The Annals of Applied Probability
2002, Vol. 12, No. 4, 1390–1418

REPRESENTATION THEOREMS FOR BACKWARD STOCHASTIC DIFFERENTIAL EQUATIONS

BY JIN MA[1] AND JIANFENG ZHANG

Purdue University and University of Minnesota

In this paper we investigate a class of backward stochastic differential equations (BSDE) whose terminal values are allowed to depend on the history of a forward diffusion. We first establish a probabilistic representation for the spatial gradient of the viscosity solution to a quasilinear parabolic PDE in the spirit of the Feynman–Kac formula, without using the derivatives of the coefficients of the corresponding BSDE. Such a representation then leads to a closed-form representation of the *martingale integrand* of a BSDE, under only a standard Lipschitz condition on the coefficients. As a direct consequence we prove that the paths of the martingale integrand of such BSDEs are at least càdlàg, which not only extends the existing path regularity results for solutions to BSDEs, but contains the cases where existing methods are not applicable. The BSDEs in this paper can be considered as the nonlinear wealth processes appearing in finance models; our results could lead to efficient Monte Carlo methods for computing both price and optimal hedging strategy for options with nonsmooth, path-dependent payoffs in the situation where the wealth is possibly nonlinear.

1. Introduction. Let $(\Omega, \mathcal{F}, P; \mathbf{F})$ be a complete, filtered probability space, where $\mathbf{F} \triangleq \{\mathcal{F}_t\}_{t\geq 0}$ is assumed to be the filtration generated by a standard, d-dimensional Brownian motion $W = \{W_t;\ t \geq 0\}$. Consider the following *backward stochastic differential equation* (BSDE):

$$Y_t = \xi + \int_t^T f(r, Y_r, Z_r)\,dr - \sum_{i=1}^d \int_t^T Z_r^i\,dW_r^i, \qquad t \in [0, T], \tag{1.1}$$

where $f : \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R}^d \mapsto \mathbb{R}$ is some appropriate measurable function, called the *generator* of the BSDE. An *adapted solution* to the BSDE (1.1) is a pair of $\mathbf{F}$-adapted, $\mathbb{R} \times \mathbb{R}^d$-valued processes $(Y, (Z^1, \dots, Z^d))$ that satisfies (1.1) almost surely. In this paper we call the process $Z = (Z^1, \dots, Z^d)$ the *martingale integrand* of the BSDE (1.1), for the obvious reason. Furthermore, we should note that here the process Z is defined as a *row vector* for notational simplicity. For instance, the stochastic integral in (1.1) can now be conveniently written as $\int_t^T Z_s\,dW_s$.

Received April 2000; revised March 2002.
[1]Supported in part by NSF Grant 99-71720.
AMS 2000 subject classifications. Primary 60H10; secondary 34F05, 90A12.
Key words and phrases. Backward SDE's, adapted solutions, anticipating stochastic calculus, viscosity solutions, path regularity.

The BSDEs of this kind, initiated by Bismut [2] and later developed by Pardoux and Peng [17], have been studied extensively in the past decade. We refer the readers to the books of El Karoui and Mazliak [4], Ma and Yong [15] and Yong and Zhou [22] and the survey paper of El Karoui, Peng and Quenez [5] for the detailed accounts of both theory and application (especially in mathematical finance and stochastic control) for such equations.

A well-investigated class of BSDEs is of the following form:

$$Y_t = g(X_T) + \int_t^T f(r, X_r, Y_r, Z_r)\,dr - \int_t^T Z_r\,dW_r, \qquad t \in [0, T], \tag{1.2}$$

where g and f are deterministic functions, and X satisfies an SDE:

$$X_t = x + \int_0^t b(r, X_r)\,dr + \int_0^t \sigma(r, X_r)\,dW_r, \qquad t \in [0, T], \tag{1.3}$$

where b and σ are some measurable functions. In this case, Pardoux and Peng proved in one of their seminal works [18], among other things, that the adapted solution Y gives a probabilistic representation of the (viscosity) solution of a quasilinear parabolic PDE (suppressing variables, and with slight abuse of notation):

$$\begin{aligned} &0 = u_t + \tfrac{1}{2}\mathrm{tr}\{\sigma\sigma^T u_{xx}\} + bu_x + f(t, x, u, u_x\sigma), \\ &u(T, x) = g(x), \qquad x \in \mathbb{R}^n. \end{aligned} \tag{1.4}$$

More precisely, assume that the functions f and g are Lipschitz continuous in their spatial variables, and define

$$u(t, x) \triangleq Y_t^{t,x} = E\left\{g(X_T^{t,x}) + \int_t^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})\,dr\right\}, \tag{1.5}$$

where $(X^{t,x}, Y^{t,x}, Z^{t,x})$ denotes the adapted solution to the SDE's (1.3) and (1.2), restricted to $[t, T]$ with $X_t^{t,x} = x$, a.s. Then, as a deterministic function thanks to the Blumenthal 0–1 law, $u(\cdot, \cdot)$ is a viscosity solution of the PDE (1.4). Furthermore, under more stringent regularity conditions on the coefficients (e.g., f and g are both C^3 in their spatial variables), it is shown in [18] that the process Z has continuous paths, a very appealing property for many applications. We remark that, by viewing (1.2) and (1.3) as a special (decoupled) case of the so-called *forward–backward* SDE's (FBSDE), one can also apply the results of Ma, Protter and Yong [14] to get an explicit expression of the solution (Y, Z):

$$Y_t = u(t, X_t), \qquad Z_t = \partial_x u(t, X_t)\sigma(t, X_t), \qquad t \in [0, T], \tag{1.6}$$

where u is the *classical* solution to the quasilinear PDE (1.4), verifying the conclusions of [18]. However, the results of [14] again require rather heavy smoothness conditions of the coefficients. It is noted that the BSDEs with nonsmooth coefficients have also been studied in recent years (see, e.g., [8, 13, 19], and [9–11]), but the main focus has been the existence of the adapted solutions. To the best of our knowledge, to date there has been no discussion in the literature concerning the path regularity of the process Z when f and g are only Lipschitz continuous, even in the special cases (1.3) and (1.2).

Our goal in this paper is twofold. First we show that if the coefficients f and g are continuously differentiable, then the viscosity solution u to the PDE (1.4) will have a continuous spatial gradient $\partial_x u$ and, more important, the following probablistic representation holds:

$$\partial_x u(t,x) = E\biggl\{g(X_T^{t,x})N_T^t + \int_t^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})N_r^t\,dr\biggr\}, \tag{1.7}$$

where $N_\cdot^t$ is some process defined on $[t,T]$, depending only on the forward diffusion and solution to its *variational equation*. This representation can be thought of as a new type of nonlinear Feynman–Kac formula for the derivative of the solution, which does not seem to exist in the literature. The main significance of the formula, however, lies in that it *does not* depend on the derivatives of the coefficients of the BSDE(!), a pleasant surprise in many ways. Because of this special feature, and with the help of the identity (1.6), we can then derive a similar representation for the martingale integrand Z, under only a Lipschitz condition on f and g. This latter representation then enables us to prove the path regularity of the process Z, the second goal of this paper, even in the case where the terminal value of Y is of the form $g(X_{t_0},\dots,X_{t_n})$, where $0 \le t_0 < \cdots < t_n \le T$ is any partition of $[0,T]$, a result that does not seem to be amendable by any existing method.

We would like to mention here that the main device of our proof is an *integration by parts* formula for anticipating stochastic integrals. Such an idea was recently employed in numerical finance for computing various "Greeks" of the market (cf. [6]). This paper is, in a sense, an attempt to extend their results to the market models in which the "wealth" process of an investor could be nonlinear (e.g., models involving higher interest rates for borrowing, taxes for capital gains, large investors or combinations of these features). Our results are expected to be potentially useful for developing an efficient Monte Carlo method along the lines of so-called Δ-hedging approach (see, e.g., [1]). We should also note that, although the results of this paper apply only to some special (discrete type) path-dependent options, it is possible to extend them to more general exotic options such as lookback options and Asian options, and even to those with discontinuous payoffs such as digital options. We shall address these issues in our future publications.

The rest of the paper is organized as follows. In Section 2 we give all the necessary preparations. In Section 3 we establish the relation between the SDEs (1.3), (1.2) and the quasilinear PDE (1.4), under only the C^1 assumption of the coefficients. In Section 4 we remove the C^1 assumption and give the main representation theorem. In Section 5 we study the path regularity of the process Z.

2. Preliminaries. Throughout this paper we assume that $(\Omega, \mathcal{F}, P)$ is a complete probability space on which is defined a d-dimensional Brownian motion $W = (W_t)_{t\geq 0}$. Let $\mathbf{F} \triangleq \{\mathcal{F}_t\}_{t\geq 0}$ denote the natural filtration generated by W, augmented by the P-null sets of $\mathcal{F}$; and let $\mathcal{F} = \mathcal{F}_\infty$. We let $\mathbb{E}$ denote a generic Euclidean space (or $\mathbb{E}_1, \mathbb{E}_2, \dots,$ if different spaces are used simultaneously); regardless of their dimensions we let $\langle \cdot, \cdot \rangle$ and $|\cdot|$ denote the inner product and norm in all $\mathbb{E}$'s, respectively. Furthermore, we use the notation $\partial_t = \frac{\partial}{\partial t}$, $\partial_x = (\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_d})$ and $\partial^2 = \partial_{xx} = (\partial^2_{x_i x_j})^d_{i,j=1}$, for $(t,x) \in [0,T] \times \mathbb{R}^d$. Note that if $\psi = (\psi^1, \dots, \psi^d) : \mathbb{R}^d \mapsto \mathbb{R}^d$, then $\partial_x \psi \triangleq (\partial_{x_j} \psi^i)^d_{i,j=1}$ is a matrix. The meaning of ∂_{xy}, ∂_{yy} etc. should be clear from the context.

The following spaces will be used frequently in the sequel (let $\mathcal{X}$ denote a generic Banach space):

1. For $t \in [0,T]$, $L^0([t,T]; \mathcal{X})$ is the space of all measurable functions $\varphi : [t,T] \mapsto \mathcal{X}$.
2. For $0 \leq t \leq T$, $C([t,T]; \mathcal{X})$ is the space of all continuous functions $\varphi : [t,T] \mapsto \mathcal{X}$; further, for any $p > 0$ we write $|\varphi|^{*,p}_{t,T} \triangleq \sup_{t\leq s\leq T} \|\varphi(s)\|^p_{\mathcal{X}}$ for all $\mathcal{X}$, when the context is clear.
3. For integers k and ℓ, $C^{k,\ell}([0,T] \times \mathbb{E}; \mathbb{E}_1))$ is the space of all $\mathbb{E}_1$-valued functions $\varphi(t,e)$, $(t,e) \in [0,T] \times \mathbb{E}$, such that they are k-times continuously differentiable in t and ℓ-times continuously differentiable in e.
4. $C^{k,\ell}_b([0,T] \times \mathbb{E}; \mathbb{E}_1)$ is the space of those $\varphi \in C^{k,\ell}([0,T] \times \mathbb{E}; \mathbb{E}_1)$ such that all the partial derivatives are uniformly bounded.
5. $W^{1,\infty}(\mathbb{E}; \mathbb{E}_1)$ is the space of all measurable functions $\psi : \mathbb{E} \mapsto \mathbb{E}_1$, such that for some constant $K > 0$ it holds that $\|\psi(x) - \psi(y)\|_{\mathbb{E}_1} \leq K\|x - y\|_{\mathbb{E}}, \forall x, y \in \mathbb{E}$.
6. For any sub-σ-field $\mathcal{G} \subseteq \mathcal{F}_T$ and $0 \leq p < \infty$, $L^p(\mathcal{G}; \mathbb{E})$ denotes all $\mathbb{E}$-valued, $\mathcal{G}$-measurable random variable ξ such that $E|\xi|^p < \infty$; moreover, $\xi \in L^\infty(\mathcal{G}; \mathbb{E})$ means it is $\mathcal{G}$-measurable and bounded.
7. For $0 \leq p < \infty$, $L^p(\mathbf{F}, [0,T]; \mathcal{X})$ is the space of all $\mathcal{X}$-valued, $\mathbf{F}$-adapted processes ξ satisfying $E\int_0^T \|\xi_t\|^p_{\mathcal{X}}\, dt < \infty$; also, $\xi \in L^\infty(\mathbf{F}, [0,T]; \mathbb{R}^d)$ means it is a process uniformly bounded in (t, ω).
8. $C(\mathbf{F}, [0,T] \times \mathbb{E}; \mathbb{E}_1))$ is the space of all $\mathbb{E}_1$-valued, continuous random fields $\varphi : \Omega \times [0,T] \times \mathbb{E} \mapsto \mathbb{E}_1$, such that, for fixed $e \in \mathbb{E}$, $\varphi(\cdot, \cdot, e)$ is an $\mathbf{F}$-adapted process.

