\nabla g(X_T^{t,y}) - \nabla g(X_T^{t,x})\|^2 \cdot \|\nabla X_T^{t,x}\|^2 \right) \right. \\
& \left. + \mathbb{E} \int_t^T \|\nabla f(r, \Theta_r^{t,y}) - \nabla f(r, \Theta_r^{t,x})\|^2 \cdot \|\nabla \Theta_r^{t,x}\|^2 dr \right] = 0.
\end{aligned}$$

Plugging this and (4.6) into (4.5), we obtain

$$\lim_{y \rightarrow x} \mathbb{E} \left( \sup_{s \in [t, T]} \|\nabla Y_s^{t,y} - \nabla Y_s^{t,x}\|^2 + \int_t^T \|\nabla Z_r^{t,y} - \nabla Z_r^{t,x}\|^2 dr \right) = 0, \tag{4.7}$$

which implies the continuity of  $\nabla Y_t^{t,\cdot}$ . We therefore conclude that  $\nabla u(t, \cdot)$  is continuous and  $\nabla u(t, x) = \nabla Y_t^{t,x}$ .

Note that  $\nabla Y_t^{t,x}$  is deterministic. Then letting  $s = t$  and taking expectation in (4.2) yield (2.44).

*Step 2: Claim:* (2.45) holds. Applying the Itô formula to (4.2), we have

$$\begin{aligned} & \|\nabla Y_s^{t,x}\|^2 + \int_s^T \|\nabla Z_r^{t,x}\|^2 dr + 2 \int_s^T \langle \nabla Y_r^{t,x}, \nabla Z_r^{t,x} dW_r \rangle \\ &= \|\nabla g(X_T^{t,x}) \nabla X_T^{t,x}\|^2 + 2 \int_s^T \langle \nabla f(r, \Theta_r^{t,x}) \nabla \Theta_r^{t,x}, \nabla Y_r^{t,x} \rangle dr \\ &\leq K_1^2 \|\nabla X_T^{t,x}\|^2 + 2K_2 \int_s^T \|\nabla \Theta_r^{t,x}\| \cdot \|\nabla Y_r^{t,x}\| dr, \end{aligned}$$

where we have used the assumptions (A1) and (A2). Then it follows from the Hölder, Gronwall and B-D-G inequalities that

$$\begin{aligned} & \mathbb{E} \left( \sup_{s \in [t, T]} \|\nabla Y_s^{t,x}\|^2 + \int_t^T \|\nabla Z_r^{t,x}\|^2 dr \right) \\ & \leq C_{K_1, K_2, T} \left( \mathbb{E} \|\nabla X_T^{t,x}\|^2 + \int_t^T \mathbb{E} \|\nabla X_r^{t,x}\|^2 dr \right) \end{aligned}$$

for some constant  $C_{K_1, K_2, T} > 0$ . Combining this with  $\nabla u(t, x) = \nabla Y_t^{t,x}$  and recalling that  $\nabla Y_t^{t,x}$  is deterministic, we obtain that there exists some constant  $q_1 > 0$  such that

$$\begin{aligned} \|\nabla u(t, x)\| &\leq C_{K_1, K_2, T} \left[ \left( \mathbb{E} \|\nabla X_T^{t,x}\|^2 \right)^{\frac{1}{2}} + \left( \int_t^T \mathbb{E} \|\nabla X_r^{t,x}\|^2 dr \right)^{\frac{1}{2}} \right] \\ &\leq C_{K_1, K_2, k_3, T} (1 + |x|^{q_1}), \quad x \in \mathbb{R}^d, \end{aligned}$$

where the last inequality is due to (2.31) (with  $k_1 = 2$  and  $q_1 = k_3$ ) of (C1). This yields the desired assertion.

