SOME FINE PROPERTIES OF BACKWARD STOCHASTIC DIFFERENTIAL EQUATIONS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Jianfeng Zhang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2001

To My Family

ACKNOWLEDGMENTS

I am greatly indebted to my advisors, Professor Jin Ma and Professor Jim Douglas, for their guidance and suggestions during my four years of study at Purdue University. In particular, I enjoyed and appreciate endless care, help, encouragement and support from Professor Ma, much more than what one could have expected from an advisor.

I would like to express my special thanks to Professor Philip Protter, one of my committee members, for all his help and encouragement in various aspects. I sincerely appreciate Professor Jaksa Cvitanić for his help on my research, Professor Frederi Viens for serving on my committee, and Dr. Rita Saerens for her supervision on my teaching. Many thanks are also due to Professors Xiaoyun Zhao, Yinnan Zhang and Xianliang Shi for guiding me to the area of Mathematics.

Our secretaries Nancy Eberle, Lil Reifel and Jackie Oswalt deserve grateful appreciation for their help in many ways. I particularly appreciate a special group of colleagues with whom I enjoyed wonderful collaboration during my study, among them: Xiang Long, Xiaodong Sun and Liqing Yan. I would also like to thank all my good friends, especially Rong Li and Tianyu Yang, for their friendship.

As always, my parents and my sister's family well deserve my love and thanks.

TABLE OF CONTENTS

Page

ABSTRACT

Jianfeng Zhang. Ph.D., Purdue University, August 2001. Some Fine Properties of Backward Stochastic Differential Equations. Major Advisors: Jin Ma and Jim Douglas.

In this thesis we investigate various properties of the martingale part, usually denoted by Z, of the solution to a class of Backward Stochastic Differential Equations (BSDEs, for short), with path-dependent terminals.

We first establish some Feynman-Kac type representation formulae for the process Z for BSDEs with simple terminals. The main feature of these formulae is that they do not involve the derivatives of the coefficients of the BSDEs, and the main device is the Malliavin Calculus. We also provide a probabilistic approach towards the classical solution to a nonlinear PDE.

By extending our representation formulae and using some approximating techniques, we prove that, for a large class of BSDEs with path-dependent terminals, the process Z is pathwisely càdlàg (right continuous with left limits), or even continuous. Our proof of convergence relies heavily on the Meyer-Zheng tightness criterion.

Based on the above results, we propose a "two-step scheme" to numerically solve BSDEs with path-dependent terminals. Our scheme (strongly) converges in L^2, under mild conditions, with rate of convergence $\sqrt{\frac{\log n}{n}}$.

Finally, with the same spirit but different techniques, we extend our representation formulae and path regularity results to models driven by Lévy processes, motivated by questions arising in financial asset pricing theory where the market is incomplete.

CHAPTER 1. INTRODUCTION

1.1 Problem Description

The motivation for studying *Backward Stochastic Differential Equations* (BSDEs, for short) comes originally from stochastic optimal control theory, and the theory can be traced back to Bismut [6] (1973) who studied the linear case. In 1990, Pardoux-Peng [43] proved the well-posedness for nonlinear BSDEs. Since then, BSDEs have been extensively studied and used in many applied and theoretical areas, particularly in Mathematical Finance. Moreover, initiated by Antonelli [1] (1993), Forward-Backward Stochastic Differential Equations (FBSDEs) are also investigated systematically, especially Ma-Protter-Yong [35] established the *Four Step Scheme*, and Hu-Peng [26] and Yong [50] established the *Method of Continuation*. For the theory and application of BSDEs and FBSDEs, we refer the readers to the books of El Karoui-Mazliak [18], Ma-Yong [37], as well as Yong-Zhou [51]; and the survey paper of El Karoui-Peng-Quenez [19].

Now let us turn into our specific subject. We begin with the following set-up. Unless otherwise specified, we let $T > 0$ be a *fixed* terminal time, and $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ be a complete filtered probability space on which is defined a d-dimensional standard Brownian motion W, such that $\mathbf{F} = \{\mathcal{F}_t\}_{0\le t\le T}$ is the natural filtration of W, augmented by all the $\mathbf{P}$-null sets.

A well-investigated class of decoupled FBSDEs is of the following form:

$$\begin{cases} X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s, \\ Y_t = g(X_T) + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T \langle Z_s, dW_s \rangle, \end{cases} \tag{1.1}$$

where all the coefficient functions are deterministic, and $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^d$. In one of their seminar works [44] (1992), Pardoux-Peng proved that, among other things, the adapted solution to the BSDE in (1.1) provided a probabilistic representation to the *viscosity* solution to some nonlinear parabolic PDE, in the spirit of the well known Feynman-Kac formula; On the other hand, they proved that Y_t and Z_t can be expressed as *deterministic* functions of (t, X_t). This point of view is essential in this thesis.

Although BSDEs have received intensive attention during the past decade, people have much less knowledge of the process Z than that of Y. The main difficulty for studying Z lies in the fact that it is the "derivative" of Y, either in the sense of Feynman-Kac formula (cf. e.g., [44] or [35]) or in the sense of Clark-Ocone formula (cf. e.g., [42]). Consequently, Z does not behave as nicely as Y does. In practice, however, Z is very useful. It is interpreted as the hedging strategy in Mathematical Finance theory, for example. Due to the practical significance and the theoretical challenge, the process Z plays a special role in BSDEs theory.

My work mainly focuses on various properties of the process Z, in particular when the terminal value of the BSDE is a true functional of the forward diffusion, motivated by exotic options in Finance theory. Our goal is to find an efficient numerical method for quite general BSDEs. Along the way getting there we also obtain some fine properties of the process Z, in particular some new representation formulae and some path regularity results. These results are interesting in their own rights.

We begin by establishing some nonlinear Feynman-Kac type representation formulae for the process Z, when the terminal of the BSDE is of the form $g(X_T)$ for some function g. Although Z is the "derivative" of Y, as mentioned above, our new formulae *do not* involve the derivatives of the coefficients of the BSDE, and thus hold true even when those coefficients are not differentiable. The main device of our proof is an integration by parts formula of Malliavin Calculus. Such an idea was recently employed in numerical finance for computing various "greeks" of the market (cf. Fournié, et al. [21]). Using the same idea we also provide a probabilistic representation formula for the second order derivatives of the viscosity solution to a nonlinear PDE, and prove that, under certain conditions, that viscosity solution is in fact the classical solution to the PDE.

We further extend our formulae to the case that the terminal of the BSDE depends on finite number of values of the forward diffusion X. Then by using some delicate approximating techniques, especially the Meyer-Zheng pseudo-path topology as the key device for proof of convergence, we may study the path regularity of Z, in the case that the terminal of the BSDE is a true *functional* of X. We show that the process Z admits a càdlàg version if the terminal of the BSDE is an "L^∞-Lipschitz" functional of the forward diffusion, and Z is even continuous if the terminal is an "L^1-Lipschitz" functional.

It is worth pointing out here that the path regularity of the process Z has been an open problem in BSDE literature for about a decade. In general the martingale part of the solution to a BSDE is not regular in a pathwise manner. For example, if we let $\xi = \int_0^T h_s dW_s$ for any $\mathbf{F}$-predictable process h, such that $E\{\int_0^T |h_s|^2 ds\} < \infty$, and let (Y, Z) be the $\mathbf{F}$-adapted solution to the following simple BSDE:

$$Y_t = \xi - \int_t^T \langle Z_s, dW_s \rangle, \quad t \in [0, T],$$

then by the uniqueness of martingale representation theorem (or by the uniqueness of the adapted solution to a BSDE) we have $Z = h$, which has no path regularity in general. But as we will prove in this thesis, the process Z does have path regularity when the terminal of the BSDE is a "nice" functional of the forward diffusion X. Our results will enable one to put the solution pair (Y, Z) in a *canonical* path space, such as the well-known $\mathbb{D}$-space with Skorohod topology, which opens the door to many further studies on BSDEs, especially to those concerning the solutions in a *weak* sense, both theoretically and numerically.

The path regularity of Z also contributes greatly to our numerical scheme for BSDEs with path-dependent terminals. To my knowledge, to date there are few numerical methods for BSDEs in publication, and all the existing methods either require high regularity of the coefficients (thus Z is "nice", see [44] or [35], for example) or lack a good rate of convergence (because Z is not "nice"). Relying heavily on our regularity results of Z, our scheme improves in both aspects. To be more specific, it converges strongly in L^2 under certain mild conditions, with a rate of convergence $\sqrt{\frac{\log n}{n}}$ for L^∞-Lipschitz terminals, and $\frac{1}{\sqrt{n}}$ for L^1-Lipschitz terminals. Another feature of this new method is that both Y and Z are approximated by step processes, which is also important in practice. But, although it is already an improvement compared with some other methods, our scheme still faces the "high dimension" problem. That is, in high dimensional case it also requires high computational cost.

Finally we investigate the case where the forward diffusion is driven by Lévy processes, which leads to an incomplete financial market. In the same spirit but with different techniques from the Brownian case, we establish some representation formulae for the risking minimizing hedging strategy (corresponding to the process Z) when the terminal payoff is a "discrete functional" of the underlying assets prices, and furthermore prove that that hedging strategy admits a càglàd version for a quite general class of path-dependent payoffs.

The rest of the thesis is organized as follows. In next section we present some preliminaries. In Chapter 2 some nonlinear Feynman-Kac type representation formulae are established and the relationship between PDEs and BSDEs is reinvestigated. In Chapter 3 we prove the path regularity results, and in Chapter 4 we propose a "two-step" scheme to numerically solve BSDEs with path-dependent terminals. The last chapter deals with the Lévy case.

1.2 Preliminaries

In this section we list the notations used throughout the thesis and present some useful results which are either standard or slight variations of the well-known results in their own literature. We give only the statements for ready references.

1.2.1 Definitions and Notations

Recall that $T > 0$ is a fixed time and that $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ is a complete probability space on which is defined a d-dimensional Brownian motion W. Let $\mathcal{X}$ denote a generic Banach space. In particular, $\mathcal{X}$ denotes a generic Euclidean space when derivatives are involved, and regardless of their dimensions we denote $\langle \cdot, \cdot \rangle$ and $|\cdot|$ to be the inner product and norm in all $\mathcal{X}$'s, respectively. The following spaces will be frequently used in the sequel:

- for $0 \le s < t \le T$, $\mathbb{D}[s,t]$ is the space of all càdlàg (right continuous with left limits) vector functions $\varphi : [s,t] \mapsto \mathbb{R}^k$, where k can be specified from the context. In particular, $\mathbb{D} = \mathbb{D}[0,T]$.

- for $0 \le p < \infty$, $L^p([0,T]; \mathcal{X})$ is the space of all measurable functions $\varphi : [0,T] \mapsto \mathcal{X}$ such that $\int_0^T |\varphi(t)|^p dt\} < \infty$; and also, $\xi \in L^\infty(\mathbf{F}, [0,T]; \mathcal{X})$ means it is a process uniformly bounded in (t, ω);

- for integers ℓ, $C^\ell([0,T], \mathcal{X}_1; \mathcal{X}_2)$ is the space of all continuous functions $\varphi : [0,T] \times \mathcal{X}_1 \mapsto \mathcal{X}_2$, such that φ is $\ell-$times differentiable with respect to the spatial variables. In particular, $C([0,T], \mathcal{X}_1; \mathcal{X}_2) \stackrel{\triangle}{=} C^0([0,T], \mathcal{X}_1; \mathcal{X}_2)$.

- for integers ℓ, $C_b^\ell([0,T], \mathcal{X}_1; \mathcal{X}_2)$ is the subspace of $C^\ell([0,T], \mathcal{X}_1; \mathcal{X}_2)$ such that its element φ has uniformly bounded partial derivatives for all orders up to ℓ.

- $C_L([0,T], \mathcal{X}_1; \mathcal{X}_2)$ is the space of all continuous functions $\varphi : [0,T] \times \mathcal{X}_1 \mapsto \mathcal{X}_2$, such that φ is uniformly Lipschitz continuous with respect to the spatial variables.

- $C_L^{\frac{1}{2}}([0,T], \mathcal{X}_1; \mathcal{X}_2)$ is the subspace of $C_L([0,T], \mathcal{X}_1; \mathcal{X}_2)$ such that its element φ is Hölder-$\frac{1}{2}$ continuous with respect to time, with uniformly bounded Lipschitz and Hölder constants.

- for $0 \le p < \infty$, $L^p(\mathbf{F}, [0,T]; \mathcal{X})$ is the space of all $\mathcal{X}$-valued, $\mathbf{F}$-adapted processes ξ satisfying $E\{\int_0^T \|\xi_t\|_{\mathcal{X}}^p dt\} < \infty$.

- $C(\mathbf{F},[0,T],\mathcal{X}_1;\mathcal{X}_2)$ is the space of all continuous $\mathbf{F}$-adapted processes $\varphi:\Omega\times[0,T]\times\mathcal{X}_1\mapsto\mathcal{X}_2$;

- $C_L(\mathbf{F},[0,T],\mathcal{X}_1;\mathcal{X}_2)$ is the space of all $\mathbf{F}$-adapted processes $\varphi:\Omega\times[0,T]\times\mathcal{X}_1\mapsto\mathcal{X}_2$, such that for a.s. ω, $\varphi(\omega,.)\in C_L([0,T],\mathcal{X}_1;\mathcal{X}_2)$ and the Lipschitz constants are uniformly bounded for all ω.

When the context is clear, for notational simplicity we often omit the time and space parameters. For example, we write C_L instead of $C_L([0,T],\mathcal{X}_1;\mathcal{X}_2)$, and $C_L(\mathbf{F})$ instead of $C_L(\mathbf{F},[0,T],\mathcal{X}_1;\mathcal{X}_2)$, etc..

The main object of this thesis is the following decoupled FBSDE:

$$\begin{cases} X_t = x+\int_0^t b(s,X_s)ds+\int_0^t \sigma(s,X_s)dW_s, \\ Y_t = \Phi(X)+\int_t^T f(s,X_s,Y_s,Z_s)ds-\int_t^T Z_s dW_s, \end{cases} \tag{1.2}$$

where b,σ, and f are deterministic functions, taking values in $\mathbb{R}^{d_1},\mathbb{R}^{d_1\times d}$, and $\mathbb{R}$, respectively, and $\Phi:\mathbb{D}\mapsto\mathbb{R}$ is a deterministic *functional*. We denote the solution to (1.2) by $\Theta\stackrel{\triangle}{=}(X,Y,Z)$, where X is of dimension d_1, Y is scalar, and Z is of dimension d. For notational simplicity, we take the convention that X and b are *column* vectors while Z is a *row* vector. Furthermore, for a scalar function φ on $\mathbb{R}^n$, we denote $\partial_x\varphi=(\frac{\partial\varphi}{\partial x_1},\cdots,\frac{\partial\varphi}{\partial x_n})$ as a *row* vector. Note that if $\varphi=(\varphi^1,\cdots,\varphi^m)^T:\mathbb{R}^n\mapsto\mathbb{R}^m$, then $\partial_x\varphi$ is a matrix whose j-th row is $\partial_x\varphi^j$.

In this thesis, we content ourselves with functionals Φ of the following four types.

Definition 1.2.1 A functional $\Phi:\mathbb{D}\mapsto\mathbb{R}$ is called

- *simple*, if $\Phi(\mathbf{x})=g(\mathbf{x}(T))$, for $\forall\mathbf{x}\in\mathbb{D}$;

- *discrete*, if $\Phi(\mathbf{x})=g(\mathbf{x}(t_0),\cdots,\mathbf{x}(t_n))$, for $\forall\mathbf{x}\in\mathbb{D}$, where $0=t_0<\cdots<t_n=T$;

- L^∞*-Lipschitz*, if there exists a constant K such that

$$|\Phi(\mathbf{x}_1)-\Phi(\mathbf{x}_2)|\le K\sup_{0\le t\le T}|\mathbf{x}_1(t)-\mathbf{x}_2(t)|,\quad\forall\mathbf{x}_1,\mathbf{x}_2\in\mathbb{D}; \tag{1.3}$$

- L^1-*Lipschitz*, if Φ satisfies the following estimate

$$|\Phi(\mathbf{x}_1) - \Phi(\mathbf{x}_2)| \le K \int_0^T |\mathbf{x}_1(t) - \mathbf{x}_2(t)|dt, \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{D}. \tag{1.4}$$

Remark 1.2.2 L^1-Lipschitz implies L^∞-Lipschitz. If Φ is simple or discrete, then Φ is L^∞-Lipschitz if and only if the corresponding g is a Lipschitz function.

1.2.2 Some Results of BSDEs

We begin with the well-posedness of a BSDE.

Theorem 1.2.3 [43] For $\forall \xi \in L^2(\mathcal{F}_T)$ and $\forall f \in C_L(\mathbf{F})$, the following BSDE:

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dW_s,$$

has a unique solution $(Y, Z) \in C(\mathbf{F}) \times L^2(\mathbf{F})$.

To see the relation between PDEs and BSDEs, we consider the following parabolic PDE:

$$\begin{cases} u_t + \frac{1}{2}tr\{\sigma\sigma^T u_{xx}\} + bu_x + f(t, x, x, u_x\sigma) = 0; \\ u(T, x) = g(x), \end{cases} \tag{1.5}$$

and decoupled FBSDE:

$$\begin{cases} X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dW_r, \\ Y_s^{t,x} = g(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr - \int_s^T Z_r^{t,x} dW_r, \end{cases} \quad \forall s \in [t, T]. \tag{1.6}$$

Here the superscript ^{t,x} indicates the dependence of the solution on the initial date and value (t, x), and it will be omitted when the context is clear. Our result is:

Theorem 1.2.4 [44] If the PDE (1.5) has a classical solution u, then $Y_s^{t,x} \triangleq u(s, X_s^{t,x})$ and $Z_s^{t,x} \triangleq u_x(s, X_s^{t,x})\sigma(s, X_s^{t,x})$ solve the BSDE in (1.6). On the other hand, if $b, \sigma, f, g \in C_L$, then (1.6) has an $\mathbf{F}$-adapted solution and the function $u(t, x) \triangleq Y_t^{t,x}$ is the unique viscosity solution to the PDE (1.5).

The following two estimates are very useful in our future discussion.

Lemma 1.2.5 [23] Suppose that $\widetilde{b}, \widetilde{\sigma} : \Omega \times [0,T] \times \mathcal{X}_1 \mapsto \mathcal{X}_2$ are $\mathbf{F}$-adapted processes such that they are uniformly Lipschitz continuous with respect to $x \in \mathcal{X}_1$, with a common Lipschitz constant $K > 0$. We assume further that $\widetilde{b}(t,0), \widetilde{\sigma}(t,0) \in L^2(\mathbf{F})$. Let X be the solution to the following SDE:

$$X_t = x + \int_0^t \widetilde{b}(s, X_s)ds + \int_0^t \widetilde{\sigma}(s, X_s)dW_s. \tag{1.7}$$

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

$$E\Big\{ \sup_{0 \leq t \leq T} |X_t|^p \Big\} \leq C_p E\Big\{ |x|^p + \int_0^T \Big[|\widetilde{b}(t,0)|^p + |\widetilde{\sigma}(t,0)|^p \Big] dt \Big\};$$

and

$$E\{|X_t - X_s|^p\} \leq C_p E\Big\{ |x|^p + \sup_{0 \leq t \leq T} |\widetilde{b}(t,0)|^p + \sup_{0 \leq t \leq T} |\widetilde{\sigma}(t,0)|^p \Big\} |t-s|^{\frac{p}{2}}.$$

Lemma 1.2.6 [19] Assume that $\widetilde{f} : \Omega \times [0,T] \times \mathcal{X}_1 \mapsto \mathcal{X}_2$ is an $\mathbf{F}$-adapted process such that it is uniformly Lipschitz continuous with respect to $x \in \mathcal{X}_1$, with a common Lipschitz constant $K > 0$. We assume further that $\widetilde{f}(t,0,0) \in L^2(\mathbf{F})$. For any $\xi \in L^2(\mathcal{F}_T; \mathbb{R})$, let (Y, Z) be the adapted solution to the BSDE:

$$Y_t = \xi + \int_t^T \widetilde{f}(s, Y_s, Z_s)ds - \int_t^T Z_s dW_s. \tag{1.8}$$

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

$$E\Big\{ \sup_{0 \leq t \leq T} |Y_t|^p + \Big(\int_0^T |Z_t|^2 dt \Big)^{\frac{p}{2}} \Big\} \leq C_p E\Big\{ |\xi|^p + \int_0^T |\widetilde{f}(t,0,0)|^p dt \Big\};$$

and

$$E\{|Y_t - Y_s|^p\} \leq C_p E\left\{ \Big[|\xi|^p + \sup_{0 \leq t \leq T} |\widetilde{f}(t,0,0)|^p \Big] |t-s|^{p-1} + \Big(\int_s^t |Z_r|^2 dr \Big)^{\frac{p}{2}} \right\}.$$

We conclude this subsection with a stability result.

Lemma 1.2.7 Let $\Theta = (X, Y, Z)$ denote the solution to FBSDE (1.7) and (1.8), and $\Theta^\varepsilon = (X^\varepsilon, Y^\varepsilon, Z^\varepsilon)$ denote the solution to a corresponding FBSDE with parameters $x^\varepsilon, \widetilde{b}^\varepsilon, \widetilde{\sigma}^\varepsilon, \widetilde{f}^\varepsilon$ and ξ^ε. Assume all the coefficients involved satisfy the conditions in Lemmas 1.2.5 and 1.2.6. If for $\forall (t, x_0)$,

$$\lim_{\epsilon \to 0} E\Big\{|x^\varepsilon - x|^2 + |\widetilde{b}^\varepsilon(t, x_0) - \widetilde{b}(t, x_0)|^2 + |\widetilde{\sigma}^\varepsilon(t, x_0) - \widetilde{\sigma}(t, x_0)|^2\Big\} = 0,$$

then

$$\lim_{\epsilon \to 0} E\Big\{\sup_{0 \le t \le T} |X_t^\varepsilon - X_t|^2\Big\} = 0.$$

If we assume further that, for $\forall (y_0, z_0)$,

$$\lim_{\epsilon \to 0} E\Big\{|\xi^\varepsilon - \xi|^2 + \int_0^T |\widetilde{f}^\varepsilon(t, y_0, z_0) - \widetilde{f}(t, y_0, z_0)|^2 dt\Big\} = 0,$$

then

$$\lim_{\epsilon \to 0} E\Big\{\sup_{0 \le t \le T} |Y_t^\varepsilon - Y_t|^2 + \int_0^T |Z_t^\varepsilon - Z_t|^2 dt\Big\} = 0.$$

1.2.3 Motivation in Finance

Along its theoretical development, BSDEs have been applied to various areas, especially in Mathematical Finance. In a financial derivative market, FBSDE (1.2) has the following interpretation. The process X represents the underlying assets price; $\Phi(X)$ is the terminal payoff; Y is the wealth process, or the security price; and Z is the hedging strategy, or stock portfolio.

To see the connection between SDEs and finance more specifically, let us consider the standard Black-Schole market model:

$$\begin{cases} dS_t^0 &= S_t^0 r_t dt, & \text{(Bond/Money Market)} \\ S_0^0 &= s^0, & \\ dS_t^i &= S_t^i \{b_t^i dt + \sum_{j=1}^d \sigma_t^{ij} dW_t^j\}, & \text{(Stocks)} \\ S_0^i &= s^i, \qquad i = 1, \cdots, d. & \end{cases}$$

We adopt the following notations:

- S_t^0, S_t^i—prices of bond/(i-th) stocks (per share) at time t;

- r_t—interest rate at time t;
- $\{b_t^i\}_{i=1}^N$—appreciation rates of the market at time t;
- $[\sigma_t^{ij}]$—volatility matrix of the market at time t;
- V_t—dollar amount of the total wealth of an investor at time t;
- π_t^i—dollars invested in i-th stock at time t, $i = 1, \cdots, N$;
- $V_t - \sum_{i=1}^{N} \pi_t^i$—dollars in the bond at time t;
- C_t—cumulated consumption up to time t.

Then, V satisfies the SDE: for $t \in [0, T]$,

$$\begin{cases} dV_t = [r_t V_t + \pi_t(b_t - r_t \mathbf{1})]dt + \pi_t \sigma_t dW_t - dC_t, \\ V_0 = v. \end{cases} \tag{1.9}$$

where $\mathbf{1} = (1, \cdots, 1)^T$, π is a row vector and b is a column vector.

Note that (1.9) is a forward SDE, since the initial investment V_0 is given. But in derivative markets, for example in option markets, we do not know V_0; instead, we know the terminal payoff $V_T = \Phi(S)$. Following are some typical options: Let q denote the strike price,

- $\Phi(S) = (S_T - q)^+$—European option;
- $\Phi(S) = (\frac{1}{T}\int_0^T S_t dt - q)^+$—Asian option;
- $\Phi(S) = \max_{0 \le t \le T} S_t$—Look-back option;
- $\xi = 1_{\{S_T > q\}}$—Digital option;
- $\xi = (S_T - q)^+ 1_{\{\max_{0 \le t \le T} S_t \le h\}}$— European barrier option;
- $\xi = (S_\tau - q)^+$—American option (τ-stopping time).

Among them, European options are extensively studied. My work is mainly motivated by Asian options and Look-back options which are path dependent, but are continuous in the sense of L^1-Lipschitz and L^∞-Lipschitz, respectively.

Given the terminal payoff $\Phi(S)$, the calculation of the option price and stock portfolio leads to backward SDEs. To illustrate it, we take European option as an example. Define the "fair price" of an option to be

$$p = \inf\{v : \exists(\pi, C), \text{such that } \ V_T^{v,\pi,C} \geq \Phi(S)\}.$$

El Karoui-Peng-Quenez [19] showed that the fair price p of an European option and the corresponding "hedging strategy" (π, C) can be determined as follows:

- $C \equiv 0$;
- $p = V_0$, and $\pi_t = (\pi_t^1, \cdots, \pi_t^N)$, where (V, π) solves the BSDE:

$$V_t = \xi - \int_t^T \Big[r_s V_s + \pi_s(b_s - r_s \mathbf{1})\Big] ds - \int_t^T \pi_s \sigma_s dW_s. \tag{1.10}$$

While obviously we may identify (X, Y) with (S, V), from (1.10) we see that Z represents $\pi\sigma$. Since σ is given by the underlying assets market, and we assume that σ is nondegenerate, we know that, in order to find π, it is equivalent to solve for Z.

1.2.4 Basics of Malliavin Calculus

In this subsection we review some basic facts of Malliavin calculus (also known as anticipating stochastic calculus), especially those related to the SDEs. We refer the readers to Nualart [42] for the basic theory and to Pardoux-Peng [44] for the results related to BSDEs. To begin with, let $\mathcal{S}$ be the space of all random variables of the form

$$\xi = F\Big(\int_0^T \varphi_1(t) dW_t, \cdots, \int_0^T \varphi_n(t) dW_t\Big),$$

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

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

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

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

$$\|\xi\|_{1,2}^2 = E|\xi|^2 + E\int_0^T |D_t\xi|^2 dt, \quad \forall \xi \in \mathcal{S},$$

and we denote $\mathbb{D}^{1,2}$ to be the completion of $\mathcal{S}$ in $L^2(\Omega)$ under $\|\cdot\|_{1,2}$. It can be shown (see, e.g., [42]) 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 Malliavin calculus to SDEs, we consider FBSDE (1.6) on the subinterval $[t,T] \subseteq [0,T]$. The following *variational equation* of (1.6) will play an important role in this thesis: for $i = 1, \cdots, d_1$,

$$\begin{cases} \nabla_i X_s = e_i + \displaystyle\int_t^s \partial_x b(r, X_r)\nabla_i X_r dr + \sum_{j=1}^d \int_t^s [\partial_x \sigma^j(r, X_r)]\nabla_i X_r dW_r^j, \\ \nabla_i Y_s = \partial_x g(X_T)\nabla_i X_T + \displaystyle\int_s^T [\partial_x f(r,\Theta_r)\nabla_i X_r + \partial_y f(r,\Theta_r)\nabla_i Y_r \\ \qquad\qquad + \langle\, \partial_z f(r,\Theta_r), \nabla_i Z_r \,\rangle]dr - \displaystyle\int_s^T \nabla_i Z_r dW_r, \end{cases} \tag{1.11}$$

where $e_i = (0, \cdots, \overset{i}{1}, \cdots, 0)^T$ is the i-th coordinate vector of $\mathbb{R}^{d_1}$; $\sigma^j(\cdot)$ is the j-th column of the matrix $\sigma(\cdot)$. Note that we omitted the superscript ^{t,x} in (1.11). Further, we denote

$$\nabla X = (\nabla_1 X, \cdots, \nabla_{d_1} X),\ \nabla Y = (\nabla_1 Y, \cdots, \nabla_{d_1} Y),\ \nabla Z = ([\nabla_1 Z]^T, \cdots, [\nabla_{d_1} Z]^T).$$

Then, $(\nabla X, \nabla Y, \nabla Z) \in C(\mathbf{F};\mathbb{R}^{d_1\times d_1}) \times C(\mathbf{F};\mathbb{R}^{d_1}) \times L^2(\mathbf{F};\mathbb{R}^{d\times d_1})$, and satisfy the following FBSDE:

$$\begin{cases} \nabla X_s = I_{d_1} + \displaystyle\int_t^s \partial_x b(r, X_r)\nabla X_r dr + \sum_{j=1}^d \int_t^s [\partial_x \sigma^j(r, X_r)]\nabla X_r dW_r^j, \\ \nabla Y_s = \partial_x g(X_T)\nabla X_T + \displaystyle\int_s^T [\partial_x f(r,\Theta_r)\nabla X_r + \partial_y f(r,\Theta_r)\nabla Y_r \\ \qquad\qquad + \partial_z f(r,\Theta_r)\nabla Z_r]dr - \Big(\displaystyle\int_s^T [\nabla Z_r]^T dW_r\Big)^T. \end{cases} \tag{1.12}$$

Note that the $d_1 \times d_1$-matrix-valued process ∇X satisfies the linear SDE above and $\nabla X_t = I$, thus $[\nabla X_s]^{-1}$ exists for all $s \in [t,T]$, P-a.s. Moreover, by applying Itô's formula we may easily prove that $[\nabla X_s]^{-1}$ is the solution to the following SDE:

$$\begin{aligned} [\nabla X_s]^{-1} &= I_{d_1} - \int_t^s [\nabla X_r]^{-1}\Big(\partial_x b - \sum_{j=1}^d (\partial_x \sigma^j)^2\Big)(r, X_r)dr \\ &\quad - \sum_{j=1}^d \int_t^s [\nabla X_r]^{-1}\partial_x \sigma^j(r, X_r)dW_r^j. \end{aligned} \tag{1.13}$$

The following lemma concerns the Malliavin derivatives of the solution (X,Y,Z) to (1.6). Since the proof is standard and can be found in, e.g., Nualart [42] and Pardoux-Peng [44], we omit it.

Lemma 1.2.8 Assume that $b, \sigma, f, g \in C_b^1$. Then $(X,Y,Z) \in L^2([0,T]; \mathbb{D}^{1,2})$, and there exists a version of (D_sX_r, D_sY_r, D_sZ_r) that satisfies

$$\begin{cases} D_sX_r = \nabla X_r[\nabla X_s]^{-1}\sigma(s,X_s)1_{\{s\le r\}}; \\ D_sY_r = \nabla Y_r[\nabla X_s]^{-1}\sigma(s,X_s)1_{\{s\le r\}}; \\ D_sZ_r = \nabla Z_r[\nabla X_s]^{-1}\sigma(s,X_s)1_{\{s\le r\}}, \end{cases} \tag{1.14}$$

where

$$D_sX_r \triangleq \begin{bmatrix} D_sX_r^1 \\ \vdots \\ D_sX_r^{d_1} \end{bmatrix}, \qquad \text{and} \qquad D_sZ_r \triangleq \begin{bmatrix} D_sZ_r^1 \\ \vdots \\ D_sZ_r^d \end{bmatrix}.$$

To conclude this subsection let us introduce the notion of *Skorohod integral* (or *Hitsuda-Skorohod integral*) which will be one of the key devices in this thesis. 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

$$E\{F\delta(u)\} = E\int_0^T D_tFu_tdt, \qquad \forall F \in \mathbb{D}^{1,2}, \quad \forall u \in \mathrm{Dom}(\delta), \tag{1.15}$$

where $\mathrm{Dom}(\delta) \triangleq \Big\{u \in L^2(\Omega \times [0,T]; \overset{d}{\mathbb{R}}) : |E\{\int_0^T D_tFu_tdt\}| \le C\|F\|_{1,2},\ \forall F \in \mathbb{D}^{1,2}\Big\}$. The operator δ is then called the Skorohod integral of the process u, and by a slight abuse of notation, we still denote it as

$$\delta(u) = \int_0^T \langle\, u_t, dW_t \,\rangle, \qquad u \in \mathrm{Dom}(\delta). \tag{1.16}$$

One should keep in mind that, if in the sequel the integrand of a stochastic integral is not **F**-adapted, then it should always be understood as a Skorohod integral. On the other hand, it can be shown that, if $u \in L^2(\mathbf{F}, [0,T]; \mathbb{R}^d)$, then $u \in$Dom(δ), and the Skorohod integral (1.16) coincides with the usual Itô integral. Furthermore, we have the following important properties of such integrals (cf. [42]).

Lemma 1.2.9 Suppose that $F \in \mathbb{D}^{1,2}$. Then

(i) (Integration by parts formula): for any $u = (u^1, \cdots, u^n) \in (\text{Dom}(\delta))^n$ and $F \in L^2(\mathcal{F}_T)$ such that $Fu \in L^2([0,T] \times \Omega; \mathbb{R}^{n\times d})$, one has $Fu \in (\text{Dom}(\delta))^n$, and it holds that

$$\Big(\int_0^T F u_t^T dW_t\Big)^T = \delta(Fu) = \Big(F\int_0^T u_t^T dW_t\Big)^T - \int_0^T D_t F u_t dt;$$

(ii) (Clark-Haussmann-Ocone formula):

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

1.2.5 Meyer-Zheng Tightness Criterion

In this subsection we introduce the notions of *pseudo-path topology* and *quasi-martingales* (cf. Dellacherie-Meyer [15] or Meyer-Zheng [41]), adjusted to our setting. To begin with, note that $\mathbb{D} \subset L^0([0,T])$. For any $w \in L^0([0,T])$, we define the *pseudo-path* of w to be a probability measure on $[0,T] \times \bar{\mathbb{R}}$:

$$P^w(A) \triangleq \frac{1}{T}\int_0^T 1_A(t, w(t))dt, \qquad \forall A \in \mathcal{B}([0,T] \times \bar{\mathbb{R}}). \tag{1.17}$$

It can be shown that the mapping $\psi : w \mapsto P^w$ is 1-1 on $L^0([0,T])$. Thus we can identify all $w \in L^0([0,T])$ with its pseudo-path; and we denote all pseudo-paths by Ψ. In particular, using the mapping ψ the space $\mathbb{D}$ can then be embedded into the compact space $\overline{\mathcal{P}}$ of all probability laws on the compact space $[0,T] \times \bar{\mathbb{R}}$ (with the Prohorov metric). Clearly, in this sense

$$\mathbb{D} \subset \Psi \subset \overline{\mathcal{P}}. \tag{1.18}$$

The induced topology on Ψ and $\mathbb{D}$ are known as the *pseudo-path topology* or sometimes called *Meyer-Zheng topology*. The following characterization of the Meyer-Zheng topology is worthing noting.