To simplify notation we often write $C([0,T] \times \mathbb{E}; \mathbb{E}_1) = C^{0,0}([0,T] \times \mathbb{E}; \mathbb{E}_1)$; and if $\mathbb{E}_1 = \mathbb{R}$, then we often suppress $\mathbb{E}_1$ for simplicity [e.g., $C^{k,\ell}([0,T] \times \mathbb{E}; \mathbb{R}) = C^{k,\ell}([0,T] \times \mathbb{E})$, $C(\mathbf{F}, [0,T] \times \mathbb{E}; \mathbb{R}) = C(\mathbf{F}, [0,T] \times \mathbb{E}), \dots,$ etc.]. Finally, unless otherwise specified (such as process Z mentioned in Section 1), all vectors in the paper will be regarded as *column* vectors.

Throughout this paper we shall make use of the following *standing assumptions*:

(A1) $n = d$. The functions $\sigma \in C_b^{0,1}([0,T] \times \mathbb{R}^d; \mathbb{R}^{d\times d})$ and $b \in C_b^{0,1}([0,T] \times \mathbb{R}^d; \mathbb{R}^d)$ and all the partial derivatives of b and σ (with respect to x) are uniformly bounded by a common constant $K > 0$. Further, there exists constant $c > 0$, such that

$$\xi^T \sigma(t,x)\sigma^T(t,x)\xi \geq c|\xi|^2 \qquad \forall x, \xi \in \mathbb{R}^d;\ t \in [0,T]. \tag{2.1}$$

(A2) The functions $f \in C([0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d) \cap L^0([0,T]; W^{1,\infty}(\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d))$ and $g \in W^{1,\infty}(\mathbb{R}^d)$. Furthermore, we denote the Lipschitz constants of f and g by a common one $K > 0$ as in (A1); and we assume that

$$\sup_{0\leq t\leq T} \{|b(t,0)| + |\sigma(t,0)| + |f(t,0,0,0)|\} + |g(0)| \leq K. \tag{2.2}$$

The following results are either standard or slight variations of the well-known results in the SDE and the backward SDE literature; we give only the statements for ready reference.

LEMMA 2.1. *Suppose that* $\widetilde{b} \in C(\mathbf{F}, [0,T] \times \mathbb{R}^d; \mathbb{R}^d) \cap L^0(\mathbf{F}, [0,T]; W^{1,\infty}(\mathbb{R}^d; \mathbb{R}^d))$, $\widetilde{\sigma} \in C(\mathbf{F}, [0,T] \times \mathbb{R}^d; \mathbb{R}^{d\times d}) \cap L^0(\mathbf{F}, [0,T]; W^{1,\infty}(\mathbb{R}^d; \mathbb{R}^{d\times d}))$, *with a common Lipschitz constant* $K > 0$. *Suppose also that* $\widetilde{b}(t,0) = 0$ *and* $\widetilde{\sigma}(t,0) = 0$, *P-a.s. For any* $h^0 \in L^2(\mathbf{F}, [0,T]; \mathbb{R}^d)$ *and* $h^1 \in L^2(\mathbf{F}, [0,T]; \mathbb{R}^{d\times d})$, *let* X *be the solution of the following SDE*:

$$X_t = x + \int_0^t [\widetilde{b}(s, X_s) + h_s^0]\, ds + \int_0^t [\widetilde{\sigma}(s, X_s) + h_s^1]\, dW_s. \tag{2.3}$$

Then, for any $p \geq 2$, *there exists a constant* $C > 0$ *depending only on* p, T *and* K, *such that*

$$E|X|_{t,T}^{*,p} \leq C\biggl\{|x|^p + E\int_0^T [|h_t^0|^p + |h_t^1|^p]\, dt\biggr\}. \tag{2.4}$$

LEMMA 2.2. *Assume that* $\widetilde{f} \in C(\mathbf{F}, [0,T] \times \mathbb{R} \times \mathbb{R}^d) \cap L^0(\mathbf{F}, [0,T]; W^{1,\infty}(\mathbb{R} \times \mathbb{R}^d))$ *with a uniform Lipschitz constant* $K > 0$, *and that* $\widetilde{f}(\omega, s, 0, 0) = 0$, P*-a.e.* $\omega \in \Omega$. *For any* $\xi \in L^2(\mathcal{F}_T; \mathbb{R})$ *and* $h \in L^2(\mathbf{F}, [0,T]; \mathbb{R})$, *let* (Y, Z) *be the adapted solution to the BSDE*

$$Y_t = \xi + \int_t^T [\widetilde{f}(s, Y_s, Z_s) + h_s]\, ds - \int_t^T Z_s\, dW_s. \tag{2.5}$$

Then there exists a constant $C > 0$ *depending only on* T *and the Lipschitz constant of* $\widetilde{f}$, *such that*

$$E\int_0^T |Z_t|^2\, dt \leq CE\biggl\{|\xi|^2 + \int_0^T |h_t|^2\, dt\biggr\}. \tag{2.6}$$

Moreover, for all $p \geq 2$, *there exists a constant* $C_p > 0$, *such that*

$$E|Y|_{t,T}^{*,p} \leq C_p E\biggl\{|\xi|^p + \int_0^T |h_t|^p\, dt\biggr\}. \tag{2.7}$$

REMARK 2.3. We should note that in Lemmas 2.1 and 2.2 we assume only that the processes $h^0 \in L^2(\mathbf{F}, [0, T]; \mathbb{R}^d)$ and $h^1 \in L^2(\mathbf{F}, [0, T]; \mathbb{R}^{d\times d})$ guarantee the solvability of the FSDE (2.3) and the BSDE (2.5). However, the estimates (2.4) and (2.7) obviously hold if $p \geq 2$ is such that the right-hand side is ∞. The proof of Lemma 2.1 can be found in, for example, [7], while the estimates (2.6) and (2.7) are a slight modification of Theorem 5.1 of El Karoui, Peng and Quenez [5].

We now review some basic facts of the anticipating stochastic calculus, especially those related to the SDEs. We refer the readers to Nualart [16] for the basic theory and to Pardoux and Peng [18] for the results related to BSDEs. To begin with, let $\mathcal{S}$ be the space of all random variables of the form

$$\xi = F\biggl(\int_0^T \varphi_1(t)\, dW_t, \dots, \int_0^T \varphi_n(t)\, dW_t\biggr),$$

where $F \in C_b^\infty(\mathbb{R}^n)$ and $\varphi_1, \dots, \varphi_n \in L^2([0, T]; \mathbb{R}^d)$. To simplify notation later, we make the convention here that all φ_i's are *row vectors*.

We call a mapping $D : \mathcal{S} \mapsto L^2([0, T] \times \Omega)$ the *derivative operator* if, for each $\xi \in \mathcal{S}$ and $t \in [0, T]$,

$$D_t\xi = \sum_{i=1}^n \frac{\partial F}{\partial x_i}\biggl(\int_0^T \varphi_1(t)\, dW_t, \dots, \int_0^T \varphi_n(t)\, dW_t\biggr)\varphi_i(t).$$

Next, we introduce a norm on $\mathcal{S}$:

$$\|\xi\|_{1,2}^2 = E|\xi|^2 + E\int_0^T |D_r\xi|^2\, dr \qquad \forall \xi \in \mathcal{S},$$

and we let $\mathbb{D}^{1,2}$ denote the completion of $\mathcal{S}$ in $L^2(\Omega)$ under $\|\cdot\|_{1,2}$. It can be shown (see, e.g., [16]) that D is a densely defined, closed linear operator from $\mathbb{D}^{1,2}$ to $L^2(\Omega \times [0, T])$ with domain $\mathbb{D}^{1,2}$.

To apply the anticipating stochastic calculus to SDEs (1.2) and (1.3), we consider these equations on the subinterval $[t, T] \subseteq [0, T]$: for $s \in [t, T]$,

$$\begin{aligned} X_s^{t,x} &= x + \int_t^s b(r, X_r^{t,x})\, dr + \int_t^s \sigma(r, X_r^{t,x})\, dW_r, \\ Y_s^{t,x} &= g(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})\, dr - \int_s^T Z_r^{t,x}\, dW_r. \end{aligned} \tag{2.8}$$

Here the superscript ^{t,x} indicates the dependence of the solution on the initial date (t, x), and it will be omitted when the context is clear. The following *variational* *equation* of (2.8) will play an important role in this paper: for $i = 1, \dots, d$,

$$\begin{aligned}
\nabla_i X_s &= e_i + \int_t^s \partial_x b(X_r)\nabla_i X_r\,dr + \sum_{j=1}^d \int_t^s [\partial_x \sigma^j(X_r)]\nabla_i X_r\,dW_r^j,\\
\nabla_i Y_s &= \partial_x g(X_T)\nabla_i X_T\\
&\qquad + \int_s^T \big[\partial_x f\big(r,\Theta(r)\big)\nabla_i X_r + \partial_y f\big(r,\Theta(r)\big)\nabla_i Y_r\\
&\qquad\qquad + \big\langle \partial_z f\big(r,\Theta(r)\big), \nabla_i Z_r\big\rangle\big]\,dr - \int_s^T \nabla_i Z_r\,dW_r,
\end{aligned}\tag{2.9}$$

where $e_i = (0, \dots, \overset{i}{1}, \dots, 0)^T$ is the ith coordinate vector of $\mathbb{R}^d$; $\sigma^j(\cdot)$ is the jth column of the matrix $\sigma(\cdot)$; $\Theta(r)$ denotes (X_r, Y_r, Z_r). Further, we denote

$$\nabla X = (\nabla_1 X, \dots, \nabla_d X), \qquad \nabla Y = (\nabla_1 Y, \dots, \nabla_d Y), \qquad \nabla Z = \begin{bmatrix} \nabla_1 Z \\ \vdots \\ \nabla_d Z \end{bmatrix}^T.$$

Then $(\nabla X, \nabla Y, \nabla Z) \in L^2(\mathbf{F}; C([0,T];\mathbb{R}^{d\times d}) \times C([0,T];\mathbb{R}^d) \times L^2([0,T];\mathbb{R}^{d\times d}))$.

Note that the $d \times d$-matrix-valued process ∇X satisfies a linear SDE and $\nabla X_t = I$, thus $[\nabla X_s]^{-1}$ exists for all $s \in [t,T]$, P-a.s. The following lemma concerns the anticipating (Malliavin) derivatives of the solution (X,Y,Z) to (2.8). Since the proof is standard and can be found in, for example, Nualart [16] and Pardoux and Peng [18], we omit it.

LEMMA 2.4. *Assume that* (A1) *holds, and suppose that* $f \in C_b^{0,1}([0,T]\times\mathbb{R}^{2d+1})$ *and that* $g \in C_b^1(\mathbb{R}^d)$. *Then* $(X,Y,Z) \in L^2([0,T];\mathbb{D}^{1,2}(\mathbb{R}^{2d+1}))$, *and there exists a version of* (D_sX_r, D_sY_r, D_sZ_r) *that satisfies*

$$\begin{aligned}
D_s X_r &= \nabla X_r (\nabla X_s)^{-1}\sigma(s, X_s)1_{\{s\le r\}},\\
D_s Y_r &= \nabla Y_r (\nabla X_s)^{-1}\sigma(s, X_s)1_{\{s\le r\}}, \qquad t \le s, r \le T,\\
D_s Z_r &= \nabla Z_r (\nabla X_s)^{-1}\sigma(s, X_s)1_{\{s\le r\}},
\end{aligned}\tag{2.10}$$

where

$$D_s X_r \triangleq \begin{bmatrix} D_s X_r^1 \\ \vdots \\ D_s X_r^d \end{bmatrix} \quad \textit{and} \quad D_s Z_r \triangleq \begin{bmatrix} D_s Z_r^1 \\ \vdots \\ D_s Z_r^d \end{bmatrix}.$$

To conclude this section let us introduce the notion of *anticipating stochastic integral* (also known as *Skorohod integral* or *Hitsuda–Skorohod integral*), which will be one of the key devices in this paper. Recall the derivative operator D is a closed, densely defined operator from $L^2(\Omega)$ to $L^2(\Omega \times [0, T])$; we can define its *adjoint operator* $\delta : \mathrm{Dom}(\delta) \subset L^2(\Omega \times [0, T]; \mathbb{R}^d) \mapsto L^2(\Omega; \mathbb{R})$ by

$$E\{F\delta(u)\} = E\int_0^T D_t F u_t \, dt \qquad \forall F \in \mathbb{D}^{1,2},\ \forall u \in \mathrm{Dom}(\delta), \tag{2.11}$$

where $\mathrm{Dom}(\delta) \triangleq \{u \in L^2(\Omega \times [0, T]) : |E\int_0^T D_t F u_t \, dt| \le C\|F\|_{1,2},\ \forall F \in \mathbb{D}^{1,2}\}$. The operator δ is then called the anticipating stochastic integral of the process u. The definition can be extended in an obvious way to the case when u is vector-valued; and by a slight abuse of notation, we still denote it as

$$\delta(u) = \int_0^T \langle u_t, dW_t\rangle, \qquad u \in \mathrm{Dom}(\delta). \tag{2.12}$$

One should keep in mind that, if in the sequel the integrand of a stochastic integral is not $\mathbf{F}$-adaped, then it should always be understood as an anticipating stochastic integral. On the other hand, it can be shown that, if $u \in L^2(\mathbf{F}, [0, T]; \mathbb{R}^d)$, then $u \in \mathrm{Dom}(\delta)$, and the anticipating stochastic integral (2.12) coincides with the usual Itô integral. Furthermore, we have the following important properties of such integrals (cf. [16]).