*Step 3: Claim:*  $Y_s^{n,t,x} = u(s, X_s^{n,t,x})$  and  $Z_s^{n,t,x} = \nabla u(s, X_s^{n,t,x}) \sigma(s, X_s^{n,t,x})$ ,  $ds \otimes d\mathbb{P}$ -a.e. For  $x_n$  and  $(X_s^{n,t,x_n}, \nabla X_s^{n,t,x_n})_{s \in [t, T]}$  as in (C2), let  $(Y_s^{n,t,x_n}, Z_s^{n,t,x_n})_{s \in [t, T]}$  be the solution of the following equation:

$$Y_s^{n,t,x_n} = g(X_T^{n,t,x_n}) + \int_s^T f(r, \Theta_r^{n,t,x_n}) dr - \int_s^T Z_r^{n,t,x_n} dW_r,$$

where  $\Theta^{n,t,x_n} := (X^{n,t,x_n}, Y^{n,t,x_n}, Z^{n,t,x_n})$ . Since  $g(\cdot)$  and  $f(r, \cdot, \cdot, \cdot)$  are continuously differentiable, we arrive at

$$\nabla Y_s^{n,t,x_n} = \nabla g(X_T^{n,t,x_n}) \nabla X_T^{n,t,x_n} + \int_s^T \nabla f(r, \Theta_r^{n,t,x_n}) \nabla \Theta_r^{n,t,x_n} dr - \int_s^T \nabla Z_r^{n,t,x_n} dW_r.$$

Using (2.35) and (2.36) of (C2) and following the same arguments as for (4.4) and (4.7), we have

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left( \sup_{s \in [t, T]} |Y_s^{n,t,x_n} - Y_s^{t,x}|^2 + \int_t^T \|Z_r^{n,t,x_n} - Z_r^{t,x}\|^2 dr \right) = 0 \quad (4.8)$$

and

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left( \sup_{s \in [t, T]} \|\nabla Y_s^{n,t,x_n} - \nabla Y_s^{t,x}\|^2 + \int_t^T \|\nabla Z_r^{n,t,x_n} - \nabla Z_r^{t,x}\|^2 dr \right) = 0. \quad (4.9)$$

We now set  $u_n(t, x_n) := Y_t^{n,t,x_n}$ . According to [36, Theorem 3.1 and Remark 3.3], we obtain that  $u_n \in C^{0,1}([0, T] \times \mathbb{R}^d; \mathbb{R}^l)$  and

$$Y_s^{n,t,x_n} = u_n(s, X_s^{n,t,x_n}), \quad Z_s^{n,t,x_n} = \nabla u_n(s, X_s^{n,t,x_n}) \sigma_n(s, X_s^{n,t,x_n}), \quad s \in [t, T]. \quad (4.10)$$

Then by (4.8), (4.9) and Step 1, we have that for any  $t \in [0, T]$ ,

$$\lim_{n \rightarrow +\infty} u_n(t, x_n) = \lim_{n \rightarrow +\infty} Y_t^{n,t,x_n} = Y_t^{t,x} = u(t, x) \quad (4.11)$$

and

$$\lim_{n \rightarrow +\infty} \nabla u_n(t, x_n) = \lim_{n \rightarrow +\infty} \nabla Y_t^{n,t,x_n} = \nabla Y_t^{t,x} = \nabla u(t, x). \quad (4.12)$$

Observe that from (2.35) of (C2), there exists a subsequence of  $X_s^{n,t,x_n}$ , which we still denote by  $X_s^{n,t,x_n}$ , such that

$$\lim_{n \rightarrow +\infty} X_s^{n,t,x_n} = X_s^{t,x}, \quad ds \otimes d\mathbb{P}\text{-a.e.},$$

and moreover from (2.37) of (C2), it follows that

$$\lim_{n \rightarrow +\infty} \sigma_n(s, X_s^{n,t,x_n}) = \sigma(s, X_s^{t,x}), \quad ds \otimes d\mathbb{P}\text{-a.e.}$$

Consequently, by (4.11) and (4.12) we derive that  $ds \otimes d\mathbb{P}\text{-a.e.}$ ,

$$\lim_{n \rightarrow +\infty} u_n(s, X_s^{n,t,x_n}) = u(s, X_s^{t,x})$$

and

$$\lim_{n \rightarrow +\infty} \nabla u_n(s, X_s^{n,t,x_n}) \sigma_n(s, X_s^{n,t,x_n}) = \nabla u(s, X_s^{t,x}) \sigma(s, X_s^{t,x}).$$