Lemma 1.2.10 *(Meyer-Zheng [41, Lemma 1])* The pseudo-path topology on Ψ is equivalent to the convergence in measure.

Furthermore, it is known that (see, e.g., [41]) that Ψ is a Polish space; and $\mathbb{D}$ is a Borel set in $\overline{\mathcal{P}}$. Consequently, we have

$$\mathcal{B}(\mathbb{D}) = \mathbb{D} \cap \mathcal{B}(\Psi) \stackrel{\triangle}{=} \{A \cap \mathbb{D} : \ A \in \mathcal{B}(\Psi)\}.$$

We now make the following observation. Denote $\mathcal{M}(\mathbb{D})$ to be the space of all probability measures on $\mathbb{D}$, and $\mathcal{M}(\Psi)$ be that of Ψ. Then, any probability measure $P \in \mathcal{M}(\mathbb{D})$ induces a probability measure $\widehat{P} \in \mathcal{M}(\Psi)$ by:

$$\widehat{P}(A) = P(A \cap \mathbb{D}), \qquad \forall A \in \mathcal{B}(\Psi). \tag{1.19}$$

In this sense we then have $\mathcal{M}(\mathbb{D}) \subset \mathcal{M}(\Psi)$.

The most significant application of the Meyer-Zheng topology is a tightness result for *quasimartingales*, which we now briefly describe. Let X be an $\mathbf{F}$-adapted, càdlàg process defined on $[0,T]$, such that $E|X_t| < \infty$ for all $t \geq 0$. For any partition $\pi : 0 = t_0 < t_1 < \cdots < t_n \leq T$, let us define

$$V_T^\pi(X) \stackrel{\triangle}{=} \sum_{0 \leq i < n} E\{|E\{X_{t_{i+1}} - X_{t_i} | \mathcal{F}_{t_i}\}|\} + E\{|X_{t_n}|\}, \tag{1.20}$$

and define the *conditional variation* of X by $V_T(X) \stackrel{\triangle}{=} \sup_\pi V_T^\pi(X)$. If $V_T(X) < \infty$, then X is called a *quasimartingale* [1]. We have the following result.

Lemma 1.2.11 *(Meyer-Zheng [41])* Let $\{P_n\}_{n\geq 1} \subset \mathcal{M}(\mathbb{D})$, such that under each P_n the coordinate process $X_t(\omega) = \omega(t)$, $t \in [0,T]$, $\omega \in \mathbb{D}$, is a quasimartingale. Assume

[1] We should note that the quasimartingale in [41] is defined on $[0,\infty]$. However, it is fairly easy to check that if X is a quasimartingale on $[0,T]$ as is defined above, then the process $\widehat{X}_t = X_t 1_{[0,T)}(t) + X_T 1_{[T,\infty)}(t)$, $t \in [0,\infty]$ is a quasimartingale in the sense of [41]. Furthermore, the conditional variation $V_T(X)$ defined here, although looks slightly different, is exactly the same as $V(\widehat{X})$ defined in [41]. In other words, our quasimartingale is a "local" version of that in [41].

that $V_n(X)$, $n \geq 1$, the conditional variation of X under P_n's, are uniformly bounded in n. Then there exists a subsequence $\{P_{n_k}\}$ which converges weakly on $\mathbb{D}$ to a law $P^* \in \mathcal{M}(\mathbb{D})$, and X is a quasimartingale under P^*.

CHAPTER 2. SOME NEW REPRESENTATION FORMULAE

In a recent paper [21] Fournié-Lasry-Lebuchoux-Lions-Touzi employed Malliavin calculus in numerical finance for computing various "greeks" of the market. In this chapter we extend their idea to the nonlinear FBSDE (1.6) and establish a representation formula for Z without using the derivatives of f or g. In light of Theorem 1.2.4, this formula also represents the derivatives of the viscosity solution to PDE (1.5). We extend the formula further for the second order derivatives. The results are presented in §2 and §3, respectively. In §1 we prove some estimates involving X and two auxiliary processes N and R. In §4, we conclude that actually we have obtained, under certain conditions, the classical solution to PDE (1.5). Finally in §5, we extend the formula in §2 to the case that the terminal of the BSDE is a discrete functional of X.

Throughout the chapter, we shall often make use of the following Assumption:

Assumption 2.0.12 (i) The functions $b, \sigma \in C_b^1$. We use a common constant $K > 0$ to denote all the Lipschitz constants, and assume

$$\sup_{0 \le t \le T} \Big\{ |b(t,0)| + |\sigma(t,0)| \Big\} \le K.$$

(ii) $d_1 = d$. Moreover, we assume that σ satisfies:

$$\sigma(t,x)\sigma^T(t,x) \ge \frac{1}{K} I_d, \qquad \forall (t,x) \in [0,T] \times \mathbb{R}^d.$$

(iii) The functions $f, g \in C_L$. We use the same constant K to denote all the Lipschitz constants, and assume

$$\sup_{0 \le t \le T} |f(t,0,0,0)| + |g(0)| \le K.$$

Throughout the thesis, unless otherwise specified, we shall use the following conventions: (i) In what follows we denote $C > 0$ to be a generic constant depending only on constants T and K, which is allowed to vary from line to line; (ii) Although most of our results hold true in high dimensional case (as will be stated in the theorems), to simplify the presentation we will often prove the results only in the case that $d_1 = d = 1$.

2.1 Some Estimates for the Forward Diffusion

In this section we establish some useful estimates involving only the forward diffusion X and two useful auxiliary processes N and R. Readers may consider them as prerequisites for later sections. First we introduce N and R.

2.1.1 Processes N and R

For $\forall(t,x) \in [0,T) \times \mathbb{R}^d$, our objective process here is the forward diffusion $X^{t,x}$:

$$X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dW_r, \tag{2.1}$$

and its gradient process $\nabla X^{t,x}$ defined as in (1.12). As usual, we omit the supscript ^{t,x} when there is no confusion.

We assume all the coefficients are smooth enough and define the following processes on $(t,T]$:

$$\Lambda_s^{t,x} \triangleq \sigma^{-1}(s, X_s)\nabla X_s; \quad N_r^{t,x} \triangleq \frac{1}{r-t}\Big(\int_t^r \Lambda_s^T dW_s\Big)^T; \tag{2.2}$$

and

$$R_r^{t,x} \triangleq N_r^T N_r - \frac{1}{r-t}\int_t^r D_\tau N_r^T \Lambda_\tau d\tau + \nabla N_r^T, \tag{2.3}$$

where D_τ is the Malliavin derivative operator and ∇N_r is the gradient of N with respect to x.

Note that in (2.3) $D_\tau N_r$ and ∇N_r are somewhat complicated. We aim to express them by other simpler terms. To simplify the presentation, we assume $d = 1$, consequently (1.12) becomes:

$$\nabla X_s^{t,x} = 1 + \int_t^s \partial_x b(r, X_r^{t,x}) \nabla X_r^{t,x} dr + \int_t^s [\partial_x \sigma(r, X_r^{t,x})] \nabla X_r^{t,x} dW_r. \tag{2.4}$$

Recalling (2.2) we know that

$$\nabla N_r = \frac{1}{r-t} \int_t^r \nabla \Lambda_s dW_s; \quad D_\tau N_r = \frac{1}{r-t} \Big[\int_\tau^r D_\tau \Lambda_s dW_s + \Lambda_\tau \Big]. \tag{2.5}$$

Applying the Itô's formula on $\Big(\int_t^r \Lambda_\tau dW_\tau \Big)^2$ we have

$$N_r^2 - \frac{1}{(r-t)^2} \int_t^r \Lambda_\tau^2 d\tau = \frac{2}{(r-t)^2} \int_t^r (\tau - t) N_\tau \Lambda_\tau dW_\tau. \tag{2.6}$$

Using (2.5) and (2.6) we may rewrite (2.3) as:

$$\begin{aligned} R_r &= \frac{1}{(r-t)^2} \Big[2 \int_t^r (\tau - t) N_\tau \Lambda_\tau dW_\tau - \int_t^r \Lambda_\tau \Big(\int_\tau^r D_\tau \Lambda_s dW_s \Big) d\tau \Big] \\ &\quad + \frac{1}{r-t} \int_t^r \nabla \Lambda_s dW_s \\ &= \frac{1}{(r-t)^2} \Big[2 \int_t^r (\tau - t) N_\tau \Lambda_\tau dW_\tau - \int_t^r \Big(\int_t^s \Lambda_\tau D_\tau \Lambda_s d\tau \Big) dW_s \Big] \\ &\quad + \frac{1}{r-t} \int_t^r \nabla \Lambda_s dW_s, \end{aligned} \tag{2.7}$$

thanks to Fubini's theorem.

So it suffices to simplify $\Lambda_\tau D_\tau \Lambda_s$ and $\nabla \Lambda_s$. To this end we apply the chain rule on $\Lambda_s = \sigma^{-1}(s, X_s) \nabla X_s$. Then

$$\begin{aligned} D_\tau \Lambda_s &= (\sigma^{-2} \sigma_x)(s, X_s) (\nabla X_s)^2 \Lambda_\tau^{-1} + \sigma^{-1}(s, X_s) D_\tau \nabla X_s \\ &= -\Lambda_s^2 \sigma_x(s, X_s) \Lambda_\tau^{-1} + \sigma^{-1}(s, X_s) D_\tau \nabla X_s, \end{aligned} \tag{2.8}$$

and

$$\nabla \Lambda_s = -(\sigma^{-2} \sigma_x)(s, X_s)(\nabla X_s)^2 + \sigma^{-1}(s, X_s) \nabla^2 X_s = -\Lambda_s^2 \sigma_x(s, X_s) + \sigma^{-1}(s, X_s) \nabla^2 X_s. \tag{2.9}$$

By (2.4) we know that $D_\tau \nabla X_s$ satisfies the following SDE:

$$\begin{aligned} d(D_\tau \nabla X_s) &= \Big[b_x(s, X_s) D_\tau \nabla X_s + b_{xx}(s, X_s)(\nabla X_s)^2 \Lambda_\tau^{-1} \Big] ds \\ &+ \Big[\sigma_x(s, X_s) D_\tau \nabla X_s + \sigma_{xx}(s, X_s)(\nabla X_s)^2 \Lambda_\tau^{-1} \Big] dW_s. \end{aligned} \tag{2.10}$$

Then by recalling (1.13) and applying the Itô's formula we have

$$d\Big(D_\tau \nabla X_s[\nabla X_s]^{-1}\Big) = \nabla X_s \Lambda_\tau^{-1}\Big[(b_{xx} - \sigma_x\sigma_{xx})(s, X_s)ds + \sigma_{xx}(s, X_s)dW_s\Big]. \quad (2.11)$$

Define

$$\Gamma_r \triangleq \int_t^r \nabla X_s\Big[(b_{xx} - \sigma_x\sigma_{xx})(s, X_s)ds + \sigma_{xx}(s, X_s)dW_s\Big]. \quad (2.12)$$

Note that $D_\tau \nabla X_\tau = \sigma_x(\tau, X_\tau)\nabla X_\tau$, (2.11) leads to

$$D_\tau \nabla X_s[\nabla X_s]^{-1} = \sigma_x(\tau, X_\tau) + \Lambda_\tau^{-1}(\Gamma_s - \Gamma_\tau). \quad (2.13)$$

Therefore we may rewrite (2.8) as

$$D_\tau \Lambda_s = -\Lambda_s^2 \sigma_x(s, X_s)\Lambda_\tau^{-1} + \Lambda_s\sigma_x(\tau, X_\tau) + \Lambda_s\Lambda_\tau^{-1}(\Gamma_s - \Gamma_\tau). \quad (2.14)$$

Moreover, since $\nabla^2 X_t = 0$ and $\nabla^2 X_s$ satisfies the following SDE:

$$\begin{aligned} d(\nabla^2 X_s) &= \Big[b_x(s, X_s)\nabla^2 X_s + b_{xx}(s, X_s)(\nabla X_s)^2\Big]ds \\ &+ \Big[\sigma_x(s, X_s)\nabla^2 X_s + \sigma_{xx}(s, X_s)(\nabla X_s)^2\Big]dW_s; \end{aligned} \quad (2.15)$$

Using analogous arguments one can show that

$$\nabla^2 X_s[\nabla X_s]^{-1} = \Gamma_s, \quad (2.16)$$

and

$$\nabla \Lambda_s = -\Lambda_s^2\sigma_x(s, X_s) + \Lambda_s\Gamma_s. \quad (2.17)$$

Combined with (2.14) and (2.17), now (2.7) can be rewritten as

$$\begin{aligned} R_r &= \frac{1}{(r-t)^2}\Big[2\int_t^r (\tau - t)N_\tau\Lambda_\tau dW_\tau \\ &- \int_t^r \Lambda_s\Big[-\Lambda_s\sigma_x(s, X_s)(s-t) + \int_t^s \sigma_x(\tau, X_\tau)\Lambda_\tau d\tau + \Gamma_s(s-t) - \int_t^s \Gamma_\tau d\tau\Big]dW_s\Big] \\ &+ \frac{1}{r-t}\int_t^r \Lambda_s\Big[-\Lambda_s\sigma_x(s, X_s) + \Gamma_s\Big]dW_s \end{aligned}$$

Thus we have

$$R_r = \frac{1}{(r-t)^2}\int_t^r \Lambda_s\Big[2(s-t)N_s + (r-s)\bar{\Gamma}_s + \int_t^s \bar{\Gamma}_\tau d\tau\Big]dW_s, \quad (2.18)$$

where

$$\bar{\Gamma}_s \triangleq -\Lambda_s\sigma_x(s, X_s) + \Gamma_s, \quad (2.19)$$

and Γ, N and Λ are as defined in (2.12) and (2.2).

2.1.2 L^p-norm Estimates for N and R

In this subsection we give the L^p-norm estimates for processes N and R. To facilitate our proof, we need a seemingly simple technical lemma.

Lemma 2.1.1 Let α and β be two **F**-predictable process with good enough integrability. Then for $\forall p \geq 1$, their exists a constant C_p, depending only on p, such that

$$\| \int_t^r \alpha_s ds \|_p \leq (r-t) \sup_{t \leq s \leq r} \|\alpha_s\|_p; \tag{2.20}$$

and

$$\| \int_t^r \alpha_s dW_s \|_p \leq C_p \sqrt{r-t} \sup_{t \leq s \leq r} \|\alpha_s\|_p. \tag{2.21}$$

Consequently, we have

$$\begin{aligned} \| \int_t^r \alpha_s \beta_s ds \|_p &\leq (r-t) \sup_{t \leq s \leq r} [\|\alpha_s\|_{2p} \|\beta_s\|_{2p}]; \\ \| \int_t^r \alpha_s \beta_s dW_s \|_p &\leq C_p \sqrt{r-t} \sup_{t \leq s \leq r} [\|\alpha_s\|_{2p} \|\beta_s\|_{2p}]. \end{aligned} \tag{2.22}$$

Proof. By Hölder inequality we have

$$| \int_t^r \alpha_s ds | \leq \Big(\int_t^r |\alpha_s|^p ds \Big)^{\frac{1}{p}} (r-t)^{1-\frac{1}{p}}.$$

Thus

$$E\Big\{ \Big| \int_t^r \alpha_s ds \Big|^p \Big\} \leq (r-t)^{p-1} \int_t^r E\{|\alpha_s|^p\} ds \leq (r-t)^p \sup_{t \leq s \leq r} \|\alpha_s\|_p^p,$$

which clearly implies (2.20).

Moreover, applying the Burkholder-Davis-Gundy inequality and then using (2.20) we get

$$E\Big\{ \Big| \int_t^r \alpha_s dW_s \Big|^p \Big\} \leq C_p E\Big\{ \Big(\int_t^r |\alpha_s|^2 ds \Big)^{\frac{p}{2}} \Big\} \leq C_p (r-t)^{\frac{p}{2}} \sup_{t \leq s \leq r} \|\alpha_s\|_p^{\frac{p}{2}},$$

which proves (2.21).

Finally, note that $\|\alpha_s \beta_s\|_p \leq \|\alpha_s\|_{2p} \|\beta_s\|_{2p}$, we infer (2.22) from (2.20) and (2.21). ∎

Next theorem gives the L^p estimates for all the processes involved in (2.18).

Theorem 2.1.2 Assume that (i) and (ii) of Assumption 2.0.12 hold. Let $\Lambda, N, \Gamma, \bar{\Gamma}$ and R be defined as in previous section. Then for $\forall p \geq 2$, there exists a constant C_p, depending only on T, K and p, such that, for $\forall r \in (t, T]$,

$$\begin{array}{ll} \sup_{t\leq s\leq T} \|X_s\|_p \leq C_p(1+|x|); & \sup_{t\leq s\leq T} \|\nabla X_s\|_p \leq C_p; \\ \sup_{t\leq s\leq T} \|\Lambda_s\|_p \leq C_p; & \|N_r\|_p \leq \dfrac{C_p}{\sqrt{r-t}}. \end{array} \tag{2.23}$$

Moreover, if we assume $b, \sigma \in C_b^2$, then it holds that

$$\|\Gamma_r\|_p \leq C_p\sqrt{r-t}; \quad \|\bar{\Gamma}_r\|_p \leq C_p; \quad \|R_r\|_p \leq \frac{C_p}{r-t}. \tag{2.24}$$

Proof. First applying Lemma 1.2.5 on (2.1) and (2.4) we easily get the estimates for X and ∇X, and then that for Λ in virtue of the assumption that $\sigma\sigma^T \geq \frac{1}{K} I_d$. Applying Lemma 2.1.1 we prove the estimate for N.

If $b, \sigma \in C_b^2$, applying Lemma 2.1.1 again we get

$$\begin{aligned} \|\Gamma_r\|_p &\leq C_p\sqrt{r-t}\sup_{t\leq s\leq r}\|\nabla X_s\|_p \leq C_p\sqrt{r-t}; \\ \|\bar{\Gamma}_r\|_p &\leq \|\Lambda_r\|_p + \|\Gamma_r\|_p \leq C_p; \\ \|R_r\|_p &\leq \frac{C_p}{(r-t)^2}\sqrt{r-t}\sup_{t\leq s\leq r}\|\Lambda_s\|_{2p}\Big[(s-t)\|N_s\|_{2p} \\ &\quad +(r-s)\|\bar{\Gamma}_s\|_{2p} + (s-t)\sup_{t\leq\tau\leq s}\|\bar{\Gamma}_\tau\|_{2p}\Big] \\ &\leq \frac{C_p}{(r-t)^{\frac{3}{2}}}\sup_{t\leq s\leq r}\Big[(s-t)\frac{1}{\sqrt{s-t}} + (r-s) + (s-t)\Big] \leq \frac{C_p}{r-t}, \end{aligned}$$

which completes the proof. ∎

2.1.3 Sensitivity of N and R on (t, x)

Now given $0 \leq t_1 < t_2 < T$ and $x_1, x_2 \in \mathbb{R}^d$, denote $\Delta\varphi \triangleq \varphi^1 - \varphi^2$ and $\varphi^i \triangleq \varphi^{t_i, x_i}$, $i = 1, 2$, for any process φ with parameters (t, x). We have the following estimates.

Theorem 2.1.3 Assume (i) and (ii) of Assumption 2.0.12 hold. Then for $\forall p \geq 1$, there exists a constant C_p, depending only on T, K and p, such that

$$\sup_{t_2\leq s\leq T}\|\Delta X_s\|_p \leq C_p\Big[(1+|x_1|)\sqrt{t_2-t_1} + |x_1 - x_2|\Big], \tag{2.25}$$

and

$$\|\Delta N_r\|_p \le \frac{C_p}{\sqrt{r-t_2}}\Big[(1+|x_1|)\sqrt{\frac{t_2-t_1}{r-t_1}}+|x_1-x_2|+\|\Delta b_x\|_{2p}+\Delta\sigma_x\|_{2p}\|\Big], \quad (2.26)$$

where for $\varphi = b_x, \sigma_x$,

$$\|\Delta\varphi\|_{2p} = \Big(E\Big\{\int_{t_2}^T |\varphi(r,X_r^1)-\varphi(r,X_r^2)|^{2p}dr\Big\}\Big)^{\frac{1}{2p}}.$$

Moreover, if $b,\sigma \in C_b^2$, then

$$\|\Delta N_r\|_p \le \frac{C_p}{\sqrt{r-t_2}}\Big[(1+|x_1|)\sqrt{\frac{t_2-t_1}{r-t_1}}+|x_1-x_2|\Big], \quad (2.27)$$

and

$$\|\Delta R_r\|_p \le \frac{C_p}{r-t_2}\Big[(1+|x_1|)\sqrt{\frac{t_2-t_1}{r-t_1}}+|x_1-x_2|+\|\Delta b_{xx}\|_{4p}+\|\Delta\sigma_{xx}\|_{4p}\Big]. \quad (2.28)$$

Proof. First, by SDE (2.1) we know that

$$\Delta X_s = X_{t_2}^1 - x_2 + \int_{t_2}^s \tilde{b}_r \Delta X_r dr + \int_{t_2}^s \tilde{\sigma}_r \Delta X_r dW_r, \quad (2.29)$$

where $\tilde{\varphi} \triangleq \frac{\Delta\varphi}{\Delta X}$ for $\varphi = b,\sigma$. Since b and σ are Lipschitz continuous, we know that $|\tilde{b}|+|\tilde{\sigma}| \le K$. Applying Lemma 1.2.5 twice we infer from (2.29) that

$$\begin{aligned}\|\Delta X_s\|_p &\le C_p\|X_{t_2}^1-x_2\|_p \le C_p\Big[\|X_{t_2}^1-X_{t_1}^1\|_p+|x_1-x_2|\Big]\\ &\le C_p\Big[(1+|x_1|)\sqrt{t_2-t_1}+|x_1-x_2|\Big],\end{aligned}$$

which proves (2.25).

Next, to prove (2.26) we establish some estimates for $\Delta\nabla X$ and $\Delta\Lambda$ first. By (2.4) we have

$$\begin{aligned}\Delta\nabla X_s &= \nabla X_{t_2}^1 - 1 + \int_{t_2}^s \Big[b_x(r,X_r^1)\Delta\nabla X_r + \Delta b_x(r)\nabla X_r^2\Big]dr\\ &+ \int_{t_2}^s \Big[\sigma_x(r,X_r^1)\Delta\nabla X_r + \Delta\sigma_x(r)\nabla X_r^2\Big]dW_r. \quad (2.30)\end{aligned}$$

Applying Lemma 1.2.5 three times we get from (2.30) that

$$\begin{aligned}\|\Delta\nabla X_s\|_p^p &\le C_p E\Big\{|\nabla X_{t_2}^1-1|^p + \int_{t_2}^s \Big[|\Delta b_x(r)\nabla X_r^2|^p + |\Delta\sigma_x(r)\nabla X_r^2|^p\Big]dr\Big\}\\ &\le C_p\Big[(t_2-t_1)^{\frac{p}{2}} + \Big(\|\Delta b_x\|_{2p}^p + \|\Delta\sigma_x\|_{2p}^p\Big)\Big(\int_{t_2}^T E|\nabla X_r^2|^{2p}dr\Big)^{\frac{1}{2}}\Big]\\ &\le C_p\Big[(t_2-t_1)^{\frac{p}{2}} + \|\Delta b_x\|_{2p}^p + \|\Delta\sigma_x\|_{2p}^p\Big],\end{aligned}$$

which implies that

$$\|\Delta\nabla X_s\|_p \le C_p\Big[\sqrt{t_2-t_1}+\|\Delta b_x\|_{2p}+\|\Delta\sigma_x\|_{2p}\Big]. \tag{2.31}$$

Moreover, note that

$$\Delta\Lambda_s=\widetilde{\sigma^{-1}}(s)(\Delta X_s)\nabla X_s^1+\sigma^{-1}(s,X_s^2)\Delta\nabla X_s.$$

Using (2.23, (2.25) and (2.31) we have

$$\begin{aligned}\|\Delta\Lambda_s\|_p &\le C_p\Big[\|\Delta X_s\|_{2p}\|\nabla X_s^1\|_{2p}+\|\Delta\nabla X_s\|_p\Big]\\ &\le C_p\Big[(1+|x_1|)\sqrt{t_2-t_1}+|x_1-x_2|+\|\Delta b_x\|_{2p}+\|\Delta\sigma_x\|_{2p}\Big].\end{aligned} \tag{2.32}$$

Now we are ready to prove (2.26). To this end we first note that

$$\begin{aligned}\Delta N_r &= \frac{1}{r-t_1}\int_{t_1}^r\Lambda_s^1dW_s-\frac{1}{r-t_2}\int_{t_2}^r\Lambda_s^2dW_s\\ &=\frac{1}{r-t_1}\int_{t_1}^{t_2}\Lambda_s^1dW_s+\Big(\frac{1}{r-t_1}-\frac{1}{r-t_2}\Big)\int_{t_2}^r\Lambda_s^1dW_s+\frac{1}{r-t_2}\int_{t_2}^r\Delta\Lambda_sdW_s.\end{aligned}$$

Apply Lemma 2.1.1, we have

$$\begin{aligned}\|\Delta N_r\|_p \;\le\; & C_p\Big[\frac{1}{r-t_1}\sqrt{t_2-t_1}\sup_{t_1\le s\le t_2}\|\Lambda_s^1\|_p+\frac{t_2-t_1}{(r-t_1)(r-t_2)}\sqrt{r-t_2}\sup_{t_2\le s\le r}\|\Lambda_s^1\|\Big]\\ &+\frac{1}{r-t_2}\sqrt{r-t_2}\sup_{t_2\le s\le r}\|\Delta\Lambda_s\|_p\Big]\end{aligned}$$

which, combined with (2.23) and (2.32), implies (2.26).

Now we assume that $b,\sigma\in C_b^2$. Then

$$|\Delta b_x(s)|+|\Delta\sigma_x(s)|\le 2K|\Delta X_s|, \tag{2.33}$$

thus combining (2.33) and (2.25) we can easily get (2.27).

It remains to prove (2.28). We shall prove the estimates for Γ and $\bar\Gamma$ first. Applying Lemma 2.1.1 and recalling (2.33), by some direct calculation we have

$$\begin{aligned}\|\Delta\Gamma_r\|\le C_p\Big[&\sqrt{t_2-t_1}\sup_{t_1\le s\le t_2}\|\nabla X_s^1\|_p+\sqrt{r-t_2}\sup_{t_2\le s\le r}\|\Delta\nabla X_s\|_p\\ &+\sqrt{r-t_2}\sup_{t_2\le s\le r}\|\nabla X_s^1\|_{2p}\Big(\|\Delta b_{xx}\|_{2p}+\|\Delta\sigma_{xx}\|_{2p}+\|\Delta\sigma_x\|_{2p}\Big)\Big]\end{aligned}$$

$$
\begin{aligned}
&\leq C_p\Big[\sqrt{t_2-t_1}+\sqrt{r-t_2}\Big(\sqrt{t_2-t_1}+\|\Delta b_x\|_{2p}+\|\Delta\sigma_x\|_{2p}\Big) \\
&\qquad +\sqrt{r-t_2}\Big(\|\Delta b_{xx}\|_{2p}+\|\Delta\sigma_{xx}\|_{2p}\Big)\Big] \\
&\leq C_p\Big[\sqrt{t_2-t_1}+\sqrt{r-t_2}\Big(\sup_{t_2\leq s\leq T}\|\Delta X_s\|_{2p}+\|\Delta b_{xx}\|_{2p}+\|\Delta\sigma_{xx}\|_{2p}\Big)\Big] \\
&\leq C_p\Big[(1+|x_1|)\sqrt{t_2-t_1}+\sqrt{r-t_2}\Big(|x_1-x_2|+\|\Delta b_{xx}\|_{2p}+\|\Delta\sigma_{xx}\|_{2p}\Big)\Big].
\end{aligned}
\tag{2.34}
$$

Moreover, recalling (2.19) we have

$$
\begin{aligned}
&\|\Delta\bar{\Gamma}_s\|_p\leq\|\Delta\Lambda_s\|_p+\|\Lambda^1_s\|_{2p}\|\Delta\sigma_x(s)\|_{2p}+\|\Delta\Gamma_s\|_p \\
&\leq C_p\Big[(1+|x_1|)\sqrt{t_2-t_1}+|x_1-x_2|+\sqrt{r-t_2}\Big(\|\Delta b_{xx}\|_{2p}+\|\Delta\sigma_{xx}\|_{2p}\Big)\Big].
\end{aligned}
\tag{2.35}
$$

Finally, we analyze ΔR. Recall (2.18), it holds that

$$
\begin{aligned}
&\Delta R_r=R^1_r-R^2_r \\
&=\frac{1}{(r-t_1)^2}\Big[\int_{t_1}^{t_2}\Lambda^1_s\Big[2(s-t_1)N^1_s+(r-s)\bar{\Gamma}^1_s+\int_{t_1}^{s}\bar{\Gamma}^1_\tau d\tau\Big]dW_s \\
&\quad+\int_{t_2}^{r}\Lambda^1_s\Big[2(t_2-t_1)N^1_s+\int_{t_1}^{t_2}\bar{\Gamma}^1_\tau d\tau\Big]dW_s \\
&\quad+\int_{t_2}^{r}\Lambda^1_s\Big[2(s-t_2)N^1_s+(r-s)\bar{\Gamma}^1_s+\int_{t_2}^{s}\bar{\Gamma}^1_\tau d\tau\Big]dW_s\Big] \\
&\quad-\frac{1}{(r-t_2)^2}\int_{t_2}^{r}\Lambda^2_s\Big[2(s-t_2)N^2_s+(r-s)\bar{\Gamma}^2_s+\int_{t_2}^{s}\bar{\Gamma}^2_\tau d\tau\Big]dW_s\Big] \\
&=\frac{1}{(r-t_1)^2}\int_{t_1}^{t_2}\Lambda^1_s\Big[2(s-t_1)N^1_s+(r-s)\bar{\Gamma}^1_s+\int_{t_1}^{s}\bar{\Gamma}^1_\tau d\tau\Big]dW_s \\
&\quad+\frac{1}{(r-t_1)^2}\int_{t_2}^{r}\Lambda^1_s\Big[2(t_2-t_1)N^1_s+\int_{t_1}^{t_2}\bar{\Gamma}^1_\tau d\tau\Big]dW_s \\
&\quad+\Big(\frac{1}{(r-t_1)^2}-\frac{1}{(r-t_2)^2}\Big)\int_{t_2}^{r}\Lambda^1_s\Big[2(s-t_2)N^1_s+(r-s)\bar{\Gamma}^1_s+\int_{t_2}^{s}\bar{\Gamma}^1_\tau d\tau\Big]dW_s \\
&\quad+\frac{1}{(r-t_2)^2}\int_{t_2}^{r}\Delta\Lambda_s\Big[2(s-t_2)N^1_s+(r-s)\bar{\Gamma}^1_s+\int_{t_2}^{s}\bar{\Gamma}^1_\tau d\tau\Big]dW_s\Big] \\
&\quad\frac{1}{(r-t_2)^2}\int_{t_2}^{r}\Lambda^2_s\Big[2(s-t_2)\Delta N_s+(r-s)\Delta\bar{\Gamma}_s+\int_{t_2}^{s}\Delta\bar{\Gamma}_\tau d\tau\Big]dW_s
\end{aligned}
$$

Applying Lemma 2.1.1 several times and using the estimates established at above we have

$$
\begin{aligned}
&\|\Delta R_r\|_p\leq\frac{C_p}{(r-t_1)^2}\sqrt{t_2-t_1}\sup_{t_1\leq s\leq t_2}\|\Lambda^1_s\|_{2p}\Big[(s-t_1)\|N^1_s\|_{2p} \\
&+(r-s)\|\bar{\Gamma}^1_s\|_{2p}+(s-t_1)\sup_{t_1\leq\tau\leq s}\|\bar{\Gamma}^1_\tau\|_{2p}\Big]
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{C_p}{(r-t_1)^2}\sqrt{r-t_2}\sup_{t_2\le s\le r}\|\Lambda^1_s\|_{2p}\Big[(t_2-t_1)\|N^1_s\|_{2p}+(t_2-t_1)\sup_{t_1\le\tau\le t_2}\|\bar\Gamma^1_\tau\|_{2p}\Big]\\
&+\frac{C_p(t_2-t_1)}{(r-t_1)(r-t_2)^2}\sqrt{r-t_2}\sup_{t_2\le s\le r}\|\Lambda^1_s\|_{2p}\Big[(s-t_2)\|N^1_s\|_{2p}\\
&+(r-s)\|\bar\Gamma^1_s\|_{2p}+(s-t_2)\sup_{t_2\le\tau\le s}\|\bar\Gamma^1_\tau\|_{2p}\Big]\\
&+\frac{C_p}{(r-t_2)^2}\sqrt{r-t_2}\sup_{t_2\le s\le r}\|\Delta\Lambda_s\|_{2p}\Big[(s-t_2)\|N^1_s\|_{2p}\\
&+(r-s)\|\bar\Gamma^1_s\|_{2p}+(s-t_2)\sup_{t_2\le\tau\le s}\|\bar\Gamma^1_\tau\|_{2p}\Big]\\
&+\frac{C_p}{(r-t_2)^2}\sqrt{r-t_2}\sup_{t_2\le s\le r}\|\Lambda^2_s\|_{2p}\Big[(s-t_2)\|\Delta N_s\|_{2p}\\
&+(r-s)\|\Delta\bar\Gamma_s\|_{2p}+(s-t_2)\sup_{t_2\le\tau\le s}\|\Delta\bar\Gamma_\tau\|_{2p}\Big]\\
\le\;& C_p\Big\{\frac{\sqrt{t_2-t_1}}{(r-t_1)^2}\Big[\sqrt{t_2-t_1}+r-t_1\Big]+\frac{\sqrt{r-t_2}}{(r-t_1)^2}\Big[\sqrt{t_2-t_1}+(t_2-t_1)\Big]\\
&+\frac{t_2-t_1}{(r-t_1)(r-t_2)^{\frac32}}\Big[\frac{r-t_2}{\sqrt{t_2-t_1}}+r-t_2\Big]\\
&+\frac{1}{(r-t_2)^{\frac32}}\sup_{t_2\le s\le r}\|\Delta\Lambda_s\|_{2p}\Big[\sqrt{r-t_2}+r-t_2\Big]\\
&+\frac{1}{(r-t_2)^{\frac32}}(r-t_2)\sup_{t_2\le s\le r}\Big[\|\Delta N_s\|_{2p}+\|\Delta\bar\Gamma_s\|_{2p}\Big]\Big\}\\
\le\;& \frac{C_p}{r-t_2}\Big[\sqrt{\frac{t_2-t_1}{r-t_1}}+\sup_{t_2\le s\le r}\Big(\|\Delta\Lambda_s\|_{2p}+\|\Delta N_s\|_{2p}+\|\Delta\bar\Gamma_s\|_{2p}\Big)\Big]\\
\le\;& \frac{C_p}{r-t_2}\Big[(1+|x_1|)\sqrt{\frac{t_2-t_1}{r-t_1}}+|x_1-x_2|+\|\Delta b_{xx}\|_{4p}+\|\Delta\sigma_{xx}\|_{4p}\Big],
\end{aligned}
$$

which completes the proof for (2.28), whence the theorem. ■

2.2 First Order Representation Formulae

Since $b,\sigma\in C^1_b$, we may define

$$
\begin{cases}
u(t,x)\triangleq E\Big\{g(X^{t,x}_T)+\int_t^T f(r,\Theta^{t,x}_r)dr\Big\};\\
v(t,x)\triangleq E\Big\{g(X^{t,x}_T)N^{t,x}_T+\int_t^T f(r,\Theta^{t,x}_r)N^{t,x}_r dr\Big\}.
\end{cases}
\tag{2.36}
$$

Recalling (1.6) obviously it holds that

$$u(t,x)=Y^{t,x}_t. \tag{2.37}$$

To see the relation between Z and v, let us first assume f and g are also smooth.