LEMMA 2.5. *Suppose that $F \in \mathbb{D}^{1,2}$. Then the following hold*:

(i) (*integration by parts formula*) *for any $u \in \mathrm{Dom}(\delta)$ such that $Fu \in L^2([0, T] \times \Omega; \mathbb{R}^d)$, one has $Fu \in \mathrm{Dom}(\delta)$, and it holds that*

$$\int_0^T \langle Fu_t, dW_t\rangle = \delta(Fu) = F\int_0^T \langle u_t, dW_t\rangle - \int_0^T D_t F u_t \, dt;$$

(ii) (*Clark–Haussmann–Ocone formula*)

$$F = E(F) + \int_0^T E\{D_t F | \mathcal{F}_t\} \, dW_t.$$

3. Relations to PDEs revisited. In this section we prove the relation (1.6) between the FBSDE (2.8) and the quasilinear parabolic PDE (1.4), under the condition that the coefficients f and g are only continuously differentiable. We should note that such a relation is well understood when the coefficients are regular enough [e.g., f and g are both C^3 in (x, y, z); see, e.g., [18] or [14]). On the other hand, in the case when f and g are only Lipschitz continuous, it is known that $u(t, x) \triangleq Y_t^{t,x}$ is a *viscosity solution* of (1.4) (see, e.g., [18] or [15]). However, in that case the second relation in (1.6) becomes questionable. The following result is therefore interesting in its own right, and to the best of our knowledge it is new.

THEOREM 3.1. *Assume* (A1) *and suppose that $f \in C_b^{0,1}([0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d)$ and $g \in C_b^1(\mathbb{R}^d)$. Let $(X^{t,x}, Y^{t,x}, Z^{t,x})$ be the adapted solution to the FBSDE* (2.8), *and define $u(t, x) = Y_t^{t,x}$. Then the following hold*:

(i) $\partial_x u$ *exists for all* $(t,x) \in [0,T] \times \mathbb{R}^d$; *and for each* (t,x) *and* $i = 1, \dots, d$, *the following representation holds*:

$$\begin{aligned}\partial_{x_i} u(t,x) = E\bigg\{&\partial_x g(X_T^{t,x})\nabla_i X_T \\ &+ \int_t^T \big[\partial_x f\big(r, \Theta^{t,x}(r)\big)\nabla_i X_r \\ &\qquad + \partial_y f\big(r, \Theta^{t,x}(r)\big)\nabla_i Y_r + \big\langle \partial_z f\big(r, \Theta^{t,x}(r)\big), \nabla_i Z_r\big\rangle\big]\,dr\bigg\},\end{aligned} \tag{3.1}$$

where $\Theta^{t,x}(r) \triangleq (X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})$, *and* $(\nabla X, \nabla Y, \nabla Z)$ *is the solution to the variational equation* (2.9);

(ii) $\partial_x u$ *is continuous on* $[0,T] \times \mathbb{R}^d$;

(iii) $Z_s^{t,x} = \partial_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x})$, $\forall s \in [t,T]$, P-*a.s.*

PROOF. To simplify presentation, we shall prove only the case when $d = 1$, as the higher dimensional case can be treated in the same way without substantial difficulty. Also in what follows we use the simpler notation g_x and (f_x, f_y, f_z) for the partial derivatives of g and f.

We first prove (i). Let $(t,x) \in [0,T] \times \mathbb{R}^d$ be fixed. For any $h \neq 0$, define

$$\nabla X^h = \frac{X_s^{t,x+h} - X_s^{t,x}}{h}, \qquad \nabla Y^h = \frac{Y_s^{t,x+h} - Y_s^{t,x}}{h}, \qquad \nabla Z^h = \frac{Z_s^{t,x+h} - Z_s^{t,x}}{h}.$$

It is standard (see, e.g., [7]) to show that

$$E|\nabla X^h - \nabla X|_{t,T}^{*,2} \to 0 \qquad \text{as } h \to 0. \tag{3.2}$$

To check the limits of ∇Y^h and ∇Z^h we note that, for each $s \in [0,T]$,

$$\begin{aligned}\nabla Y_s^h = \widetilde{g}_x^h(T)\nabla X_T^h &+ \int_s^T \big[\widetilde{f}_x^h(r)\nabla X_r^h + \widetilde{f}_y^h(r)\nabla Y_r^h + \widetilde{f}_z^h(r)\nabla Z_r^h\big]\,dr \\ &- \int_s^T \nabla Z_r^h\,dW_r,\end{aligned} \tag{3.3}$$

where, with $\Theta^{t,x} = (X^{t,x}, Y^{t,x}, Z^{t,x})$,

$$\begin{aligned}
\widetilde{g}_x^h(T) &= \int_0^1 g_x(X_T^{t,x} + \theta(X_T^{t,x+h} - X_T^{t,x}))\,d\theta, &\qquad \widetilde{g}_x^0(T) &= g_x(X_T^{t,x}),\\
\widetilde{f}_x^h(r) &= \int_0^1 f_x(r, \Theta_r^{t,x} + \theta(\Theta_r^{t,x+h} - \Theta_r^{t,x}))\,d\theta, &\qquad \widetilde{f}_x^0(r) &= f_x(r, \Theta_r^{t,x}),\\
\widetilde{f}_y^h(r) &= \int_0^1 f_y(r, \Theta_r^{t,x} + \theta(\Theta_r^{t,x+h} - \Theta_r^{t,x}))\,d\theta, &\qquad \widetilde{f}_y^0(r) &= f_y(r, \Theta_r^{t,x}),\\
\widetilde{f}_z^h(r) &= \int_0^1 f_z(r, \Theta_r^{t,x} + \theta(\Theta_r^{t,x+h} - \Theta_r^{t,x}))\,d\theta, &\qquad \widetilde{f}_z^0(r) &= f_z(r, \Theta_r^{t,x}).
\end{aligned}$$

Applying Lemma 2.2 to (3.3) we see that for all $p \geq 2$ one has

$$E\left\{|\nabla Y^h|_{t,T}^{*,p} + \int_t^T |\nabla Z_s^h|^2\,ds\right\} \leq C,$$

where $C > 0$ is some constant independent of h. Therefore

$$\lim_{h\to 0} E\left\{|Y^{t,x+h} - Y^{t,x}|_{t,T}^{*,p} + \int_t^T |Z_s^{t,x+h} - Z_s^{t,x}|^2\,ds\right\} = 0 \qquad \forall\, p \geq 2. \tag{3.4}$$

Consequently, for all $p \geq 2$,

$$\begin{aligned} &E|\widetilde{g}_x^h(T) - \widetilde{g}_x^0(T)|^p \to 0,\\ &E\int_0^T |\widetilde{\varphi}^h(r) - \widetilde{\varphi}^0(r)|^p\,dr \to 0 \qquad \text{as } h \to 0, \end{aligned} \tag{3.5}$$

for $\varphi = f_x, f_y, f_z$, respectively. We now show that

$$E\left\{|\nabla Y^h - \nabla Y|_{t,T}^{*,2} + \int_t^T |\nabla Z_s^h - \nabla Z_s|^2\,ds\right\} \to 0 \qquad \text{as } h \to 0, \tag{3.6}$$

where $(\nabla Y, \nabla Z)$ is the solution to the (backward) variational equation in (2.9). To do this we define $\Delta X_s^h \triangleq \nabla X_s^h - \nabla X_s$, $\Delta Y_s^h \triangleq \nabla Y_s^h - \nabla Y_s$ and $\Delta Z_s^h = \nabla Z_s^h - \nabla Z_s$. Combining (3.3) and (2.9) we have

$$\begin{aligned} \Delta Y_s^h &= \widetilde{g}^h(T)\Delta X_T^h + \big(\widetilde{g}_x^h(T) - \widetilde{g}_x^0(T)\big)\nabla X_T\\ &\quad + \int_s^T \big[\widetilde{f}_x^h(r)\Delta X_r^h + \widetilde{f}_y^h(r)\Delta Y_r^h + \widetilde{f}_z^h(r)\Delta Z_r^h + \varepsilon^h(r)\big]\,dr\\ &\quad - \int_s^T \Delta Z_r^h\,dW_r, \end{aligned} \tag{3.7}$$

for $s \in [t, T]$, where

$$\begin{aligned} \varepsilon^h(r) \triangleq{}& [\widetilde{f}_x^h(r) - \widetilde{f}_x^0(r)]\nabla X_r + [\widetilde{f}_y^h(r) - \widetilde{f}_y^0(r)]\nabla Y_r\\ &+ [\widetilde{f}_z^h(r) - \widetilde{f}_z^0(r)]\nabla Z_r. \end{aligned} \tag{3.8}$$

Next, applying Lemma 2.2 to (3.7) we have

$$\begin{aligned} &E\left\{|\Delta Y^h|_{t,T}^{*,2} + \int_t^T |\Delta Z_s^h|^2\,ds\right\}\\ &\qquad \leq CE\left\{|\Delta X_T^h|^2 + |\widetilde{g}_x^h(T) - \widetilde{g}_x^0(T)|^2|\nabla X_T|^2 + \int_t^T \big(|\Delta X_r^h|^2 + |\varepsilon^h(r)|^2\big)\,dr\right\}; \end{aligned}$$

thus (3.6) follows from (3.2), (3.5), definition (3.8) and the dominated convergence theorem. In particular, since f and g are deterministic, the processes $Y^{t,x}$, $Y^{t,x+h}$, ∇Y^h and ΔY^h are all adapted to the filtration $\{\mathcal{F}_s^t\}_{s\geq t}$, where $\mathcal{F}_s^t = \sigma\{W_u - W_t;\ t \leq u \leq s\}$, with the usual P-augmentation. Thus by the Blumenthal 0–1 law (see, e.g., [12]), the quantities $u(t,x) = Y_t^{t,x}$, $u(t,x+h) = Y_t^{t,x+h}$, $\nabla_i Y_t^h = \frac{1}{h}[u(t,x+h) - u(t,x)]$ and ΔY_t^h are all deterministic. Therefore, we conclude from the above that $\partial_x u$ exists, and $\partial_x u(t,x) = \nabla Y_t$, for all (t,x). Finally, taking the expectation on both sides of (2.9) at $s = t$ we obtain the representation (3.1).

We now prove (ii). Let $(t_i, x_i) \in [0,T] \times \mathbb{R}^d$, $i = 1,2$. We first assume that $t_1 < t_2$. To simplify notation we write, for $i = 1,2$ and $r \in [0,T]$,

$$
\begin{aligned}
\Theta^i &= (X^i, Y^i, Z^i) = (X^{t_i,x_i}, Y^{t_i,x_i}, Z^{t_i,x_i}),\\
\nabla\Theta^i &= (\nabla X^i, \nabla Y^i, \nabla Z^i),\\
f_x^i(r) &= \partial_x f\big(r, \Theta^i(r)\big), \qquad f_y^i(r) = \partial_y f\big(r, \Theta^i(r)\big),\\
f_z^i(r) &= \partial_z f\big(r, \Theta^i(r)\big),\\
g_x^i &= \partial_x g(X_T^i), \qquad b_x^i(r) = \partial_x b\big(r, X^i(r)\big),\\
\sigma_x^i(r) &= \partial_x \sigma\big(r, X^i(r)\big).
\end{aligned}
\tag{3.9}
$$

Further, we write $\widetilde{\Delta} X_r = \nabla X_r^1 - \nabla X_r^2$, $\widetilde{\Delta} Y_r = \nabla Y_r^1 - \nabla Y_r^2$, $\widetilde{\Delta} Z_r = \nabla Z_r^1 - \nabla Z_r^2$, and for any function φ we let $\widetilde{\Delta}_{12}[\varphi] = \varphi^1 - \varphi^2$. Also, to simplify notation in what follows we let $C > 0$ denote a generic constant depending only on K [in (A1), (A2)] and T, and we allow it to vary from line to line. Recalling (3.1) we have

$$
\begin{aligned}
&|\partial_x u(t_1,x_1) - \partial_x u(t_2,x_2)|\\
&\quad\le E\{|g_x^1 \nabla X_T^1 - g_x^2 \nabla X_T^2|\}\\
&\qquad + E\left\{\int_{t_1}^{t_2} \big[|f_x^1(r)|\,|\nabla X_r^1| + |f_y^1(r)|\,|\nabla Y_r^1| + |f_z^1(r)|\,|\nabla Z_r^1|\big]\,dr\right\}\\
&\qquad + E\left\{\int_{t_2}^{T} \big[|\widetilde{\Delta}_{12}[f_x \nabla X.](r)| + |\widetilde{\Delta}_{12}[f_y \nabla Y.](r)| + |\widetilde{\Delta}_{12}[f_z \nabla Z.](r)|\big]\,dr\right\}\\
&\quad\le CE\bigg\{|\widetilde{\Delta} X_T| + |(\nabla X_T^2)|\,|\widetilde{\Delta}_{12}[g_x]| + \int_{t_1}^{t_2} \big[|\nabla X_r^1| + |\nabla Y_r^1| + |\nabla Z_r^1|\big]\,dr\\
&\qquad\qquad + \int_{t_2}^{T} [|\widetilde{\Delta} X_r| + |\widetilde{\Delta} Y_r| + |\widetilde{\Delta} Z_r|]\,dr\\
&\qquad\qquad + \int_{t_2}^{T} \big[|\widetilde{\Delta}_{12}[f_x](r) \nabla X_r^2| + |\widetilde{\Delta}_{12}[f_y](r) \nabla Y_r^2|\\
&\qquad\qquad\qquad\qquad + |\widetilde{\Delta}_{12}[f_z](r) \nabla Z_r^2|\big]\,dr\bigg\}.
\end{aligned}
\tag{3.10}
$$