Note that by (4.8) and selecting a subsequence if necessary, we obtain that for each  $s \in [t, T]$ ,  $\lim_{n \rightarrow +\infty} Y_s^{n,t,x_n} = Y_s^{t,x}$ ,  $\mathbb{P}$ -a.s. and  $\lim_{n \rightarrow +\infty} Z_s^{n,t,x_n} = Z_s^{t,x}$ ,  $ds \otimes d\mathbb{P}$ -a.e. Hence, because of (4.10), it is easy to see that the desired relations hold. This completes the proof.  $\square$

## 4.2 Proof of Lemma 2.6

*Proof.* Note that for each  $0 \leq t \leq s < r \leq T$ ,  $\tilde{W}([s, r])$  is independent of  $\mathcal{F}_s^t$  and  $X_s^{t,x} \in \mathcal{F}_s^t$ . Then by (2.73), we have for any  $\psi \in \mathcal{B}_b(\mathbb{R}^d)$ ,

$$\begin{aligned} & \mathbb{E} [\psi(X_r^{t,x}) M_r^s(X_s^{t,x}) | \mathcal{F}_s^t] \\ &= \mathbb{E} \left[ \psi(X_r^{t,x}) \int_s^r H_{s,r} \left( \theta, X_s^{t,x}, \tilde{W}([s, r]), X_\theta^{s, X_s^{t,x}}, \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right) dW_\theta \Big| \mathcal{F}_s^t \right] \\ &= \mathbb{E} \left[ \psi(X_r^{s, X_s^{t,x}}) \int_s^r H_{s,r} \left( \theta, X_s^{t,x}, \tilde{W}([s, r]), X_\theta^{s, X_s^{t,x}}, \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right) dW_\theta \Big| \mathcal{F}_s^t \right] \end{aligned}$$

$$\begin{aligned}
&= \left[ E \left( \psi(X_r^{s,y}) \int_s^r H_{s,r}(\theta, y, \tilde{W}([s, r]), X_\theta^{s,y}, \nabla X_\theta^{s,y}) dW_\theta \right) \right] \Big|_{y=X_s^{t,x}} \\
&= [\mathbb{E}(\psi(X_r^{s,y}) M_r^s(y))] \Big|_{y=X_s^{t,x}}.
\end{aligned}$$

Observe that all equalities above still hold true for  $|M_r^s(X_s^{t,x})|$  replacing  $M_r^s(X_s^{t,x})$ . Hence (2.6) holds for each  $\psi \in \mathcal{B}_b(\mathbb{R}^d)$   $\mathbb{P}$ -a.s.

Let us now investigate the function  $K(t, s, r, \cdot)$ . According to (2.73) and (2.74), we obtain

$$\begin{aligned}
K^2(t, s, r, x) &= \mathbb{E} |M_r^s(X_s^{t,x})|^2 \\
&= \mathbb{E} \left| \int_s^r H_{s,r} \left( \theta, X_s^{t,x}, \tilde{W}([s, r]), X_\theta^{s, X_s^{t,x}}, \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right) dW_\theta \right|^2 \\
&= \mathbb{E} \int_s^r \left\| H_{s,r} \left( \theta, X_s^{t,x}, \tilde{W}([s, r]), X_\theta^{t,x}, \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right) \right\|^2 d\theta \\
&\leq 3C_{q_1, q_2, T}^2 \mathbb{E} \int_s^r \left( K_\theta^2(t, s, r, x) + |X_\theta^{t,x}|^{2q_1} + \left\| \nabla X_\theta^{s,y} \Big|_{y=X_s^{t,x}} \right\|^{2q_2} \right) d\theta \\
&\leq C_{q_1, q_2, k_2, k_3, T} \left( \mathbb{E} \int_s^r K_\theta^2(t, s, r, x) d\theta + 1 + |x|^{2q_1 k_2} + \mathbb{E} |X_s^{t,x}|^{2q_2 k_3} \right) \\
&\leq C_{q_1, q_2, k_2, k_3, T} \left( \mathbb{E} \int_s^r K_\theta^2(t, s, r, x) d\theta + 1 + |x|^{2q_1 k_2} + |x|^{2q_2 k_3 k_2} \right). \quad (4.13)
\end{aligned}$$