Theorem 2.2.1 Assume Assumption 2.0.12 holds and that $f, g \in C_b^1$. Let $\Theta^{t,x} = (X^{t,x}, Y^{t,x}, Z^{t,x})$ denote the solution to FBSDE (1.6). Then $u \in C^1$, and for $\forall (t,x) \in [0,T] \times \mathbb{R}^d$ and $\forall s \in [t,T)$, it holds that

(1) $|v(t,x)| \le C$, where C depends on T and K;

(2) $Z_s^{t,x} = v(s, X_s^{t,x})\sigma(s, X_s^{t,x})$;

(3) $u_x(t,x) = v(t,x)$.

Proof. As usual, we prove only the case $d = 1$, and we shall drop the superscript ^{t,x} when there is no confusion.

Note that by Lemma 1.2.8, for any $r \in (t,T)$, $X_r, Y_r, Z_r \in \mathbb{D}^{1,2}$ and $D_s Y_r = 0$ whenever $s > r$. Since

$$Y_t = g(X_T) + \int_t^T f(r, \Theta_r)dr - \int_t^T Z_r dW_r,$$

applying D_s to the both sides for $s > t$ we get

$$0 = \partial_x g D_s X_T + \int_s^T [f_x(r) D_s X_r + f_y(r) D_s Y_r + f_z(r) D_s Z_r]\, dr - Z_s - \int_s^T D_s Z_r dW_r,$$

which, combined with Lemma 1.2.8, implies that

$$Z_s = \Big[\partial_x g \nabla X_T + \int_s^T [f_x(r)\nabla X_r + f_y(r)\nabla Y_r + f_z(r)\nabla Z_r]dr - \int_s^T \nabla Z_r dW_r\Big]\Lambda_s^{-1}.$$

Taking conditional expectation $E\{\cdot\ |\mathcal{F}_s\}$ on both sides we then get

$$Z_s = E^{s,X_s}\Big\{\partial_x g \nabla X_T + \int_s^T [f_x(r)\nabla X_r + f_y(r)\nabla Y_r + f_z(r)\nabla Z_r]\, dr\Big\}\Lambda_s^{-1}. \tag{2.38}$$

Denote

$$\bar{v}(t,x) \triangleq E^{t,x}\Big\{\partial_x g \nabla X_T + \int_s^T [f_x(r)\nabla X_r + f_y(r)\nabla Y_r + f_z(r)\nabla Z_r]\, dr. \tag{2.39}$$

By Lemmas 1.2.5 and 1.2.6 we know that $|\bar{v}(t,x)| \le C$. By Markovian property, we may easily check that $\nabla\Theta_r^{s,X_s^{t,x}} = \nabla\Theta_r^{t,x}[\nabla X_s^{t,x}]^{-1}$. Thus (2.38) leads to that $Z_s^{t,x} = \bar{v}(s, X_s^{t,x})\sigma(s, X_s^{t,x})$. Consequently, to prove the theorem it suffices to show that $v = \bar{v} = u_x$.

To this end we note that the arguments above imply that

$$|Z_s| \le C(1+|X_s|). \tag{2.40}$$

By chain rule for the Malliavin derivative operator and Lemma 1.2.8, for any $s<\tau\le r$ one has

$$D_\tau f(r,\Theta_r) = [f_x(r)\nabla X_r + f_y(r)\nabla Y_r + f_z(r)\nabla Z_r]\Lambda_\tau^{-1},$$

which implies that

$$f_x(r)\nabla X_r + f_y(r)\nabla Y_r + f_z(r)\nabla Z_r = \frac{1}{r-s}\int_s^r D_\tau f(r,\Theta_r)\Lambda_\tau d\tau. \tag{2.41}$$

Recalling (2.40) and Lemma 1.2.6, one may easily check that

$$\sup_{t\le s\le T}\|\Theta_s^{t,x}\|_4 \le C, \tag{2.42}$$

which, combined with (2.23), clearly implies that $E\Big\{\int_s^T |f(r,\Theta_r)N_r^s|dr\Big\}<\infty$. Therefore, using the integration by parts formula in Lemma 1.16 and recalling the definition of process N (2.2) we derive from (2.41) that, for $t<s<r<T$,

$$\begin{aligned}
&E^{s,X_s}\left\{f_x(r)\nabla X_r + f_y(r)\nabla Y_r + f_z(r)\nabla Z_r\right\} = E^{s,X_s}\left\{\frac{1}{r-s}\int_s^r D_\tau f(r,\Theta_r)\Lambda_\tau d\tau\right\}\\
&= E^{s,X_s}\left\{\frac{1}{r-s}f(r,\Theta_r)\int_s^r \Lambda_\tau dW_\tau\Big|\mathcal{F}_s\right\} = E^{s,X_s}\left\{f(r,\Theta_r)N_r^s\right\}\nabla X_s.
\end{aligned} \tag{2.43}$$

Similarly we may obtain

$$E^{s,X_s}\Big\{\partial_x g\nabla X_T\Big\} = E^{s,X_s}\Big\{g(X_T)N_T^s\Big\}\nabla X_s. \tag{2.44}$$

Combining (2.38), (2.43) and (2.44) we prove that $\bar v = v$, whence (1) and (2).

It remains to prove (3). Recalling (2.37) and (1.11) we know that u is differentiable with respect to x, and

$$\begin{aligned}
u_x(t,x) &= \nabla Y_t^{t,x} = E^{t,x}\Big\{\partial_x g\nabla X_T\\
&\quad + \int_t^T [f_x(r)\nabla X_r + f_y(r)\nabla Y_r + f_z(r)\nabla Z_r]\,dr\Big\} = \bar v(t,x) = v(t,x).
\end{aligned}$$

That proves (3), whence the theorem. ∎

Now we assume that f and g are Lipschitz continuous only. Let f^ε and g^ε be smooth molifiers of f and g, respectively, such that the first derivatives of f^ε and g^ε are uniformly bounded by K. Let $(Y^\varepsilon, Z^\varepsilon)$ be the solution to the corresponding BSDE, and define $(u^\varepsilon, v^\varepsilon)$ in a similar manner. Then we have

Lemma 2.2.2 Assume Assumption 2.0.12 holds. Then for (u, v) and $(u^\varepsilon, v^\varepsilon)$ defined as above and for $\forall(t, x)$, we have

$$\lim_{\epsilon\to 0} u^\varepsilon(t,x) = u(t,x); \quad \lim_{\epsilon\to 0} v^\varepsilon(t,x) = v(t,x).$$

Proof. Note that $u(t,x) = Y_t^{t,x}$, then Lemma 1.2.7 obviously implies that

$$\lim_{\epsilon\to 0} u^\varepsilon(t,x) = u(t,x). \tag{2.45}$$

To prove the convergence of v^ε, recalling (2.23) and applying the Hölder inequality we get

$$\begin{aligned}
&|v^\varepsilon(t,x) - v(t,x)| = \Big| E\Big\{[g^\varepsilon(X_T) - g(X_T)]N_T + \int_t^T [f^\varepsilon(r,\Theta_r^\varepsilon) - f(r,\Theta_r)]N_r dr\Big\}\Big| \\
&\le \|g^\varepsilon - g\|_\infty \|N_T\|_1 + \|f^\varepsilon - f\|_\infty \int_t^T \|N_r\|_1 dr + C\int_t^T \|\Theta_r^\varepsilon - \Theta_r\|_2 \|N_r\|_2 dr \\
&\le C\Big[\|g^\varepsilon - g\|_\infty + \|f^\varepsilon - f\|_\infty + \int_t^T \frac{\|\Theta_r^\varepsilon - \Theta_r\|_2}{\sqrt{r-t}} dr\Big].
\end{aligned} \tag{2.46}$$

Clearly it holds that

$$\lim_{\epsilon\to 0}\Big[\|g^\varepsilon - g\|_\infty + \|f^\varepsilon - f\|_\infty\Big] = 0. \tag{2.47}$$

By Lemma 1.2.7 we know $\lim_{\epsilon\to 0}\int_t^T \|\Theta_r^\varepsilon - \Theta_r\|_2 dr = 0$, thus, for dt-a.s $r \in [t, T]$,

$$\lim_{\epsilon\to 0} \|\Theta_r^\varepsilon - \Theta_r\|_2 = 0. \tag{2.48}$$

By (2.42) we know that $\sup_{t\le r\le T} \|\Theta_r^\varepsilon\|_2$ is uniformly bounded, which further implies that $\operatorname*{esssup}_{t\le r\le T} \|\Theta_r\|_2 < \infty$. So we may apply the Dominated Convergence Theorem and get

$$\lim_{\epsilon\to 0}\int_t^T \frac{\|\Theta_r^\varepsilon - \Theta_r\|_2}{\sqrt{r-t}} dr = 0,$$

which, combined with (2.46) and (2.47), proves the Lemma. ■

Applying Lemma 1.2.5 and (1) and (3) of Theorem 2.2.4, we get the following Corollary.

Corollary 2.2.3 Assume Assumption 2.0.12 holds. Then for $\forall(r,x) \in [t,T] \times \mathbb{R}^d$ and $\forall p \geq 2$, there exists a constant C_p, depending only on T, K and p such that

$$\|Z_r^{t,x}\|_p \leq C_p(1+|x|).$$

Our main result of this section is the following theorem.

Theorem 2.2.4 Assume Assumption 2.0.12 holds. Then for $\forall(t,x) \in [0,T) \times \mathbb{R}^d$, we have

(1) $|v(t,x)| \leq C$, where C depends on T and K;

(2) v is continuous;

(3) $Z_s^{t,x} = v(s, X_s^{t,x})\sigma(s, X_s^{t,x})$;

(4) $u_x(t,x) = v(t,x)$;

(5) if we assume further that $g \in C_b^1$, then 1)-4) hold true on $[0,T] \times R^d$, and $v(T,x) = \partial_x g(x)$.

Proof. As in Lemma 2.2.2, we have $(f^\varepsilon, g^\varepsilon)$ and define $(u^\varepsilon, v^\varepsilon)$.

(1) The result follows directly from (1) of Theorem 2.2.1 and Lemma 2.2.2.

(2) Let $0 \leq t_1 < t_2 < T$ and $x_1, x_2 \in \mathbb{R}^d$. For $\eta = X, \Theta, N$, denote $\eta^i \triangleq \eta^{t_i,x_i}$ for $i = 1,2$, and $\Delta\eta_r \triangleq \eta_r^1 - \eta_r^2$ for $r \in [t_2, T]$. Then by (2.23), Corollary 2.2.3 and Lemma 1.2.6, we have

$$\begin{aligned}
&|v(t_1,x_1) - v(t_2,x_2)| \\
&= \Big|E\Big\{g(X_T^1)N_T^1 + \int_{t_1}^T f(r,\Theta_r^1)N_r^1 dr - g(X_T^2)N_T^2 - \int_{t_2}^T f(r,\Theta_r^2)N_r^2 dr\Big\}\Big| \\
&\leq E\Big\{|g(X_T^1) - g(X_T^2)||N_T^2| + |g(X_T^1)||\Delta N_T| + \int_{t_1}^{t_2} |f(r,\Theta_r^1)||N_r^1| dr \\
&\quad + \int_{t_2}^T \Big[|f(r,\Theta_r^1) - f(r,\Theta_r^2)||N_r^2| + |f(r,\Theta_r^1)||\Delta N_r|\Big] dr\Big\} \qquad (2.49) \\
&\leq C\Big\{\|\Delta X_T\|_2\|N_T^2\|_2 + \|g(X_T^1)\|_2\|\Delta N_T\|_2 + \int_{t_1}^{t_2} \|f(r,\Theta_r^1)\|_2\|N_r^1\|_2 dr \\
&\quad + \int_{t_2}^T \Big[\|\Delta\Theta_r\|_2\|N_r^2\|_2 + \|f(r,\Theta_r^1)\|_2\|\Delta N_r\|_2\Big] dr\Big\} \\
&\leq C\Big[\frac{\|\Delta X_T\|_2}{\sqrt{T-t_2}} + \int_{t_1}^{t_2} \frac{dr}{\sqrt{r-t_1}} + \int_{t_2}^T \frac{\|\Delta\Theta_r\|_2}{\sqrt{r-t_2}} dr + \|\Delta N_T\|_2 + \int_{t_2}^T \|\Delta N_r\|_2 dr\Big].
\end{aligned}$$

Recalling FBSDE (1.6) we subtract two equations and know that, for $s \in [t_2, T]$, $\Delta\Theta$ satisfies the following BSDE:

$$\Delta Y_s = \tilde{g}\Delta X_T + \int_s^T \Big[\tilde{f}_x(r)\Delta X_r + \tilde{f}_y(r)\Delta Y_r + \tilde{f}_z(r)\Delta Z_r\Big]dr - \int_s^T \Delta Z_r dW_r, \quad (2.50)$$

where $\widetilde{\partial}\varphi$ are the corresponding difference quotients, for $\varphi = b, \sigma, f$ and g. Recalling (2.25) and applying Lemma 1.2.6 we have

$$\sup_{t_2 \le s \le T}[\|\Delta X_s\|_2 + \|\Delta Y_s\|_2] + \int_{t_2}^T \|\Delta Z_s\|^2 ds \le C\Big[(1+|x_1|^2)|t_2 - t_1| + |x_1 - x_2|^2\Big]. \quad (2.51)$$

Again, since $\|\Theta\|_2$, whence $\|\Delta\Theta\|_2$, is uniformly bounded, applying the Dominated Convergence Theorem we see that, as $|(t_1, x_1) - (t_2, x_2)| \to 0$,

$$\frac{\|\Delta X_T\|_2}{\sqrt{T - t_2}} + \int_{t_1}^{t_2} \frac{dr}{\sqrt{r - t_1}} + \int_{t_2}^T \frac{\|\Delta\Theta_r\|_2}{\sqrt{r - t_2}} dr \to 0. \quad (2.52)$$

Moreover, as $|(t_1, x_1) - (t_2, x_2)| \to 0$, since $\|\Delta b_x\|_4 + \|\Delta\sigma_x\|_4 \to 0$, applying the Dominated Convergence Theorem again we get from (2.26) that

$$\|\Delta N_T\|_2 + \int_{t_2}^T \|\Delta N_r\|_2 dr \to 0,$$

which, combined with (2.49) and (2.52), implies that

$$|v(t_1, x_1) - v(t_2, x_2)| \to 0,$$

whence (2).

(3) By Theorem 2.2.1, we have

$$Z_s^\varepsilon = v^\varepsilon(s, X_s)\sigma(s, X_s).$$

Then (2.48) and Lemma 2.2.2 imply that

$$Z_s = v(s, X_s)\sigma(s, X_s); \quad ds \times dP \quad a.e.$$

That is, $v(s, X_s)\sigma(s, X_s)$ is a version of Z. So we may choose Z as this version and then the result follows.

(4) Apply Theorem 2.2.1 again, we have

$$u^\varepsilon(t,x) = u^\varepsilon(t,0) + \int_0^x v^\varepsilon(t,y)dy.$$

Using Lemma 2.2.2 and applying the Dominated Convergence Theorem, thanks to (1), we obtain that

$$u(t,x) = u(t,0) + \int_0^x v(t,y)dy,$$

which proves (4) due to (2).

5) If $g \in C_b^1$, then

$$v(t,x) = E\left\{\partial_x g(X_T^{t,x})\nabla X_T^{t,x} + \int_t^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})N_r^{t,x} dr\right\}. \tag{2.53}$$

Since $X_T^{T,x} = x$ and $\nabla X_T^{T,x} = 1$, we have

$$\lim_{t\uparrow T} v(t,x) = \partial_x g(x).$$

What left is trivial now. ■

2.3 Second Order Representation Formulae

In this section we shall establish a probabilistic formula for u_{xx}. Recalling (2.18) we define

$$w(t,x) \triangleq E\Big\{g(X_T^{t,x})R_T^{t,x} + \int_t^T [f(r,\Theta_r^{t,x}) - f(r,\theta_r^{t,x})]R_r^{t,x} dr\Big\}; \tag{2.54}$$

where

$$\theta_r \triangleq (x, u(r,x), v(r,x)\sigma(r,x)); \tag{2.55}$$

We aim to show that $u_{xx} = w(t,x)$. Recalling (2.24) we note that (2.54) involves some singularity. To estimate the integrand we need a technical lemma.

Lemma 2.3.1 Assume that Assumption 2.0.12 holds and that $b, \sigma \in C_b^2$. Then for $\forall \alpha \in (0, \frac{1}{2})$, $\forall t \in [0,T)$, v is α-Hölder continuous with respect to x. To be precise, there exists a constant C_α, depending only on T, K and α, such that

$$|v(t,x+\delta) - v(t,x)| \le \frac{C_\alpha(1+|x|)}{\sqrt{T-t}}|\delta|^\alpha.$$

Proof. For $\forall p > 2$, let $q < 2$ be its conjugate. Denote $\Delta^\delta \varphi_r \triangleq \varphi_r^{x+\delta} - \varphi_r^x$ for $\varphi = X, N, \Theta$. Recall (2.36) and apply Hölder inequality, we get

$$\begin{aligned}
&|v(t,x+\delta) - v(t,x)| \\
&\le E\Big\{|\Delta^\delta X_T||N_T^x| + |g(X_T^{x+\delta})||\Delta^\delta N_T| \\
&\quad + \int_t^T \Big[|\Delta^\delta \Theta_r||N_r| + |f(r,\Theta_r^{x+\delta})||\Delta^\delta N_r|\Big]dr\Big\} \\
&\le E\Big\{\|\Delta^\delta X_T\|_p \|N_T^x\|_q + \|g(X_T^{x+\delta})\|_p \|\Delta^\delta N_T\|_q \\
&\quad + \Big(E\Big\{\int_t^T |\Delta^\delta \Theta_r|^p dr\Big\}\Big)^{\frac{1}{p}} \Big(E\Big\{\int_t^T |N_r|^q dr\Big\}\Big)^{\frac{1}{q}} \\
&\quad + \Big(E\Big\{\int_t^T |f(r,\Theta_r^{x+\delta})|^p dr\Big\}\Big)^{\frac{1}{p}} \Big(E\Big\{\int_t^T |\Delta^\delta N_r|^q dr\Big\}\Big)^{\frac{1}{q}}\Big\}.
\end{aligned} \tag{2.56}$$

Apply Corollary 2.2.3, Lemmas (1.2.5) and (1.2.6), we know that $\|\Theta_r^{x+\delta}\|_p \le C_p(1 + |x|)$. Recalling (2.23) we infer from (2.56) that

$$\begin{aligned}
|v(t,x+\delta) - v(t,x)| &\le C_p\Big[\frac{1}{\sqrt{T-t}}\|\Delta^\delta X_T\|_p + (1+|x|)\|\Delta^\delta N_T\|_q \\
&+ \Big(E\Big\{\int_t^T |\Delta^\delta \Theta_r|^p dr\Big\}\Big)^{\frac{1}{p}} + (1+|x|)\Big(E\Big\{\int_t^T |\Delta^\delta N_r|^q dr\Big\}\Big)^{\frac{1}{q}}\Big].
\end{aligned} \tag{2.57}$$

Note that

$$\Delta^\delta \varphi = \delta \int_0^1 \nabla \varphi^{t,x+\lambda\delta} d\lambda. \tag{2.58}$$

Applying Lemmas (1.2.5) and (1.2.6) we can easily get

$$\sup_{t \le r \le T} \Big[\|\nabla X_r\|_p + \|\nabla Y_r\|_p\Big] + \Big(E\Big\{\int_t^T |\nabla Z_r|^2 dr\Big\}\Big)^{\frac{1}{2}} \le C_p;$$

Thus by applying Lemma 2.1.1 we get

$$\sup_{t \le r \le T} \Big[\|\Delta^\delta X_r\|_p + \|\Delta^\delta Y_r\|_p\Big] + \Big(E\Big\{\int_t^T |\Delta^\delta Z_r|^2 dr\Big\}\Big)^{\frac{1}{2}} \le C_p \delta, \tag{2.59}$$

which, combined with Corollary 2.2.3, implies that

$$\begin{aligned}
&E\Big\{\int_t^T |\Delta^\delta Z_r|^p dr\Big\} = E\Big\{\int_t^T |\Delta^\delta Z_r||\Delta^\delta Z_r|^{p-1} dr\Big\} \\
&\le \Big(E\Big\{\int_t^T |\Delta^\delta Z_r|^2 dr\Big\}\Big)^{\frac{1}{2}} \Big(E\Big\{\int_t^T |\Delta^\delta Z_r|^{2(p-1)} dr\Big\}\Big)^{\frac{1}{2}} \le C_p(1+|x|^{p-1})\delta.
\end{aligned} \tag{2.60}$$

Moreover, recalling (2.5) and (2.17) and applying Lemma 2.1.1 we have

$$\|\nabla N_r\|_q \le \frac{C_p}{\sqrt{r-t}} \sup_{t\le s\le r} \Big[\|\Lambda_s^2\|_q + \|\Lambda_s\|_{2q}\|\Gamma_s\|_{2q}\Big] \le \frac{C_p}{\sqrt{r-t}}, \tag{2.61}$$

which, Combined with (2.58), implies that

$$\|\Delta^\delta N_r\|_q \le \frac{C_p\delta}{\sqrt{r-t}}. \tag{2.62}$$

Now combine (2.59), (2.60) and (2.62), and note that $q < 2$, we deduce from (2.57) that

$$\begin{aligned}
&|v(t,x+\delta) - v(t,x)| \\
&\le C_p\Big[\frac{\delta}{\sqrt{T-t}} + \frac{(1+|x|)\delta}{\sqrt{T-t}} + (1+|x|^{\frac{p-1}{p}})\delta^{\frac{1}{p}} + (1+|x|)\delta\Big(\int_t^T \frac{dr}{(r-t)^{\frac{q}{2}}}\Big)^{\frac{1}{q}}\Big] \\
&\le \frac{C_p(1+|x|)}{\sqrt{T-t}}\delta^{\frac{1}{p}}.
\end{aligned}$$

That is, v is $\frac{1}{p}-$Hölder continuous with respect to x for $\forall p > 2$. ■

Next result gives the estimate for w.

Lemma 2.3.2 Assume that Assumption 2.0.12 holds and that $b, \sigma \in C_b^2$. Then for $\forall \alpha > 0$, there exists a constant C_α, depending only on T, K and α, such that

$$|w(t,x)| \le \frac{C_\alpha(1+|x|^{2+\alpha})}{\sqrt{T-t}}.$$

Proof. Without lose of generality, we assume $\alpha < \frac{1}{2}$.

First, recall that $E\{R_T\} = 0$, we have

$$\Big|E\{g(X_T^{t,x})R_T^{t,x}\}\Big| = \Big|E\Big\{[g(X_T^{t,x}) - g(x)]R_T^{t,x}\Big\}\Big| \le C\|X_T^{t,x} - x\|_2\|R_T^{t,x}\|_2. \tag{2.63}$$

Next, applying Theorem 2.2.4 and Lemma 2.3.1 we get

$$\begin{aligned}
&|f(r,\Theta_r) - f(r,\theta_r)| \le C\Big[|X_r - x| + |Y_r - u(r,x)| + |Z_r - v(r,x)\sigma(r,x)|\Big] \\
&= C\Big[|X_r - x| + |u(r,X_r) - u(r,x)| + |v(r,X_r)\sigma(r,X_r) - v(r,x)\sigma(r,x)|\Big] \\
&\le C_\alpha\Big[|X_r - x| + |X_r - x| + \frac{1+|x|}{\sqrt{T-r}}|\sigma(r,x)||X_r - x|^\alpha + |v(r,X_r)||X_r - x|\Big] \\
&\le C_\alpha\Big[|X_r - x| + \frac{1+|x|^2}{\sqrt{T-r}}|X_r - x|^\alpha\Big].
\end{aligned} \tag{2.64}$$

Thus

$$\Big|E\Big\{[f(r,\Theta_r)-f(r,\theta_r)]R_r\Big\}\Big| \le C_\alpha\Big[\|X_r-x\|_2+\frac{1+|x|^2}{\sqrt{T-r}}\|X_r-x\|_2^{\frac{\alpha}{2}}\Big]\|R_r\|_2. \quad (2.65)$$

Finally, apply Lemma 1.2.5 and recall (2.24), (2.63) and (2.65) lead to that

$$|w(t,x)| \le C_a\Big[(1+|x|)\frac{\sqrt{T-t}}{T-t}+\int_t^T\Big[(1+|x|)\sqrt{r-t}+\frac{1+|x|^{2+\alpha}}{\sqrt{T-r}}(r-t)^{\frac{\alpha}{2}}\Big]\frac{dr}{r-t}$$
$$\le C_\alpha(1+|x|^{2+\alpha})\Big[\frac{1}{\sqrt{T-t}}+\int_t^T\frac{dr}{\sqrt{T-r}(r-t)^{1-\frac{\alpha}{2}}}\Big]\le\frac{C_\alpha(1+|x|^{2+\alpha})}{\sqrt{T-t}},$$

which proves the lemma. ∎

Now we establish the formula for FBSDEs with smooth coefficients .

Theorem 2.3.3 Assume that Assumption 2.0.12 holds, and that $b,\sigma\in C_b^2$ and $f,g\in C_b^1$. If we assume further that w is continuous with respect to x, then we have

$$w(t,x)=v_x(t,x)=u_{xx}(t,x).$$

Proof. Recall that

$$v(t,x)=E\{g(X_T)N_T\}+\int_t^T E\{f(r,\Theta_r)N_r\}dr.$$

Since all the coefficients are smooth, we may differentiate the right side *formally* with respect to x and denote

$$\begin{aligned}\bar w(t,x) &= E\Big\{g_x(X_T)\nabla X_T N_T+g(X_T)\nabla N_T\Big\} \qquad (2.66)\\ &\quad+\int_t^T E\Big\{[(f_x\nabla X_r+f_y\nabla Y_r+f_z\nabla Z_r)N_r+f(r,\Theta_r)\nabla N_r]\Big\}dr.\end{aligned}$$

We should mention here that the integrand in the right side of (2.66) involves some singularity. We shall prove that it is indeed integrable and that $w=\bar w=v_x$.

To this end we use the Malliavin calculus again. For $t\le\tau\le r$, the chain rule implies that

$$D_\tau(f(r,\Theta_r)N_r)=[f_xD_\tau X_r+f_yD_\tau Y_r+f_zD_\tau Z_r]N_r+f(r,\Theta_r)D_\tau N_r,$$

which, combined with (1.14), implies that

$$[f_x\nabla X_r+f_y\nabla Y_r+f_z\nabla Z_r]N_r=D_\tau(f(r,\Theta_r)N_r)\Lambda_\tau-f(r,\Theta_r)D_\tau N_r\Lambda_\tau.$$

Integrating both sides over $[t,r]$ and dividing by $r-t$ we get

$$[f_x\nabla X_r + f_y\nabla Y_r + f_z\nabla Z_r]N_r = \frac{1}{r-t}\int_t^r [D_\tau(f(r,\Theta_r)N_r)\Lambda_\tau - f(r,\Theta_r)D_\tau N_r\Lambda_\tau]d\tau,$$

Now using the integration by parts formula in Lemma 1.16 we obtain that

$$\begin{aligned}
&E\{[f_x\nabla X_r + f_y\nabla Y_r + f_z\nabla Z_r]N_r\}\\
&= \frac{1}{r-t}E\left\{f(r,\Theta_r)N_r\int_t^r \Lambda_\tau dW_\tau - f(r,\Theta_r)\int_t^r D_\tau N_r\Lambda_\tau d\tau\right\},
\end{aligned}$$

which, combined with (2.3), implies that

$$\begin{aligned}
&E\{[f_x\nabla X_r + f_y\nabla Y_r + f_z\nabla Z_r]N_r + f(r,\Theta_r)\nabla N_r\}\\
&= E\Big\{f(r,\Theta_r)\Big[N_r^2 - \frac{1}{r-t}\int_t^r D_\tau N_r\Lambda_\tau d\tau + \nabla N_r\Big]\Big\}\\
&= E\{f(r,\Theta_r)R_r\}.
\end{aligned} \tag{2.67}$$

Similarly, we have

$$E\{\partial_x g(X_T)\nabla X_T N_T + g(X_T)\nabla N_T\} = E\{g(X_T)R_T\}. \tag{2.68}$$

Note that $E\{R_r\}=0$, combining (2.66), (2.67) and (2.68) we obtain

$$\bar{w}(t,x) = E\{g(X_T)R_T\} + \int_t^T E\{[f(r,\Theta_r) - f(r,\theta_r)]R_r\}dr. \tag{2.69}$$

Recalling (2.65) we know the integrand in the right side of (2.69) is absolutely integrable. Consequently, by Fubini's theorem we know that $\bar{w}=w$.

Moreover, by the definition of $\bar{w}$ we have

$$\begin{aligned}
&v(t,x) - v(t,0)\\
&= E\Big\{g(X_T^x)N_T^x - g(X_T^0)N_T^0\Big\} + \int_t^T E\Big\{f(r,\Theta_r^x)N_r^x - f(r,\Theta_r^0)N_r^0\Big\}dr\\
&= E\Big\{\int_0^x [g_x(X_T^y)N_T^y + g(X_T^y)\nabla N_T^y]dy\Big\} + \int_t^T E\Big\{\int_0^x \Big[f(r,\Theta_r^y)\nabla N_r^y\\
&+\Big(f_x(r,\Theta_r^y)\nabla X_r^y + f_y(r,\Theta_r^y)\nabla Y_r^y + f_z(r,\Theta_r^y)\nabla Z_r^y\Big)N_r^y\Big]dy\Big\}dr.
\end{aligned} \tag{2.70}$$

Recalling (2.63), (2.65) and applying the Fubini's theorem we get

$$v(t,x) - v(t,0) = \int_0^x \bar{w}(t,y)dy = \int_0^x w(t,y)dy.$$

Since we assume w is continuous with respect to x, we conclude that $v_x = \bar{w} = w$, which completes the proof. ∎

In order to investigate the Lipschitz case, we need the following convergence result of w.

Lemma 2.3.4 Assume that Assumption 2.0.12 holds and that $b, \sigma \in C_b^2$. Let $(f^\varepsilon, g^\varepsilon)$ be smooth molifiers of (f, g). Define all the involving terms for ε in a similar manner. Then we have

$$\lim_{\epsilon \to 0} w^\varepsilon(t,x) = w(t,x).$$

Proof. First by recalling (2.24) we know that

$$\begin{aligned}
&|w^\varepsilon(t,x) - w(t,x)| \le E\Big\{|g^\varepsilon(X_T) - g(X_T)||R_T| \\
&\quad + \int_t^T |f^\varepsilon(r,\Theta_r^\varepsilon) - f^\varepsilon(r,\theta_r^\varepsilon) - f(r,\Theta_r) + f(r,\theta_r)||R_r|dr\Big\} \qquad (2.71)\\
&\le C\Big[\frac{\|g^\varepsilon - g\|_\infty}{T-t} + \int_t^T \frac{1}{r-t}\|f^\varepsilon(r,\Theta_r^\varepsilon) - f^\varepsilon(r,\theta_r^\varepsilon) - f(r,\Theta_r) + f(r,\theta_r)\|_2 dr\Big]
\end{aligned}$$

Take $\alpha = \frac{1}{4}$, by (2.64) and Lemma 1.2.5 we get

$$\|f^\varepsilon(r,\Theta_r^\varepsilon) - f^\varepsilon(r,\theta_r^\varepsilon) - f(r,\Theta_r) + f(r,\theta_r)\|_2 \le C(1+|x|^{\frac{9}{4}})\frac{(r-t)^{\frac{1}{8}}}{\sqrt{T-r}}. \qquad (2.72)$$

Note that

$$\int_t^T \frac{1}{r-t}\frac{(r-t)^{\frac{1}{8}}}{\sqrt{T-r}}dr < \infty.$$

Apply Lemma 1.2.7 and the Dominated Convergence Theorem, we prove the result from (2.71). ∎

Theorem 2.3.5 Assume that Assumption 2.0.12 holds, and that $b, \sigma \in C_b^2$. Then for $\forall \alpha > 0$, there exists a constant C_α, depending only on T, K and α, such that the following results hold true in $[0,T) \times \mathbb{R}^d$:

(1) $|w(t,x)| \le C_\alpha \frac{1+|x|^{2+\alpha}}{\sqrt{T-t}}$;

(2) w is continuous;

(3) $u_{xx}(t,x) = v_x(t,x) = w(t,x)$;

(4) If furthermore we assume that $g \in C_b^2$, then (2) and (3) hold true on $[0,T] \times R$; $|w(t,x)| \le C_\alpha(1 + |x|^{4+\alpha})$ and $w(T,x) = g_{xx}(x)$.