To estimate the right-hand side of (3.10) we note that the process $(\widetilde{\Delta}X, \widetilde{\Delta}Y, \widetilde{\Delta}Z)$ satisfies the following FBSDE (for $s \in [t_2, T]$):

$$
\begin{aligned}
\widetilde{\Delta}X_s &= (\nabla X^1_{t_2} - 1) + \int_{t_2}^s \bigl[b^1_x(r)\widetilde{\Delta}X_r + \widetilde{\Delta}_{12}[b_x](r)\nabla X^2_r\bigr]\,dr \\
&\quad + \int_{t_2}^s \bigl[\sigma^1_x(r)\widetilde{\Delta}X_r + \widetilde{\Delta}_{12}[\sigma_x](r)\nabla X^2_r\bigr]\,dW_r, \\
\widetilde{\Delta}Y_s &= g^1_x\widetilde{\Delta}X_T + \widetilde{\Delta}_{12}[g_x]\nabla X^2_T \\
&\quad + \int_s^T \bigl[f^1_x(r)\widetilde{\Delta}X_r + f^1_y(r)\widetilde{\Delta}Y_r + f^1_z(r)\widetilde{\Delta}Z_r + \varepsilon(r)\bigr] \\
&\quad - \int_s^T \widetilde{\Delta}Z_r\,dW_r,
\end{aligned}
\tag{3.11}
$$

where $\varepsilon(r) = \widetilde{\Delta}_{12}[f_x](r)\nabla X^2_r + \widetilde{\Delta}_{12}[f_y](r)\nabla Y^2_r + \widetilde{\Delta}_{12}[f_z](r)\nabla Z^2_r$, $r \in [s, T]$.

Now let $G_{t_1,t_2}(\cdot)$ denote a generic $\mathbf{F}$-adapted, continuous process that is uniformly bounded and satisfies $\lim_{t_1 \uparrow t_2} G_{t_1,t_2}(r) = 0$, $\forall r \in [t_2, T]$, P-a.s. Again, we allow it to vary from line to line (e.g., all $\widetilde{\Delta}_{12}[\varphi](\cdot)$, where $\varphi = b_x, \sigma_x, f_x, f_y, f_z$ can be denoted as such). Applying Lemma 2.1 and recalling the convention on the constant C and the assumptions on b and σ we get

$$
\begin{aligned}
E|\widetilde{\Delta}X|^{*,2}_{t,T} &\le CE\biggl\{|\nabla X^1_{t_2} - 1|^2 + \int_{t_2}^T \bigl[|\widetilde{\Delta}_{12}[b_x](s)|^2 + |\widetilde{\Delta}_{12}[\sigma_x](s)|^2\bigr]|\nabla X^2_s|^2\,ds\biggr\} \\
&\le CE\biggl\{|\nabla X^1_{t_2} - 1|^2 + \int_{t_2}^T G_{t_1,t_2}(s)|\nabla X^2_s|^2\,ds\biggr\}.
\end{aligned}
\tag{3.12}
$$

Combining (3.12) with Lemma 2.2 we have

$$
\begin{aligned}
&E\biggl\{|\widetilde{\Delta}Y|^{*,2}_{t,T} + \int_{t_2}^T |\widetilde{\Delta}Z_s|^2\,ds\biggr\} \\
&\qquad \le CE\biggl\{|\widetilde{\Delta}X_T|^2 + |\widetilde{\Delta}_{12}[g_x]|^2|\nabla X^2_T|^2 \\
&\qquad\qquad + \int_{t_2}^T \bigl[|\widetilde{\Delta}X_r|^2 + \bigl(|\widetilde{\Delta}_{12}[f_x](r)|^2 + |\widetilde{\Delta}_{12}[f_y](r)|^2 \\
&\qquad\qquad\qquad + |\widetilde{\Delta}_{12}[f_z](r)|^2\bigr)|\nabla\Theta^2_r|^2\bigr]\,dr\biggr\} \\
&\qquad \le CE\biggl\{|\widetilde{\Delta}_{12}[g_x]|^2|\nabla X^2_T|^2 + |\nabla X^1_{t_2} - 1|^2 + \int_{t_2}^T G_{t_1,t_2}(r)|\nabla\Theta^2_r|^2\,dr\biggr\}.
\end{aligned}
\tag{3.13}
$$

Plugging (3.12) and (3.13) into (3.10) we obtain that

$$\begin{aligned}
&|\partial_x u(t_1,x_1)-\partial_x u(t_2,x_2)|^2 \\
&\quad \le CE\biggl\{|\nabla X^1_{t_2}-1|^2+|\widetilde{\Delta}_{12}[g_x]|^2|\nabla X^2_T|^2 \\
&\qquad\qquad + (t_2-t_1)\int_{t_1}^{t_2}|\nabla\Theta^1_r|^2\,dr+\int_{t_2}^{T}G_{t_1,t_2}(r)|\nabla\Theta^2_r|^2\,dr\biggr\}.
\end{aligned}\tag{3.14}$$

Now for fixed (t_2,x_2), by the dominated convergence theorem we easily derive that

$$\lim_{\substack{t_1\uparrow t_2\\ x_1\to x_2}}|\partial_x u(t_1,x_1)-\partial_x u(t_2,x_2)|^2=0.$$

Similarly we can show that, for fixed (t_1,x_1),

$$\lim_{\substack{t_2\downarrow t_1\\ x_2\to x_1}}|\partial_x u(t_1,x_1)-\partial_x u(t_2,x_2)|^2=0.$$

This proves (ii).

It remains to prove (iii). For a continuous function φ, let $\{\varphi^\varepsilon\}_{\varepsilon>0}$ denote the family of C_0^∞ functions that converge to φ uniformly. Since b,σ,f,g are all uniformly Lipschitz continuous, we may assume that the first order partial derivatives of $b^\varepsilon,\sigma^\varepsilon,f^\varepsilon,g^\varepsilon$ are all uniformly bounded, uniformly in $\varepsilon>0$. To simplify notation in what follows we drop the superscript ^{t,x} from the solution $\Theta^{t,x}$. Consider the family of FBSDEs parametrized by $\varepsilon>0$:

$$\begin{aligned}
X_s&=x+\int_t^s b^\varepsilon(r,X_r)\,dr+\int_s^t\sigma^\varepsilon(r,X_r)\,dW_r,\\
Y_s&=g^\varepsilon(X_T)+\int_s^T f^\varepsilon(r,X_r,Y_r,Z_r)\,dr-\int_s^T Z_r\,dW_r.
\end{aligned}\tag{3.15}$$

Let us denote the solution by $(X^{t,x}(\varepsilon),Y^{t,x}(\varepsilon),Z^{t,x}(\varepsilon))$ and define $u^\varepsilon(t,x)=Y^{t,x}_t(\varepsilon)$. Applying Theorem 3.2 of [18] we see that u^ε is the classical solution of a PDE (suppressing all variables):

$$\begin{aligned}
&u^\varepsilon_t+\tfrac12\mathrm{tr}\{\sigma^\varepsilon(\sigma^\varepsilon)^T\partial^2_{xx}u^\varepsilon\}+\langle\partial_x u^\varepsilon,b^\varepsilon\rangle+f^\varepsilon(t,x,u^\varepsilon,\partial_x u^\varepsilon\sigma^\varepsilon)=0,\\
&\qquad\qquad\qquad\qquad\qquad\qquad u^\varepsilon(T,x)=g^\varepsilon(x).
\end{aligned}\tag{3.16}$$

Now for any $\{x^\varepsilon\}\subset\mathbb{R}^n$ such that $x^\varepsilon\to x$ as $\varepsilon\to0$, define $(X^\varepsilon,Y^\varepsilon,Z^\varepsilon)\triangleq(X^{t,x^\varepsilon}(\varepsilon),Y^{t,x^\varepsilon}(\varepsilon),Z^{t,x^\varepsilon}(\varepsilon))$. Then it is known [14] that

$$Y^\varepsilon_s=u^\varepsilon(s,X^\varepsilon_s),\qquad Z^\varepsilon_s=\partial_x u^\varepsilon(s,X^\varepsilon_s)\sigma^\varepsilon(s,X^\varepsilon_s)\qquad\forall s\in[t,T],\ P\text{-a.s.}\tag{3.17}$$

Now by Lemmas 2.1 and 2.2 we know that for all $p\ge2$ it holds that

$$E\biggl\{|X^\varepsilon-X|^{*,p}_{t,T}+|Y^\varepsilon-Y|^{*,p}_{t,T}+\int_t^T|Z^\varepsilon_s-Z_s|^2\,ds\biggr\}\to0,\tag{3.18}$$

as $\varepsilon \to 0$. Moreover, recall that

$$\nabla X_s^\varepsilon = 1 + \int_t^s \partial_x b^\varepsilon(r, X_r^\varepsilon)\nabla X_r^\varepsilon\,dr + \int_t^s \sigma_x^\varepsilon(r, X_r^\varepsilon)\nabla X_r^\varepsilon\,dW_r;$$

$$\begin{aligned}\nabla Y_s^\varepsilon = g_x^\varepsilon(X_T^\varepsilon)\nabla X_T^\varepsilon + \int_s^T \big[f_x^\varepsilon\big(r, \Theta^\varepsilon(r)\big)\nabla X_r^\varepsilon + f_y^\varepsilon\big(r, \Theta^\varepsilon(r)\big)\nabla Y_r^\varepsilon \\ + f_z\big(r, \Theta^\varepsilon(r)\big)\nabla Z_r^\varepsilon\big]\,dr - \int_s^T \nabla Z_r^\varepsilon\,dW_r,\end{aligned} \tag{3.19}$$

Applying Lemmas 2.1 and 2.2 again we obtain that, for any $p \ge 2$,

$$E\left\{|\nabla X^\varepsilon - \nabla X|_{t,T}^{*,p} + |\nabla Y^\varepsilon - \nabla Y|_{t,T}^{*,p} + \int_t^T |Z_s^\varepsilon - Z_s|^2\,ds\right\} \to 0, \tag{3.20}$$

as $\varepsilon \to 0$. Thus, using the dominated convergence theorem one derives that

$$\begin{aligned}\lim_{\varepsilon\to 0}\partial_x u^\varepsilon(t, x^\varepsilon) &= E\left\{g_x^\varepsilon(X_T^\varepsilon)\nabla X_T^\varepsilon + \int_t^T [f_x^\varepsilon\nabla X_r^\varepsilon + f_y^\varepsilon\nabla Y_r^\varepsilon + f_z^\varepsilon\nabla Z_r^\varepsilon]\,dr\right\} \\ &= E\left\{g_x(X_T)\nabla X_T + \int_t^T [f_x\nabla X_r + f_y\nabla Y_r + f_z\nabla Z_r]\,dr\right\} \\ &= \partial_x u(t, x) \qquad \forall\,(t, x).\end{aligned}$$

Consequently, possibly along a subsequence, we have

$$\begin{aligned}Z_s = \lim_{\epsilon\to 0} Z_s^\varepsilon &= \lim_{\varepsilon\to 0}\partial_x u^\varepsilon(s, X_s^\varepsilon)\sigma^\varepsilon(s, X_s^\varepsilon) \\ &= \partial_x u(s, X_s)\sigma(s, X_s), \qquad ds \times dP\text{-a.e.}\end{aligned}$$

Since $\partial_x u(\cdot,\cdot)$ and X are both continuous, we see that the equalities above actually holds for all $s \in [t, T]$, P-a.s., proving (iii), whence the theorem. □

A direct consequence of Theorem 3.1 is the following corollary. Recall again the convention on the generic constant $C > 0$.