Here, we have used the moment conditions (2.7) and (2.31) in the last two inequalities. Due to the assumption that the term  $\mathbb{E} \int_s^r K_\theta^2(t, s, r, \cdot) d\theta$  is of polynomial growth, so is  $K(t, s, r, \cdot)$ . Finally, it is easily seen that (2.39) follows from (4.13) and (2.75), which completes the proof.  $\square$

### Acknowledgement

X. Fan and M. Röckner are grateful to the financial support by the DFG through the CRC 1283 Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications. X. Fan is partially supported by the Natural Science Foundation of Anhui Province (No. 2008085MA10). S.-Q. Zhang is supported in part by the National Natural Science Foundation of China (No. 11771326, 11901604).

### References

- [1] O. Aboura and S. Bourguin, *Density estimates for solutions to one dimensional backward SDE's*, Potential Anal. 38 (2013), 573–587.
- [2] S. Albeverio, Y. G. Kondratiev and M. Röckner, *Differential geometry of Poisson spaces*, C. R. Acad. Sci. Paris Sér. I Math. 323 (1996), 1129–1134.
- [3] F. Antonelli and A. Kohatsu-Higa, *Densities of one-dimensional backward SDEs*, Potential Anal. 22 (2005), 263–287.
- [4] D. Bakry, P. Cattiaux and A. Guillin, *Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré*, J. Funct. Anal. 254 (2008), 727–759.

- [5] V. Bally, E. Pardoux and L. Stoica, *Backward stochastic differential equations associated to a symmetric Markov process*, Potential Anal. 22 (2005), 17–60.
- [6] R. Buckdahn, B. Djehiche, J. Li and S. Peng, *Mean-field backward stochastic differential equations: a limit approach*, Ann. Probab. 37 (2009), 1524–1565.
- [7] R. Buckdahn, J. Li, S. Peng and C. Rainer, *Mean-field stochastic differential equations and associated PDEs*, Ann. Probab. 2 (2017), 824–878.
- [8] P. Cardaliaguet, *Notes on mean field games*, P.-L. Lions lectures at Collège de France, <https://www.ceremade.dauphine.fr/cardaliaguet/MFG20130420.pdf>, 2013.
- [9] R. Carmona and F. Delarue, *Forward-backward stochastic differential equations and controlled McKean-Vlasov dynamics*, Ann. Probab. 43 (2015), 2647–2700.
- [10] J.-F. Chassagneux, D. Crisan and F. Delarue, *A probabilistic approach to classical solutions of the master equation for large population equilibria*, arXiv:1411.3009v2.
- [11] D. L. Cohn, *Measure Theory* (Second Edition), Birkhäuser, Boston, Mass., 2013.
- [12] D. Crisan and F. Delarue, *Sharp derivative bounds for solutions of degenerate semi-linear partial differential equations*, J. Funct. Anal. 263 (2012), 3024–3101.
- [13] D. Crisan and E. McMurray, *Smoothing properties of McKean-Vlasov SDEs*, Probab. Theory Related Fields 171 (2018), 97–148.
- [14] K. D. Elworthy and X.-M. Li, *Formulae for the derivatives of heat semigroups*, J. Funct. Anal. 125 (1994), 252–286.
- [15] X. Fan and J.-L. Wu, *Density estimates for the solutions of backward stochastic differential equations driven by Gaussian processes*, Potential Anal. 54 (2021), 483–501.
- [16] M. Fuhrman and G. Tessitore, *The Bismut-Elworthy formula for backward SDEs and applications to nonlinear Kolmogorov equations and control in infinite dimensional spaces*, Stoch. Stoch. Rep. 74 (2002), 429–464.
- [17] A. Guillin and F.-Y. Wang, *Degenerate Fokker-Planck equations: Bismut formula, gradient estimate and Harnack inequality*, J. Differential Equations 253 (2012), 20–40.
- [18] M. Kac, *Foundations of kinetic theory*, in: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, University California Press, Berkeley, 1956, 171–197.
- [19] N. El Karoui, S. Peng and M. C. Quenez, *Backward stochastic differential equations in finance*, Math. Finance 7 (1997), 1–71.
- [20] J. Li, *Mean-field forward and backward SDEs with jumps and associated nonlocal quasi-linear integral-PDEs*, Stochastic Process. Appl. 128 (2018), 3118–3180.