Proof. (1) follows directly from Lemma 2.3.2.

(2) For $0 \le t_1 < t_2 < T$, denote

$$\widetilde{\theta}^1_r \triangleq \Big(X^1_{t_2}, u(r, X^1_{t_2}), v(r, X^1_{t_2})\sigma(r, X^1_{t_2})\Big).$$

In light of (2.64) we can analogously prove that, for $t_2 \le r < T$,

$$\|\widetilde{\theta}^1_r - \theta^1_r\|_2 \le C(1+|x_1|^{2+\alpha})\frac{(t_2-t_1)^{\frac{\alpha}{2}}}{\sqrt{T-r}}; \quad \|\Theta^1_r - \widetilde{\theta}^1_r\|_2 \le C(1+|x_1|^{2+\alpha})\frac{(r-t_2)^{\frac{\alpha}{2}}}{\sqrt{T-r}}. \tag{2.73}$$

Then, by denoting $\varphi^i \triangleq \varphi^{t_i,x_i}$, $i = 1,2$ and $\Delta\varphi \triangleq \varphi^1 - \varphi^2$ for $\varphi = \Theta, R$, we have

$$\begin{aligned}
|w(t_1,x_1) - w(t_2,x_2)| &\le E\Big\{|g(X^1_T)R^1_T - g(X^2_T)R^2_T| \\
&\quad + \int_{t_1}^{t_2} |f(r,\Theta^1_r) - f(r,\theta^1_r)||R^1_r|dr + \int_{t_2}^{T} |f(r,\widetilde{\theta}^1_r) - f(r,\theta^1_r)||R^1_r|dr \\
&\quad + \int_{t_2}^{T} \Big|[f(r,\Theta^1_r) - f(r,\widetilde{\theta}^1_r]R^1_r - [f(r,\Theta^2_r) - f(r,\theta^2_r)]R^2_r\Big|dr\Big\} \\
&\le CE\Big\{|\Delta X_T||R^2_T| + |g(X^1_T)||\Delta R_T| + \int_{t_1}^{t_2} |\Theta^1_r - \theta^1_r||R^1_r|dr \\
&\quad + \int_{t_2}^{T} |\widetilde{\theta}^1_r - \theta^1_r||R^1_r|dr + \int_{t_2}^{T} |f(r,\Theta^1_r) - f(r,\widetilde{\theta}^1_r)||\Delta R_r|dr \\
&\quad + \int_{t_2}^{T} |f(r,\Theta^1_r) - f(r,\widetilde{\theta}^1_r) - f(r,\Theta^2_r) + f(\theta^2_r)||R^2_r|dr\Big\} \\
&\le C\Big[\|\Delta X_T\|_2\|R^2_T\|_2 + \|g(X^1_T)\|_2\|\Delta R_T\|_2 + \int_{t_1}^{t_2} \|\Theta^1_r - \theta^1_r\|_2\|R^1_r\|_2 \\
&\quad + \int_{t_2}^{T} \|\widetilde{\theta}^1_r - \theta^1_r\|_2\|R^1_r\|_2 dr + \int_{t_2}^{T} \|\Theta^1_r - \widetilde{\theta}^1_r\|_2\|\Delta R_r\|_2 dr \\
&\quad + \int_{t_2}^{T} \|f(r,\Theta^1_r) - f(r,\widetilde{\theta}^1_r) - f(r,\Theta^2_r) + f(r,\theta^2_r)\|_2\|R^2_r\|_2 dr\Big]
\end{aligned} \tag{2.74}$$

Now choose $\alpha = \frac{1}{4}$ and let $C(x)$ be a constant depending on T, K and x, which is allowed to vary line by line. Recalling (2.24), (2.23), (2.64) and (2.73), (2.74) leads to that

$$\begin{aligned}
|w(t_1,x_1) - w(t_2,x_2)| &\le C(x_1)\Big[\frac{\|\Delta X_T\|_2}{T-t_2} + \|\Delta R_T\|_2 + \int_{t_1}^{t_2} \frac{(r-t_1)^{\frac{1}{8}}}{\sqrt{T-r}}\frac{dr}{r-t_1} \\
&\quad + \int_{t_2}^{T} \frac{(t_2-t_1)^{\frac{1}{8}}}{\sqrt{T-r}}\frac{dr}{r-t_1} + \int_{t_2}^{T} \frac{(r-t_2)^{\frac{1}{8}}}{\sqrt{T-r}}\|\Delta R_r\|_2 dr \\
&\quad + \int_{t_2}^{T} \frac{1}{r-t_2}\|f(r,\Theta^1_r) - f(r,\widetilde{\theta}^1_r) - f(r,\Theta^2_r) + f(r,\theta^2_r)\|_2 dr\Big]
\end{aligned}$$

$$\leq \frac{C(x_1)}{T-t_2}\Big[\|\Delta X_T\|_2 + \|\Delta R_T\|_2 + (t_2-t_1)^{\frac{1}{8}} \tag{2.75}$$
$$+(t_2-t_1)^{\frac{1}{8}}|\log(t_2-t_1)| + \int_{t_2}^{T} \frac{(r-t_2)^{\frac{1}{8}}}{\sqrt{T-r}}\|\Delta R_r\|_2 dr$$
$$+\int_{t_2}^{T} \|f(r,\Theta_r^1) - f(r,\widetilde{\theta}_r^1) - f(r,\Theta_r^2) + f(r,\theta_r^2)\|_2 \frac{dr}{r-t_2}\Big].$$

Let $|(t_1,x_1)-(t_2,x_2)| \to 0$, recalling (2.25) we have

$$\Delta X_T \to 0. \tag{2.76}$$

Then by Lemma 1.2.7 we can easily prove that

$$\|f(r,\Theta_r^1) - f(r,\widetilde{\theta}_r^1) - f(r,\Theta_r^2) + f(r,\theta_r^2)\|_2 \leq C\Big[\|\Delta\Theta_r\|_2 + \|\widetilde{\theta}_r^1 - \theta_r^2\|_2\Big] \to 0. \tag{2.77}$$

Moreover, recalling (2.28) we have

$$\|\Delta R_T\|_2 + \int_{t_2}^{T} \frac{(r-t_2)^{\frac{1}{8}}}{\sqrt{T-r}}\|\Delta R_r\|_2 dr \to 0. \tag{2.78}$$

Finally, note that

$$\|f(r,\Theta_r^1) - f(r,\widetilde{\theta}_r^1) - f(r,\Theta_r^2) + f(r,\theta_r^2)\|_2 \leq C[\|\Theta_r^1 - \widetilde{\theta}_r^1\|_2 + \|\Theta_r^2 - \theta_r^2\|_2] \leq C\frac{(r-t_2)^{\frac{1}{8}}}{\sqrt{T-r}},$$

then, combined with (2.77), the Dominated Convergence Theorem implies that

$$\int_{t_2}^{T} \frac{1}{r-t_2}\|f(r,\Theta_r^1) - f(r,\widetilde{\theta}_r^1) - f(r,\Theta_r^2) + f(r,\theta_r^2)\|_2 dr\Big) \to 0. \tag{2.79}$$

Combining (2.76), (2.78) and (2.79), we conclude from (2.75) that w is continuous.

(3) For any fixed $t < T$, by (2) and Theorem 2.3.3 we know that $v_x^\varepsilon(t,x) = w^\varepsilon(t,x)$. Applying Lemmas 2.2.2 and 2.3.4, we get

$$v(t,x) = \lim_{\epsilon\to 0} v^\varepsilon(t,x) = \lim_{\epsilon\to 0}[v^\varepsilon(t,0) + \int_0^x w^\varepsilon(t,y)dy] = v(t,0) + \int_0^x w(t,y)dy;$$

thanks to (1) and the Dominated Convergence Theorem. Since w is continuous, we get $v_x(t,x) = w(t,x)$.

(4) If $g \in C_b^2$, then we can easily prove that, for $t < T$,

$$w(t,x) = E\Big\{g_{xx}(X_T^{t,x})(\nabla X_T^{t,x})^2 + g_x(X_T^{t,x})\nabla^2 X_T^{t,x} + \int_t^T \Big[f(r,\Theta_r^{t,x}) - f(r,\theta_r^{t,x})\Big]R_r^{t,x} dr\Big\}. \tag{2.80}$$

(2.16) obviously implies that

$$\lim_{t\uparrow T} E\Big\{g_{xx}(X_T^{t,x})(\nabla X_T^{t,x})^2 + g_x(X_T^{t,x})\nabla^2 X_T^{t,x}\Big\} = g_{xx}(x).$$

Therefore, to prove $\lim_{t\uparrow T} w(t,x) = g_{xx}(x)$, it suffices to show that

$$\lim_{t\uparrow T} E\Big\{\int_t^T \Big[f(r,\Theta_r^{t,x}) - f(r,\theta_r^{t,x})\Big] R_r^{t,x} dr\Big\} = 0. \tag{2.81}$$

To this end we note that, by (1) and (3) of this theorem, we may improve the result of Lemma 2.3.1 now. Actually we have

$$|v(t,x+\delta) - v(t,x)| \le \frac{C_\alpha(1+|x|^{2+\alpha})}{\sqrt{T-t}}\delta.$$

Then by similar arguments as those in (2.64) we get

$$\|f(r,\Theta_r^{t,x}) - f(r,\theta_r^{t,x})\|_2 \le C_\alpha(1+|x|^{4+\alpha})\sqrt{\frac{r-t}{T-r}},$$

where C_α is independent of t. Thus we have

$$\begin{aligned}
&\int_t^T \|f(r,\Theta_r^{t,x}) - f(r,\theta_r^{t,x})\|_2 \|R_r^{t,x}\|_2 dr \le C_\alpha(1+|x|^{4+\alpha})\int_t^T \sqrt{\frac{r-t}{T-r}}\frac{dr}{r-t} \\
&= C_\alpha(1+|x|^{4+\alpha})\int_0^1 \frac{dr}{\sqrt{r(1-r)}} = C_\alpha(1+|x|^{4+\alpha}) < \infty.
\end{aligned}$$

Then using (2.80) we can easily get

$$|w(t,x)| \le C_\alpha(1+|x|^{4+\alpha}). \tag{2.82}$$

By (2.82) we may repeat the above procedure again and get the following estimates:

$$\begin{aligned}
&|v(t,x+\delta) - v(t,x)| \le C_\alpha(1+|x|^{4+\alpha})\delta; \\
&\|f(r,\Theta_r^{t,x}) - f(r,\theta_r^{t,x})\|_2 \le C_\alpha(1+|x|^{5+\alpha})\sqrt{r-t}; \\
&\int_t^T \|f(r,\Theta_r^{t,x}) - f(r,\theta_r^{t,x})\|_2 \|R_r^{t,x}\|_2 dr \\
&\le C_\alpha(1+|x|^{5+\alpha})\int_t^T \frac{\sqrt{r-t}}{r-t} dr = C_\alpha(1+|x|^{5+\alpha})\sqrt{T-t}.
\end{aligned}$$

Now (2.81) follows directly from the last inequality at above. ∎

2.4 BSDEs and Quasi-linear Parabolic PDEs

As a corollary to the results of the previous sections, we establish the following theorem relating SDEs with PDEs:

Theorem 2.4.1 Assume that Assumption 2.0.12 holds and that $b, \sigma \in C_b^1$. Then the viscosity solution u to PDE (1.5) is in C_b^1 and $|u_x| \leq C$ for some constant C depending only on T and K. If we assume further that $b, \sigma \in C_b^2$, then u is a classical solution to (1.5) such that $|u_{xx}(t,x)| \leq \frac{C_\alpha}{\sqrt{T-t}}(1+|x|^{2+\alpha})$ for $\forall \alpha > 0$, where C_α depends only on T, K and α. Furthermore, if $g \in C_b^2$, then $|u_{xx}(t,x)| \leq C_a(1+|x|^{4+\alpha})$.

Proof. Let $u(t,x) = Y_t^{t,x}$. By Theorem 1.2.4 we know that u is the viscosity solution to (1.5). If $b, \sigma \in C_b^1$, applying Theorem 2.2.4 we know $u \in C_b^1$ and $|u_x| \leq C$.

If $b, \sigma \in C_b^2$, by Theorem 2.3.5, it suffices to prove that u_t exists and satisfies PDE (1.5). Denote

$$\zeta(t,x) \triangleq \frac{1}{2}w(t,x)\sigma^2(t,x) + v(t,x)b(t,x) + f(t,x,u(t,x),v(t,x)\sigma(t,x)).$$

Then ζ is continuous, and for $\alpha > 0$, we have

$$|\zeta(t,x)| \leq C_\alpha\Big[\frac{1+|x|^{2+\alpha}}{T-t}(1+|x|^2) + (1+|x|) + (1+|x|)\Big] \leq C\frac{1+|x|^{4+\alpha}}{T-t}.$$

Let f^ε and g^ε be smooth molifiers of f and g, respectively. As $\varepsilon \to 0$, we have $\phi^\varepsilon(t,x) \to \phi(t,x)$ for $\phi = u, v, w$, whence for $\phi = \zeta$. By Pardoux-Peng [44], u^ε is the classical solution of the corresponding PDE, that is

$$u_t^\varepsilon + \zeta^\varepsilon = 0.$$

Therefore,

$$u^\varepsilon(t,x) = u^\varepsilon(0,x) - \int_0^t \zeta^\varepsilon(s,x)ds.$$

Let $\varepsilon \to 0$, since

$$|\zeta^\varepsilon(s,x)| \leq C_\alpha\frac{1+|x|^{4+\alpha}}{T-s} \leq C_\alpha\frac{1+|x|^{4+\alpha}}{T-t},$$

applying the Dominated Convergence Theorem we have

$$u(t,x) = u(0,x) - \int_0^t \zeta(s,x)ds.$$

Note that ζ is continuous, so u is differentiable with respect to t, and

$$u_t(t,x) = -\zeta(t,x).$$

That completes the proof. ■

Remark 2.4.2 The existence of the classical solution under such conditions seems classical in the literature of PDEs, but I cannot find a simple proof of it. We are presenting here a nice probabilistic proof, of intrinsic interest.

2.5 Extension to Discrete Functional Case

In this section we extend our representation formula, (3) of theorem 2.2.4, to BSDEs with discrete functional terminals. We shall prove a new representation theorem for the process Z, and we will extend the path regularity result to such a case. This is a key step towards our study of BSDEs with L^∞-Lipschitz functional terminals. Our main result is the following.

Theorem 2.5.1 Let $\Theta = (X, Y, Z)$ be the solution to the FBSDE:

$$\begin{cases} X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \\ Y_t = g(X_{t_0}, \cdots, X_{t_n}) + \int_t^T f(s, \Theta_s)ds - \int_t^T Z_s dW_s, \end{cases} \tag{2.83}$$

where $0 = t_0 < \cdots < t_n = T$ is a partition of $[0, T]$. Assume Assumption 2.0.12 holds. Then in each interval (t_{i-1}, t_i), $i = 1, \cdots, n$, the following identity holds for $\forall s \in (t_{i-1}, t_i)$:

$$Z_s = E\Big\{g(X_{t_0}, ..., X_{t_n})N^s_{t_i} + \int_s^T f(r, \Theta_r)N^s_{r\wedge t_i}dr \Big| \mathcal{F}_s\Big\}\sigma(s, X_s), \tag{2.84}$$

where

$$N^s_t = \frac{1}{t-s}\left(\int_s^t [\sigma^{-1}(r, X_s)\nabla X_r]^T dW_r\right)^T.$$

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), $i = 1, \cdots, n$;

(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 $\forall p > 2$, there exists a constant $C_p > 0$ depending only on T, K and p such that

$$E\left\{|\Delta Z_{t_i}|^p\right\} \le C_p(1 + |x|^p) < \infty. \tag{2.85}$$

Consequently, the process Z has both càdlàg and càglàd versions, with discontinuities $t_0, \cdots, t_n$ and jump sizes satisfying (2.85).

Proof. As before we will consider only the case $d = 1$; and we assume first that $f, g \in C_b^1$.

Let us first establish the identity (2.84). 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 the proof of Theorem 2.2.1 or in Pardoux-Peng [44], 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 $t_{i-1} < s \le \tau < t_i$,

$$D_sY_\tau = \sum_{j\ge i} \partial_j g D_s X_{t_j} + \int_\tau^T \left[f_x(r)D_sX_r + f_y(r)D_sY_r + f_z(r)D_sZ_r\right]dr - \int_\tau^T D_sZ_r dW_r, \tag{2.86}$$

where $\partial_j g \triangleq \partial_{x_j} g(X_{t_0}, ..., X_{t_n})$, $j = 1, \cdots, n$; 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 1.2.8 and the uniqueness of the adapted solution to BSDEs we have, for $t_{i-1} < s \le \tau < t_i$,

$$D_sX_\tau = \nabla X_\tau \Lambda_s^{-1}; \quad D_sY_\tau = \nabla^i Y_\tau \Lambda_s^{-1}; \quad D_sZ_\tau = \nabla^i Z_\tau \Lambda_s^{-1}, \tag{2.87}$$

where Λ is as defined in (2.2) and $(\nabla^i Y, \nabla^i Z)$ denotes the adapted solution to the following BSDE for $\tau \in [t_{i-1}, T]$:

$$\nabla^i Y_\tau = \sum_{j\ge i} \partial_j g \nabla X_{t_j} + \int_\tau^T \Big[f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r\Big]dr - \int_\tau^T \nabla^i Z_r dW_r. \tag{2.88}$$

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

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

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

$$\begin{aligned} 0 &= \sum_{j\geq i} \partial_j g D_s X_{t_j} + \int_s^T \left[f_x(r)D_sX_r + f_y(r)D_sY_r + f_z(r)D_sZ_r\right] dr \\ &\quad -Z_s - \int_s^T D_sZ_r dW_r. \end{aligned} \tag{2.89}$$

Combining (2.89) with (2.87) and (2.88) we obtain

$$\begin{aligned} Z_s &= \sum_{j\geq i} \partial_j g D_s X_{t_j} + \int_s^T [f_x(r)D_sX_r + f_y(r)D_sY_r + f_z(r)D_sZ_r]dr - \int_s^T D_sZ_r dW_r \\ &= \Big\{ \sum_{j\geq i} \partial_j g \nabla X_{t_j} + \int_s^T [f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r]dr \\ &\quad - \int_s^T \nabla^i Z_r dW_r \Big\} \Lambda_s^{-1} \qquad (2.90) \\ &= \nabla^i Y_s \Lambda_s^{-1}, \qquad t_{i-1} < s < t_i. \end{aligned}$$

Taking conditional expectation $E\{\cdot\,|\mathcal{F}_s\}$ on the both sides of (2.90) we then get

$$Z_s = E\Big\{ \sum_{j\geq i} \partial_j g \nabla X_{t_j} + \int_s^T \Big[f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r\Big] dr \Big| \mathcal{F}_s \Big\} \Lambda_s^{-1}. \tag{2.91}$$

The rest of the proof is similar to that of Theorem 2.2.4. First we note that by chain rule for the anticipating derivative operator and the relation (2.87), for any $t_{i-1} < \tau \leq t_i$ and $\tau < r$ one has

$$D_\tau f(X_r, Y_r, Z_r) = [f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r][\nabla X_\tau]^{-1}\sigma(\tau, X_\tau). \tag{2.92}$$

We consider the following two cases:

(a) $t_{i-1} < r \leq t_i$. In this case we derive from (2.92) that

$$f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r = \frac{1}{r-s}\int_s^r D_\tau f(r, \Theta_r)\sigma^{-1}(\tau, X_\tau)\nabla X_\tau d\tau, \tag{2.93}$$

for $t_{i-1} < s < r \leq t_i$. Therefore, using the integration by parts formula for anticipating integrals and recalling the definition of process N (2.2) we have

$$E\left\{ f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r \Big| \mathcal{F}_s \right\}$$

$$
\begin{aligned}
&= E\left\{\frac{1}{r-s}\int_s^r f(r,\Theta_r)\sigma^{-1}(\tau,X_\tau)\nabla X_\tau D_\tau d\tau\Big|\mathcal{F}_s\right\}\\
&= E\left\{\frac{1}{r-s}f(r,\Theta_r)\int_s^r \sigma^{-1}(\tau,X_\tau)\nabla X_\tau dW_\tau\Big|\mathcal{F}_s\right\} \qquad (2.94)\\
&= E\left\{f(r,\Theta_r)N_r^s\Big|\mathcal{F}_s\right\}\nabla X_s.
\end{aligned}
$$

(b) $t_i < r$. In this case we see that (2.92) is still true, but (2.93) should be replaced by

$$
f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r = \frac{1}{t_i - s}\int_s^{t_i} D_\tau f(r,\Theta_r)\sigma^{-1}(\tau,X_\tau)\nabla X_\tau d\tau,
$$

for $t_{i-1} < s < t_i < r$. Consequently, (2.94) is changed to

$$
\begin{aligned}
&E\Big\{f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r\Big|\mathcal{F}_s\Big\}\\
= \;& E\Big\{\frac{1}{t_i - s}\int_s^{t_i}\sigma^{-1}(\tau,X_\tau)\nabla X_\tau D_\tau f(r,\Theta_r)d\tau\Big|\mathcal{F}_s\Big\}\\
= \;& E\Big\{f(r,\Theta_r)N_{t_i}^s\Big|\mathcal{F}_s\Big\}\nabla X_s. \qquad (2.95)
\end{aligned}
$$

Combining (2.94) and (2.95) we see that for all $s \in (t_{i-1}, t_i)$ it holds that

$$
E\Big\{\int_s^T [f_x(r)\nabla X_r + f_y(r)\nabla^i Y_r + f_z(r)\nabla^i Z_r]dr\Big|\mathcal{F}_s\Big\} = E\Big\{\int_s^T f(r,\Theta_r)N_{r\wedge t_i}^s dr\Big|\mathcal{F}_s\Big\}\nabla X_s. \qquad (2.96)
$$

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

$$
D_\tau g(X_{t_0},\cdots,X_{t_n}) = \sum_{j\ge i}\partial_j g D_\tau X_{t_j} = \Big\{\sum_{j\ge i}\partial_j g\nabla X_{t_j}\Big\}\Lambda_\tau^{-1},
$$

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

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

Thus, using integration by parts again we have

$$
\begin{aligned}
E\Big\{\sum_{j\ge i}\partial_j g\nabla X_{t_j}|\mathcal{F}_s\Big\} &= E\Big\{\frac{1}{t_i - s}g(X_{t_0},\cdots,X_{t_n})\int_s^{t_i}\Lambda_\tau dW_\tau\Big|\mathcal{F}_s\Big\}\\
&= E\{g(X_{t_0},\cdots,X_{t_n})N_{t_i}^s|\mathcal{F}_s\}\nabla X_s. \qquad (2.97)
\end{aligned}
$$

Plugging (2.96) and (2.97) into (2.91) we obtain (2.84) 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) and (ii) are obvious, in light of Theorem 2.2.4, with T there being replaced by t_i, for each i, and thanks to the representation (2.84). Therefore we shall only check (iii).

To this end, let $\Delta Z_{t_i} = Z_{t_i+} - Z_{t_i-}$. From (2.90) it is easily seen that

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

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

$$\Delta Z_{t_i} = (\nabla^{i+1}Y_{t_i} - \nabla^i Y_{t_i})[\nabla X_{t_i}]^{-1}\sigma(t_i, X_{t_i}) = -\alpha^i_{t_i}[\nabla X_{t_i}]^{-1}\sigma(t_i, X_{t_i}), \tag{2.98}$$

Further, let us denote $\beta^i_s \stackrel{\triangle}{=} -(\nabla^{i+1}Z_s - \nabla^i Z_s)$. Then (2.88) leads to that

$$\alpha^i_s = \partial_i g \nabla X_{t_i} + \int_s^T \left[f_y(r)\alpha^i_r + f_z(r)\beta^i_r\right] dr - \int_s^T \beta^i_r dW_r, \quad s \in [t, T]. \tag{2.99}$$

In other words, (α^i, β^i) is the adapted solution to the linear BSDE (2.99). Therefore, by Lemma 1.2.6 we know that $\forall p > 0$ there exists a $C_p > 0$ such that $E\left\{|\alpha^i_{t_i}|^p\right\} \le C_p$. Recall (2.23), it is readily seen that (2.85) follows from (2.98). This proves (iii).

Finally, we note that when f and g are only Lipschitz, (2.84) still holds, modulo a standard approximation same as that in Theorem 2.2.4. Thus the properties (i) and (ii) are obvious. To see (iii) we should note that the standard approximation yields that $\Delta Z^\varepsilon_{t_i} \to \Delta Z_{t_i}$ a.s. So if (2.85) holds for $\Delta Z^\varepsilon_{t_i}$, then letting $\varepsilon \to 0$ we see that (2.85) remains true for ΔZ_{t_i}, thanks to the Fatou's lemma. The proof is now complete. ∎

CHAPTER 3. PATH REGULARITY FOR SOLUTIONS OF BSDES

In this chapter we shall study FBSDE (1.2), where the terminal Φ of the BSDE is a true functional, and prove the path regularity of the process Z. More precisely, we show that if the terminal Φ is an L^∞-Lipschitz functional, then the process Z admits a càdlàg version. Moreover, if the functional Φ is L^1-Lipschitz, then Z will admit even an almost surely continuous version. To ensure our results hold true, we shall use the following assumption similar to Assumption 2.0.12:

Assumption 3.0.2 (i) The functions $b, \sigma \in C^1_b$ and $f \in C_L$. We use a common constant $K > 0$ to denote all the Lipschitz constants, and assume

$$\sup_{0\le t\le T}\Big\{|b(t,0)| + |\sigma(t,0)| + |f(t,0,0,0)|\Big\} + |\Phi(\mathbf{0})| \le K.$$

(ii) $d_1 = d$. Moreover, we assume that σ satisfies:

$$\sigma(t,x)\sigma^T(t,x) \ge \frac{1}{K} I_d, \qquad \forall (t,x) \in [0,T]\times \mathbb{R}^d.$$

The rest of the chapter is organized as follows. In §1 we study the case that Φ is a discrete functional and establish a crucial estimate on the conditional variation of Z. In §2 we study the case that Φ is an L^∞-Lipschitz functional; and in §3 we extend the result to the L^1-Lipschitz case and prove a much stronger result. Finally in §4 we establish an L^2 type regularity result for the process Z, which plays an important role in the numerical scheme proposed in next chapter.

3.1 Discrete Functional Case Revisited

Before we begin our investigation, let us first recall a path regularity result we derived in Theorem 2.5.1. The main purpose of this chapter is to generalize that result to the case where the terminal value $\Phi(X)$ of the BSDE is an L^∞-Lipschitz functional of X. Our plan is to approximate an L^∞-Lipschitz functional Φ by a sequence of discrete functionals, and try to prove that the paths of the martingale part of the solution under study is a limit of the sequence corresponding solutions of BSDEs with discrete functional terminal, on the path space(!), from which the path regularity will follow. In this section we shall establish some further properties of the adapted solution to BSDEs with discrete functional terminal values, which is critical for proving the convergence of the sequence corresponding solutions of BSDEs.

To this end, we fix a partition $\pi : 0 = t_0 < t_1 < \cdots < t_n = T$ and assume $\Phi(X) = g(X_{t_0}, \cdots, X_{t_n})$. Recalling (2.88) we denote

$$\nabla^\pi Y_s = \sum_{i=1}^n \nabla^i Y_s 1_{[t_{i-1},t_i)}(s) + \nabla^n Y_{T-} 1_{\{T\}}(s), \qquad s \in [0,T]; \tag{3.1}$$

Then $\nabla^\pi Y$ is a càdlàg process.

For application convenience, we shall rewrite $\nabla^\pi Y$ in another form. Note that for each i (2.88) is linear. Let (γ^0, ζ^0) and (γ^j, ζ^j), $j = 1, \cdots, n$ be the adapted solutions of the BSDEs

$$\begin{aligned} \gamma_t^0 &= \int_t^T \Big[f_x(r)\nabla X_r + f_y(r)\gamma_r^0 + f_z(r)\zeta_r^0 \Big] dr - \Big\{ \int_t^T \zeta_r^0 dW_r \Big\}^T; \\ \gamma_t^j &= \partial_j g \nabla X_{t_j} + \int_t^T \Big[f_y(r)\gamma_r^j + f_z(r)\zeta_r^j \Big] dr - \Big\{ \int_t^T \zeta_r^j dW_r \Big\}^T, \end{aligned} \tag{3.2}$$

respectively, we have the following decomposition:

$$\nabla^i Y_s = \gamma_s^0 + \sum_{j \ge i} \gamma_s^j, \qquad s \in [t_{i-1}, t_i). \tag{3.3}$$

We may simplify (3.3) further. Let us define, for any $\eta \in L^1(\mathbf{F}, [0,T])$ and $\theta \in L^2(\mathbf{F}, [0,T]; \mathbb{R}^d)$ (viewed as *row vector*!),

$$\Lambda_t^s(\eta) \ \stackrel{\triangle}{=}\ \exp\Big\{ \int_s^t \eta(r) dr \Big\}; \quad s,t \in [0,T]; \tag{3.4}$$

$$\mathcal{E}_t^s(\theta) \ \stackrel{\triangle}{=}\ \exp\Big\{ \int_s^t \theta(r) dW_r - \frac{1}{2}\int_t^s |\theta(r)|^2 dr \Big\}; \quad s,t \in [0,T]. \tag{3.5}$$

($\mathcal{E}^t(\theta)$ is known as the *Doléan-Dade stochastic exponential* of θ.) Then it is easily checked that, for any $p > 0$, one has

$$[\mathcal{E}_t^s(\theta)]^p = \mathcal{E}_t^s(p\theta)\Lambda_t^s\Big(\frac{p(p-1)}{2}|\theta|^2\Big); \tag{3.6}$$

and

$$[\mathcal{E}_t^s(\theta)]^{-1} = \mathcal{E}_t^s(-\theta)\Lambda_t^s\Big(|\theta|^2\Big). \tag{3.7}$$

In particular, we denote, for $s, t \in [0, T]$,

$$\Lambda_t^s = \Lambda_t^s(-f_y); \quad M_t^s = \mathcal{E}_t^s(f_z); \tag{3.8}$$

and if there is no danger of confusion, we denote $\Lambda_\cdot = \Lambda_\cdot^0$ and $M_\cdot = M_\cdot^0$. Since f_z is uniformly bounded, by Girsanov's Theorem (see, e.g., [32]) we know that M is a P-martingale on $[0,T]$, and $\widetilde{W}_t \triangleq W_t - \int_0^t f_z(r)dr$, $t \in [0,T]$ is an $\mathbf{F}$-Brownian motion on the new probability space $(\Omega, \mathcal{F}, \widetilde{P})$, where $\widetilde{P}$ is defined by $\frac{d\widetilde{P}}{dP} = M_T$. Moreover, noting that f_y and f_z are uniformly bounded, by virtue of (3.6) and (3.7) one can deduce easily from (3.8) that, for $\forall p \geq 1$, there exists a constant C_p depending only on T, K and p, such that

$$\begin{cases} \sup\limits_{0\le t\le T}[|\Lambda_t|^p + |\Lambda_t^{-1}|^p] \le C_p; \quad E\{\sup\limits_{0\le t\le T}[|M_t|^p + |M_t^{-1}|^p]\} \le C_p; \\ |\Lambda_t - \Lambda_s|^p + |\Lambda_t^{-1} - \Lambda_s^{-1}|^p \le C_p|t-s|^p; \\ E\{|M_t - M_s|^p + |M_t^{-1} - M_s^{-1}|^p\} \le C_p|t-s|^{\frac{p}{2}}. \end{cases} \tag{3.9}$$

Now we define

$$\begin{cases} \widetilde{\xi}^0 \triangleq \int_0^T f_x(r)\nabla X_r \Lambda_r^{-1}dr; \quad \widetilde{\zeta}_t^0 \triangleq \zeta_t^0\Lambda_t^{-1}; \quad \widetilde{\gamma}_t^0 \triangleq \gamma_t^0\Lambda_t^{-1} + \int_0^t f_x(r)\nabla X_r\Lambda_r^{-1}dr; \\ \widetilde{\xi}^i \triangleq \partial_i g^\pi \nabla X_{s_i}\Lambda_T^{-1}; \qquad \widetilde{\zeta}_t^i \triangleq \zeta_t^i\Lambda_t^{-1}; \quad \widetilde{\gamma}_t^i \triangleq \gamma_t^i\Lambda_t^{-1}. \end{cases} \tag{3.10}$$

Then, using integration by parts and equation (3.2) we have, for $i = 0, 1, \cdots, n$,

$$\widetilde{\gamma}_t^i = \widetilde{\xi}^i - \Big\{\int_t^T \widetilde{\zeta}_r^i d\widetilde{W}_r\Big\}^T, \qquad t \in [0,T]. \tag{3.11}$$

Therefore, by the Bayes rule (see, e.g., [32, Lemma 3.5.3]) we have, for $t \in [0,T]$ and $i = 1, \cdots, n$,

$$\begin{cases} \gamma_t^i = \widetilde{\gamma}_t^i\Lambda_t = \widetilde{E}\{\widetilde{\xi}^i|\mathcal{F}_t\}\Lambda_t = E\{M_T\widetilde{\xi}^i|\mathcal{F}_t\}M_t^{-1}\Lambda_t = \xi_t^i M_t^{-1}\Lambda_t; \\ \gamma_t^0 = \widetilde{\gamma}_t^0\Lambda_t - \int_0^t f_x(r)\nabla X_r\Lambda_t^{-1}dr\Lambda_t = \xi_t^0 M_t^{-1}\Lambda_t - \int_0^t f_x(r)\nabla X_r\Lambda_r^{-1}dr\Lambda_t, \end{cases} \tag{3.12}$$

where for $i = 0, 1, \cdots, n$,

$$\xi_t^i \triangleq E\{M_T\widetilde{\xi}^i|\mathcal{F}_t\} = E\{M_T\widetilde{\xi}^i\} + \int_0^t \eta_r^i dW_r. \tag{3.13}$$

Note that the boundedness of f_z and (3.6) imply that $M_T \in L^p(\Omega)$ and $\nabla X \in L^p(\mathbf{F}; C([0,T]; \mathbb{R}^{d\times d}))$ for all $p \geq 2$. Therefore for each $p \geq 1$, (3.15) leads to

$$E\Big\{\sum_{j=1}^{n} |M_T \widetilde{\xi}^j|\Big\}^p \leq CE\Big\{|M_T|^p \sup_{0\leq t\leq T} |\nabla X_t|^p\Big\} \leq C. \tag{3.14}$$

In particular, for each j, $M_T\widetilde{\xi}^j \in L^2(\mathcal{F}_T)$. So (3.13) makes sense.

Note that we have already proved the following lemma.