COROLLARY 3.2. *Assume that the same conditions of Theorem* 3.1 *hold*, *and denote the solution of FBSDE* (2.8) *by* (X, Y, Z). *Then there exists a constant* $C > 0$ *depending only on* K *and* T, *such that*

$$|\partial_x u(t, x)| \le C \qquad \forall\,(t, x) \in [0, T] \times \mathbb{R}^d. \tag{3.21}$$

Consequently, *one has*

$$|Z_s| \le C(1 + |X_s|) \qquad \forall\, s \in [t, T],\ P\text{-a.s.} \tag{3.22}$$

Furthermore, *for* $\forall\, p > 0$, *there exists a constant* $C_p > 0$, *depending on* K, T *and* p *such that*

$$E\{|X|_{t,T}^{*,p} + |Y|_{t,T}^{*,p} + |Z_s|_{t,T}^{*,p}\} \le C_p(1 + |x|^p). \tag{3.23}$$

PROOF. We first assume that $p \geq 2$. By Lemmas 2.1 and 2.2, we can find a constant $C > 0$ so that

$$E\left\{|\nabla X|_{t,T}^{*,2} + |\nabla Y|_{t,T}^{*,2} + \int_t^T |\nabla Z_r|^2\,dr\right\} \leq C. \tag{3.24}$$

Then from the representation (3.1) we deduce immediately that $|\partial_x u(t,x)| \leq C$, for all $(t,x) \in [0,T] \times \mathbb{R}$; and Theorem 3.1(iii) implies that

$$|Z_s| \leq C(1 + |X_s|) \qquad \forall s \in [t,T],\ \ P\text{-a.s.},$$

proving (3.22). Now, applying Lemmas 2.1 and 2.2 for $p \geq 2$ and using (3.22) we derive (3.23).

The case for $0 < p < 2$ follows easily from the Hölder inequality. □

To conclude this section, we would like to point out that in Theorem 3.1 the functions f and g are assumed to be continuously differentiable in all spatial variables with uniformly bounded partial derivatives, which is much stronger than standing assumption (A2). The following theorem reduces the requirement on f and g to only uniformly Lipschitz continuous, which will be important in our future discussion.

THEOREM 3.3. *Assume* (A1) *and* (A2), *and let* (X, Y, Z) *be the solution to the FBSDE* (2.8). *Then, for* $\forall\, p > 0$, *there exists a constant* $C_p > 0$ *such that*

$$E\left\{|X|_{t,T}^{*,p} + |Y|_{t,T}^{*,p} + \operatorname*{ess\,sup}_{t\leq s\leq T}|Z_s|^p\right\} \leq C_p(1 + |x|^p). \tag{3.25}$$

PROOF. In light of the proof of Corollary 3.2, we need only consider $p \geq 2$. By Lemmas 2.1 and 2.2 we see that for any $p > 0$ there exists $C_p > 0$ such that

$$E\{|X|_{t,T}^{*,p} + |Y|_{t,T}^{*,p}\} \leq C_p(1 + |x|^p). \tag{3.26}$$

Next, we repeat the argument in the proof of Theorem 3.1(iii) to get two sequences of smooth functions $\{f^\varepsilon\}_{\varepsilon>0}$ and $\{g^\varepsilon\}_{\varepsilon>0}$ such that

$$\lim_{\varepsilon\to 0}\left\{\sup_{(t,x,y,z)}|f^\varepsilon(t,x,y,z) - f(t,x,y,z)| + \sup_x|g^\varepsilon(x) - g(x)|\right\} = 0;$$

and that the first order partial derivatives of f^ε's and g^ε's in (x, y, z) are uniformly bounded, uniformly in t and ε. Letting $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon)$ denote the solution of the corresponding FBSDEs and applying Corollary 3.2, we can find for any $p \geq 2$ a constant $C_p > 0$, independent of ε, such that

$$E|Z^{\varepsilon_n}|_{t,T}^{*,p} \leq C_p(1 + |x|^p). \tag{3.27}$$

Furthermore, by (3.18) we know that $E\{\int_t^T |Z_s^\varepsilon - Z_s|^2\} \to 0$, as $\varepsilon \to 0$. Thus, possibly along a sequence, say $\{\varepsilon_n\}_{n\ge 1}$, we have $\lim_{n\to\infty} Z^{\varepsilon_n} = Z$, $ds \times dP$-a.s. Applying Fatou's lemma and using (3.27) we then have

$$E\left\{\operatorname*{ess\,sup}_{t\le s\le T} |Z_s|^p\right\} \le C_p(1+|x|^p),$$

which leads to (3.26), as desired. □

4. The representation theorem. In this section we shall prove the first main theorem of the paper. This theorem can be regarded as an extension of the nonlinear Feynman–Kac formula of Pardoux and Peng [18], as it gives a probabilistic representation of the gradient (rather than the solution itself) of the viscosity solution, whenever it exists, to a quasilinear parabolic PDE. We should point out here that the main feature of our representation, however, lies in that it *does not* depend on the partial derivatives of the functions f and g as we saw in (3.1). Such a representation then paves the way for us to study the path regularity of the process Z in the BSDE with nonsmooth coefficients. Again, we shall drop the superscript ^{t,x} from the solution (X, Y, Z) of FBSDE (2.8) for notational simplicity.

To begin with, let us introduce a stochastic integral that will play a key role in the representation: for $t < r_1 < r_2 < T$,

$$M_{r_2}^{r_1} = \int_{r_1}^{r_2} [\sigma^{-1}(\tau, X_\tau)\nabla X_\tau]^T \, dW_\tau, \tag{4.1}$$

where $\nabla X = (\nabla_1 X, \dots, \nabla_d X)$ is the solution of the variational equation (2.9). Clearly, for fixed r_1, the process M^{r_1} is a martingale. By the Burkholder–Davis–Gundy inequality, for any $p \ge 1$ one has

$$\begin{aligned} E|M_{r_2}^{r_1}|^{2p} &\le C_p E\left[\int_{r_1}^{r_2} |\sigma^{-1}(\tau, X_\tau)\nabla X_\tau|^2 \, d\tau\right]^p \\ &\le C_p (r_2 - r_1)^p E\{|\nabla X_\tau|_{r_1,r_2}^{*,2p}\} \le C_p (r_2 - r_1)^p, \end{aligned}$$

where $C_p > 0$ is a generic constant depending only on constants K and c in (A1) and (A2), the time duration T and $p \ge 1$. Thus the following estimate is not surprising:

$$E|M_{r_2}^{r_1}|^{2p} \le C_p (r_2 - r_1)^p. \tag{4.2}$$

To study further the (two parameter) process M, we denote, for any $0 \le t < s \le T$, the σ-field $\mathcal{F}_s^t \triangleq \sigma\{W_u - W_t : t \le u \le s\}$, with the usual P-augmentation; and $\mathbf{F}^t = \{\mathcal{F}_s^t\}_{s\ge t}$. The following result is important in our future discussion.

LEMMA 4.1. *Let* $t \in [0, T)$ *be fixed. Then for any* $H \in L^{p_0}(\mathbf{F}^t, [0, T]; \mathbb{R})$, *with* $p_0 > 2$, *one has*:

(i) $E|\int_s^T [1/(r-s)]H_r M_r^s\,dr| < \infty$;

(ii) *for P-a.e. $\omega \in \Omega$, the mapping $s \mapsto \int_s^T [1/(r-s)]H_r(\omega)M_r^s(\omega)\,dr$ is Hölder-$([p_0-2]/[p_0(p_0+2)])$ continuous on $[t,T]$*;

(iii) *for P-a.e. $\omega \in \Omega$, the mapping $s \mapsto E\{\int_s^T [1/(r-s)]H_r M_r^s\,dr|\mathcal{F}_s^t\}(\omega)$ is continuous on $[t,T]$.*

PROOF. First, for any $t < \tau \le T$ we denote

$$A_s^\tau = \begin{cases} \displaystyle\int_s^\tau \frac{1}{r-s} H_r M_r^s\,dr, & \text{for } t \le s < \tau, \\ 0, & \text{if } s = \tau. \end{cases} \tag{4.3}$$

To simplify notation, when $\tau = T$ we denote $A_s^T = A_s$.

Further, let $q_0 > 0$ be such that $\frac{1}{p_0} + \frac{1}{q_0} = 1$, and define $\beta = \frac{p_0}{2+p_0}$ and $\alpha = 1 - \beta$. It is readily seen that $\beta < \frac{1}{q_0}$ and $\alpha < \frac{1}{2}$. Consider the random variable

$$M^* = \sup_{t \le t_1 < t_2 \le T} \frac{|M_{t_2}^{t_1}|}{(t_2-t_1)^\alpha}. \tag{4.4}$$

then by (4.2) and Theorem 2.1 of Revuz and Yor [21], Chapter 1, we see that $E[M^*]^2 < \infty$.

To prove (i) we note that for any $t \le s < \tau \le T$ by Hölder's inequality one has

$$\begin{aligned} |A_s^\tau| &= \left| \int_s^\tau \frac{H_r}{(r-s)^\beta} \cdot \frac{M_r^s}{(r-s)^a}\,dr \right| \le \int_s^\tau \left| \frac{H_r}{(r-s)^\beta} \right| dr\, M^* \\ &\le \left\{ \int_s^\tau \frac{dr}{(r-s)^{\beta q_0}} \right\}^{1/q_0} \|H\|_{p_0} M^* = C(\tau-s)^{(1/q_0)-\beta} \|H\|_{p_0} M^*, \end{aligned} \tag{4.5}$$

where $\|\cdot\|_{p_0}$ denotes the norm of $L^{p_0}([0,T])$. Again letting $C > 0$ be a generic constant depending only on p_0 and T, and noting that $p_0 > 2$ we have

$$\begin{aligned} E|A_s^\tau| &\le C\{E\|H\|_{p_0}^2\}^{1/2}\{E(M^*)^2\}^{1/2} \\ &\le C\|H\|_{L^{p_0}([0,T]\times\Omega)} \|M^*\|_{L^2(\Omega)} < \infty. \end{aligned} \tag{4.6}$$

Setting $\tau = T$ in the above we proved (i).

To prove (ii) we let $\tau = T$ and observe that, for $t \le s_1 < s_2 < T$,

$$\begin{aligned} A_{s_1} - A_{s_2} &= \int_{s_1}^{s_2} \frac{1}{r-s_1} H_r M_r^{s_1}\,dr + \int_{s_2}^T \frac{1}{r-s_1} H_r M_{s_2}^{s_1}\,dr \\ &\quad + \int_{s_2}^T \left(\frac{1}{r-s_1} - \frac{1}{r-s_2} \right) H_r M_r^{s_2}\,dr \triangleq \Gamma_1 + \Gamma_2 + \Gamma_3, \end{aligned} \tag{4.7}$$

where Γ_i's are defined in an obvious way. Comparing to (4.3) it is easily seen that $\Gamma_1 = A_{s_1}^{s_2}$. Thus (4.5) shows that

$$|\Gamma_1| \le C(s_2 - s_1)^{(1/q_0)-\beta} \|H\|_{p_0} M^*. \tag{4.8}$$

Further, by definition (4.4) we see that

$$
\begin{aligned}
|\Gamma_2| &= \left| \int_{s_2}^T \frac{H_r}{r-s_1} \cdot \frac{M_{s_2}^{s_1}}{(s_1-s_2)^\alpha} dr \right| |s_1-s_2|^\alpha \\
&\le (s_2-s_1)^\alpha \int_{s_2}^T \frac{|H_r|}{r-s_1} dr M^* \le (s_2-s_1)^\alpha \left\{ \int_{s_2}^T \frac{dr}{(r-s_1)^{q_0}} \right\}^{1/q_0} \|H\|_{p_0} M^* \qquad (4.9) \\
&\le C(s_2-s_1)^{(1/q_0)-\beta} \|H\|_{p_0} M^*.
\end{aligned}
$$

Finally

$$
\begin{aligned}
|\Gamma_3| &= \left| \int_{s_2}^T \frac{s_2-s_1}{(r-s_2)(r-s_1)} H_r M_r^{s_2}\, dr \right| \\
&\le (s_2-s_1) \left\{ \int_{s_2}^T \frac{|H_r|}{(r-s_2)^\beta (r-s_1)}\, dr \right\} M^* \qquad (4.10) \\
&\le (s_2-s_1) \left\{ \int_{s_2}^T \frac{dr}{(r-s_2)^{\beta q_0} (r-s_1)^{q_0}} \right\}^{1/q_0} \|H\|_{p_0} M^*.
\end{aligned}
$$

Since

$$
\begin{aligned}
&\int_{s_2}^T \frac{1}{(r-s_2)^{\beta q_0}(r-s_1)^{q_0}}\, dr \\
&\quad = \int_0^{T-s_2} \frac{1}{r^{\beta q_0}(r+s_2-s_1)^{q_0}}\, dr \\
&\quad = \int_0^{(T-s_2)/(s_2-s_1)} \frac{(s_2-s_1)}{(s_2-s_1)^{\beta q_0} r^{\beta q_0} (s_2-s_1)^{q_0} (r+1)^{q_0}}\, dr \\
&\quad \le (s_2-s_1)^{1-(\beta+1)q_0} \int_0^\infty \frac{1}{r^{\beta q_0}(r+1)^{q_0}}\, dr \\
&\quad = C(s_2-s_1)^{1-(\beta+1)q_0},
\end{aligned}
$$

plugging this into (4.10) we have

$$
\begin{aligned}
|\Gamma_3| &\le C(s_2-s_1)(s_2-s_1)^{(1/q_0)-(\beta+1)} \|H\|_{p_0} M^* \qquad (4.11) \\
&= C(s_2-s_1)^{(1/q_0)-\beta} \|H\|_{p_0} M^*.
\end{aligned}
$$

Combining (4.8)–(4.11) we obtain that

$$
|A_{s_1} - A_{s_2}| \le C(s_2-s_1)^{(1/q_0)-\beta} \|H\|_{p_0} M^*. \qquad (4.12)
$$

We should note that by (4.5) with $\tau = T$ we see that (4.12) holds true even when $s_2 = T$. This, together with the fact $\frac{1}{q_0} - \beta = \frac{p_0-2}{p_0(p_0+2)}$, proves (ii).