- [21] J. Ma, P. Protter and J. Yong, *Solving forward-backward stochastic differential equations explicitly—a four step scheme*, Probab. Theory Related Fields 98 (1994), 339–359.
- [22] J. Ma and J. Yong, *Forward-Backward Stochastic Differential Equations and Their Applications*, Lecture Notes in Math., 1702, Springer, Berlin, 1999.
- [23] J. Ma and J. Zhang, *Representation theorems for backward stochastic differential equations*, Ann. Appl. Probab. 12 (2002), 1390–1418.
- [24] F. Masiero, *A Bismut-Elworthy formula for quadratic BSDEs*, Stochastic Process. Appl. 125 (2015), 1945–1979.
- [25] T. Mastrolia, *Density analysis of non-Markovian BSDEs and applications to biology and finance*, Stochastic Process. Appl. 128 (2018), 897–938.
- [26] T. Mastrolia, D. Possamaï and A. Réveillac, *Density analysis of BSDEs*, Ann. Probab. 44 (2016), 2817–2857.
- [27] C. Olivera and E. Shamarova, *Gaussian density estimate for solutions of fully coupled forward-backward SDEs*, Math. Nachr. 293 (2020), 1554–1564.
- [28] E. Pardoux, *BSDE, weak convergence and homogenization of semilinear PDEs*, in: Nonlinear Analysis, Differential Equations and Control, eds. F. H. Clarke and R. J. Stern, Kluwer Academic Publishers, 1999, 505–549.
- [29] E. Pardoux and S. Peng, *Adapted solution of a backward stochastic differential equation*, Systems Control Lett. 14 (1990), 55–61.
- [30] E. Pardoux and S. Peng, *Backward stochastic differential equations and quasilinear parabolic partial differential equations and their applications*, in: Stochastic Partial Differential Equations and Their Applications, Lecture Notes in Control and Inform. Sci., 176, Springer, Berlin, 1992, 200–217.
- [31] P. Ren and F.-Y. Wang, *Bismut formula for Lions derivative of distribution dependent SDEs and applications*, J. Differential Equations 267 (2019), 4745–4777.
- [32] F.-Y. Wang, *Harnack Inequalities for Stochastic Partial Differential Equations*, Springer, Berlin, 2013.
- [33] F.-Y. Wang, *Derivative formula and gradient estimates for Grushin type semi-groups*, J. Theoret. Probab. 27 (2014), 80–95.
- [34] L. Wu, *Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems*, Stochastic Process. Appl. 91 (2001), 205–238.
- [35] J. Yong and X. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, Springer, Berlin, 1999.
- [36] J. Zhang, *Representation of solutions to BSDEs associated with a degenerate FSDE*, Ann. Appl. Probab. 15 (2005), 1798–1831.



- [37] J. Zhang, *Backward Stochastic Differential Equations—From Linear to Fully Non-linear Theory*, Springer, New York, 2017.
- [38] R. Zhu, *BSDE associated with generalized Dirichlet forms: finite dimensional case*, *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* 15 (2012), 1250022 (40 pages).
- [39] R. Zhu, *BSDE and generalized Dirichlet forms: the infinite dimensional case*, *Forum Math.* 27 (2015), 201–253.