Lemma 3.1.1 Assume $\Phi(X) = g(X_{t_0}, \cdots, X_{t_n})$ in (1.2) for some partition $\pi : 0 = t_0 < \cdots < t_n$. Assume (i) of Assumption 3.0.2 holds, and that $f, g \in C_b^1$, then we have

$$\nabla^i Y_t = (\xi_t^0 + \sum_{j\geq i} \xi_t^j) M_t^{-1} \Lambda_t - \int_0^t f_x(r) \nabla X_r \Lambda_r^{-1} dr \Lambda_t.$$

The following result is essential.

Theorem 3.1.2 Assume all the conditions of Lemma 3.1.1 hold; and that for the same constant $K > 0$, it holds for all $x = (x_0, \cdots, x_n) \in \mathbb{R}^{d_1(n+1)}$ that

$$\sum_{i=0}^{n} |\partial_{x_i} g(x)| \leq K. \tag{3.15}$$

Then, there exists a constant $C > 0$, depending on T and K, but independent of the partition π, such that

$$\sum_{i=1}^{n} E\left\{\left|E\left\{\nabla^\pi Y_{t_{i-1}} - \nabla^\pi Y_{t_i} \middle| \mathcal{F}_{t_{i-1}}\right\}\right|\right\} + E\{|\nabla^\pi Y_T|\} \leq C. \tag{3.16}$$

Proof. Again we assume $d_1 = d = 1$ for simplicity.

First note that (3.15) implies that $|\partial_{x_n} g| \leq K$. Thus by (3.1) we have

$$E\{|\nabla^\pi Y_T|\} = E\{|\partial_n g \nabla X_T|\} \leq KE\{|\nabla X_T|\} \leq C.$$

Next, using (3.1) and (3.3) we see that for each i,

$$\begin{aligned} \nabla^\pi Y_{t_{i-1}} - \nabla^\pi Y_{t_i} &= \nabla^i Y_{t_{i-1}} - \nabla^{i+1} Y_{t_i} = \Big(\gamma_{t_{i-1}}^0 + \sum_{j\geq i} \gamma_{t_{i-1}}^j\Big) - \Big(\gamma_{t_i}^0 + \sum_{j\geq i+1} \gamma_{t_i}^j\Big) \\ &= [\gamma_{t_{i-1}}^0 - \gamma_{t_i}^0] + \gamma_{t_i}^i + \sum_{j\geq i} \Big[\gamma_{t_{i-1}}^j - \gamma_{t_i}^j\Big]. \end{aligned} \tag{3.17}$$

Now let us denote the first term of the left hand side of (3.16) by I and show that $I \le C$ as well. First note that

$$\begin{aligned} I &\le E\Big\{\sum_{i=1}^n \Big|E\Big\{\gamma^0_{t_{i-1}} - \gamma^0_{t_i}\Big|\mathcal{F}_{t_{i-1}}\Big\}\Big|\Big\} + E\Big\{\sum_{i=1}^n \Big|E\Big\{\gamma^i_{t_i}\Big|\mathcal{F}_{t_{i-1}}\Big\}\Big|\Big\} \\ &\quad + E\Big\{\sum_{i=1}^n \Big|\sum_{j\ge i} E\Big\{\gamma^j_{t_{i-1}} - \gamma^j_{t_i}\Big|\mathcal{F}_{t_{i-1}}\Big\}\Big|\Big\} \qquad (3.18) \\ &= I_1 + I_2 + I_3, \end{aligned}$$

where I_i, $i = 1, 2, 3$ are defined in the obvious way. We now estimate I_1–I_3 separately. First, by definition (3.2) we have

$$\begin{aligned} I_1 &= E\Big\{\sum_{i=1}^n \Big|E\Big\{\int_{t_{i-1}}^{t_i} \Big[f_x(r)\nabla X_r + f_y(r)\gamma^0_r + f_z(r)\zeta^0_r\Big]dr\Big|\mathcal{F}_{t_{i-1}}\Big\}\Big|\Big\} \\ &\le \sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i} |f_x(r)\nabla X_r + f_y(r)\gamma^0_r + f_z(r)\zeta^0_r|dr\Big\} \qquad (3.19) \\ &= E\Big\{\int_0^T |f_x(r)\nabla X_r + f_y(r)\gamma^0_r + f_z(r)\zeta^0_r|dr\Big\} \\ &\le CE\Big\{\int_0^T (1 + |\nabla X_r|^2 + |\gamma^0_r|^2 + |\zeta^0_r|^2)dr\Big\}. \end{aligned}$$

Recalling (2.23) and applying Lemmas 1.2.5 and 1.2.6 to SDE (3.2) we conclude that $I_1 \le C$.

Now, recall (3.12), (3.10) and the boundedness of f_y (whence both $\Lambda_\cdot$ and $\Lambda^{-1}_\cdot$), we obtain that

$$\begin{aligned} I_2 &= \sum_{i=1}^n E\Big|E\{\gamma^i_{t_i}|\mathcal{F}_{t_{i-1}}\}\Big| = \sum_{i=1}^n E\Big|E\{\widetilde{\xi}^i M^{t_i}_T \Lambda_{t_i}|\mathcal{F}_{t_{i-1}}\}\Big| \le \sum_{i=1}^n E\{|\widetilde{\xi}^i M^{t_i}_T \Lambda_{t_i}|\} \\ &= E\Big\{\sum_{i=1}^n |\partial_i g \nabla X_{t_i}\Lambda_T^{-1} M^{t_i}_T \Lambda_{t_i}|\Big\} \le CE\Big\{\sup_{0\le t\le T} |M^t_T \nabla X_t|\Big\} \le C. \qquad (3.20) \end{aligned}$$

Here we have used the assumption (3.15), as well as the fact that both M and ∇X are continuous, square-integrable processes.

The estimate for I_3 is a little more involved. First, from (3.8), (3.7), and (3.12) we see that

$$\begin{aligned} &E\{\gamma^j_{t_{i-1}} - \gamma^j_{t_i}|\mathcal{F}_{t_{i-1}}\} = E\{\widetilde{\xi}^j[M^{t_{i-1}}_T \Lambda_{t_{i-1}} - M^{t_i}_T \Lambda_{t_i}]|\mathcal{F}_{t_{i-1}}\} \\ &= E\{\widetilde{\xi}^j M^{t_i}_T(\Lambda_{t_{i-1}} - \Lambda_{t_i})|\mathcal{F}_{t_{i-1}}\} + E\{\widetilde{\xi}^j \Lambda_{t_{i-1}}(M^{t_{i-1}}_T - M^{t_i}_T)|\mathcal{F}_{t_{i-1}}\}. \end{aligned}$$

Thus

$$\begin{aligned} I_3 &= E\Big\{\sum_{i=1}^n\Big|\sum_{j\ge i}E\left\{\gamma^j_{t_{i-1}}-\gamma^j_{t_i}\Big|\mathcal{F}_{t_{i-1}}\right\}\Big|\Big\}\\ &\le \sum_{i=1}^n\sum_{j\ge i}E\{|\widetilde{\xi}^j||M_T^{t_i}||\Lambda_{t_{i-1}}-\Lambda_{t_i}|\}+\sum_{i=1}^n E\Big\{\Big|\sum_{j\ge i}E\{\widetilde{\xi}^j\Lambda_{t_{i-1}}(M_T^{t_{i-1}}-M_T^{t_i})\Big|\mathcal{F}_{t_{i-1}}\}\Big|\Big\}\\ &= I_3^1+I_3^2, \end{aligned}$$

where I_3^1 and I_3^2 are defined in the obvious way.

Again, using the boundedness of f_y we have

$$|\Lambda_{t_{i-1}}-\Lambda_{t_i}|=|\Lambda_{t_{i-1}}|\left|\exp\Big\{-\int_{t_{i-1}}^{t_i}f_y(r)dr\Big\}-1\right|\le C\left|\int_{t_{i-1}}^{t_i}f_y(r)dr\right|\le C(t_i-t_{i-1}). \tag{3.21}$$

Moreover, by assumption (3.15) we have

$$\sum_{j\ge i}|\widetilde{\xi}^j|\le\sum_{j=0}^n|\partial_j g\nabla X_{t_j}\Lambda_T^{-1}|\le C\max_{0\le i\le n}|\nabla X_{t_i}|\le C\max_{0\le t\le T}|\nabla X_t|. \tag{3.22}$$

Thus, combining (3.21) and (3.22) we get

$$\begin{aligned} I_3^1 &= \sum_{i=1}^n E\Big\{|M_T^{t_i}||\Lambda_{t_{i-1}}-\Lambda_{t_i}|\Big(\sum_{j\ge i}|\widetilde{\xi}^j|\Big)\Big\}\\ &\le C\sum_{i=1}^n E\Big\{|M_T^{t_i}|\sup_{0\le t\le T}|\nabla X_t|\Big\}(t_i-t_{i-1}) \\ &\le CE\Big\{\sup_{0\le t\le T}|M_T^t|^2+\sup_{0\le t\le T}|\nabla X_t|^2\Big\}\sum_{i=1}^n(t_i-t_{i-1})\le C. \end{aligned} \tag{3.23}$$

We now turn to I_3^2. To this end we define $\widetilde{M}_t=\mathcal{E}_t^0(-f_z)$, $t\in[0,T]$ (compare to $M.$ in (3.8)!). Again, the boundedness of f_z renders $\widetilde{M}$ a P-martingale, and by (3.7) we have

$$M_t^{-1}=\widetilde{M}_t\Lambda_t(|f_z|^2)\stackrel{\triangle}{=}\widetilde{M}_t\widetilde{\Lambda}_t,\quad t\in[0,T], \tag{3.24}$$

where $\widetilde{\Lambda}_t\stackrel{\triangle}{=}\Lambda_t(|f_z|^2)$. Now by definition of (3.8) we have

$$M_T^{t_{i-1}}-M_T^{t_i}=M_T\{M_{t_{i-1}}^{-1}-M_{t_i}^{-1}]\}=M_T\{\widetilde{M}_{t_{i-1}}\widetilde{\Lambda}_{t_{i-1}}-\widetilde{M}_{t_i}\widetilde{\Lambda}_{t_i}\}.$$

Thus we see that the I_3^2 can be written as

$$I_3^2 = \sum_{i=1}^n E\Big\{\Big|\sum_{j\ge i}E\{\widetilde{\xi}^jM_T\Lambda_{t_{i-1}}(\widetilde{M}_{t_{i-1}}\widetilde{\Lambda}_{t_{i-1}}-\widetilde{M}_{t_i}\widetilde{\Lambda}_{t_i})|\mathcal{F}_{t_{i-1}}\}\Big|\Big\}$$

$$\leq \sum_{i=1}^n E\Big\{\Big|\sum_{j\geq i} E\{\tilde\xi^j M_T \Lambda_{t_{i-1}}\tilde\Lambda_{t_{i-1}}(\widetilde M_{t_{i-1}}-\widetilde M_{t_i})|\mathcal{F}_{t_{i-1}}\}\Big|\Big\} \qquad (3.25)$$
$$+\sum_{i=1}^n E\Big\{\Big|\sum_{j\geq i} E\{\tilde\xi^j M_T \Lambda_{t_{i-1}}\widetilde M_{t_i}(\tilde\Lambda_{t_{i-1}}-\tilde\Lambda_{t_i})|\mathcal{F}_{t_{i-1}}\}\Big|\Big\}.$$

Similar to (3.23) we can show that

$$\sum_{i=1}^n E\Big\{\Big|\sum_{j\geq i} E\{\tilde\xi^j M_T \Lambda_{t_{i-1}}\widetilde M_{t_i}(\tilde\Lambda_{t_{i-1}}-\tilde\Lambda_{t_i})|\mathcal{F}_{t_{i-1}}\}\Big|\Big\}\leq C, \qquad (3.26)$$

thanks to the boundedness of f_z. Thus it remains to prove that the first term on the right hand side of (3.25) is bounded as well. Recall (3.13) and note that $\widetilde M$ is a P-martingale as well, it is easily checked that,

$$\begin{aligned}
&\sum_{i=1}^n E\Big|\sum_{j\geq i} E\{\tilde\xi^j M_T \Lambda_{t_{i-1}}\tilde\Lambda_{t_{i-1}}(\widetilde M_{t_{i-1}}-\widetilde M_{t_i})|\mathcal{F}_{t_{i-1}}\}\Big| \\
=\ &\sum_{i=1}^n E\Big\{|\Lambda_{t_{i-1}}\tilde\Lambda_{t_{i-1}}|\Big|\sum_{j\geq i} E\{\tilde\xi^j M_T(\widetilde M_{t_{i-1}}-\widetilde M_{t_i})|\mathcal{F}_{t_{i-1}}\}\Big|\Big\} \\
\leq\ & CE\Big\{\sum_{i=1}^n\Big|E\Big\{\Big[\sum_{j\geq i}(\xi^j_{t_i}-\xi^j_{t_{i-1}})\Big](\widetilde M_{t_{i-1}}-\widetilde M_{t_i})|\mathcal{F}_{t_{i-1}}\Big\}\Big|\Big\} \qquad (3.27)\\
\leq\ & CE\Big\{\sum_{i=1}^n\Big|\sum_{j\geq i}(\xi^j_{t_i}-\xi^j_{t_{i-1}})\Big|^2+\sum_{i=1}^n(\widetilde M_{t_{i-1}}-\widetilde M_{t_i})^2\Big\}.
\end{aligned}$$

Now, using Itô's formula one shows that the exponential martingale $\widetilde M$ satisfies

$$E(\widetilde M_{t_{i-1}}-\widetilde M_{t_i})^2 = E\int_{t_{i-1}}^{t_i}|f_z(r)\widetilde M_r|^2dr\leq C(t_i-t_{i-1}),$$

and thus we have

$$\sum_{i=1}^n E|(\widetilde M_{t_{i-1}}-\widetilde M_{t_i})|^2\leq C. \qquad (3.28)$$

On the other hand, since

$$\begin{aligned}
&\sum_{i=1}^n E\Big\{\Big|\sum_{j\geq i}(\xi^j_{t_i}-\xi^j_{t_{i-1}})\Big|^2\Big\} = \sum_{i=1}^n\sum_{j_1,j_2\geq i}E\Big\{(\xi^{j_1}_{t_i}-\xi^{j_1}_{t_{i-1}})(\xi^{j_2}_{t_i}-\xi^{j_2}_{t_{i-1}})\Big\}\\
=\ &\sum_{i=1}^n\sum_{j_1,j_2\geq i}E\Big\{\int_{t_{i-1}}^{t_i}\eta^{j_1}_t\eta^{j_2}_t dt\Big\} = \sum_{j_1,j_2=1}^n\sum_{i\leq j_1\wedge j_2}E\Big\{\int_{t_{i-1}}^{t_i}\eta^{j_1}_t\eta^{j_2}_t dt\Big\} \qquad (3.29)\\
=\ &\sum_{j_1,j_2=1}^n E\Big\{\int_0^{t_{j_1\wedge j_2}}\eta^{j_1}_t\eta^{j_2}_t dt\Big\} = \sum_{j_1,j_2=1}^n E\Big\{\Big(\xi^{j_1}_{t_{j_1\wedge j_2}}-\xi^{j_1}_0\Big)\Big(\xi^{j_2}_{t_{j_1\wedge j_2}}-\xi^{j_2}_0\Big)\Big\}.
\end{aligned}$$

Let us define, for each j, a positive martingale associated to ξ^j:

$$|\xi|_t^j = E\{|M_T\widetilde{\xi}^j||\mathcal{F}_t\}, \qquad t\in[0,T],$$

then $|\xi_t^j - \xi_0^j| \le |\xi|_t^j$, $t\in[0,T]$, $j=1,\cdots,n$, such that

$$\begin{aligned}
&\sum_{j_1,j_2=1}^n E\Big\{\Big(\xi^{j_1}_{t_{j_1\wedge j_2}}-\xi^{j_1}_0\Big)\Big(\xi^{j_2}_{t_{j_1\wedge j_2}}-\xi^{j_2}_0\Big)\Big\} \le 2\sum_{j_1\le j_2} E\{|\xi|^{j_1}_{t_{j_1}}|\xi|^{j_2}_{t_{j_1}}\}\\
&= 2\sum_{j_1=1}^n E\Big\{|\xi|^{j_1}_{t_{j_1}}\sum_{j_2=j_1}^n|\xi|^{j_2}_{t_{j_1}}\Big\} \le 2\sum_{j_1=1}^n E\Big\{|\xi|^{j_1}_{t_{j_1}}\sum_{j_2=1}^n|\xi|^{j_2}_{t_{j_1}}\Big\}\\
&= 2\sum_{j_1=1}^n E\Big\{|M_T\xi_{j_1}|\sum_{j_2=1}^n|\xi|^{j_2}_{t_{j_1}}\Big\} \le 2\sum_{j_1=1}^n E\{|M_T\xi_{j_1}|\sup_{0\le t\le T}\sum_{j_2=1}^n|\xi|^{j_2}_t\} \qquad (3.30)\\
&\le E\Big\{\Big(\sum_{j=1}^n|M_T\widetilde{\xi}^j|\Big)^2+\Big(\sup_{0\le t\le T}\sum_{j=1}^n|\xi|^j_t\Big)^2\Big\} \le CE\Big\{\sum_{j=1}^n|M_T\widetilde{\xi}^j|\Big\}^2.
\end{aligned}$$

Here for the last inequality above we used Doob's inequality (applied to to the martingale $\sum_{j=1}^n|\xi|^j$). Now by (3.14) we see that (3.30) and (3.29) yield that

$$\sum_{i=1}^n E\Big|\sum_{j\ge i}^n(\xi^j_{t_i}-\xi^j_{t_{i-1}})\Big|^2 \le CE\Big\{\sum_{j=1}^n|M_T\widetilde{\xi}^j|\Big\}^2 \le C. \qquad (3.31)$$

Plugging (3.31) and (3.28) into (3.27), then combining with (3.26) and (3.25) we obtain that $I_3^2\le C$. This, together with (3.23), shows that $I_3\le C$, and hence $I\le C$. The proof is now complete. ■

Remark 3.1.3 We should point out that the generic constant C in (3.16) is independent of n and the choice of the partition π. This will be crucial in our future discussion.

3.2 L^∞-Lipschitz Functional Case

We are now ready to study the path regularity of the adapted solution to the FBSDE (1.2) where Φ is a functional on $\mathbb{D}$ satisfying the L^∞-Lipschitz condition (1.3).

Our first step is to approximate a functional Φ satisfying (1.3) by a sequence of discrete functionals satisfying (3.15). We proceed as follows. For any partition π :

$0 = t_0 < t_1 < ... < t_n = T$, we define a mapping $\varphi_\pi : C([0,T]; \mathbb{R}^{d_1}) \mapsto C([0,T]; \mathbb{R}^{d_1})$ by $\mathbf{x} \mapsto \varphi_\pi(\mathbf{x}) \stackrel{\triangle}{=} \mathbf{x}_\pi$, where

$$\mathbf{x}_\pi(t) \stackrel{\triangle}{=} \frac{1}{t_i - t_{i-1}}[(t_i - t)\mathbf{x}(t_{i-1}) + (t - t_{i-1})\mathbf{x}(t_i)], \quad t \in [t_{i-1}, t_i]. \tag{3.32}$$

Denote $|\pi| = \max_{1 \le i \le n} |t_i - t_{i-1}|$ to be the mesh size of the partition π. Then, using the uniform continuity of $\mathbf{x}$ it is easy to see that $\lim_{|\pi| \to 0} \sup_{0 \le t \le T} |\mathbf{x}_\pi(t) - \mathbf{x}(t)| = 0$. Next, for the given functional Φ we define a new functional Φ_π as

$$\Phi_\pi(\mathbf{x}) \stackrel{\triangle}{=} \Phi(\mathbf{x}_\pi), \qquad \forall \mathbf{x} \in C([0,T]; \mathbb{R}^{d_1}). \tag{3.33}$$

then by assumption (1.3), one has

$$\lim_{|\pi| \to 0} |\Phi_\pi(\mathbf{x}) - \Phi(\mathbf{x})| \le K \lim_{|\pi| \to 0} \sup_{0 \le t \le T} |\mathbf{x}_\pi(t) - \mathbf{x}(t)| = 0, \quad \forall \mathbf{x} \in C([0,T]; \mathbb{R}^{d_1}). \tag{3.34}$$

Now let X be the forward part of the solution to (1.2), and denote $\xi^\pi \stackrel{\triangle}{=} \Phi_\pi(X)$. Then (3.34) implies that $\xi^\pi \to \Phi(X)$, P-a.s., as $|\pi| \to 0$. Moreover, if we denote $X^\pi_\cdot(\omega) \stackrel{\triangle}{=} \varphi_\pi(X).(\omega)$, then (1.3) leads to

$$|\Phi_\pi(X)| \le C\Big\{|\Phi(\mathbf{0})| + \sup_{0 \le s \le T} |X^\pi_s|\Big\} \le C\Big\{|\Phi(\mathbf{0})| + \sup_{0 \le s \le T} |X_s|\Big\}.$$

Thus, by the Dominated Convergence Theorem we see that

$$\lim_{|\pi| \to 0} E|\Phi_\pi(X) - \Phi(X)|^2 = 0. \tag{3.35}$$

Consequently, Theorem 1.2.7 tells us that, if one denotes (Y^π, Z^π) to be the backward part of the adapted solution to the FBSDE (1.2) with the terminal $\Phi(X)$ being replaced by $\Phi_\pi(X)$, then it holds that

$$E\Big\{\sup_{0 \le t \le T} |Y^\pi_t - Y_t|^2 + \int_0^T |Z^\pi_t - Z_t|^2 dt\Big\} \to 0, \quad \text{ as } |\pi| \to 0. \tag{3.36}$$

To construct the desired family of discrete functional, we make a further reduction. For the given partition π we define a mapping $\psi_\pi : C([0,T]; \mathbb{R}^{d_1}) \mapsto \mathbb{R}^{d_1(n+1)}$ by

$$\psi_\pi(\mathbf{x}) = (\mathbf{x}(t_0), \mathbf{x}(t_1), \cdots, \mathbf{x}(t_n)), \qquad \forall \mathbf{x} \in C([0,T]; \mathbb{R}^{d_1}). \tag{3.37}$$

Denote $C_\pi([0,T]; \mathbb{R}^{d_1}) = \{\mathbf{x}_\pi : \mathbf{x} \in C([0,T]; \mathbb{R}^{d_1})\}$, then $C_\pi([0,T]; \mathbb{R}^{d_1})$ is a subspace of $C([0,T]; \mathbb{R}^{d_1})$, and ψ_π is a $1-1$ correspondence between $C_\pi([0,T]; \mathbb{R}^{d_1})$ and $\mathbb{R}^{d_1(n+1)}$. We have the following lemma.

Lemma 3.2.1 Suppose that Φ is an L^∞-Lipschitz functional satisfying condition (1.3). Let $\Pi = \{\pi\}$ be a family of partitions of $[0,T]$. Then there exists a family of discrete functionals $\{g_\pi : \pi \in \Pi\}$ such that

(i) for each $\pi \in \Pi$, denoting $n = \#\pi - 1$ (where $\#\pi$ denotes the number of partition points in π), $g_\pi \in C_b^\infty(\mathbb{R}^{d_1(n+1)})$, and satisfies (3.15), with constant K being the same as that in (1.3).

(ii) for any $\mathbf{x} \in C([0,T];\mathbb{R}^{d_1})$, it holds that

$$\lim_{|\pi|\to 0} |g_\pi(\psi_\pi(\mathbf{x})) - \Phi_\pi(\mathbf{x})| = 0. \tag{3.38}$$

Proof. Let Φ and $\pi \in \Pi$ be given. Define $G_\pi \stackrel{\triangle}{=} \Phi \circ \psi_\pi^{-1}$, and denote $n = \#\pi - 1$. Then it is easily checked that G_π is a mapping from $\mathbb{R}^{d_1(n+1)}$ to $\mathbb{R}$, such that

$$G_\pi(x(t_0), x(t_1), \cdots, x(t_n)) = G_\pi(\psi_\pi(\mathbf{x})) = \Phi(\mathbf{x}_\pi), \qquad \forall \mathbf{x} \in C([0,T];\mathbb{R}^{d_1}). \tag{3.39}$$

Since Φ satisfies (1.3), one can easily check that

$$|G_\pi(x_0, x_1, \cdots, x_n) - G_\pi(y_0, y_1, \cdots, y_n)| \le K \max_{0\le i\le n} |x_i - y_i|.$$

That is, G_π is (uniform) Lipschitz continuous with Lipschitz constant K being the same as that in (1.3).

Now let $\phi \in C_0^\infty(\mathbb{R}^{d_1(n+1)})$ be such that $\phi \ge 0$ and $\int_{\mathbb{R}^{d_1(n+1)}} \phi(z)dz = 1$. For fixed π and $\varepsilon > 0$ we define

$$G_\pi^\varepsilon(x) = \int_{\mathbb{R}^{d_1(n+1)}} G_\pi(x - \varepsilon z)\phi(z)dz,$$

Then $G_\pi^\varepsilon \in C_b^\infty(\mathbb{R}^{d_1(n+1)})$, such that $\sup_{x\in\mathbb{R}^{d_1(n+1)}} |G_\pi^\varepsilon(x) - G_\pi(x)| \to 0$, as $\varepsilon \to 0$. Next, for each $\pi \in \Pi$ choose $\varepsilon(\pi)$ such that

$$\sup_{(x_0,x_1,\cdots,x_n)} |G_\pi^{\varepsilon(\pi)}(x_0, x_1, \cdots, x_n) - G_\pi(x_0, x_1, \cdots, x_n)| < |\pi|, \tag{3.40}$$

and define $g_\pi = G_\pi^{\varepsilon(\pi)}$. Then, clearly $g_\pi \in C_b^\infty(\mathbb{R}^{d_1(n+1)})$; and by definitions of $G_\pi^{\varepsilon(\pi)}$, G_π, and (3.39), for any $\mathbf{x} \in C([0,T];\mathbb{R}^{d_1})$ we have

$$|g_\pi(\psi_\pi(\mathbf{x})) - \Phi_\pi(\mathbf{x})| = |G_\pi^{\varepsilon(\pi)}(\psi_\pi(\mathbf{x})) - G_\pi(\psi_\pi(\mathbf{x}))| \le \sup_{x\in\mathbb{R}^{d_1(n+1)}} |G_\pi^{\varepsilon(\pi)}(x) - G_\pi(x)| \le |\pi|.$$

Namely (3.38) holds, proving (ii).

We now show that g_π satisfies (3.15). Indeed, denoting $\delta_j(x) = \operatorname{sgn}(\partial_j g_\pi(x))$, $x = (x_0, x_1, \cdots, x_n)$ (same for $z \in \mathbb{R}^{d_1(n+1)}$), we have

$$
\begin{aligned}
\sum_{j=0}^{n} |\partial_j g_\pi(x_0, ..., x_n)| &= \sum_{j=0}^{n} \partial_j g_\pi(x_0, ..., x_n)\delta_j = \lim_{h\to 0} \frac{1}{h} \left(g_\pi(x + h\delta) - g_\pi(x)\right) \\
&= \lim_{h\to 0} \frac{1}{h} \int_{\mathbb{R}^{d_1(n+1)}} \Big\{G_\pi(x - \varepsilon(\pi)z + h\delta) - G_\pi(x - \varepsilon(\pi)z)\Big\}\phi(z)dz \\
&= \lim_{h\to 0} \frac{1}{h} \int_{\mathbb{R}^{d_1(n+1)}} \Big[\Phi \circ \psi_\pi^{-1}(x - \varepsilon(\pi)z + h\delta) - \Phi \circ \psi_\pi^{-1}(x - \varepsilon(\pi)z)\Big]\phi(z)dz \quad (3.41) \\
&\le \varliminf_{h\to 0} \frac{1}{h} \int_{\mathbb{R}^{d_1(n+1)}} K \sup_{0\le s\le T} |[\psi_\pi^{-1}(x - \varepsilon(\pi)z + h\delta) - \psi_\pi^{-1}(x - \varepsilon(\pi)z)](s)|\phi(z)dz \\
&= \varliminf_{h\to 0} \frac{1}{h} \int_{\mathbb{R}^{d_1(n+1)}} K \max_{0\le j\le n} |h\delta_j|\phi(z)dz = K.
\end{aligned}
$$

This proves (i), whence the lemma. ∎

We now give the main result of this section.

Theorem 3.2.2 Assume Assumption 3.0.2 holds and that Φ satisfies the L^∞ Lipschitz condition (1.3). Let (X, Y, Z) be the (unique) adapted solution of (1.2), then Z admits a càdlàg version.

Proof. Let $\Pi = \{\pi\}$ be the family of all partitions of $[0, T]$; and let $\{g_\pi, \pi \in \Pi\}$ be the family of discrete functionals constructed in Lemma 3.2.1. Further, let $\{f_\varepsilon, \varepsilon > 0\}$ be a family of molifiers of f, that is $f_\varepsilon \in C_b^1$ such that

$$
\sup_{(t,x,y,z)} |f_\varepsilon(t, x, y, z) - f(t, x, y, z)| \to 0, \qquad \text{as } \varepsilon \to 0. \tag{3.42}
$$

Let $\varepsilon(\pi)$ be the one chosen in (3.40), and define $f_\pi = f_{\varepsilon(\pi)}$. Then it is clear that $f_\pi(t, x, y, z) \to f(t, x, y, z)$, as $|\pi| \to 0$, uniformly in (t, x, y, z).

Now let us consider the following FBSDE: for each $\pi : 0 = t_0 < t_1 < \cdots < t_n = T$,

$$
Y_t^\pi = \xi^\pi + \int_t^T f_\pi(r, X_r, Y_r^\pi, Z_r^\pi)dr - \int_t^T Z_r^\pi dW_r, \quad t \in [0, T], \tag{3.43}
$$

where $\xi^\pi = g_\pi(X_{t_0}, ..., X_{t_n})$. Since each g_π satisfies (3.15) with the same constant $K > 0$ of (1.3), we see that

$$
|\sigma^\pi|^2 = |g_\pi(X_{t_0}, \cdots, X_{t_n}|^2 \le 2\Big(|\Phi(\mathbf{0})|^2 + K^2 \max_{0\le i\le n} |X_{t_i}|^2 \le C\Big(1 + \sup_{0\le t\le T} |X_t|^2\Big).
$$

Further, similar to the derivation of (3.35) (recall the notations there), we now apply Lemma 3.2.1 to get, P-almost surely,

$$|\xi^\pi - \Phi(X)| \le |g_\pi(\psi_\pi(X)) - \Phi_\pi(X)| + |\Phi_\pi(X) - \Phi(X)| \to 0,$$

as $|\pi| \to 0$. Therefore, the Dominated Convergence Theorem leads to $E\{\xi^\pi - \Phi(X)|^2\} \to 0$, as $|\pi| \to 0$. Now taking (3.42) into account and applying Theorem 1.2.7, we have

$$E\left\{\sup_{0\le t\le T}\sup_{0\le t\le T}|Y_t^\pi - Y_t|^2 + \int_0^T |Z_t^\pi - Z_t|^2 dt\right\} \to 0, \quad \text{as } |\pi| \to 0. \tag{3.44}$$

We now analyze the family $\{Z^\pi\}$. First, recall from Theorem 2.5.1 we know that each Z^π has a càdlàg version. We shall always take such version from now on. Second, each Z^π has the following representation:

$$Z_t^\pi = \nabla^\pi Y_t^\pi [\nabla X_t]^{-1}\sigma(t, X_t), \quad t \in [0,T], \tag{3.45}$$

where $\nabla^\pi Y^\pi$ is defined similar to (3.1), with Y being replaced by Y^π, and $\nabla^i Y^\pi$ satisfies the following BSDE:

$$\begin{aligned} \nabla^i Y_t^\pi \;\; &= \;\; \sum_{j\ge i}\partial_j g_\pi \nabla X_{t_j} + \int_t^T \Big[\partial_x f_\pi(r)\nabla X_r + \partial_y f_\pi(r)\nabla^i Y_r^\pi + \partial_z f_\pi(r)\nabla^i Z_r^\pi\Big]\,dr \\ &\quad -\Big\{\int_t^T \nabla^i Z_r^\pi dW_r\Big\}^T, \quad t \in [t_{i-1}, T], \quad i = 1, \cdots, n. \end{aligned} \tag{3.46}$$

We shall prove that the family $\{\nabla^\pi Y^\pi\}$ is tight. To this end, fix $\pi \in \Pi$, and let $\widehat{\pi} : 0 = s_0 < \ldots < s_m = T$ be any partition of $[0,T]$. We shall estimate the *conditional variation* of $\nabla^\pi Y^\pi$ (see (1.20)):

$$V_{\widehat{\pi}}(\nabla^\pi Y^\pi) = \sum_{i=1}^m E\{|E\{\nabla^\pi Y_{s_i}^\pi - \nabla^\pi Y_{s_{i-1}}^\pi | \mathcal{F}_{s_{i-1}}\}|\} + E\{|\nabla^\pi Y_T^\pi|\}. \tag{3.47}$$

To begin with, we note that for any process A, $V_\pi(A) \le V_{\pi'}(A)$ if $\pi \subseteq \pi'$, here the inclusion means all partition points of π are contained in π'. Indeed, for any $r_1 < r_2 < r_3$ one has

$$E\{|E\{A_{r_3} - A_{r_1}|\mathcal{F}_{r_1}\}|\}$$

$$
\begin{aligned}
&\le\; E\{|E\{A_{r_3}-A_{r_2}|\mathcal{F}_{r_1}\}|\}+E\{|E\{A_{r_2}-A_{r_1}|\mathcal{F}_{r_1}\}|\}\\
&=\; E\{|E\{E\{A_{r_3}-A_{r_2}|\mathcal{F}_{r_2}\}|\mathcal{F}_{r_1}\}|\}+E\{|E\{A_{r_2}-A_{r_1}|\mathcal{F}_{r_1}\}|\}\\
&\le\; E\{|E\{A_{r_3}-A_{r_2}|\mathcal{F}_{r_2}\}|\}+E\{|E\{A_{r_2}-A_{r_1}|\mathcal{F}_{r_1}\}|\}.
\end{aligned}
$$

Namely, the conditional variation increases as the partition gets finer. Therefore without loss of generality we may assume that $\pi\subseteq\widehat{\pi}$ (otherwise we simply consider $\pi\cup\widehat{\pi}$). To be more precise, let us assume that $t_i=s_{\ell_i}$, for $i=0,...,n$. We now recast the BSDE (3.43) as follows. Define a discrete functional $\widetilde{g}_{\widehat{\pi}}:\mathbb{R}^{d_1(m+1)}\mapsto\mathbb{R}$ by

$$
\widetilde{g}_{\widehat{\pi}}(x_0,x_1...,x_m)=g_\pi(x_{\ell_0},....,x_{\ell_n}),
$$

then $\widetilde{g}_{\widehat{\pi}}(X_{s_0},...,X_{s_m})=g_\pi(X_{t_0},...,X_{t_n})=\xi^\pi$, and (Y^π,Z^π) can be viewed as the solution of the BSDE

$$
Y^\pi_s=\widetilde{g}_{\widehat{\pi}}(X_{s_0},...,X_{s_m})+\int_s^T f_\pi(r,X_r,Y^\pi_r,Z^\pi_r)dr-\int_s^T Z^\pi_r dW_r. \tag{3.48}
$$

Furthermore, since $\partial_{x_j}\widetilde{g}_{\widehat{\pi}}(x_0,\cdots,x_m)=0$, if $j\notin\{\ell_0,\cdots,\ell_n\}$, we see that

$$
\sum_{k=0}^m|\partial_k\widetilde{g}_{\widehat{\pi}}(x_0,\cdots,x_m)|=\sum_{i=0}^n|\partial_i g_\pi(x_{\ell_0},\cdots,x_{\ell_n})|\le K,
$$

thanks to Lemma 3.2.1. Thus $\widetilde{g}_{\widehat{\pi}}$ satisfies (3.15) as well. We now apply Theorem 3.1.2 to (Y^π,Z^π) (regarded as the solution to BSDE (3.48)!) to get that

$$
\sum_{k=1}^m E\Big\{\Big|E\Big\{\nabla^{\widehat{\pi}}Y^\pi_{s_k}-\nabla^{\widehat{\pi}}Y^\pi_{s_{k-1}}\Big|\mathcal{F}_{s_{k-1}}\Big\}\Big|\Big\}+E|\nabla^{\widehat{\pi}}Y^\pi_T|\le C, \tag{3.49}
$$

where $C>0$ is a constant independent of the choice of partition $\widehat{\pi}$,

$$
\nabla^{\widehat{\pi}}Y^\pi_t=\sum_{k=1}^m\widehat{\nabla}^kY^\pi_t 1_{[s_{k-1},s_k)}(t)+\partial_{x_m}\widetilde{g}_{\widehat{\pi}}(X_{s_0},\cdots,X_{s_m})\nabla X_T 1_{\{T\}}(t),\quad t\in[0,T],
$$

and

$$
\begin{aligned}
\widehat{\nabla}^kY^\pi_t=&\sum_{j\ge k}\partial_j\widetilde{g}_{\widehat{\pi}}\nabla X_{s_j}+\int_t^T\Big[\partial_x f_\pi(r)\nabla X_r+\partial_y f_\pi(r)\widehat{\nabla}^kY^\pi_r+\partial_z f_\pi(r)\widehat{\nabla}^kZ^\pi_r\Big]dr\\
&-\Big\{\int_t^T\widehat{\nabla}^kZ^\pi_r dW_r\Big\}^T,\quad t\in[s_{k-1},T],\quad k=1,\cdots,m.
\end{aligned} \tag{3.50}
$$

Now note that $\partial_j \widetilde{g}_{\widehat{\pi}}(x_0, \cdots, x_m) = 0$ for $j \notin \{l_0, \cdots, l_n\}$. For any $[s_{k-1}, s_k) \subseteq [t_{i-1}, t_i)$ we have

$$\sum_{j \geq k} \partial_j \widetilde{g}_{\widehat{\pi}} \nabla X_{s_j} = \sum_{j \geq i} \partial_j g_\pi \nabla X_{t_j}.$$

Thus, by the uniqueness of the solution to BSDE (3.46) we have $\widehat{\nabla}^k Y^\pi_s = \nabla^i Y^\pi_s$, $\forall s \in [s_{k-1}, s_k) \subseteq [t_{i-1}, t_i)$. In other words, we have $\nabla^{\widehat{\pi}} Y^\pi_t = \nabla^\pi Y^\pi_t$, $t \in [0, T]$, and (3.49) becomes $V_{\widehat{\pi}}(\nabla^\pi Y^\pi) \leq C$. Since C is independent of $\widehat{\pi}$ and π, and both π and $\widehat{\pi}$ are arbitrarily chosen, we obtain that $\sup_{\pi \in \Pi} V(\nabla^\pi Y^\pi) \leq C$. Consequently, all $\nabla^\pi Y^\pi$'s are quasi-martingales; and the family $\{\nabla^\pi Y^\pi\}$ is tight, thanks to Lemma 1.2.11.