It remains to prove (iii). To this end, we note that the right-hand side of inequality (4.5) (with $\tau = T$) is clearly in L^1; thus it is easy to check that the process A is uniformly integrable. Therefore, applying Theorem VI-47 and Remarks VI-50(f) of Dellacherie and Meyer [3], we see that the $\mathbf{F}^t$-optional projection of A, denoted by ${}^oA_s = E\{A_s|\mathcal{F}_s^t\}$, $s \in [t, T]$, has càdlàg paths. To show that the paths are actually continuous, we note that the filtration $\mathbf{F}^t$ is Brownian, whence quasi-left-continuous. Thus every $\mathbf{F}^t$-stopping time $\tau > t$ is accessible. That is, there exists a sequence of $\mathbf{F}^t$-stopping times $\{\tau_k\}_{k\geq 0}$ such that $\tau_k < \tau$, $\forall k$, P-a.s., and that $\tau_k \uparrow \tau$, as $k \to \infty$. Note that

$$\begin{aligned} {}^oA_{\tau_k} - {}^oA_\tau &= E\{A_{\tau_k}|\mathcal{F}_{\tau_k}\} - E\{A_\tau|\mathcal{F}_\tau\} \\ &= E\{A_{\tau_k} - A_\tau|\mathcal{F}_{\tau_k}\} + \big(E\{A_\tau|\mathcal{F}_{\tau_k}\} - E\{A_\tau|\mathcal{F}_\tau\}\big). \end{aligned} \tag{4.13}$$

Letting $k \to \infty$ we see that $E\{A_{\tau_k} - A_\tau|\mathcal{F}_{\tau_k}\} \to 0$, thanks to the quasi-left-continuity of $\mathbf{F}^t$; and $E\{A_\tau|\mathcal{F}_{\tau_k}\} - E\{A_\tau|\mathcal{F}_\tau\} \to 0$, thanks to (4.12). Thus ${}^oA_{\tau-} = {}^oA_\tau$, P-a.s. Since ${}^oA$ is càdlàg and τ is arbitrary, we conclude that ${}^oA$ is in fact continuous on $[t, T]$, almost surely. This proves (iii), whence the lemma. □

Next, recall that the variational equation (2.9) is a linear ($d \times d$-matrix-valued) SDE with $\nabla X_t = I$; thus ∇X_s is invertible for all $s \in [t, T]$, thanks to the Doléan–Dade stochastic exponential formula (see, e.g., [20]). Define

$$N_r^s = \frac{1}{r-s}(M_r^s)^T[\nabla X_s]^{-1}, \qquad 0 \leq t \leq s < r \leq T. \tag{4.14}$$

(Note that N_r^s, is a *row vector*.) We now prove the main representation theorem.

THEOREM 4.2. *Assume that assumptions* (A1) *and* (A2) *hold*, *and let* (X, Y, Z) *be the adapted solution to FBSDE* (2.8). *Then the following hold*:

(i) *the following identity holds* P*-almost surely*:

$$Z_s = E\left\{g(X_T)N_T^s + \int_s^T f(r, X_r, Y_r, Z_r)N_r^s\,dr\,\middle|\,\mathcal{F}_s^t\right\}\sigma(s, X_s) \qquad \forall s \in [t, T); \tag{4.15}$$

(ii) *there exists a version of* Z *such that for* P*-a.e.* $\omega \in \Omega$, *the mapping* $s \mapsto Z_s(\omega)$ *is continuous*;

(iii) *if in addition the functions* f *and* g *satisfy the assumptions of Theorem* 3.1, *then for all* $(t, x) \in [0, T) \times \mathbb{R}^d$ *it holds that*

$$\partial_x u(t, x) = E\left\{g(X_T)N_T^t + \int_t^T f(r, X_r, Y_r, Z_r)N_r^t\,dr\right\}. \tag{4.16}$$

PROOF. Again we shall consider only the case $d = 1$. Let us first assume that $g \in C_b^1(\mathbb{R})$ and $f \in C_b^{0,1}([0, T] \times \mathbb{R}^3)$. Applying the nonlinear Feynman–Kac formula of Pardoux and Peng [18] we have, for $t \leq s \leq T$, that

$$u(s, X_s) = Y_s = E\left\{g(X_T) + \int_s^T f(r, X_r, Y_r, Z_r)\,dr\,\middle|\,\mathcal{F}_s^t\right\}. \tag{4.17}$$

First formally differentiating (4.17) and then following a line-by-line analogue of Lemma 3.1 one can show that, for $t \le s \le T$,

$$\begin{aligned}
&\partial_x u(s, X_s)\nabla X_s \\
&= E\biggl\{g_x(X_T)\nabla X_T + \int_s^T \bigl[f_x\bigl(r, \Theta(r)\bigr)\nabla X_r \\
&\qquad + f_y\bigl(r, \Theta(r)\bigr)\nabla Y_r + f_z\bigl(r, \Theta(r)\bigr)\nabla Z_r\bigr]\,dr\Bigm|\mathcal{F}_s^t\biggr\},
\end{aligned}\tag{4.18}$$

where g_x, (f_x, f_y, f_z) denote the partial derivatives of g and f, as in the proof of Theorem 3.1. [Note that if $s = t$, then (4.18) reduces to (3.1).]

Our proof depends on the following observation. By Lemma 2.4, we know that if b, σ, f and g are all C^1 in (x, y, z), then the adapted solution (X, Y, Z) of the FBSDEs (2.8) all belong to $\mathbb{D}^{1,2}$. Thus, using the *chain rule* (cf., e.g., [16]) and the relation (2.10) in Lemma 2.4 we see that, for $t < \tau < r$,

$$\begin{aligned}
D_\tau f\bigl(r, \Theta(r)\bigr) &= f_x\bigl(r, \Theta(r)\bigr)D_\tau X_r + f_y\bigl(r, \Theta(r)\bigr)D_\tau Y_r + f_z\bigl(r, \Theta(r)\bigr)D_\tau Z_r \\
&= \bigl[f_x\bigl(r, \Theta(r)\bigr)\nabla X_r + f_y\bigl(r, \Theta(r)\bigr)\nabla Y_r + f_z\bigl(r, \Theta(r)\bigr)\nabla Z_r\bigr] \\
&\quad \times (\nabla X_\tau)^{-1}\sigma(\tau, X_\tau),
\end{aligned}$$

or, equivalently,

$$\begin{aligned}
&f_x\bigl(r, \Theta(r)\bigr)\nabla X_r + f_y\bigl(r, \Theta(r)\bigr)\nabla Y_r + f_z\bigl(r, \Theta(r)\bigr)\nabla Z_r \\
&\quad = D_\tau f\bigl(r, \Theta(r)\bigr)\sigma^{-1}(\tau, X_\tau)\nabla X_\tau.
\end{aligned}$$

Note that the left-hand side above is independent of τ; integrating both sides from $\tau = s \ge t$ to $\tau = r > s$ and then dividing by $r - s$, we obtain that

$$\begin{aligned}
&f_x\bigl(r, \Theta(r)\bigr)\nabla X_r + f_y\bigl(r, \Theta(r)\bigr)\nabla Y_r + f_z\bigl(r, \Theta(r)\bigr)\nabla Z_r \\
&\quad = \frac{1}{r - s}\int_s^r D_\tau f\bigl(r, \Theta(r)\bigr)\sigma^{-1}(\tau, X_\tau)\nabla X_\tau\,d\tau.
\end{aligned}\tag{4.19}$$

Since σ^{-1} is bounded by (2.1), the process $\sigma^{-1}(\cdot, X)\nabla X \in L^2(\mathbf{F}^t, [0, T])$ and therefore it belongs to $\mathrm{Dom}(\delta)$ (see Section 2). Further, it can be checked that

$$E\biggl\{\bigl|f\bigl(r, \Theta(r)\bigr)\bigr|^2\int_t^T |\sigma^{-1}(\tau, X_\tau)\nabla X_\tau|^2\,d\tau\biggr\} < \infty \qquad \forall r \in [t, T],\tag{4.20}$$

thanks to Lemmas 2.1 and 2.2 and Corollary 3.2. Thus, by Lemma 2.5(i) we have (integration by parts)

$$\begin{aligned}
&\int_s^r D_\tau f\bigl(r, \Theta(r)\bigr)\sigma^{-1}(\tau, X_\tau)\nabla X_\tau\,d\tau \\
&\quad = f\bigl(r, \Theta(r)\bigr)\int_s^r \sigma^{-1}(\tau, X_\tau)\nabla X_\tau\,dW_\tau \\
&\qquad - \int_s^r f\bigl(r, \Theta(r)\bigr)\sigma^{-1}(\tau, X_\tau)\nabla X_\tau\,dW_\tau,
\end{aligned}\tag{4.21}$$

where the second integral on the right-hand side should be understood as an anticipating stochastic integral. We claim that its conditional expectation $E\{\cdot|\mathcal{F}_s^t\}$ is zero. Indeed, let $\alpha \in \mathcal{S}$ be any smooth functional (see Section 2) such that it is bounded and $\mathcal{F}_s^t$-measurable. Then $D_\tau\alpha = 0$ for all $\tau > s$. Thus, if we write

$$\eta_\tau^r = f\big(r, \Theta(r)\big)\sigma^{-1}(\tau, X_\tau)\nabla X_\tau, \qquad \tau \in [s, r],$$

then it can be checked that $\eta \in \mathbb{D}^{1,2}$; by using the definition of the anticipating integral one derives

$$E\left\{\alpha \int_s^r \eta_\tau^r \, dW_\tau\right\} = E\int_s^r (D_\tau\alpha)\eta_\tau^r \, d\tau = 0 \qquad \forall r \in [s, T], \text{ a.s.,}$$

and the claim follows. Next, we plug (4.21) into the right-hand side of (4.19) and then take the conditional expectation $E\{\cdot|\mathcal{F}_s^t\}$ on both sides to get

$$\begin{aligned}
&E\big\{f_x\big(r, \Theta(r)\big)\nabla X_r + f_y\big(r, \Theta(r)\big)\nabla Y_r + f_z\big(r, \Theta(r)\big)\nabla Z_r \big| \mathcal{F}_s^t\big\} \\
&\qquad = \frac{1}{r-s} E\left\{\int_s^r \eta_\tau^r \, dW_\tau \Big| \mathcal{F}_s^t\right\} = \frac{1}{r-s} E\big\{f\big(r, \Theta(r)\big)M_r^s \big| \mathcal{F}_s^t\big\},
\end{aligned} \tag{4.22}$$

for all $s \in [t, T]$, where M is defined by (4.1). Using similar arguments one shows that

$$E\{g_x(X_T)\nabla X_T|\mathcal{F}_s^t\} = E\left\{\frac{1}{T-s} g(X_T)M_T^s \Big| \mathcal{F}_s^t\right\}. \tag{4.23}$$

Finally, plugging (4.22) and (4.23) into (4.18), and applying Corollary 3.2 and Lemma 4.1, we derive

$$\begin{aligned}
&\partial_x u(s, X_s)\nabla X_s \\
&\qquad = E\left\{\frac{1}{T-s} g(X_T)M_T^s + \int_s^T \frac{1}{r-s} f(r, X_r, Y_r, Z_r)M_r^s \, dr \Big| \mathcal{F}_s^t\right\}.
\end{aligned} \tag{4.24}$$

Recalling the process N [see (4.14)] we can rewrite (4.24) as

$$\partial_x u(s, X_s) = E\left\{g(X_T)N_T^s + \int_s^T f(r, X_r, Y_r, Z_r)N_r^s \, dr \Big| \mathcal{F}_s^t\right\}. \tag{4.25}$$

In particular, setting $s = t$ we obtain (4.16); this proves (iii).