Note that so far we have not used (ii) of Assumption 3.0.2 yet. Now we shall use it to obtain a (strong) L^2 limit of $\{\nabla^\pi Y^\pi\}$. Denote $\widetilde{Z}^\pi = \nabla^\pi Y^\pi$ and $\widetilde{Z}_t = Z_t \sigma^{-1}(t, X_t) \nabla X_t$. Since $\widetilde{Z}^\pi$ satisfies (3.45) and $\nabla X \in L^p(\mathbf{F}; C([0, T]; \mathbb{R}^{d_1}))$ for all $p \geq 2$, using (3.44) and the Hölder inequality we have, for any $1 < q < 2$,

$$E\Big\{ \int_0^T |\nabla^\pi Y^\pi_t - \widetilde{Z}_t|^q dt \Big\} \leq E\Big\{ \int_0^T |(Z^\pi_t - Z_t) \sigma^{-1}(t, X_t) \nabla X_t|^q dt \Big\} \to 0, \qquad (3.51)$$

as $|\pi| \to 0$. Therefore, we can find a sequence $\{\pi_k\}$ such that outside an exceptional P-null set, for all $\omega \in \Omega$, one has $\int_0^T |\widetilde{Z}^{\pi_k}_t(\omega) - \widetilde{Z}_t(\omega)|^q dt \to 0$, as $k \to \infty$. Thus, as functions in $L^0([0, T])$, $\widetilde{Z}^{\pi_k}(\omega)$ converges to $\widetilde{Z}(\omega)$ in measure. Applying Lemma 1.2.10, we see that, as Ψ-valued random variables $\widetilde{Z}^{\pi_k}$ converges to $\widetilde{Z}$ in Meyer-Zheng pseudo-path topology, P-a.s., and hence convergence in law. Denote the law of $\widetilde{Z}^{\pi_k}$ by P^k, and that of $\widetilde{Z}$ by P^0.

On the other hand, since $\{\widetilde{Z}^{\pi_k}\}$ are quasimartingales with uniformly bounded conditional variations, by Lemma 1.2.11 we know that, possibly along a subsequence,P^k converges weakly to a probability law $P^* \in \mathcal{M}(\mathbb{D})$. Let $\widehat{P}^* \in \mathcal{M}(\Psi)$ be the extension of P^* in the sense of (1.19). The uniqueness of the weak limit then implies that $\widehat{P}^*(A) = P^0(A)$, $\forall A \in \mathcal{B}(\Psi)$. Since $\mathbb{D} \in \mathcal{B}(\Psi)$, from (1.19), the definition of P^*, and the equality above we see that

$$1 = P^*(\mathbb{D}) = \widehat{P}^*(\mathbb{D}) = P^0(\mathbb{D}) = P\{\widetilde{Z} \in \mathbb{D}\}.$$

In other words, $\widetilde{Z}$, whence Z, has paths in $\mathbb{D}$ almost surely. This proves the theorem. ■

Remark 3.2.3 Note that in the proof we required σ to be invertible only because we need a strong limit of $\widetilde{Z}^\pi$. If we assume that σ is Lipschitz with respect to t, then in light of Lemma 3.4.4 as we will see in §4, one can prove directly that $\{Z^\pi\}$ itself is tight under Meyer-Zheng pseudo-path topology. So the result of Theorem 3.2.2 still holds true if we replace (ii) of Assumption 3.0.2 by (ii'): σ is uniformly Lipschitz with respect to time t. ■

The following corollary is an extension of Corollary 2.2.3.

Corollary 3.2.4 Assume (i) of Assumption 3.0.2 holds and that Φ satisfies the L^∞-Lipschitz condition (1.3). Then for $\forall p \geq 1$, there exists a constant C_p depending only on T, K and p, such that

$$\sup_{0\leq t\leq T} \|Z_t\|_p \leq C_p(1+|x|).$$

Proof. First applying Lemma 1.2.6 on (3.46) we have, for $\forall t \in [t_{i-1}, t_i)$,

$$E\Big\{|\nabla^i Y_t^\pi|^p\Big\} \leq C_p E\Big\{|\sum_{j\geq i} \partial_j g_\pi \nabla X_{t_j}|^p + \int_t^T |\partial_x f_\pi(r)\nabla X_r|^p dr\Big\}.$$

Note that g_π satisfies (3.15), recalling (2.23) we have

$$E\Big\{|\nabla^i Y_t^\pi|^p\Big\} \leq C_p(1+|x|^p)E\Big\{\sup_{0\leq r\leq T} |\nabla X_r|^p\Big\} \leq C_p(1+|x|^p),$$

which, combined with (3.45), implies that

$$\|Z_t^\pi\|_p \leq C_p(1+|x|); \qquad \forall t \in [0,T]. \tag{3.52}$$

By (3.44), we know that for dt-a.s. t, $Z_t^\pi \to Z_t$, P-a.s.. Applying the Fatou's lemma on (3.52) we get that

$$\|Z_t\|_p \leq C_p(1+|x|); \qquad dt-a.s.t \in [0,T]. \tag{3.53}$$

By virtue of Theorem 3.2.2, Z is a.s. càdlàg , then by applying the Fatou's lemma again (3.53) proves the result. ■

3.3 L^1-Lipschitz Functional Case

In this section we consider the FBSDE (1.2) where Φ is an L^1-Lipschitz functional. Note that an L^1-Lipschitz functional is always L^∞-Lipschitz. Therefore if (X, Y, Z) is the solution to (1.2) with Φ satisfying (1.4), then Z at least has a càdlàg version. The main purpose of this section is to prove the following stronger result.

Theorem 3.3.1 Assume Assumption 3.0.2 holds true; and that Φ satisfies the L^1-Lipschitz condition (1.4). Let (X, Y, Z) be the (unique) adapted solution of the FBSDE (1.2), then Z has a continuous version.

The proof of Theorem 3.3.1 is quite lengthy, we shall split it into several lemmas. We begin with some preparations. Let $\Pi = \{\pi\}$ be a family of partitions of $[0, T]$. For a given partition π, let (Y^π, Z^π) be the solution to the BSDE (3.43). Recall the process $\widetilde{Z}_t \stackrel{\triangle}{=} Z_t \sigma^{-1}(t, X_t) \nabla X_t$. Since Z has a continuous version if and only if $\widetilde{Z}$ does, it would suffice to prove that $\widetilde{Z}$ has a continuous version, which we shall do in the sequel.

Let us first give a lemma which is a refinement of Lemma 3.2.1, under the condition (1.4). Recall the mappings φ_π (or $\mathbf{x}_\pi$) and ψ_π defined by (3.32) and (3.37), respectively, for a given partition $\pi \in \Pi$.

Lemma 3.3.2 Assume that $\Phi : C([0, T]; \mathbb{R}^{d_1}) \mapsto \mathbb{R}$ satisfies the L^1-Lipschitz condition (1.4). Then there exists a family of discrete functionals $\{g_\pi : \pi \in \Pi\}$ such that

(i) for each $\pi \in \Pi$, $g_\pi \in C_b^\infty(\mathbb{R}^{d_1(n+1)})$, where $n = \#\pi - 1$;

(ii) for each $\pi \in \Pi$, $0 \le s_1 < s_2 \le T$, and $\mathbf{x} \in C([0, T]; \mathbb{R}^{d_1})$, it holds that

$$\sum_{\substack{s_1 < t_j \le s_2 \\ t_j \in \pi}} |\partial_j g_\pi(\psi_\pi(\mathbf{x}_\pi))| \le 2K(|s_2 - s_1| + |\pi|); \tag{3.54}$$

where K is the constant in (1.4);

(iii) for any $\mathbf{x} \in C([0, T]; \mathbb{R}^{d_1})$, it holds that

$$\lim_{|\pi| \to 0} |g_\pi(\psi_\pi(\mathbf{x}_\pi)) - \Phi(\mathbf{x}_\pi)| = 0. \tag{3.55}$$

Proof. Let Φ and $\pi \in \Pi$ be given. We construct the family $\{g_\pi\}$ as the same as those in Lemma 3.2.1. That is, $G_\pi \stackrel{\triangle}{=} \Phi \circ \psi_\pi^{-1}$, and $g_\pi = G_\pi^{\varepsilon(\pi)}$, where $G_\pi^\varepsilon \in C_b^\infty(\mathbb{R}^{d(n+1)})$ is the molifier of G_π, and $\varepsilon(\pi)$ is chosen so that (3.55) holds.

Since the condition (1.4) implies (1.3), (i) and (iii) follow from Lemma 3.2.1, and we need only check (ii). To do this let $0 \le s_1 < s_2 \le T$, and $\mathbf{x}, \mathbf{y} \in C([0,T];\mathbb{R}^{d_1})$ be given. Assume that $s_1 \in [t_{j_1-1}, t_{j_1})$ and $s_2 \in [t_{j_2-1}, t_{j_2})$, for some $1 \le j_1 \le j_2 \le n$. Thus

$$|t_{j_2} - t_{j_1-1}| \le (|s_2 - s_1| + 2|\pi|) \le 2(|s_2 - s_1| + |\pi|). \tag{3.56}$$

Next, for each j we denote

$$\delta_j = \begin{cases} \operatorname{sgn}[\partial_j g_\pi(\psi_\pi(\mathbf{x}_\pi))] & j_1 \le j < j_2; \\ 0 & \text{otherwise}, \end{cases} \tag{3.57}$$

and $\delta_\pi \stackrel{\triangle}{=} (\delta_0, \cdots, \delta_n)$, $\bar{\delta}_\pi \stackrel{\triangle}{=} \psi_\pi^{-1}(\delta_\pi)$. Notice that both ψ_π (whence ψ_π^{-1}) and φ_π are linear mappings, and that $\bar{\delta}_\pi(s) = 0$, for $s \notin [t_{j_1-1}, t_{j_2}]$, then similar to (3.41) we have (with $\varepsilon = \varepsilon(\pi)$)

$$\begin{aligned} \sum_{s_1 < t_j \le s_2} &|\partial_j g_\pi(\psi_\pi(\mathbf{x}_\pi))| = \sum_{j=0}^n \partial_j g_\pi(\psi_\pi(\mathbf{x}_\pi))\delta_j \\ =\ & \lim_{h\to 0} \frac{1}{h}\left(g_\pi(\psi_\pi(\mathbf{x}_\pi) + h\delta_\pi) - g_\pi(\psi_\pi(\mathbf{x}_\pi))\right) \\ =\ & \lim_{h\to 0} \frac{1}{h}\int_{\mathbb{R}^{d_1(n+1)}} \left[\Phi(\mathbf{x}_\pi - \varepsilon\psi_\pi^{-1}(z) + h\bar{\delta}_\pi) - \Phi(\mathbf{x}_\pi - \varepsilon\psi_\pi^{-1}(z))\right]\phi(z)dz \\ \le\ & \varliminf_{h\to 0} \frac{1}{h}\int_{\mathbb{R}^{d_1(n+1)}} K\int_0^T \left[h|\bar{\delta}_\pi(s)|\right] ds\phi(z)dz \\ =\ & \int_{\mathbb{R}^{d_1(n+1)}} K\int_{t_{j_1-1}}^{t_{j_2}} |\psi_\pi^{-1}(\delta_\pi)(s)| ds\phi(z)dz \le K\max_j |\delta_j||t_{j_2} - t_{j_1-1}| \\ \le\ & 2K(|s_2 - s_1| + |\pi|), \end{aligned}$$

thanks to (3.56). This proves (ii), whence the lemma. ∎

We now take a closer look at the process $\widetilde{Z}^\pi = \nabla^\pi Y^\pi$. Let us introduce some notations similar to those used in §2. Define

$$\begin{cases} \gamma_s^{\pi,0} = \displaystyle\int_s^T [\partial_x f_\pi(r)\nabla X_r + \partial_y f_\pi(r)\gamma_r^{\pi,0} + \partial_z f_\pi \zeta_r^{\pi,0}]dr - \Big\{\int_s^T \zeta_r^{\pi,0} dW_r\Big\}^T; \\ \gamma_s^{\pi,i} = \partial_{x_i} g_\pi \nabla X_{t_i} + \displaystyle\int_s^T [\partial_y f_\pi(r)\gamma_r^{\pi,i} + \partial_z f_\pi(r)\zeta_r^{\pi,i}]dr - \Big\{\int_s^T \zeta_r^{\pi,i} dW_r\Big\}^T. \end{cases} \tag{3.58}$$

Then, using the linearity of the BSDE (3.1) we see that $\nabla^i Y^\pi$ can be written as

$$\nabla^i Y^\pi_s = \gamma^{\pi,0}_s + \sum_{j\ge i} \gamma^{\pi,j}_s, \qquad s\in[t_{i-1},T],\ i=1,\cdots,n. \tag{3.59}$$

Now let us recall the "exponentials" Λ and $\mathcal{E}$ defined by (3.4) and (3.5). We denote, for a given $\pi\in\Pi$, $\Lambda^\pi_t \triangleq \Lambda^0_t(-\partial_y f_\pi)$ and $M^\pi_t \triangleq \mathcal{E}^0_t(\partial_z f_\pi)$, $t\in[0,T]$. Since $f_\pi \in C^{0,1}_b([0,T]\times\mathbb{R}^{d_1}\times\mathbb{R}\times\mathbb{R}^d)$, we may assume without loss of generality that $(\partial_x f_\pi, \partial_y f_\pi, \partial_z f_\pi)$ is uniformly bounded. Thus by the Girsanov theorem, M^π is a martingale; and the process $W^\pi_t = W_t - \int_0^t [\partial_z f_\pi(r)]^T dr$, $t\in[0,T]$, is a Brownian motion on the new probability space $(\Omega,\mathcal{F},P^\pi)$, where $\frac{dP^\pi}{dP} = M^\pi_T$.

Now, using integration by parts we have

$$\begin{aligned}
[\Lambda^\pi_t]^{-1}\gamma^{\pi,0}_t &= \int_t^T [\Lambda^\pi_r]^{-1}\partial_x f_\pi(r)\nabla X_r dr - \Big\{\int_t^T [\Lambda^\pi_r]^{-1}\zeta^{\pi,0}_r dW_r\Big\}^T \\
&\triangleq \Gamma^{\pi,0}_T - \Gamma^{\pi,0}_t - \Big\{\int_t^T [\Lambda^\pi_r]^{-1}\zeta^{\pi,0} dW_r\Big\}^T, \quad t\in[0,T],
\end{aligned} \tag{3.60}$$

where

$$\Gamma^{\pi,0}_t \triangleq \int_0^t [\Lambda^\pi_r]^{-1}\partial_x f_\pi(r)\nabla X_r dr, \qquad t\in[0,T]. \tag{3.61}$$

Taking conditional expectation $E^\pi\{\cdot|\mathcal{F}_t\}$ on both sides of (3.60) and using the Bayes rule we have

$$[\Lambda^\pi_t]^{-1}\gamma^{\pi,0}_t = E^\pi\{\Gamma^{\pi,0}_T|\mathcal{F}_t\} - \Gamma^{\pi,0}_t = E\{M^\pi_T\Gamma^{\pi,0}_T|\mathcal{F}_t\}[M^\pi_t]^{-1} - \Gamma^{\pi,0}_t. \tag{3.62}$$

Similarly, for each i we have

$$[\Lambda^\pi_t]^{-1}\gamma^{\pi,i}_t = E^\pi\{[\Lambda^\pi_T]^{-1}\partial_i g_\pi \nabla X_{t_i}|\mathcal{F}_t\} = E\{M^\pi_T \Gamma^{\pi,i}_{t_i}|\mathcal{F}_t\}[M^\pi_t]^{-1}, \tag{3.63}$$

where

$$\Gamma^{\pi,i}_t \triangleq [\Lambda^\pi_T]^{-1}\partial_i g_\pi \nabla X_t. \tag{3.64}$$

Therefore (3.59) can be written as

$$\begin{aligned}
\nabla^i Y^\pi_t &= \Lambda^\pi_t[[\Lambda^\pi_t]^{-1}\gamma^{\pi,0}_t + \sum_{j\ge i}[\Lambda^\pi_t]^{-1}\gamma^{\pi,j}_t] \\
&= \Lambda^\pi_t\Big\{E\{M^\pi_T[\Gamma^{\pi,0}_T + \sum_{j\ge i}\Gamma^{\pi,j}_{t_j}]|\mathcal{F}_t\}[M^\pi_t]^{-1} - \Gamma^{\pi,0}_t\Big\} \\
&= E\{\Xi^\pi_i|\mathcal{F}_t\}\Lambda^\pi_t[M^\pi_t]^{-1} - \Lambda^\pi_t\Gamma^{\pi,0}_t, \quad t\in[t_{i-1},T],
\end{aligned} \tag{3.65}$$

where

$$\Xi_i^\pi \stackrel{\triangle}{=} M_T^\pi[\Gamma_T^{\pi,0} + \sum_{j\ge i}\Gamma_{t_j}^{\pi,j}], \quad i = 0, 1, \cdots, n.$$

Consequently, we have

$$\begin{aligned}\nabla^\pi Y_t^\pi &= E\Big\{\sum_{i=1}^n \Xi_i^\pi 1_{[t_{i-1},t_i)}(t)\Big|\mathcal{F}_t\Big\}\Lambda_t^\pi[M_t^\pi]^{-1} - \Lambda_t^\pi\Gamma_t^{\pi,0}\\ &= E\Big\{\Xi_t^\pi\Big|\mathcal{F}_t\Big\}\Lambda_t^\pi[M_t^\pi]^{-1} - \Lambda_t^\pi\Gamma_t^{\pi,0}, \quad t\in[0,T),\end{aligned} \tag{3.66}$$

where, for $t \in [0,T)$,

$$\begin{aligned}&\Xi_t^\pi \stackrel{\triangle}{=} \sum_{i=1}^n \Xi_i^\pi 1_{[t_{i-1},t_i)}(t) = M_T^\pi[\Gamma_T^{\pi,0} + \sum_{i=1}^n\sum_{j\ge i}\Gamma_{t_j}^{\pi,j}1_{[t_{i-1},t_i)}(t)]\\ &= M_T^\pi[\Gamma_T^{\pi,0} + \sum_{j=1}^n\sum_{i\le j}\Gamma_{t_j}^{\pi,j}1_{[t_{i-1},t_i)}(t)] = M_T^\pi[\Gamma_T^{\pi,0} + \sum_{j=1}^n\Gamma_{t_j}^{\pi,j}1_{[0,t_j)}(t)].\end{aligned}$$

For notational convenience we shall denote $\{\chi_\pi : \pi\in\Pi\}$ be a family of generic random variables that may depend on the partition π, such that for all $p\ge 2$,

$$\sup_{\pi\in\Pi} E|\chi_\pi|^p \le C_p, \tag{3.67}$$

for some constant $C_p > 0$. Note that all C, C_p, and χ are allowed to vary from line to line. Moreover, from now on we shall fix a sequence $\{\pi_n\}\subset\Pi$ such that $\lim_{n\to\infty}|\pi_n|\to 0$; and denote $\Psi^n \stackrel{\triangle}{=} \Psi^{\pi_n}$, where $\Psi = \Lambda, M, \widetilde{M}, \Xi$,...,etc. Furthermore, we denote $\widetilde{Z}^n = \nabla^{\pi_n}Y^{\pi_n}$, $f_n = f_{\pi_n}$, $\Gamma^{n,0} = \Gamma^{\pi_n,0}$, $\Gamma^{n,i} = \Gamma^{\pi_n,i}$, and $\chi_n = \chi_{\pi_n}$. We have the following lemma.

Lemma 3.3.3 There exists a family of positive random variables $\{\chi_n\}_{n\ge1}$ satisfying (3.67) such that for all stopping time $\bar\tau\in[0,T]$, it holds that $[M_{\bar\tau}^n]^{-1}\le\chi_n$ and $|\Xi_{\bar\tau}^n|\le\chi_n$, $n\ge 1$, P-a.s. Furthermore, for all $0\le s_1<s_2\le T$,

$$|[M_{s_1}^n]^{-1} - [M_{s_2}^n]^{-1}| + |\Xi_{s_1}^n - \Xi_{s_2}^n| \le \chi_n(|s_1-s_2|^{\frac13} + |\pi_n|), \quad n\ge 1. \tag{3.68}$$

Proof. First, note that (3.9) implies that $[M_{\bar\tau}^n]^{-1}\le\chi_n$. Second, for each n and any $p\ge 2$, by (3.61) and (3.64),

$$\begin{cases} E|\Gamma_T^{n,0}|^p \le C_pE\Big\{\displaystyle\int_0^T|[\Lambda_r^n]^{-1}\partial_xf_n(r)\nabla X_r|^pdr\Big\} \le C_pE\Big\{\sup_{0\le t\le T}|\nabla X_t|^p\Big\};\\ E\Big\{\displaystyle\sup_{t\in[0,T]}\Big|\sum_{j\ge1}\Gamma_{t_j^n}^{n,j}1_{[0,t_j^n)}(t)\Big|^p\Big\} \le C_pE\Big\{\sum_{j\ge1}|\partial_jg_\pi\nabla X_{t_j^n}|\Big\}^p \le C_pE\Big\{\sup_{t\in[0,T]}|\nabla X_t|^p\Big\},\end{cases} \tag{3.69}$$

thanks to the assumption (1.4) (whence (2.97)). Combining this with (3.9) we see that (3.67) yields $|\Xi^n_{\bar{\tau}}| \leq \chi_n$.

To estimate $|[M^n_{s_1}]^{-1} - [M^n_{s_2}]^{-1}|$ we recall from (3.24) that

$$[M^n_t]^{-1} = \widetilde{M}^n_t \Lambda^n_t(-|\partial_z f_n|^2),$$

where $\widetilde{M}^n_\cdot \stackrel{\triangle}{=} \mathcal{E}^0_\cdot(-\partial_z f_n)$ is a P-martingale, thanks to the boundedness of $\partial_z f_n$. Now define

$$\widetilde{M}^*_n = \sup_{0\leq r_1 < r_2 \leq T} \frac{|\widetilde{M}^n_{r_2} - \widetilde{M}^n_{r_1}|}{(r_2 - r_1)^{\frac{1}{3}}}.$$

Using (3.6) it is easy to show that the exponential martingale $\widetilde{M}^n$ satisfies, for any $p \geq 1$, that $E\{|\widetilde{M}^n_{r_2} - \widetilde{M}^n_{r_1}|^{2p}\} \leq C|r_2 - r_1|^p$. Therefore, applying Theorem 1.2.1 of [47] one shows that $\widetilde{M}^*_n \in L^p(\Omega)$ for all $p \geq 1$. Consequently,

$$\begin{aligned}
&|[M^n_{s_1}]^{-1} - [M^n_{s_2}]^{-1}| \\
&\leq |[\Lambda^n_{s_1}(-|\partial_z f_n|^2) - \Lambda^n_{s_2}(-|\partial_z f_n|^2)]\widetilde{M}^n_{s_1}| + |\Lambda^n_{s_2}(-|\partial_z f_n|^2)[\widetilde{M}^n_{s_1} - \widetilde{M}^n_{s_2}]| \\
&\leq \chi_n[|s_1 - s_2| + |s_1 - s_2|^{\frac{1}{3}}] \leq \chi_n |s_1 - s_2|^{\frac{1}{3}}.
\end{aligned}$$

Next, recalling the definitions (3.64) and (3.67), we apply Lemma 3.3.2 to get

$$\begin{aligned}
|\Xi^n_{s_1} - \Xi^n_{s_2}| &\leq |M^n_T| \sum_{j=1}^n |[\Lambda^n_T]^{-1}\partial_j g_n \nabla X_{t_j}| 1_{(s_1,s_2]}(t_j)| \\
&\leq \chi_n \sup_{0\leq t\leq T} |\nabla X_t|(|s_1 - s_2| + |\pi_n|) \leq \chi_n(|s_1 - s_2| + |\pi_n|).
\end{aligned}$$

Combining the above we derive (3.68). ■

Finally, we give a seemingly simple lemma to facilitate our argument in the proof of Theorem 3.3.1.

Lemma 3.3.4 Let $\{\xi_n\}_{n\geq 1}, \{\eta_n\}_{n\geq 1} \subset L^1(\Omega)$ be two sequences such that

(i) $|\xi_n| \leq \eta_n$, $\forall n$, P-a.s.;

(ii) $\lim_{n\to\infty} \xi_n = \xi$ and $\lim_{n\to\infty} \eta_n = \eta$, both weakly in $L^1(\Omega)$.

Then it holds P-almost surely that $|\xi| \leq \eta$.

Proof. Denote $D \stackrel{\triangle}{=} \{\omega : |\xi| - \eta > 0\}$ and $\rho \stackrel{\triangle}{=} \mathrm{sgn}\,\{\xi\}$. Then $\rho 1_D \in L^\infty(\Omega)$, and

$$\begin{aligned}&E\{|\xi|1_D\} = E\{\xi\rho 1_D\} = \lim_{n\to\infty} E\{\xi_n \rho 1_D\}\\&\le \varliminf_{n\to\infty} E\{|\xi_n|1_D\} \le \lim_{n\to\infty} E\{\eta_n 1_D\} = E\{\eta 1_D\}.\end{aligned}$$

That is, $E\{[|\xi| - \eta]1_D\} \le 0$. By definition of the set D we see that $P(D) = 0$ must hold, proving the lemma. ■

Proof of Theorem 3.3.1. As we pointed out before, we need only show that $\widetilde{Z}$ has a continuous version on $[0, T]$. Note that Z has a càdlàg version, so does $\widetilde{Z}$. We will take such a version of $\widetilde{Z}$ from now on.

We first prove that $\widetilde{Z}$ is a.s. continuous on $[0, T_1]$, for all $T_1 < T$. Since $\widetilde{Z}$ is already càdlàg , we need only show that for all stopping time $\tau \in (0, T_1]$, it holds that $\widetilde{Z}_{\tau-} = \widetilde{Z}_\tau$ (cf. [48] or [15]). To this end, we first recall that (3.51) implies that for all $1 < q < 2$,

$$\int_0^T E\left\{|\widetilde{Z}^n_r - \widetilde{Z}_r|^q\right\} dr \to 0, \qquad \text{as } n \to \infty.$$

thus for any stopping time $\bar\tau$ such that $0 < \bar\tau \le T_1$, a.s., we have

$$\begin{aligned}&E\Big\{\int_0^{T-T_1} |\widetilde{Z}^n_{\bar\tau+r} - \widetilde{Z}_{\bar\tau+r}|^q dr\Big\} = E\Big\{\int_{\bar\tau}^{T-(T_1-\bar\tau)} |\widetilde{Z}^n_r - \widetilde{Z}_r|^q dr\Big\}\\&\le E\Big\{\int_0^T |\widetilde{Z}^n_r - \widetilde{Z}_r|^q dr\Big\} \to 0, \qquad \text{as } n \to \infty.\end{aligned}$$

In other words, for a.e. $r \in [0, T - T_1]$, one has

$$E\Big\{|\widetilde{Z}^n_{\bar\tau+r} - \widetilde{Z}_{\bar\tau+r}|^q\Big\} \to 0, \qquad \text{as } n \to \infty. \tag{3.70}$$

Next, we note that $\mathbf{F}$ is a Brownian filtration, whence quasi-left continuous. Thus every stopping time $\tau > 0$ is accessible. To wit, there exists a sequence of stopping times $\{\tau_k\}$ such that $\tau_k < \tau$, and $\tau_k \uparrow \tau$, as $k \to \infty$. Now setting $\bar\tau = \tau_0 \stackrel{\triangle}{=} \tau$ and τ_k, $k = 1, 2, \cdots$, respectively. Taking away a countable union of null sets, we see that (3.70) should hold for τ_k, $k = 0, 1, \cdots$, for a.e. $r \in [0, T - T_1]$. Now let us choose $r_m \downarrow 0$ such that (3.70) holds for all k, m.