We now consider the general case. First we fix $s \in [t, T]$. For $\varphi = f, g$, let $\varphi^\varepsilon \in C^\infty$, $\varepsilon > 0$, be the mollifiers of φ, and let $(Y^\varepsilon, Z^\varepsilon)$ be the solution of the BSDE in (2.8) with coefficients $(f^\varepsilon, g^\varepsilon)$. Then, for each $\varepsilon > 0$, (4.15) holds true. That is,

$$Z_s^\varepsilon = E\left\{g^\varepsilon(X_T)N_T^s + \int_s^T f^\varepsilon(r, X_r, Y_r^\varepsilon, Z_r^\varepsilon)N_r^s \, dr \Big| \mathcal{F}_s^t\right\}\sigma(s, X_s). \tag{4.26}$$

Note that Lemma 2.2 implies that

$$E\left\{|Y^\varepsilon - Y|_{t,T}^{*,4} + \int_t^T |Z_s^\varepsilon - Z_s|^2 \, ds\right\} \to 0 \qquad \text{as } \varepsilon \to 0. \tag{4.27}$$

Thus it is readily seen that $E\{g^\varepsilon(X_T))N_T^s\,dr|\mathcal{F}_s^t\} \to E\{g(X_T)N_T^s\,dr|\mathcal{F}_s^t\}$, P-a.s., as $\varepsilon \to 0$. Furthermore, note that

$$
\begin{aligned}
&E\bigg|E\bigg\{\int_s^T f^\varepsilon(r, X_r, Y_r^\varepsilon, Z_r^\varepsilon)N_r^s\,dr\bigg|\mathcal{F}_s^t\bigg\} - E\bigg\{\int_s^T f(r, X_r, Y_r, Z_r)N_r^s\,dr\bigg|\mathcal{F}_s^t\bigg\}\bigg| \\
&\quad\le E\int_s^T \big|f^\varepsilon(r, X_r, Y_r^\varepsilon, Z_r^\varepsilon) - f(r, X_r, Y_r, Z_r)\big|\,|N_r^s|\,dr \\
&\quad\le E\int_s^T |\delta_1^\varepsilon f(r)|\,|N_r^s|\,dr + E\int_s^T |\delta_2^\varepsilon f(r)|\,|N_r^s|\,dr \qquad (4.28)\\
&\qquad + E\int_s^T |\delta_3^\varepsilon f(r)|\,|N_r^s|\,dr \\
&\quad= \Delta_1^\varepsilon + \Delta_2^\varepsilon + \Delta_3^\varepsilon,
\end{aligned}
$$

where $\Delta_i^\varepsilon \triangleq E\int_s^T |\delta_i^\varepsilon f(r)||N_r^s|\,dr$, $i = 1, 2, 3$, and

$$
\begin{aligned}
\delta_1^\varepsilon f(r) &\triangleq f^\varepsilon(r, X_r, Y_r^\varepsilon, Z_r^\varepsilon) - f(r, X_r, Y_r^\varepsilon, Z_r^\varepsilon),\\
\delta_2^\varepsilon f(r) &\triangleq f(r, X_r, Y_r^\varepsilon, Z_r^\varepsilon) - f(r, X_r, Y_r, Z_r^\varepsilon),\\
\delta_3^\varepsilon f(r) &\triangleq f(r, X_r, Y_r, Z_r^\varepsilon) - f(r, X_r, Y_r, Z_r).
\end{aligned}
$$

Since $E|Y^\varepsilon - Y|_{t,T}^{*,4} \to 0$, similarly to Lemma 4.1, one shows that

$$\Delta_2^\varepsilon = E\int_s^T |\delta_2^\varepsilon f(r)|\,|N_r^s|\,dr \le KE\int_s^T |Y_r^\varepsilon - Y_r|\,|N_r^s|\,dr \to 0.$$

Next, by Corollary 3.2 and the dominated convergence theorem we have

$$\Delta_3^\varepsilon = E\int_s^T |\delta_3^\varepsilon f(r)|\,|N_r^s|\,dr \to 0 \qquad \text{as } \varepsilon \to 0.$$

Finally, since $f^\varepsilon \to f$ in $C([0, T] \times \mathbb{R}^3)$, we have

$$
\begin{aligned}
\Delta_1^\varepsilon &= E\int_s^T |\delta^\varepsilon f(r)|\,|N_r^s|\,dr \\
&\le \|f^\varepsilon - f\|_{C([0,T]\times\mathbb{R}^3)} E\int_s^T |N_r^s|\,dr \to 0 \qquad \text{as } \varepsilon \to 0
\end{aligned}
$$

as well. Consequently, letting $\varepsilon \to 0$ in (4.26), we see that (4.15) holds P-a.s., for each fixed $s \in [0, T]$.

We should note that to prove part (i) we still need to show that (4.15) actually holds for all $s \in [0, T]$, P-a.s., but it is easy to see that this will follow from part (ii); that is, the process Z has a continuous version. Thus we need only prove (ii).

To do this we first note that Lemma 4.1 implies that the mapping

$$s \longmapsto E\bigg\{\int_s^T f(r, X_r, Y_r, Z_r)N_r^s\,dr\bigg|\mathcal{F}_s^t\bigg\}$$

is a.s. continuous on $[t, T]$. Next, since $g(\cdot)$ is uniformly Lipschitz, it is known (see, e.g., [16, Proposition 1.2.3]) that there exists $\xi \in L^2(\Omega)$, such that

$$D_r g(X_T) = \xi D_r X_T = \xi \nabla X_T (\nabla X_r)^{-1} \sigma(r, X_r) \qquad \forall r \in [s, T].$$

Applying the integration by parts formula (Lemma 2.5) again we have

$$\begin{aligned} E\{g(X_T) N_T^s | \mathcal{F}_s^t\} &= \frac{1}{T-s} E\left\{ g(X_T) \int_s^T \sigma^{-1}(r, X_r) \nabla X_r \, dW_r \middle| \mathcal{F}_s^t \right\} (\nabla X_s)^{-1} \\ &= \frac{1}{T-s} E\left\{ \int_s^T [D_r g(X_T)] \sigma^{-1}(r, X_r) \nabla X_r \, dr \middle| \mathcal{F}_s^t \right\} (\nabla X_s)^{-1} \\ &= E\{\xi \nabla X_T | \mathcal{F}_s^t\} (\nabla X_s)^{-1}. \end{aligned}$$

Thus the mapping $s \mapsto E\{g(X_T) N_T^s \big| \mathcal{F}_s^t\}$ is also continuous on $[t, T]$, thanks to the quasi-left-continuity of the Brownian filtration $\mathbf{F}^t$ again. Consequently, the right-hand side of (4.15) is a.s. continuous on $[t, T]$, and hence (4.15) holds for all $s \in [0, T]$, P-a.s., proving (ii), whence the theorem. □

REMARK. A direct consequence of Theorem 4.2 that might be useful in applications is the following improvement of Theorem 3.3: assume that (A1) and (A2) hold; then for $\forall\, p > 0$, there exists a constant $C_p > 0$ depending only on T, K and p such that

$$E\{|X|_{t,T}^{*,p} + |Y|_{t,T}^{*,p} + |Z|_{t,T}^{*,p}\} \le C_p(1 + |x|^p). \tag{4.29}$$

Indeed, since by Theorem 4.2, Z has a continuous version, (3.25) becomes (4.29).

5. Path regularity of process Z.

We have proved in Theorem 4.2(ii) that the process Z in the solution to the FBSDE (2.8) has continuous paths, under the condition that the coefficients f and g are only uniformly Lipschitz continuous. While such a result is already an improvement of that of Pardoux and Peng [18], it is still within the paradigm of the standard FBSDE in the literature, to wit, the terminal condition of the BSDE is of the form $g(X_T)$ (cf., e.g., [14] or [15]). In this section we shall consider a class of BSDEs whose terminal conditions are *path dependent*. More precisely, we assume that the terminal condition of the BSDE is of the form $\xi = g(X_{t_1}, X_{t_2}, \dots, X_{t_n})$, where $t \le t_1 < t_2 < \cdots < t_n \le T$ is any partition of $[t, T]$. We shall prove a new representation theorem for the process Z, and we will extend the path regularity result to such a case.

Our main result is the following.

THEOREM 5.1. *Assume that* (A1) *holds*; *and that in* (A2) *one has* $g \in W^{1,\infty} \times (\mathbb{R}^{d(n+1)})$. *Let* $t = t_0 < t_1 < \cdots < t_n = T$ *be a partition of* $[t, T]$, *and let* (X, Y, Z)

be the unique adapted solution to the following FBSDE:

$$
\begin{aligned}
X_s &= x + \int_t^s b(r, X_r)\,dr + \int_t^s \sigma(r, X_r)\,dW_r, \\
Y_s &= g(X_{t_0}, \dots, X_{t_n}) + \int_s^T f(r, X_r, Y_r, Z_r)\,dr - \int_s^T Z_r\,dW_r.
\end{aligned} \tag{5.1}
$$

Then on each interval (t_{i-1}, t_i), $i = 1, \dots, n$, *the following identity holds*:

$$
\begin{aligned}
Z_s = E\biggl\{ & g(X_{t_0}, \dots, X_{t_n}) N^s_{t_i} \\
& + \int_s^T f(r, X_r, Y_r, Z_r) N^s_{r \wedge t_i}\,dr \Big| \mathcal{F}_s \biggr\} \sigma(s, X_s), \qquad s \in (t_{i-1}, t_i).
\end{aligned} \tag{5.2}
$$

Furthermore, *there exists a version of process* Z *that enjoys the following properties*:

(i) *the mapping* $s \mapsto Z_s$ *is a.s. continuous on each interval* (t_{i-1}, t_i), $i = 1, \dots, n$;

(ii) *both limits* $Z_{t_i-} \triangleq \lim_{s \uparrow t_i} Z_s$ *and* $Z_{t_i+} \triangleq \lim_{s \downarrow t_i} Z_s$ *exist*;

(iii) *for* $\forall\, p > 0$, *there exists a constant* $C_p > 0$ *depending only on* T, K *and* p *such that*

$$
E|\Delta Z_{t_i}|^p \le C_p(1 + |x|^p) < \infty. \tag{5.3}
$$

Consequently, *the process* Z *has both càdlàg and càglàd versions*, *with discontinuities* $t_0, \dots, t_n$ *and jump sizes satisfying* (5.3).

PROOF. As before we will consider only the case $d = 1$, and we assume first that $f, g \in C^1_b$.

Let us first establish the identity (5.2). We fix an arbitrary index i and consider the interval (t_{i-1}, t_i). Note that, by using the similar arguments as those in Pardoux and Peng [18], it can be verified that, for any $\tau \in (t_{i-1}, t_i)$, $Y_\tau, Z_\tau \in \mathbb{D}^{1,2}$ and, for all $t_{i-1} < s \le \tau < t_i$,

$$
\begin{aligned}
D_s Y_\tau = {} & \sum_{j \ge i} \partial_j g D_s X_{t_j} + \int_\tau^T [f_x(r) D_s X_r + f_y(r) D_s Y_r + f_z(r) D_s Z_r]\,dr \\
& - \int_\tau^T D_s Z_r\,dW_r,
\end{aligned} \tag{5.4}
$$

where $\partial_j g \triangleq \partial_{x_j} g(X_{t_0}, \dots, X_{t_n})$, $j = 1, \dots, n$; and D is the Malliavin derivative operator. For notational simplicity here and in the sequel we denote $\varphi(r) = \varphi(r, \Theta(r))$ for $\varphi = f_x, f_y, f_z$, respectively.

Next, by virtue of Lemma 2.4 and the uniqueness of the adapted solution to BSDEs we have

$$
\begin{aligned}
D_s X_\tau &= \nabla X_\tau [\nabla X_s]^{-1} \sigma(s, X_s),\\
D_s Y_\tau &= \nabla^i Y_\tau [\nabla X_s]^{-1} \sigma(s, X_s),\\
D_s Z_\tau &= \nabla^i Z_\tau [\nabla X_s]^{-1} \sigma(s, X_s), \qquad t_{i-1} < s \le \tau < t_i,
\end{aligned}
\tag{5.5}
$$

where $(\nabla^i Y, \nabla^i Z)$ denotes the adapted solution to the following BSDE [cf. (2.9)] for $\tau \in [t_{i-1}, T]$:

$$
\begin{aligned}
\nabla^i Y_\tau = {} & \sum_{j \ge i} \partial_j g \nabla X_{t_j} + \int_\tau^T [f_x(r) \nabla X_r + f_y(r) \nabla^i Y_r + f_z(r) \nabla^i Z_r]\, dr\\
& - \int_\tau^T \nabla^i Z_r \, dW_r,
\end{aligned}
\tag{5.6}
$$

On the other hand, since $D_s Y_\tau = 0$ whenever $s > \tau$ and

$$
Y_{t_{i-1}} = g(X_{t_0}, \dots, X_{t_n}) + \int_{t_{i-1}}^T f(r, X_r, Y_r, Z_r)\, dr - \int_{t_{i-1}}^T Z_r \, dW_r,
$$

applying D_s to both sides for $s > t_{i-1}$ we get

$$
\begin{aligned}
0 = {} & \sum_{j \ge i} \partial_j g D_s X_{t_j} + \int_s^T [f_x(r) D_s X_r + f_y(r) D_s Y_r + f_z(r) D_s Z_r]\, dr\\
& - Z_s - \int_s^T D_s Z_r \, dW_r.
\end{aligned}
\tag{5.7}
$$