Since $\partial_y f_{\pi_n}$ and $\partial_z f_{\pi_n}$ are bounded, using definitions of Λ^n and $\Gamma^{n,0}$ one derives that, for all k and m,

$$\begin{cases} |\Gamma^{n,0}_{\tau_k+r_m} - \Gamma^{n,0}_{\tau+r_m}| \le C \sup_{0\le t\le T} |\nabla X_t||\tau-\tau_k|; \\ |\Lambda^n_{\tau_k+r_m} - \Lambda^n_{\tau+r_m}| \le C|\tau_k - \tau|. \end{cases}$$

Thus, denoting $\rho(\eta,t) \stackrel{\triangle}{=} \eta + E\{\eta|\mathcal{F}_t\}$, $(\eta,t)\in L^2(\Omega)\times[0,T]$, $\widetilde{\Xi}^n_s \stackrel{\triangle}{=} \Xi^n_s[\Lambda^n_s M^n_s]^{-1}$, and applying Lemma 3.3.3 we derive from (3.66) that

$$\begin{aligned} &\Big|[\widetilde{Z}^n_{\tau_k+r_m} - \widetilde{Z}^n_{\tau+r_m}] - [E\{\widetilde{\Xi}^n_{\tau+r_m}|\mathcal{F}_{\tau_k+r_m}\} - E\{\widetilde{\Xi}^n_{\tau+r_m}|\mathcal{F}_{\tau+r_m}\}]\Big| \\ &\le E\left\{|\widetilde{\Xi}^n_{\tau_k+r_m} - \widetilde{\Xi}^n_{\tau+r_m}|\Big|\mathcal{F}_{\tau_k+r_m}\right\} + |\Lambda^n_{\tau_k+r_m}\Gamma^{n,0}_{\tau_k+r_m} - \Lambda^n_{\tau+r_m}\Gamma^{n,0}_{\tau+r_m}| \\ &\le E\left\{|\Xi^n_{\tau_k+r_m} - \Xi^n_{\tau+r_m}|\Lambda^n_{\tau_k+r_m}[M^n_{\tau_k+r_m}]^{-1}\Big|\mathcal{F}_{\tau_k+r_m}\right\} \\ &+E\left\{|\Xi^n_{\tau+r_m}||\Lambda^n_{\tau_k+r_m} - \Lambda^n_{\tau+r_m}|[M^n_{\tau_k+r_m}]^{-1}\Big|\mathcal{F}_{\tau_k+r_m}\right\} \\ &+E\left\{|\Xi^n_{\tau+r_m}|\Lambda^n_{\tau+r_m}|[M^n_{\tau_k+r_m}]^{-1} - [M^n_{\tau+r_m}]^{-1}|\Big|\mathcal{F}_{\tau_k+r_m}\right\} \\ &+|\Lambda^n_{\tau_k+r_m}||\Gamma^{n,0}_{\tau_k+r_m} - \Gamma^{n,0}_{\tau+r_m}| + |\Lambda^n_{\tau_k+r_m} - \Lambda^n_{\tau+r_m}|\Big|\mathcal{F}^{n,0}_{\tau+r_m}\Big| \\ \le\ \ &\rho(\chi_n(|\tau_k-\tau|^{\frac{1}{3}} + |\pi_n|), \tau_k + r_m). \end{aligned} \tag{3.71}$$

To analyze (3.71) we observe that Lemma 3.3.3 implies that for any stopping time $\bar{\tau}\in(0,T]$, the sequence $\{\widetilde{\Xi}^n_{\bar{\tau}}\}_{n\ge1}$ is bounded (uniformly in $\bar{\tau}$) in $L^2(\Omega)$, thus it is *weakly* relatively compact in $L^2(\Omega)$, and so is in $L^1(\Omega)$. Consequently, possibly along a subsequence, may assume itself, it holds that

$$\begin{cases} \lim_{n\to\infty} \widetilde{\Xi}^n_{\tau+r_m} = \widetilde{\Xi}_m \in L^1(\Omega), & \text{weakly in } L^1(\Omega); \\ \lim_{m\to\infty} \widetilde{\Xi}_m = \widetilde{\Xi} \in L^1(\Omega), & \text{weakly in } L^1(\Omega). \end{cases} \tag{3.72}$$

An elementary calculation then shows that, for fixed k and m,

$$\begin{cases} \lim\limits_{n\to\infty} E\{\widetilde{\Xi}^n_{\tau+r_m}|\mathcal{F}_{\tau_k+r_m}\} = E\{\widetilde{\Xi}_m|\mathcal{F}_{\tau_k+r_m}\}, & \text{weakly in} L^1(\Omega); \\ \lim\limits_{n\to\infty} E\{\widetilde{\Xi}^n_{\tau+r_m}|\mathcal{F}_{\tau+r_m}\} = E\{\widetilde{\Xi}_m|\mathcal{F}_{\tau+r_m}\}, & \text{weakly in } L^1(\Omega). \end{cases} \tag{3.73}$$

Similarly, since by (3.67) $\{\chi_n\}$ is also bounded in $L^2(\Omega)$, we can also conclude that $\chi_n \to \chi \in L^2(\Omega)$, weakly in $L^1(\Omega)$ (!), as $n\to\infty$. Therefore,

$$\begin{cases} \lim\limits_{n\to\infty} \chi_n(|\tau_k-\tau|^{\frac{1}{3}} + |\pi_n|) = \chi|\tau_k-\tau|^{\frac{1}{3}}; \\ \lim\limits_{n\to\infty} E\{\chi_n(|\tau_k-\tau|^{\frac{1}{3}} + |\pi_n|)|\mathcal{F}_{\tau_k+r_m}\} = E\{\chi|\tau_k-\tau|^{\frac{1}{3}}|\mathcal{F}_{\tau_k+r_m}\}, \end{cases}$$

both weakly in $L^1(\Omega)$. Let us now denote

$$A^n_{k,m} \stackrel{\triangle}{=} [\widetilde Z^n_{\tau_k+r_m} - \widetilde Z^n_{\tau+r_m}] - [E\{\widetilde\Xi^n_{\tau+r_m}|\mathcal{F}_{\tau_k+r_m}\} - E\{\widetilde\Xi^n_{\tau+r_m}|\mathcal{F}_{\tau+r_m}\}];$$
$$B^n_{k,m} \stackrel{\triangle}{=} \rho(\chi_n(|\tau_k-\tau|^{\frac13}+|\pi_n|),\tau_k+r_m).$$

Then (3.71) shows that $|A^n_{k,m}| \le B^n_{k,m}$, P-a.s. Further, by (3.51) (or (3.70)) and (3.73) we see that, as $n\to\infty$,

$$\begin{cases} A^n_{k,m} \to A_{k,m} \stackrel{\triangle}{=} [\widetilde Z_{\tau_k+r_m} - \widetilde Z_{\tau+r_m}] - [E\{\widetilde\Xi_m|\mathcal{F}_{\tau_k+r_m}\} - E\{\widetilde\Xi_m|\mathcal{F}_{\tau+r_m}\}]; \\ B^n_{k,m} \to B_{k,m} \stackrel{\triangle}{=} \rho(\chi|\tau_k-\tau|^{\frac13},\tau_k+r_m), \end{cases} \tag{3.74}$$

both weakly in $L^1(\Omega)$. Applying Lemma 3.3.4 we obtain that

$$|[\widetilde Z_{\tau_k+r_m} - \widetilde Z_{\tau+r_m}] - [E\{\widetilde\Xi_m|\mathcal{F}_{\tau_k+r_m}\} - E\{\widetilde\Xi_m|\mathcal{F}_{\tau+r_m}\}]| \le \rho(\chi|\tau-\tau_k|^{\frac13},\tau_k+r_m), \tag{3.75}$$

To complete the proof we need to send $m\to\infty$ in (3.75) and apply Lemma 3.3.4 again. To this end, for any $\phi\in L^\infty(\Omega)$ we let $\phi_0 = E\{\phi|\mathcal{F}_\tau\}$ and $\phi_m = E\{\phi|\mathcal{F}_{\tau+r_m}\}$. Then using the right-continuity of the filtration $\mathbf{F}$ and the Dominated Convergence Theorem one has $\|\phi_m-\phi_0\|_{L^2(\Omega)}\to 0$, as $m\to\infty$. Note that $\{\widetilde\Xi_m\}$ is bounded in $L^2(\Omega)$ and converges weakly in $L^1(\Omega)$ (see (3.72)), we see that for any $\phi\in L^\infty(\Omega)$, it holds that

$$\begin{aligned} &\Big|E\Big\{[E\{\widetilde\Xi_m|\mathcal{F}_{\tau+r_m}\} - E\{\widetilde\Xi|\mathcal{F}_\tau\}]\phi\Big\}\Big| = |E\{\widetilde\Xi_m\phi_m - \widetilde\Xi\phi_0\}| \\ &\le |E\{[\widetilde\Xi_m-\widetilde\Xi]\phi_0\}| + |E\{\widetilde\Xi_m[\phi_m-\phi_0]\}| \\ &\le |E\{[\widetilde\Xi_m-\widetilde\Xi]\phi_0\}| + \|\widetilde\Xi_m\|_{L^2(\Omega)}\|\phi_m-\phi_0\|_{L^2(\Omega)} \to 0, \qquad \text{as } m\to\infty. \end{aligned}$$

That is, $E\{\widetilde\Xi_m|\mathcal{F}_{\tau+r_m}\}\to E\{\widetilde\Xi|\mathcal{F}_\tau\}$, weakly in $L^1(\Omega)$, as $m\to\infty$. Similarly, we have $E\{\widetilde\Xi_m|\mathcal{F}_{\tau_k+r_m}\}\to E\{\widetilde\Xi|\mathcal{F}_{\tau_k}\}$, and $E\{\chi|\tau-\tau_k|^{\frac13}|\mathcal{F}_{\tau_k+r_m}\}\to E\{\chi|\tau-\tau_k|^{\frac13}|\mathcal{F}_{\tau_k}\}$, weakly in $L^1(\Omega)$, as $m\to\infty$. Furthermore, we define for each integer $\ell\ge 1$ a set

$$\Omega_\ell \stackrel{\triangle}{=} \Big\{\omega\in\Omega : \sup_{0\le r\le r_1}[|\widetilde Z_{\tau_k+r} - \widetilde Z_{\tau_k}| + |\widetilde Z_{\tau+r} - \widetilde Z_\tau|] \le \ell\Big\},$$

where $r_1 \ge r_m \downarrow 0$, as $m\to\infty$. Then $\Omega_\ell\uparrow\Omega$, as $\ell\to\infty$, modulo a P-null set; and for each ℓ, Dominated Convergence Theorem yields that

$$1_{\Omega_\ell}A_{k,m} \to 1_{\Omega_\ell}\Big\{[\widetilde Z_{\tau_k} - \widetilde Z_\tau] - [E\{\widetilde\Xi|\mathcal{F}_{\tau_k}\} - E\{\widetilde\Xi|\mathcal{F}_\tau\}]\Big\}, \qquad \text{weakly in } L^1(\Omega).$$

(see (3.74) for definition of $A_{k,m}$). Since (3.75) implies that $|1_{\Omega_\ell}A_{k,m}| \le \rho(\chi|\tau - \tau_k|^{\frac{1}{3}}, \tau_k + r_m)$, we can now send $m \to \infty$ in (3.75) and apply Lemma 3.3.4 again to get

$$\Big|1_{\Omega_\ell}\Big\{[\widetilde{Z}_{\tau_k} - \widetilde{Z}_\tau] - [E\{\widetilde{\Xi}|\mathcal{F}_{\tau_k}\} - E\{\widetilde{\Xi}|\mathcal{F}_\tau\}]\Big\}\Big| \le \rho(\chi|\tau_k - \tau|^{\frac{1}{3}}, \tau_k), \quad P\text{-a.s.} \tag{3.76}$$

Finally, first letting $\ell \to \infty$ and then taking expectation and letting $k \to \infty$ on both sides of (3.76), using the fact that $\mathbf{F}$ is *quasi-left continuous*, and applying Fatou's Lemma, we conclude that $E|\widetilde{Z}_{\tau-} - \widetilde{Z}_\tau| \le 0$. That is, $\widetilde{Z}_{\tau-} = \widetilde{Z}_\tau$, P-a.s. Since τ is arbitrary, $\widetilde{Z}$ (whence Z) is continuous on $[0, T_1]$, for all $T_1 < T$. That is, Z is continuous on $[0, T)$. Defining $Z_T = Z_{T-}$, we see that Z is continuous on $[0, T]$. The proof is complete. ∎

The following theorem is a direct consequence of Theorem 3.2.2 and Theorem 3.3.1.

Theorem 3.3.5 Assume Assumption 3.0.2 holds true; and for some $0 \le t_1 < t_2 \le T$, Φ satisfies that

$$|\Phi(x^1) - \Phi(x^2)| \le L\left(\int_{t_1}^{t_2} |x^1(t) - x^2(t)|dt + \sup_{t\in[0,T]\setminus(t_1,t_2)} |x^1(t) - x^2(t)|\right).$$

Then Z has a version that is càdlàg on $[0, T]$ and continuous in $[t_1, t_2)$.

Proof. By Theorem 3.2.2, Z is càdlàg . Restricting the stopping time τ in (t_1, t_2) and following the same argument as that of Theorem 3.3.1 one shows that Z is continuous in $[t_1, t_2)$. ∎

In particular, we have the following result proved in Theorem 2.2.4.

Corollary 3.3.6 Assume Assumption 3.0.2 holds true; and that $\Phi(X) = g(X_T)$, then Z is continuous on $[0, T]$.

Proof. By Theorem 3.3.5 we know that Z is càdlàg on $[0, T]$ and continuous in $[0, T)$. Letting $Z_T = Z_{T-}$, we see that Z is indeed continuous on $[0, T]$. ∎

3.4 L^2-Type Regularity

In this section we shall establish an L^2-type regularity result for solutions to FBSDE (1.2), which is the base for the convergence of a numerical scheme we will propose in next chapter. In the sequel we shall use the following assumption:

Assumption 3.4.1 The functions $b, \sigma, f \in C_L^{\frac{1}{2}}$. We use a common constant $K > 0$ to denote all the Lipschitz constants, and assume that

$$\sup_{0\le t\le T}\Big\{|b(t,0)| + |\sigma(t,0)| + |f(t,0,0,0)|\Big\} + |\Phi(\mathbf{0})| \le K.$$

Let $\pi_0 : 0 = t_0 < \cdots < t_n = T$ be any given partition of $[0,T]$, and denote (X,Y,Z) as the adapted solution to FBSDE (1.2). For $\forall t \in [t_{i-1}, t_i)$, we shall use $(X_{t_{i-1}}, Y_{t_{i-1}})$ to approximate (X_t, Y_t). However, we use $Z^{\pi_0,1}_{t_{i-1}}$, rather than $Z_{t_{i-1}}$ to approximate Z_t, where $Z^{\pi_0,1}_{t_{i-1}}$ is defined as follows.

$$Z^{\pi_0,1}_{t_{i-1}} \triangleq \frac{1}{\Delta t_i} E\Big\{\int_{t_{i-1}}^{t_i} Z_r dr \Big| \mathcal{F}_{t_{i-1}}\Big\}. \tag{3.77}$$

This $Z^{\pi_0,1}_{t_{i-1}}$ is special in the following sense.

Lemma 3.4.2 Let $\xi \in L^2(\mathbf{F})$, and for $0 \le s < t \le T$ define

$$\eta_0 \triangleq \frac{1}{t-s} E\Big\{\int_s^t \xi_r dr \Big| \mathcal{F}_s\Big\}.$$

Then $\eta_0 \in \mathcal{F}_{t_{i-1}}$, and for $\forall \eta \in \mathcal{F}_{t_{i-1}}$, it holds that

$$E\Big\{\int_s^t |\xi_r - \eta_0|^2 dr\Big\} \le E\Big\{\int_s^t |\xi_r - \eta|^2 dr\Big\}.$$

Proof. Denote $\bar\eta \triangleq \int_s^t \xi_r dr$. The Lemma is a direct consequence of the fact that

$$\mathrm{Var}\Big(\bar\eta \Big| \mathcal{F}_{t_{i-1}}\Big) \le E\{(\bar\eta - \eta)^2 | \mathcal{F}_s\},$$

for $\forall \eta \in L^2(\mathcal{F}_s)$. ∎

Our main result is

Theorem 3.4.3 Assume that Assumption 3.4.1 holds; and that Φ satisfies the L^∞ Lipschitz condition (1.3). Then we have the following estimate:

$$\max_{1\le i\le n}\sup_{t\in[t_{i-1},t_i)} E\Big\{|X_t-X_{t_{i-1}}|^2+|Y_t-Y_{t_{i-1}}|^2\Big\}+\sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i}|Z_t-Z^{\pi_0,1}_{t_{i-1}}|^2dt\Big\}\le C|\pi_0|, \tag{3.78}$$

where C is a constant depending only on T,K and x, but independent of the partition π_0.

To prove the theorem, we need a technical lemma.

Lemma 3.4.4 Let $\xi^j_t=\alpha_j+\int_0^t\eta^j_r dW_r$, $j=1,\cdots,n$. Assume that $\eta^j\in L^2(\mathbf{F})$, and $\Lambda\in L^2(\mathbf{F})$. Then

$$E\Big\{\sum_{i=1}^n\int_{t_{i-1}}^{t_i}|\sum_{j\ge i}\eta^j_r|^2dr\Lambda_{t_{i-1}}\Big\}\le 2E\Big\{\Big(\sup_{0\le t\le T}\sum_{j=1}^n|\xi^j_t|\Big)^2\Lambda^*_T\Big\},$$

where $\Lambda^*_t\triangleq\sup_{0\le s\le t}|\Lambda_s|$.

Proof. By Itô's formula we have

$$E\Big\{\int_{t_{i-1}}^{t_i}\eta^{j_1}_r\eta^{j_2}_r dr\Big|\mathcal{F}_{t_{i-1}}\Big\}=E\Big\{\xi^{j_1}_{t_i}\xi^{j_2}_{t_i}-\xi^{j_1}_{t_{i-1}}\xi^{j_2}_{t_{i-1}}\Big|\mathcal{F}_{t_{i-1}}\Big\}.$$

Note that $\Lambda^*_{t_{i-1}}\in\mathcal{F}_{t_{i-1}}$, it holds obviously that

$$E\Big\{\int_{t_{i-1}}^{t_i}\eta^{j_1}_r\eta^{j_2}_r dr\Lambda^*_{t_{i-1}}\Big\}=E\Big\{(\xi^{j_1}_{t_i}\xi^{j_2}_{t_i}-\xi^{j_1}_{t_{i-1}}\xi^{j_2}_{t_{i-1}})\Lambda^*_{t_{i-1}}\Big\}.$$

Since Λ^*_t is increasing, by some simple calculation and applying the Abel transformation one can show that

$$\begin{aligned}
&E\Big\{\sum_{i=1}^n\int_{t_{i-1}}^{t_i}|\sum_{j\ge i}\eta^j_r|^2dr\Lambda_{t_{i-1}}\Big\}\le E\Big\{\sum_{i=1}^n\int_{t_{i-1}}^{t_i}|\sum_{j\ge i}\eta^j_r|^2dr\Lambda^*_{t_{i-1}}\Big\}\\
&=E\Big\{\sum_{i=1}^n\sum_{j_1,j_2\ge i}\int_{t_{i-1}}^{t_i}\eta^{j_1}_r\eta^{j_2}_r dr\Lambda^*_{t_{i-1}}\Big\}=E\Big\{\sum_{i=1}^n\sum_{j_1,j_2\ge i}(\xi^{j_1}_{t_i}\xi^{j_2}_{t_i}-\xi^{j_1}_{t_{i-1}}\xi^{j_2}_{t_{i-1}})\Lambda^*_{t_{i-1}}\Big\}\\
&=E\Big\{\sum_{j_1,j_2=1}^n\sum_{i\le j_1\wedge j_2}\xi^{j_1}_{t_i}\xi^{j_2}_{t_i}(\Lambda^*_{t_{i-1}}-\Lambda^*_{t_i})+\sum_{j_1,j_2=1}^n\xi^{j_1}_T\xi^{j_2}_T\Lambda^*_T-\sum_{j_1,j_2=1}^n\alpha_{j_1}\alpha_{j_2}\Lambda^*_0\Big\}\\
&\le E\Big\{\sum_{j_1,j_2=1}^n\sum_{i=1}^n|\xi^{j_1}_{t_i}\xi^{j_2}_{t_i}|(\Lambda^*_{t_i}-\Lambda^*_{t_{i-1}})+\sum_{j_1,j_2=1}^n|\xi^{j_1}_T\xi^{j_2}_T|\Lambda^*_T+\sum_{j_1,j_2=1}^n|\xi^{j_1}_0\xi^{j_2}_0|\Lambda^*_0\Big\}
\end{aligned}$$

$$= E\Big\{\sum_{i=1}^n(\sum_{j=1}^n|\xi^j_{t_i}|)^2(\Lambda^*_{t_i}-\Lambda^*_{t_{i-1}})+(\sum_{j=1}^n|\xi^j_T|)^2\Lambda^*_T+(\sum_{j=1}^n|\xi^j_0|)^2\Lambda^*_0\Big\}$$
$$\le E\Big\{\sup_{0\le t\le T}(\sum_{j=1}^n|\xi^j_{t_i}|)^2\sum_{i=1}^n(\Lambda^*_{t_i}-\Lambda^*_{t_{i-1}})+(\sum_{j=1}^n|\xi^j_T|)^2\Lambda^*_T+(\sum_{j=1}^n|\xi^j_0|)^2\Lambda^*_0\Big\}$$
$$\le 2E\Big\{\sup_{0\le t\le T}(\sum_{j=1}^n|\xi^j_t|)^2\Lambda^*_T\Big\}.$$

This proves the lemma. ■

Let $\xi^n=\xi$ and $\xi^j=0$ for $j=1,\cdots,n-1$. Then the following result is a direct consequence of Lemma 3.4.4.

Corollary 3.4.5 If $\xi=\alpha+\int_0^t\eta_r dW_r$, and $\eta,\Lambda\in L^2(\mathbf{F})$, then

$$E\Big\{\sum_{i=1}^n\int_{t_{i-1}}^{t_i}|\eta_r|^2dr\Lambda_{t_{i-1}}\Big\}\le 2E\Big\{\sup_{0\le t\le T}|\xi_t|^2\Lambda^*_T\Big\}.$$

Proof of Theorem. For $\forall i$ and $\forall t\in[t_{i-1},t_i)$, applying Lemma 1.2.5 we have $E\{|X_t-X_{t_{i-1}}|^2\}\le C\Delta t_i$. Recalling Corollary 3.2.4 and applying Lemma 1.2.6 we get $E\{|Y_t-Y_{t_{i-1}}|^2\}\le C\Delta t_i$.

The estimate for Z is a little involved. First we assume that $b,\sigma,f\in C^1$. Let $\pi:0=s_0<\cdots<s_m=T$ be any partition of $[0,T]$ finer than π_0, and without lose of generality, we assume $t_i=s_{l_i}$, for $i=1,\cdots,n$. Since Φ satisfies the L^∞-Lipschitz condition (1.3), by virtue of Lemma 3.2.1, one can find $g^\pi\in C^1(\mathbb{R}^{m+1})$ satisfying (3.15) and (3.38). Let (Y^π,Z^π) denote the adapted solution to the following BSDE:

$$Y^\pi_t=g^\pi(X_{s_0},\cdots,X_{s_m})+\int_t^T f(r,X_r,Y^\pi_r,X^\pi_r)dr-\int_t^T Z^\pi_r dW_r. \tag{3.79}$$

Applying the Dominated Convergence Theorem one can easily show that

$$\lim_{|\pi|\to 0}E\Big\{\sup_{1\le j\le m}\sup_{t\in[s_{j-1},s_j)}|X_t-X_{s_{j-1}}|^2\Big\}=0. \tag{3.80}$$

(Actually as we will see in §4.2, the left side of (3.80) converges with a rate of convergence $|\pi|\log\frac{1}{|\pi|}$.) Now by (1.3) and (3.38), applying Theorem 1.2.7 we know that

$$\lim_{|\pi|\to 0}E\Big\{\sup_{0\le t\le T}|Y^\pi_t-Y_t|^2+\int_0^T|Z^\pi_t-Z_t|^2dt\Big\}=0. \tag{3.81}$$

Recalling (3.77) and applying Lemma 3.4.2 we have

$$\begin{aligned}&\sum_{i=1}^{n} E\Big\{\int_{t_{i-1}}^{t_i} |Z_t - Z^{\pi_0,1}_{t_{i-1}}|^2 dt\Big\} \le \sum_{i=1}^{n} E\Big\{\int_{t_{i-1}}^{t_i} |Z_t - Z^{\pi}_{t_{i-1}}|^2 dt\Big\} \\ &\le 2\sum_{i=1}^{n} E\Big\{\int_{t_{i-1}}^{t_i} [|Z_t - Z^{\pi}_t|^2 + |Z^{\pi}_t - Z^{\pi}_{t_{i-1}}|^2] dt\Big\}. \end{aligned} \tag{3.82}$$

By (3.81) and (3.82), to prove the theorem it suffices to show that

$$\sum_{i=1}^{n} E\Big\{\int_{t_{i-1}}^{t_i} |Z^{\pi}_t - Z^{\pi}_{t_{i-1}}|^2] dt\Big\} \le C|\pi_0|, \tag{3.83}$$

where C is independent of π or π_0.

To this end we apply Lemma 3.1.1 and recall the processes defined in the paragraphs before the lemma, by (3.45) we obtain

$$Z^{\pi}_t = \Big[(\xi^0_t + \sum_{j\ge i}\xi^j_t)M^{-1}_t - \int_0^t f_x(r)\nabla X_r \Lambda^{-1}_r dr\Big]\Lambda_t[\nabla X_t]^{-1}\sigma(t, X_t). \tag{3.84}$$

Therefore,

$$|Z^{\pi}_t - Z^{\pi}_{t_{i_0-1}}| \le I^1_t + I^2_t + I^3_t, \tag{3.85}$$

where

$$\begin{cases} I^1_t \triangleq \Big|[\xi^0_t + \sum_{j\ge i}\xi^j_t] - [\xi^0_{t_{i_0-1}} + \sum_{j\ge l_{i_0-1}+1}\xi^j_{t_{i_0-1}}]\Big|\times \\ \qquad \Big|M^{-1}_{t_{i_0-1}}\Lambda_{t_{i_0-1}}[\nabla X_{t_{i_0-1}}]^{-1}\sigma(t_{i_0-1}, X_{t_{i_0-1}})\Big| \\ I^2_t \triangleq \Big|\xi^0_t + \sum_{j\ge i}\xi^j_t\Big|\Big|M^{-1}_t\Lambda_t[\nabla X_t]^{-1}\sigma(t, X_t) \\ \qquad -M^{-1}_{t_{i_0-1}}\Lambda_{t_{i_0-1}}[\nabla X_{t_{i_0-1}}]^{-1}\sigma(t_{i_0-1}, X_{t_{i_0-1}})\Big| \\ I^3_t \triangleq \Big|\int_0^t f_x(r)\nabla X_r\Lambda^{-1}_r dr\Lambda_t[\nabla X_t]^{-1}\sigma(t, X_t) \\ \qquad -\int_0^{t_{i_0-1}} f_x(r)\nabla X_r\Lambda^{-1}_r dr\Lambda_{t_{i_0-1}}[\nabla X_{t_{i_0-1}}]^{-1}\sigma(t_{i_0-1}, X_{t_{i_0-1}})\Big|. \end{cases} \tag{3.86}$$

Recalling (3.9) and applying Lemmas 1.2.5 and 1.2.6 one can easily show that

$$E\{|I^3_t|^2\} \le C|\pi_0|. \tag{3.87}$$

Recalling (3.13) and (3.15) we have

$$|\xi^0_t + \sum_{j\ge i}\xi^j_t| \le CE\Big\{\sup_{0\le t\le T}|\nabla X_t|\Big|\mathcal{F}_t\Big\}.$$

Thus by using Lemmas 1.2.5 and 1.2.6 one can similarly show that

$$E\{|I_t^2|^2\} \le C|\pi_0|. \tag{3.88}$$

It remains to estimate I_t^1. To this end we denote

$$\Gamma_t \stackrel{\triangle}{=} \sup_{0\le s\le t}\{1+|X_s|+|\nabla X_s|+|[\nabla X_s]^{-1}|+|M_s^{-1}|\}.$$

Noting that Λ is bounded and that $\Gamma_{t_{i_0-1}}\in\mathcal{F}_{t_{i_0-1}}$, by (3.13) we have

$$\begin{aligned}
E\{|I_t^1|^2\} &\le CE\Big\{\Gamma^6_{t_{i_0-1}}\Big|[\xi_t^0+\sum_{j\ge i}\xi_t^j]-[\xi^0_{t_{i_0-1}}+\sum_{j\ge i}\xi^j_{t_{i_0-1}}]\Big|^2\Big\}\\
&\le CE\Big\{\Gamma^6_{t_{i_0-1}}\Big[|\xi_t^0-\xi^0_{t_{i_0-1}}|^2+|\sum_{l_{i_0-1}+1\le j<i}\xi_t^j|^2+|\sum_{j\ge l_{i_0-1}+1}(\xi_t^j-\xi^j_{t_{i_0-1}})|^2\Big]\Big\}\\
&= CE\Big\{\Gamma^6_{t_{i_0-1}}|E\{\sum_{l_{i_0-1}+1\le j<i}\xi_T^j\Big|\mathcal{F}_t\}|^2\\
&\qquad+\Gamma^6_{t_{i_0-1}}E\Big\{|\xi_t^0-\xi^0_{t_{i_0-1}}|^2+|\sum_{j\ge l_{i_0-1}+1}(\xi_t^j-\xi^j_{t_{i_0-1}})|^2\Big|\mathcal{F}_{t_{i_0-1}}\Big\}\Big\} \qquad (3.89)\\
&\le CE\Big\{\Gamma^6_{t_{i_0-1}}\Big[|\sum_{l_{i_0-1}+1\le j<i}\xi_T^j|^2+\int_{t_{i_0-1}}^t|\eta_r^0|^2dr+\int_{t_{i_0-1}}^t|\sum_{j\ge l_{i_0-1}+1}\eta_r^j|^2dr\Big]\Big\}\\
&\le CE\Big\{\Gamma^6_{t_{i_0-1}}\Big[|\sum_{l_{i_0-1}+1\le j\le l_{i_0}}\xi_T^j|^2+\int_{t_{i_0-1}}^{t_{i_0}}|\eta_r^0|^2dr+\int_{t_{i_0-1}}^{t_{i_0}}|\sum_{j\ge l_{i_0-1}+1}\eta_r^j|^2dr\Big]\Big\}.
\end{aligned}$$

Applying Lemma 3.4.4 and Corollary 3.4.5, (3.89) leads to that

$$\begin{aligned}
\sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i}|I_t^1|^2dt\Big\} &\le \sum_{i=1}^n C\Delta t_iE\Big\{\Gamma_T^6(\sum_{l_{i-1}+1\le j\le l_i}|\xi_T^j|)^2\\
&\qquad+\Gamma^6_{t_{i-1}}\Big[\int_{t_{i-1}}^{t_i}|\eta_r^0|^2dr+\int_{t_{i-1}}^{t_i}|\sum_{k\ge i}(\sum_{l_{k-1}+1\le j\le l_k}\eta_r^j)|^2dr\Big]\Big\}\\
&\le C|\pi_0|E\Big\{\Gamma_T^6\Big[\sum_{i=1}^n(\sum_{l_{i-1}+1\le j\le l_i}|\xi_T^j|)^2 \qquad (3.90)\\
&\qquad+\sup_{0\le t\le T}|\xi_t^0|^2+\Big(\sup_{0\le t\le T}\sum_{k=1}^n|\sum_{l_{k-1}+1\le j\le l_k}\xi_t^j|\Big)^2\Big]\Big\}\\
&\le C|\pi_0|E\Big\{\Gamma_T^6\Big(\sup_{0\le t\le T}\sum_{j=0}^m|\xi_t^j|\Big)^2\Big\}\le C|\pi_0|E\Big\{\Gamma_T^{12}+\Big(\sup_{0\le t\le T}\sum_{j=0}^m|\xi_t^j|\Big)^4\Big\}.
\end{aligned}$$

Recalling (3.13), (3.10) and (3.8) we have

$$\sum_{j=0}^m|\xi_t^j|\le\sum_{j=0}^m E\Big\{|M_T\widetilde{\xi}^j|\Big|\mathcal{F}_t\Big\}$$

$$\leq E\Big\{|M_T\int_0^T f_x(r)\nabla X_r\Lambda_r^{-1}dr| + |M_T\Lambda_T^{-1}|\sum_{j=1}^m |g_j^\pi \nabla X_{s_j}|\Big|\mathcal{F}_t\Big\} \tag{3.91}$$
$$\leq CE\Big\{M_T\sup_{0\leq s\leq T}|\nabla X_s|\Big|\mathcal{F}_t\Big\},$$

where the last inequality is due to (3.15) and the fact that Λ^{-1} and f_x are uniformly bounded. Applying Doob's inequality (3.91) leads to

$$E\Big\{(\sup_{0\leq t\leq T}\sum_{j=0}^m |\xi_t^j|)^4\Big\} \leq CE\Big\{\sup_{0\leq t\leq T}\Big|E\Big\{M_T\sup_{0\leq s\leq T}|\nabla X_s|\Big|\mathcal{F}_t\Big\}\Big|^4\Big\}$$
$$\leq C\sup_{0\leq t\leq T}E\Big\{\Big|E\Big\{M_T\sup_{0\leq s\leq T}|\nabla X_s|\Big|\mathcal{F}_t\Big\}\Big|^4\Big\} \leq CE\Big\{M_T^4\sup_{0\leq t\leq T}|\nabla X_t|^4\Big\}. \tag{3.92}$$

Now by Lemmas 1.2.5, 1.2.6 and (3.9) we know

$$E\Big\{\Gamma_T^{12} + M_T^4\sup_{0\leq t\leq T}|\nabla X_t|^4\Big\} \leq C,$$

which, together with (3.90), implies that

$$\sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i}|I_t^1|^2dt\Big\} \leq C|\pi_0|. \tag{3.93}$$

Combining (3.93), (3.88) and (3.87), we infer (3.83) from (3.85). This, together with (3.82) and (3.38), leads to

$$\sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i}|Z_t - Z_{t_{i-1}}^{\pi_0,1}|^2dt\Big\} \leq C|\pi_0|,$$

which ends the proof for the smooth case.

In general case, let $b^\varepsilon, \sigma^\varepsilon$ and f^ε be molifiers of b, σ and f, respectively, and let $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon)$ be the solution triple to the corresponding FBSDE (1.2) modified in an obvious way. Then by the above arguments we have

$$\sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i}|Z_t^\varepsilon - Z_{t_{i-1}}^{\varepsilon,\pi_0,1}|^2dt\Big\} \leq C|\pi_0|. \tag{3.94}$$

Therefore, using Lemma 3.4.2 we have

$$\sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i}|Z_t - Z_{t_{i-1}}^{\pi_0,1}|^2dt\Big\} \leq \sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i}|Z_t - Z_{t_{i-1}}^{\varepsilon,\pi_0,1}|^2dt\Big\} \tag{3.95}$$
$$\leq 2\sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i}[|Z_t - Z_t^\varepsilon|^2 + |Z_t^\varepsilon - Z_{t_{i-1}}^{\varepsilon,\pi_0,1}|^2]dt\Big\} \leq CE\Big\{\int_0^T|Z_t - Z_t^\varepsilon|^2dt + |\pi_0|\Big\}.$$

Applying Theorem 1.2.7 we have

$$\lim_{\epsilon \to 0} E\Big\{ \int_0^T |Z_t^\varepsilon - Z_t|^2 dt \Big\} = 0,$$

which, combined with (3.95), proves the theorem. ∎

CHAPTER 4. NUMERICAL METHODS FOR BSDES

4.1 Introduction

In this chapter we shall turn to the numerical part for the following BSDE:

$$\begin{cases} X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \\ Y_t = \xi + \int_t^T f(s, X_s, Y_s, Z_s)ds - \int_t^T Z_s dW_s. \end{cases} \tag{4.1}$$

Some numerical methods for approximating solutions to BSDEs have already been developed. Based on a four step algorithm developed by Ma-Protter-Yong [35] to solve more general FBSDEs, Douglas-Ma-Protter built a numerical scheme by solving an associated (deterministic) PDE. The rate of convergence turns out to be as good as that for the solution of a simple SDE. However, their method requires high regularity on the coefficients ($C^{1+\alpha/2,2+\alpha}$), and the terminal value ξ of the BSDE must be in the form $g(X_T)$. By using a random time discretization Bally [3] presented a scheme which allows ξ to depend fully on the history of the driving Brownian motion, and requires only Lipschitz continuity on the coefficients. But his scheme actually calculates a sequence of functions, and thus requires a further approximation to give an actual implementation. Moreover, using this scheme one needs approximate integrals of dimension equal to the partition size, which is always very high. In his Ph.D. thesis [9], Chevance proposed a scheme which can be implemented in practice. But only Y is approximated in his scheme, and the regularity assumptions are very strong (C^4). Recently, Ma-Protter-San Martín-Torres [34] developed a method which allows ξ be to quite general and requires the coefficients to be only Lipschitz continuous. However, they got only weak convergence result.