Combining (5.7) with (5.5) and (5.6) we obtain

$$
\begin{aligned}
Z_s &= \sum_{j \ge i} \partial_j g D_s X_{t_j} + \int_s^T [f_x(r) D_s X_r + f_y(r) D_s Y_r + f_z(r) D_s Z_r]\, dr\\
&\qquad - \int_s^T D_s Z_r \, dW_r\\
&= \biggl\{ \sum_{j \ge i} \partial_j g \nabla X_{t_j} + \int_s^T [f_x(r) \nabla X_r + f_y(r) \nabla^i Y_r + f_z(r) \nabla^i Z_r]\, dr\\
&\qquad - \int_s^T \nabla^i Z_r \, dW_r \biggr\} [\nabla X_s]^{-1} \sigma(s, X_s)\\
&= \nabla^i Y_s [\nabla X_s]^{-1} \sigma(s, X_s), \qquad t_{i-1} < s < t_i.
\end{aligned}
\tag{5.8}
$$

Taking conditional expectation $E\{\cdot\,|\mathcal{F}_s\}$ on both sides of (5.8) we then get

$$
\begin{aligned}
Z_s = E\bigg\{ \sum_{j\ge i} \partial_j g \nabla X_{t_j} + \int_s^T [f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r]\,dr \bigg| \mathcal{F}_s \bigg\} \\
\times [\nabla X_s]^{-1} \sigma(s, X_s).
\end{aligned} \tag{5.9}
$$

The rest of the proof is similar to that of Theorem 4.2. First we note that by the chain rule of anticipating derivative operator and relation (5.5), for any $t_{i-1} < \tau \le t_i$ and $\tau < r$ one has

$$
\begin{aligned}
&D_\tau f(X_r, Y_r, Z_r) \\
&\quad = [f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r][\nabla X_\tau]^{-1}\sigma(\tau, X_\tau).
\end{aligned} \tag{5.10}
$$

We consider the following two cases:

(a) $t_{i-1} < r \le t_i$. In this case we derive from (5.10) that

$$
\begin{aligned}
&f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r \\
&\quad = \frac{1}{r-s}\int_s^r D_\tau f(r, X_r, Y_r, Z_r)\sigma^{-1}(\tau, X_\tau)\nabla X_\tau\,d\tau, \\
&\qquad\qquad t_{i-1} < s < r \le t_i.
\end{aligned} \tag{5.11}
$$

Therefore, using the integration by parts formula for anticipating integrals and recalling the definition of process N, (4.14), we have

$$
\begin{aligned}
&\{f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r|\mathcal{F}_s\} \\
&\quad = E\bigg\{ \frac{1}{r-s}\int_s^r f(r, X_r, Y_r, Z_r)\sigma^{-1}(\tau, X_\tau)\nabla X_\tau D_\tau\,d\tau \bigg| \mathcal{F}_s \bigg\} \\
&\quad = E\bigg\{ \frac{1}{r-s} f(r, X_r, Y_r, Z_r)\int_s^r \sigma^{-1}(\tau, X_\tau)\nabla X_\tau\,dW_\tau \bigg| \mathcal{F}_s \bigg\} \\
&\quad = E\{f(r, X_r, Y_r, Z_r)N_r^s|\mathcal{F}_s\}\nabla X_s, \qquad t_{i-1} < s < r \le t_i.
\end{aligned} \tag{5.12}
$$

(b) $t_i < r$. In this case we see that (5.10) is still true, but (5.11) should be replaced by

$$
\begin{aligned}
&f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r \\
&\quad = \frac{1}{t_i-s}\int_s^{t_i} D_\tau f(r, X_r, Y_r, Z_r)\sigma^{-1}(\tau, X_\tau)\nabla X_\tau\,d\tau, \\
&\qquad\qquad t_{i-1} < s < t_i < r.
\end{aligned} \tag{5.13}
$$

Consequently, (5.12) is changed to

$$
\begin{aligned}
&E\{f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r|\mathcal{F}_s\} \\
&\quad = E\bigg\{ \frac{1}{t_i-s}\int_s^{t_i} \sigma^{-1}(\tau, X_\tau)\nabla X_\tau D_\tau f(r, X_r, Y_r, Z_r)\,d\tau \bigg| \mathcal{F}_s \bigg\} \\
&\quad = E\{f(r, X_r, Y_r, Z_r)N_{t_i}^s|\mathcal{F}_s\}\nabla X_s, \qquad t_{i-1} < s < t_i < r.
\end{aligned} \tag{5.14}
$$

Combining (5.14) and (5.12) we see that for all $s \in (t_{i-1}, t_i)$ it holds that

$$
\begin{aligned}
(5.15)\qquad & E\left\{\int_s^T [f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r]\,dr \Big| \mathcal{F}_s\right\} \\
&\quad = E\left\{\int_s^T f(r, X_r, Y_r, Z_r) N^s_{r\wedge t_i}\,dr \Big| \mathcal{F}_s\right\} \nabla X_s.
\end{aligned}
$$

On the other hand, we note that for any $\tau \in (t_{i-1}, t_i)$ it holds that

$$
D_\tau g(X_{t_0}, \dots, X_{t_n}) = \sum_{j\ge i} \partial_j g D_\tau X_{t_j} = \left\{\sum_{j\ge i} \partial_j g \nabla X_{t_j}\right\} [\nabla X_\tau]^{-1} \sigma(\tau, X_\tau),
$$

which implies that, for any $s \in (t_{i-1}, t_i)$,

$$
\sum_{j\ge i} \partial_j g \nabla X_{t_j} = \frac{1}{t_i - s} \int_s^{t_i} D_\tau g(X_{t_0}, \dots, X_{t_n}) \sigma(\tau, X_\tau)^{-1} \nabla X_\tau\,d\tau.
$$

Thus, using integration by parts again we have

$$
\begin{aligned}
& E\left\{\sum_{j\ge i} \partial_j g \nabla X_{t_j} \Big| \mathcal{F}_s\right\} \\
(5.16)\qquad &\quad = E\left\{\frac{1}{t_i - s} g(X_{t_0}, \dots, X_{t_n}) \int_s^{t_i} \sigma^{-1}(\tau, X_\tau) \nabla X_\tau\,dW_\tau \Big| \mathcal{F}_s\right\} \\
&\quad = E\{g(X_{t_0}, \dots, X_{t_n}) N^s_{t_i} | \mathcal{F}_s\} \nabla X_s.
\end{aligned}
$$

Plugging (5.15) and (5.16) into (5.9) we obtain (5.2) for $s \in (t_{i-1}, t_i)$.

It is clear now that to prove the theorem we need only prove properties (i)–(iii), which we will do. Note that (i) is obvious, in light of Theorem 4.2 and thanks to representation (5.2). Property (ii) is a slight variation of Lemma 4.1, with T there being replaced by t_i, for each i. Therefore we shall only check (iii).

To this end, let $\Delta Z_{t_i} = Z_{t_i+} - Z_{t_i-}$. From (5.8) it is easily seen that

$$
Z_{t_i-} = \nabla^i Y_{t_i} [\nabla X_{t_i}]^{-1} \sigma(t_i, X_{t_i}), \qquad Z_{t_i+} = \nabla^{i+1} Y_{t_i} [\nabla X_{t_i}]^{-1} \sigma(t_i, X_{t_i}).
$$

Denoting $\alpha^i_s \triangleq -(\nabla^{i+1} Y_s - \nabla^i Y_s)$, $i = 1, \dots, n$, we then have

$$
\Delta Z_{t_i} = (\nabla^{i+1} Y_{t_i} - \nabla^i Y_{t_i}) [\nabla X_{t_i}]^{-1} \sigma(t_i, X_{t_i}) = -\alpha^i_{t_i} [\nabla X_{t_i}]^{-1} \sigma(t_i, X_{t_i}). \tag{5.17}
$$

Further, let us denote $\beta^i_s \triangleq -(\nabla^{i+1} Z_s - \nabla^i Z_s)$. Then (5.6) leads to that

$$
\alpha^i_s = \partial_i g \nabla X_{t_i} + \int_s^T [f_y(r)\alpha^i_r + f_z(r)\beta^i_r]\,dr - \int_s^T \beta^i_r\,dW_r, \qquad s \in [t, T]. \tag{5.18}
$$

In other words, (α^i, β^i) is the adapted solution to the linear BSDE (5.18). Therefore, by Lemma 2.2 we know that $\forall\, p > 0$ there exists a $C_p > 0$ such that $E\{|\alpha^i_{t_i}|^p\} \le C_p$. Note now that the same estimate holds for $\sigma(s, X_s)$ because of assumption (A1) and Theorem 3.3; for $[\nabla X_{t_i}]^{-1}$ since the process $[\nabla X_\cdot]^{-1}$ is the solution of the following $d \times d$-matrix-valued SDE:

$$\Phi_s = I - \int_t^s \Phi_r \left[\partial_x b(X_r) - \sum_{i=1}^d [\partial_x \sigma^j(r, X_r)]^2 \right] dr$$

$$- \sum_{j=1}^d \int_t^s \Phi_r [\partial_x \sigma^j(r, X_r)] \, dW_r^j,$$

it is readily seen that (5.3) follows from (5.17). This proves (iii).

Finally, we note that when f and g are only Lipschitz, (5.2) still holds, modulo a standard approximation the same as that in Theorem 4.2. Thus properties (i) and (ii) are obvious. To see (iii) we should note that the standard approximation yields that $\Delta Z_{t_i}^\varepsilon \to \Delta Z_{t_i}$ a.s. So if (5.3) holds for $\Delta Z_{t_i}^\varepsilon$, then letting $\varepsilon \to 0$ we see that (5.3) remains true for ΔZ_{t_i}, thanks to the Fatou lemma. The proof is now complete. □

Acknowledgment. We are very grateful to the anonymous referees for their careful reading of the original manuscript and for many useful suggestions.

REFERENCES

[1] BERMIN, H.-P. (1998). Hedging options: the Malliavin calculus approach versus the Δ-hedging approach. Preprint.

[2] BISMUT, J. M. (1973). *Théorie Probabiliste du Contrôle des Diffusions*. Amer. Math. Soc., Providence, RI.

[3] DELLACHERIE, C. and MEYER, P. (1978). *Probabilities and Potential*. North-Holland, Amsterdam.

[4] EL KAROUI, N. and MAZLIAK, L. (1997). *Backward Stochastic Differential Equations*. Longman Sci. Tech., Harlow.

[5] EL KAROUI, N., PENG, S. and QUENEZ, M. C. (1997). Backward stochastic differential equations in finance. *Math. Finance* **7** 1–72.

[6] FOURNIÉ, E., LASRY, J.-M., LEBUCHOUX, J., LIONS, P.-L. and TOUZI, N. (1997). An application of Malliavin calculus to Monte Carlo methods in finance. Preprint.

[7] GIKHMAN, I. I. and SKOROHOD, A. V. (1972). *Stochastic Differential Equations*. Springer, Berlin.

[8] HAMADENE, S. (1996). Équations différentielles stochastiques rétrogrades: le cas localment Lipschitzien. *Ann. Inst. H. Poincaré* **32** 645–659.

[9] HU, Y. (1999). On the existence of solutions to one-dimensional forward–backward SDEs. Preprint.

[10] HU, Y. (1999). On the solutions of forward–backward SDEs with monotone and continuous coefficients. Preprint.

[11] HU, Y. and YONG, J. (1999). Forward–backward stochastic differential equations with nonsmooth coefficients. Preprint.

[12] KARATZAS, I. and SHREVE, S. E. (1987). *Brownian Motion and Stochastic Calculus*. Springer, Berlin.

[13] LEPELTIER, M. and SAN MARTIN, J. (1997). Backward stochastic differential equations with nonLipschitz coefficients. *Statist. Probab. Lett.* **32** 425–430.

[14] MA, J., PROTTER, P. and YONG, J. (1994). Solving forward–backward stochastic differential equations explicitly—a four step scheme. *Probab. Theory Related Fields* **98** 339–359.

[15] MA, J. and YONG, J. (1999). *Forward–Backward Stochastic Differential Equations and Their Applications. Lecture Notes in Math.* **1702**. Springer, Berlin.

[16] NUALART, D. (1995). *The Malliavin Calculus and Related Topics*. Springer, Berlin.

[17] PARDOUX, E. and PENG, S. (1990). Adapted solutions of backward stochastic equations. *Systems Control Lett.* **14** 55–61.

[18] PARDOUX, E. and PENG S. (1992). Backward stochastic differential equations and quasilinear parabolic partial differential equations. *Lecture Notes in Control and Inform. Sci.* **176** 200–217. Springer, New York.

[19] PARDOUX, E. and PENG, S. (1995). Some backward stochastic differential equations with non-Lipschitz coefficients. Preprint.

[20] PROTTER, P. (1990). *Integration and Stochastic Differential Equations*. Springer, Berlin.

[21] REVUZ, D. and YOR, M. (1991). *Brownian Motion and Continuous Martingales*. Springer, Berlin.

[22] YONG, J. and ZHOU, X. (1999). *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer, Berlin.

DEPARTMENT OF MATHEMATICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907-1395
E-MAIL: majin@math.purdue.edu

SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455
E-MAIL: jfzhang@math.umn.edu