In this chapter we shall propose a new numerical method for FBSDEs (4.1) where the terminal $\xi = \Phi(X)$ is an L^∞-Lipschitz functional. We first note that, for practical reasons, people need approximate the solutions of BSDEs by step processes (adapted and piecewise constant processes). One of the major difficulties to get such a good approximation is the regularity of the process Z. In fact, all the existing methods either require high regularity assumptions (e.g. [35], [9]) so that the process Z is "nice", or lack a good rate of convergence (e.g. [3], [34]) because Z is not "nice". Relying heavily on our new regularity results, Theorem 3.4.3, our scheme converges strongly in L^2, under mild conditions. It turns out that, if Φ is an L^∞-Lipschitz functional, then the asymptotic rate of convergence is $\sqrt{\frac{\log n}{n}}$, which is the best rate we can get in this case, and is new, to my best knowledge, in the literature of BSDEs. Moreover, if we assume that Φ is an L^1-Lipschitz functional, or is of the form $g(X_T)$, then the rate of convergence will be $\frac{1}{\sqrt{n}}$, which coincides with the result of [17].

Note that in our case the triple (X, Y, Z) is not Markovian. To actually compute the approximating step processes, we would encounter a "high dimension problem", as always seen in numerical schemes for BSDEs with path-dependent terminals (e.g. [3]). To avoid it, we assume further that Φ is *constructible*, which, in essence, adds a state variable and assumes that $(\Phi_\cdot(X), X, Y, Z)$ is Markovian, where $\Phi_t(X)$ is defined in an appropriate way so that $\Phi_T(X) = \Phi(X)$. Some similar idea was exploited by many others, see Björk [7] and Dai-Jiang [14], for example. We should point out here that in mathematical finance theory, most contingent claims (Asian, Lookback, etc.) are constructible functionals of the underlying assets price process. With this kind of Markovian property, we may try to calculate two deterministic functions u and v, as done in [35], such that $Y_t = u(t, \Phi_t(X), X_t)$ and $Z_t = v(t, \Phi_t(X), X_t)\sigma(t, X_t)$. It turns out that to obtain u and v we need only approximate integrals whose dimension is independent of the partition size.

The rest of the chapter is organized as follows. In §2 we review the Euler scheme for the forward diffusion X. In §3 we approximate (Y, Z) in a special case where the generator f is independent of z. In §4 we study the general case. In the last section we present explicitly our numerical scheme.

4.2 Euler Scheme for a Forward SDE Revisited

In this section, we briefly review the Euler scheme for the forward diffusion X. Let $\pi : 0 = t_0 < t_1 < \cdots < t_n = T$ be a partition of $[0, T]$. Define $\pi(t) \stackrel{\triangle}{=} t_{i-1}$, for $t \in [t_{i-1}, t_i)$. Let X^π be the solution of the following FSDE:

$$X_t^\pi = x + \int_0^t b(\pi(r), X_{\pi(r)}^\pi)dr + \int_0^t \sigma(\pi(r), X_{\pi(r)}^\pi)dW_r; \tag{4.2}$$

and $\hat{X}^\pi$ be the corresponding step process defined as follows:

$$\hat{X}_t^\pi \stackrel{\triangle}{=} X_{\pi(t)}^\pi. \tag{4.3}$$

By standard arguments we can easily show that:

Lemma 4.2.1 Assume that b and σ satisfy the conditions in Assumption 3.4.1. Then for X and X^π defined as in (4.1) and (4.2), respectively, there exists a constant C, depending only on T and K, such that:

$$E\{\sup_{0\le t\le T}[|X_t^\pi|^4\} \le C(1 + |x|^4); \quad E\{\sup_{0\le t\le T}[|X_t - X_t^\pi|^2\} \le C(1 + |x|^2)|\pi|.$$

Now we are ready for the estimates involving $\hat{X}^\pi$.

Theorem 4.2.2 Assume that b and σ satisfy the conditions in Assumption 3.4.1. Then then there exists a constant C, depending only on T and K, such that the following estimates hold:

$$\sup_{0\le t\le T} E\Big\{|X_t - \hat{X}_t^\pi|^2\Big\} \le C(1+|x|^2)|\pi|; \quad E\Big\{\sup_{0\le t\le T}|X_t - \hat{X}_t^\pi|^2\Big\} \le C(1+|x|^2)|\pi| \log\frac{1}{|\pi|}.$$

Proof. To simplify the presentation, without loss of generality we assume $d_1 = 1$.

First, for $t \in [t_{i-1}, t_i)$, we can rewrite (4.2) as

$$X_t^\pi = X_{t_{i-1}}^\pi + b(t_{i-1}, X_{t_{i-1}}^\pi)(t - t_{i-1}) + \sigma(t_{i-1}, X_{t_{i-1}}^\pi)(W_t - W_{t_{i-1}}). \tag{4.4}$$

Thus

$$\begin{aligned}
&E\Big\{|X_t - \hat{X}_t^\pi|^2\Big\} = E\Big\{|X_t - X_{t_{i-1}}^\pi|^2\Big\} \\
&\le 2E\Big\{|X_t - X_{t_{i-1}}|^2 + |X_{t_{i-1}} - X_{t_{i-1}}^\pi|^2\Big\} \le C(1 + |x|^2)|\pi|,
\end{aligned}$$

thanks to Lemmas 1.2.5 and 4.2.1.

Next, by (4.4) again we have

$$\begin{aligned} &\sup_{0\le t\le T}|X_t-\hat X_t^\pi| \le \sup_{0\le t\le T}|X_t-X_t^\pi| + \max_{1\le i\le n}\sup_{t_{i-1}\le t<t_i}|X_t^\pi - X_{t_{i-1}}^\pi| \\ &\le \sup_{0\le t\le T}|X_t-X_t^\pi| + \max_{1\le i\le n}|b(t_{i-1},X_{t_{i-1}}^\pi)|\Delta t_i \\ &\quad + \max_{1\le i\le n}|\sigma(t_{i-1},X_{t_{i-1}}^\pi)|\sup_{t_{i-1}\le t<t_i}|W_t-W_{t_{i-1}}|, \end{aligned} \tag{4.5}$$

Now using Lemmas 1.2.5 and 4.2.1, we infer from (4.5) that

$$\begin{aligned} &E\Big\{\sup_{0\le t\le T}|X_t-\hat X_t^\pi|^2\Big\} \le CE\Big\{\sup_{0\le t\le T}|X_t-X_t^\pi|^2 \\ &\quad +|\pi|^2\max_{1\le i\le n}|b(t_{i-1},X_{t_{i-1}}^\pi)|^2 + \max_{1\le i\le n}|\sigma(t_{i-1},X_{t_{i-1}}^\pi)|^2\max_{1\le i\le n}\sup_{t_{i-1}\le t<t_i}|W_t-W_{t_{i-1}}|^2\Big\} \\ &\le C(1+|x|^2)\Big[|\pi| + \sqrt{E\Big\{\max_{1\le i\le n}\sup_{t_{i-1}\le t<t_i}|W_t-W_{t_{i-1}}|^4\Big\}}\Big] \end{aligned} \tag{4.6}$$

Note that

$$\sup_{t_{i-1}\le t<t_i}|W_t-W_{t_{i-1}}| \le \sup_{t_{i-1}\le t<t_i}(W_t-W_{t_{i-1}}) + \sup_{t_{i-1}\le t<t_i}(W_{t_{i-1}}-W_t),$$

and that both $\sup_{t_{i-1}\le t<t_i}(W_t-W_{t_{i-1}})$ and $\sup_{t_{i-1}\le t<t_i}(W_{t_{i-1}}-W_t)$ are equal to $\sqrt{\Delta t_i}|W_1|$ in distribution (see [47], for example), then (4.6) leads to

$$E\Big\{\sup_{0\le t\le T}|X_t-\hat X_t^\pi|^2\Big\} \le C(1+|x|^2)\Big[|\pi| + \sqrt{E\Big\{\max_{1\le i\le n}(\Delta t_i)^2N_i^4\Big\}}\Big], \tag{4.7}$$

where $N_i \sim N(0,1)$ are i.i.d.. Denote $C_\varepsilon \triangleq 2\varepsilon\log\frac{1}{\varepsilon}$, and note that $\frac{C_{|\pi|}}{\Delta t_i} \ge 2\log\frac{1}{|\pi|}$, we have

$$\begin{aligned} &E\Big\{\max_{1\le i\le n}(\Delta t_i)^2N_i^4\Big\} \\ &= E\Big\{\max_{1\le i\le n}(\Delta t_i)^2N_i^4 1_{\{\max_i\Delta t_iN_i^2\le C_{|\pi|}\}}\Big\} + E\Big\{\max_{1\le i\le n}(\Delta t_i)^2N_i^4 1_{\{\max_i\Delta t_iN_i^2\ge C_{|\pi|}\}}\Big\} \\ &\le C_{|\pi|}^2 + \sum_{i=1}^n(\Delta t_i)^2E\Big\{N_i^4 1_{\{N_i^2\ge\frac{C_{|\pi|}}{\Delta t_i}\}}\Big\} \le C_{|\pi|}^2 + T|\pi|E\Big\{N^4 1_{\{N^2\ge 2\log\frac{1}{|\pi|}\}}\Big\}, \end{aligned} \tag{4.8}$$

where $N\sim N(0,1)$. By direct calculation we get, for $\forall a\ge 1$,

$$E\Big\{N^4 1_{\{N^2>a\}}\Big\} = \frac{1}{\sqrt{2\pi}}\int_{\sqrt a}^\infty y^4\exp(-\frac{y^2}{2})dy$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}}\Big[(\sqrt{a^3}+\sqrt{a})\exp(-\frac{a}{2}) + \int_{\sqrt{a}}^{\infty}\exp(-\frac{y^2}{2})dy\Big] \\
&\le C\Big[(\sqrt{a^3}+\sqrt{a})\exp(-\frac{a}{2}) + \frac{1}{\sqrt{a}}\int_{\sqrt{a}}^{\infty} y\exp(-\frac{y^2}{2})dy\Big] \qquad (4.9)\\
&= C[\sqrt{a^3}+\sqrt{a}+\frac{1}{\sqrt{a}}]\exp(-\frac{a}{2}) \le C\sqrt{a^3}\exp(-\frac{a}{2}),
\end{aligned}
$$

Take $a = 2\log\frac{1}{|\pi|}$ and assume π fine enough so that $a \ge 1$, and plug (4.9) into (4.8), we obtain that

$$E\Big\{\max_{0\le i\le n-1}(\Delta t_i)^2 N_i^4\Big\} \le C_{|\pi|}^2 + C|\pi|(\log\frac{1}{|\pi|})^{\frac{3}{2}}\exp(-\log\frac{1}{|\pi|}) \le C|\pi|^2(\log\frac{1}{|\pi|})^2,$$

which, together with (4.7), proves the theorem. ■

Remark 4.2.3 We should note here that $|\pi|\log\frac{1}{|\pi|}$ is the best asymptotic error in this case (see [2], Proposition 1, for example).

The following result is a direct consequence of Theorem 4.2.2.

Corollary 4.2.4 Assume all the conditions in Theorem 4.2.2 hold. If $\Phi : \mathbb{D} \mapsto \mathbb{R}$ satisfies the L^∞ Lipschitz condition (1.3), then

$$E\Big\{|\Phi(X) - \Phi(\hat{X}^\pi)|^2\Big\} \le C(1+|x|^2)|\pi|\log\frac{1}{|\pi|}.$$

Moreover, if Φ satisfies the L^1 Lipschitz condition (1.4), then

$$E\Big\{|\Phi(X) - \Phi(\hat{X}^\pi)|^2\Big\} \le C(1+|x|^2)|\pi|.$$

4.3 A Special Case

In this section, we consider a special case where f is independent of z, that is, the BSDE in (4.1) becomes

$$Y_t = \xi + \int_t^T f(r, X_r, Y_r)dr - \int_t^T Z_r dW_r. \qquad (4.10)$$

We propose the following discretization procedure. For any partition $\pi : 0 = t_0 < \cdots < t_n = T$, define the approximating pairs (Y^π, Z^π) inductively:

$$\begin{cases} Y_{t_n}^\pi = \xi^\pi; \\ Y_t^\pi = Y_{t_i}^\pi + f(t_i, X_{t_i}^\pi, Y_{t_i}^\pi)\Delta t_i - \displaystyle\int_t^{t_i} Z_r^\pi dW_r, \quad t \in [t_{i-1}, t_i), \end{cases} \qquad (4.11)$$

where $\xi^\pi \in L^2(\mathcal{F}_T)$. We should point out here that the family $\{(Y^\pi, Z^\pi)\}$ is *different* from that in (3.79).

Theorem 4.3.1 Assume that Assumption 3.4.1 holds, and that f is independent of z. Then the following estimate holds:

$$\sup_{0\le t\le T} E\{|Y_t - Y_t^\pi|^2\} + E\Big\{\int_0^T |Z_r - Z_r^\pi|^2 dr\Big\} \le C\Big[(1+|x|^2+E\{|\xi|^2\})|\pi| + E\{|\xi-\xi^\pi|^2\}\Big],$$

where C depends only on T and K.

Proof. As often done in BSDE theory, we shall apply the Gronwall Inequality to prove the estimate.

Denote

$$I_{i-1} \triangleq E\Big\{|Y_{t_{i-1}} - Y_{t_{i-1}}^\pi|^2 + \int_{t_{i-1}}^{t_i} |Z_r - Z_r^\pi|^2 dr\Big\}. \tag{4.12}$$

By (4.10) and (4.11) we have

$$Y_{t_{i-1}} - Y_{t_{i-1}}^\pi + \int_{t_{i-1}}^{t_i} (Z_r - Z_r^\pi) dW_r = Y_{t_i} + \int_{t_{i-1}}^{t_i} f(r, X_r, Y_r) dr - Y_{t_i}^\pi - f(t_i, X_{t_i}^\pi, Y_{t_i}^\pi)\Delta t_i.$$

Square both sides and take expectation, and note that $Y_{t_{i-1}} - Y_{t_{i-1}}^\pi$ is uncorrelated with $\int_{t_{i-1}}^{t_i} (Z_r - Z_r^\pi) dW_r$, we get

$$\begin{aligned}
I_{i-1} &= E\Big\{\Big[(Y_{t_i} - Y_{t_i}^\pi) + \int_{t_{i-1}}^{t_i} (f(r, X_r, Y_r) - f(t_i, X_{t_i}^\pi, Y_{t_i}^\pi)) dr\Big]^2\Big\} \\
&\le E\Big\{\Big[|Y_{t_i} - Y_{t_i}^\pi| + \int_{t_{i-1}}^{t_i} |f(r, X_r, Y_r) - f(t_i, X_{t_i}^\pi, Y_{t_i}^\pi)| dr\Big]^2\Big\} \\
&\le E\Big\{\Big[|Y_{t_i} - Y_{t_i}^\pi| + C\int_{t_{i-1}}^{t_i} \Big(\sqrt{t_i - r} + |X_r - X_{t_i}^\pi| + |Y_r - Y_{t_i}^\pi|\Big) dr\Big]^2\Big\} \qquad (4.13) \\
&\le E\Big\{\Big[|Y_{t_i} - Y_{t_i}^\pi| + C\int_{t_{i-1}}^{t_i} \Big(\sqrt{t_i - r} \\
&\qquad + [|X_r - X_{t_i}| + |X_{t_i} - X_{t_i}^\pi|] + [|Y_r - Y_{t_i}| + |Y_{t_i} - Y_{t_i}^\pi]\Big) dr\Big]^2\Big\} \\
&= E\Big\{\Big[(1 + C\Delta t_i)|Y_{t_i} - Y_{t_i}^\pi| \\
&\qquad + C\int_{t_{i-1}}^{t_i} \Big(\sqrt{t_i - r} + |X_r - X_{t_i}| + |X_{t_i} - X_{t_i}^\pi| + |Y_r - Y_{t_i}|\Big) dr\Big]^2\Big\}.
\end{aligned}$$

Note that, for $|\pi| \le 1$,

$$(a+b)^2 = a^2 + b^2 + 2ab \le a^2 + b^2 + \Delta t_i a^2 + \frac{1}{4\Delta t_i} b^2 \le (1 + \Delta t_i) a^2 + \frac{2}{\Delta t_i} b^2,$$

applying the Hölder Inequality (4.13) leads to

$$\begin{aligned}
I_{i-1} &\leq E\Big\{(1+C\Delta t_i)^2|Y_{t_i}-Y^\pi_{t_i}|^2 \\
&\quad +\frac{C}{\Delta t_i}\Big[\int_{t_{i-1}}^{t_i}\Big(\sqrt{t_i-r}+|X_r-X_{t_i}|+|X_{t_i}-X^\pi_{t_i}|+|Y_r-Y_{t_i}|\Big)dr\Big]^2\Big\} \\
&\leq (1+C\Delta t_i)E\{|Y_{t_i}-Y^\pi_{t_i}|^2\} \qquad (4.14)\\
&\quad +CE\Big\{\int_{t_{i-1}}^{t_i}\Big((t_i-r)+|X_{t_i}-X^\pi_{t_i}|^2+|X_r-X_{t_i}|^2+|Y_r-Y_{t_i}|^2\Big)dr\Big\} \\
&\leq (1+C\Delta t_i)E\{|Y_{t_i}-Y^\pi_{t_i}|^2\} \\
&\quad +C|\pi|\Big[(1+|x|^2+E\{|\xi|^2\})\Delta t_i+\int_{t_{i-1}}^{t_i}E\{|Z_r|^2\}dr\Big],
\end{aligned}$$

thanks to Lemmas 1.2.5, 1.2.6 and 4.2.1. Since $E\{|Y_{t_{i-1}}-Y^\pi_{t_{i-1}}|^2\}\leq I_{i-1}$, applying the Gronwall Inequality we deduce from (4.14) that

$$\max_{0\leq i\leq n}E\{|Y_{t_i}-Y^\pi_{t_i}|^2\}\leq CE\{|Y_T-Y^\pi_T|^2\}+C|\pi|\Big[1+|x|^2+E\{|\xi|^2\}+\int_0^T E|Z_r|^2dr\Big]. \qquad (4.15)$$

Since $Y_T=\xi$ and $Y^\pi_T=\xi^\pi$, applying Lemma 1.2.6 (4.15) implies that

$$\max_{0\leq i\leq n}E\{|Y_{t_i}-Y^\pi_{t_i}|^2\}\leq C\Big[(1+|x|^2+E|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big]. \qquad (4.16)$$

Then, for $t\in(t_{i-1},t_i)$, using (4.16) and Lemmas 1.2.5, 1.2.6 and 4.2.1 we get

$$\begin{aligned}
&E\{|Y_t-Y^\pi_t|^2\}\leq E\Big\{|Y_{t_i}+\int_t^{t_i}f(r,X_r,Y_r)dr-Y^\pi_{t_i}-f(t_i,X^\pi_{t_i},Y^\pi_{t_i})\Delta t_i|^2\Big\} \\
&\leq CE\Big\{|Y_{t_i}-Y^\pi_{t_i}|^2+|\Delta t_i|^2\sup_{0\leq r\leq T}|f(r,X_r,Y_r)|^2 \\
&\quad +|\Delta t_i|^2|f(t_i,X^\pi_{t_i},Y^\pi_{t_i})-f(t_i,X^\pi_{t_i},Y^\pi_{t_i})|^2\Big\} \\
&\leq CE\Big\{|Y_{t_i}-Y^\pi_{t_i}|^2+|\Delta t_i|^2\sup_{0\leq r\leq T}(1+|X_r|^2+|Y_r|^2)+|X_{t_i}-X^\pi_{t_i}|^2+|Y_{t_i}-Y^\pi_{t_i}|^2\Big\} \qquad (4.17)\\
&\leq C\Big[(1+|x|^2+E|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big].
\end{aligned}$$

Moreover, (4.14) and (4.16) imply that

$$\begin{aligned}
&E\Big\{|Y_{t_{i-1}}-Y^\pi_{t_{i-1}}|^2+\int_{t_{i-1}}^{t_i}|Z_r-Z^\pi_r|^2dr\Big\}\leq E\{|Y_{t_i}-Y^\pi_{t_i}|^2\} \\
&\quad +C\Delta t_i\Big[(1+|x|^2+E|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big]+C|\pi|\int_{t_{i-1}}^{t_i}E\{Z_r^2\}dr.
\end{aligned}$$

Summing over i from 1 to n we get

$$
\begin{aligned}
&E\Big\{\sum_{i=1}^{n}|Y_{t_{i-1}}-Y^\pi_{t_{i-1}}|^2+\int_0^T|Z_r-Z^\pi_r|^2dr\Big\}\\
&\le\sum_{i=1}^{n}E\{|Y_{t_i}-Y^\pi_{t_i}|^2\}+C|\pi|\Big[1+|x|^2+E|\xi|^2+\int_0^T E|Z_r|^2dr\Big]+CE\{|\xi-\xi^\pi|^2\}\\
&\le\sum_{i=1}^{n}E\{|Y_{t_i}-Y^\pi_{t_i}|^2\}+C\Big[(1+|x|^2+E|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big], \qquad (4.18)
\end{aligned}
$$

thanks to Lemma 1.2.6 again. Subtract both sides by $\sum_{i=1}^{n-1}E\{|Y_{t_i}-Y^\pi_{t_i}|^2\}$, (4.18) leads to

$$
\begin{aligned}
E\Big\{\int_0^T|Z_r-Z^\pi_r|^2dr\Big\} &\le E\{|Y_{t_n}-Y^\pi_{t_n}|^2\}+C\Big[(1+|x|^2+E|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big]\\
&\le C\Big[(1+|x|^2+E|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big],
\end{aligned}
$$

which, combined with (4.17), proves the theorem. ∎

Now let us define two step processes:

$$
\widehat{Y}^\pi_t\triangleq Y^\pi_{t_{i-1}};\qquad \widehat{Z}^\pi_t\triangleq\widehat{Z}^{\pi,1}_{t_{i-1}},\quad\text{for}\quad t\in[t_{i-1},t_i),
$$

where

$$
\widehat{Z}^{\pi,1}_{t_{i-1}}\triangleq\frac{1}{\Delta t_i}E\Big\{\int_{t_{i-1}}^{t_i}Z^\pi_r dr\Big|\mathcal{F}_{t_{i-1}}\Big\}. \qquad (4.19)
$$

Then we have the following theorem.

Theorem 4.3.2 Assume that all the conditions in Theorem 4.3.1 hold; and that $\xi=\Phi(X)$ where Φ satisfies the L^∞ Lipschitz condition (1.3). Then we have the following estimate:

$$
\sup_{0\le t\le T}E\{|Y_t-\widehat{Y}^\pi_t|^2\}\}+E\Big\{\int_0^T|Z_r-\widehat{Z}^\pi_r|^2dr\Big\}\le C\Big[(1+|x|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big].
$$

Proof. First, by Theorems 3.4.3 and 4.3.1 we have, for $\forall t\in[t_{i-1},t_i)$,

$$
\begin{aligned}
&E\{|Y_t-\widehat{Y}^\pi_t|^2\}\le 2E\Big\{|Y_t-Y_{t_{i-1}}|^2+|Y_{t_{i-1}}-Y^\pi_{t_{i-1}}|^2\Big\}\\
&\le C\Big[(1+|x|^2+E|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big]. \qquad (4.20)
\end{aligned}
$$

Now let us estimate the process $Z - \hat{Z}^\pi$. Recall (3.77) and apply Lemma 3.4.2, we get

$$\begin{aligned}
&E\Big\{\int_0^T |Z_t - \widehat{Z}^\pi_t|^2 dt\Big\} \le 2E\Big\{\int_0^T |Z_t - Z^\pi_t|^2 dt + \sum_{i=1}^n \int_{t_{i-1}}^{t_i} |Z^\pi_t - \widehat{Z}^{\pi,1}_{t_{i-1}}|^2 dt\Big\} \\
&\le 2E\Big\{\int_0^T |Z_t - Z^\pi_t|^2 dt + \sum_{i=1}^n \int_{t_{i-1}}^{t_i} |Z^\pi_t - Z^{\pi,1}_{t_{i-1}}|^2 dt\Big\} \\
&\le 2E\Big\{\int_0^T |Z_t - Z^\pi_t|^2 dt + \sum_{i=1}^n \int_{t_{i-1}}^{t_i} 2[|Z^\pi_t - Z_t|^2 + |Z_t - Z^{\pi,1}_{t_{i-1}}|^2] dt\Big\} \qquad (4.21) \\
&\le CE\Big\{\int_0^T |Z_t - Z^\pi_t|^2 dt + \sum_{i=1}^n \int_{t_{i-1}}^{t_i} |Z_t - Z^{\pi,1}_{t_{i-1}}|^2] dt\Big\} \\
&\le C\Big[(1 + |x|^2 + E|\xi|^2)|\pi| + E\{|\xi - \xi^\pi|^2\}\Big],
\end{aligned}$$

thanks to Theorems 4.3.1 and 3.4.3.

Finally, since Φ satisfies (1.3), applying Lemma 1.2.5 we have

$$\begin{aligned}
&E|\xi|^2 = E\{|\Phi(X)|^2\} \le 2E\{|\Phi(X) - \Phi(\mathbf{0})|^2 + |\Phi(\mathbf{0})|^2\} \\
&\le CE\{1 + \sup_{0\le t\le T} |X_t|^2\} \le C(1 + |x|^2), \qquad (4.22)
\end{aligned}$$

which, combined with (4.20) and (4.21), proves the theorem. ■

4.4 Two-Step Scheme for the General Case

In this section, we shall investigate the general case (4.1). In this case, the *one-step* discretization procedure (as that in §3) does not work. In fact, if we define, for $t \in [t_{i-1}, t_i)$,

$$Y^\pi_t = Y^\pi_{t_i} + f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^\pi_{t_i})\Delta t_i - \int_t^{t_i} Z^\pi_r dW_r,$$

where $\widetilde{Z}^\pi_{t_i}$ is some random variable determined by Z^π over $[t_i, T]$, then we will have trouble to estimate the term $E\{\int_{t_{i-1}}^{t_i} |Z_t - \widetilde{Z}^\pi_{t_i}|^2 dt\}$ involved when we try to estimate I_{i-1} as in (4.13), consequently we will not be able to apply the Gronwall Inequality. It turns out that the term $\widetilde{Z}^\pi_{t_i}$ needs to be $\mathcal{F}_{t_{i-1}}$-measurable, so that in our estimate it will contribute to I_{i-1}, rather than I_i. In light of this, we propose a *two-step* discretization procedure for the general case, which will enable us to apply the Gronwall Inequality again. To be more precise, for any partition $\pi : 0 = t_0 < \cdots < t_n = T$, we define two pairs of processes $(\widetilde{Y}^\pi, \widetilde{Z}^\pi)$ and (Y^π, Z^π) inductively:

$$\begin{cases} Y^\pi_{t_n} = \xi^\pi, \quad Z^\pi_{t_n} = 0; \\ \widetilde{Y}^\pi_t = Y^\pi_{t_i} + f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widehat{Z}^{\pi,1}_{t_i})\Delta t_i - \int_t^{t_i} \widetilde{Z}^\pi_r dW_r; \qquad \text{for} \quad t \in [t_{i-1}, t_i), \\ Y^\pi_t = Y^\pi_{t_i} + f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})\Delta t_i - \int_t^{t_i} Z^\pi_r dW_r; \end{cases} \tag{4.23}$$

where $\xi^\pi \in L^2(\mathcal{F}_T)$ and

$$\widehat{Z}^{\pi,1}_{t_n} \triangleq 0; \quad \widehat{Z}^{\pi,1}_{t_{i-1}} \triangleq \frac{1}{\Delta t_i} E\Big\{ \int_{t_{i-1}}^{t_i} Z^\pi_t dt \Big| \mathcal{F}_{t_{i-1}} \Big\}; \quad \widetilde{Z}^{\pi,1}_{t_{i-1}} \triangleq \frac{1}{\Delta t_i} E\Big\{ \int_{t_{i-1}}^{t_i} \widetilde{Z}^\pi_t dt \Big| \mathcal{F}_{t_{i-1}} \Big\}. \tag{4.24}$$

Note that in the last equation of (4.23), the term $\widetilde{Z}^\pi_{t_{i-1}}$ is $\mathcal{F}_{t_{i-1}}$-measurable. Our result is:

Theorem 4.4.1 Assume that Assumption 3.4.1 holds. Then it holds that

$$\begin{aligned} &\max_{0 \le i \le n} E\{|Y_{t_i} - Y^\pi_{t_i}|^2\}\} + E\Big\{ \int_0^T |Z_r - Z^\pi_r|^2 dr \Big\} \\ &\le C\Big[(1 + |x|^2 + E\{|\xi|^2\})|\pi| + E\{|\xi - \xi^\pi|^2\}\Big]. \end{aligned} \tag{4.25}$$

Proof. We explore some idea used to prove the well-posedness of BSDEs.

For any constant $\beta > 0$, by (4.1), (4.23), and applying the Itô's formula we have

$$\begin{aligned} d\Big[e^{\beta t}(Y_t - Y^\pi_t)^2\Big] &= \beta e^{\beta t}(Y_t - Y^\pi_t)^2 dt + e^{\beta t}(Z_t - Z^\pi_t)^2 dt \\ &\quad -2e^{\beta t}(Y_t - Y^\pi_t) f(t, \Theta_t) dt + 2e^{\beta t}(Y_t - Y^\pi_t)(Z_t - Z^\pi_t) dW_t. \end{aligned}$$

Integrate both sides over $[t_{i-1}, t_i)$, we get

$$\begin{aligned} &e^{\beta \Delta t_i}|Y_{t_i} - Y^\pi_{t_i} - f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})\Delta t_i|^2 - |Y_{t_{i-1}} - Y^\pi_{t_{i-1}}|^2 \\ &= \beta \int_{t_{i-1}}^{t_i} e^{\beta(t - t_{i-1})}(Y_t - Y^\pi_t)^2 dt + \int_{t_{i-1}}^{t_i} e^{\beta(t-t_{i-1})}(Z_t - Z^\pi_t)^2 dt \\ &-2 \int_{t_{i-1}}^{t_i} e^{\beta(t-t_{i-1})}(Y_t - Y^\pi_t) f(t, \Theta_t) dt + 2 \int_{t_{i-1}}^{t_i} e^{\beta(t-t_{i-1})}(Y_t - Y^\pi_t)(Z_t - Z^\pi_t) dW_t. \end{aligned}$$

Thus, by simply rearranging the terms we obtain

$$E\Big\{|Y_{t_{i-1}} - Y^\pi_{t_{i-1}}|^2 + \beta \int_{t_{i-1}}^{t_i} e^{\beta(t-t_{i-1})}(Y_t - Y^\pi_t)^2 dt + \int_{t_{i-1}}^{t_i} e^{\beta(t-t_{i-1})}(Z_t - Z^\pi_t)^2 dt\Big\}$$

$$
\begin{aligned}
&= E\Big\{e^{\beta\Delta t_i}|Y_{t_i} - Y^\pi_{t_i} - f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})\Delta t_i|^2 \\
&\quad +2\int_{t_{i-1}}^{t_i} e^{\beta(t-t_{i-1})}(Y_t - Y^\pi_t)f(t,\Theta_t)dt\Big\} \\
&= E\Big\{e^{\beta\Delta t_i}|Y_{t_i} - Y^\pi_{t_i}|^2 + e^{\beta\Delta t_i}|f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})|^2|\Delta t_i|^2 \qquad (4.26)\\
&-2e^{\beta\Delta t_i}(Y_{t_i} - Y^\pi_{t_i})f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})\Delta t_i + 2\int_{t_{i-1}}^{t_i} e^{\beta(t-t_{i-1})}(Y_t - Y^\pi_t)f(t,\Theta_t)dt\Big\} \\
&= E\Big\{e^{\beta\Delta t_i}|Y_{t_i} - Y^\pi_{t_i}|^2 + I_1 + I_2 + I_3\Big\},
\end{aligned}
$$

where

$$
\begin{cases}
I_1 \stackrel{\triangle}{=} e^{\beta\Delta t_i}|f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})|^2|\Delta t_i|^2; \\
I_2 \stackrel{\triangle}{=} 2\displaystyle\int_{t_{i-1}}^{t_i}[e^{\beta(t-t_{i-1})} - e^{\beta\Delta t_i}]dt(Y_{t_i} - Y^\pi_{t_i})f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}}); \\
I_3 \stackrel{\triangle}{=} 2\displaystyle\int_{t_{i-1}}^{t_i} e^{\beta(t-t_{i-1})}\Big[(Y_t - Y^\pi_t)f(t,\Theta_t) - (Y_{t_i} - Y^\pi_{t_i})f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})\Big]dt.
\end{cases}
\tag{4.27}
$$

We now estimate $I_1 - I_3$ separately. First, by using Lemmas 4.2.1, 1.2.5 and 1.2.6 we note that

$$
\begin{aligned}
&E\Big\{|f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})|^2\Big\} \le CE\Big\{1 + |X^\pi_{t_i}|^2 + |Y^\pi_{t_i}|^2 + |\widetilde{Z}^{\pi,1}_{t_{i-1}}|^2\Big\} \\
&\le CE\Big\{1 + |X^\pi_{t_i}|^2 + |Y_{t_i}|^2 + |Y_{t_i} - Y^\pi_{t_i}|^2 + |\widetilde{Z}^{\pi,1}_{t_{i-1}}|^2\Big\} \\
&\le CE\Big\{1 + |x|^2 + |\xi|^2 + |Y_{t_i} - Y^\pi_{t_i}|^2 + |\widetilde{Z}^{\pi,1}_{t_{i-1}}|^2\Big\}.
\end{aligned}
\tag{4.28}
$$

Thus we have

$$
E\{|I_1|\} \le C|\Delta t_i|^2 E\Big\{1 + |x|^2 + |\xi|^2 + |Y_{t_i} - Y^\pi_{t_i}|^2 + |\widetilde{Z}^{\pi,1}_{t_{i-1}}|^2\Big\}. \tag{4.29}
$$

Since $|e^{\beta(t-t_{i-1})} - e^{\beta\Delta t_i}| \le e^{\beta\Delta t_i}\Delta t_i$, we get

$$
\begin{aligned}
&E\{|I_2|\} \le CE\Big\{e^{\beta\Delta t_i}|\Delta t_i|^2|Y_{t_i} - Y^\pi_{t_i}||f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})|\Big\} \\
&\le Ce^{\beta\Delta t_i}|\Delta t_i|^2 E\Big\{|Y_{t_i} - Y^\pi_{t_i}|^2 + |f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})|^2\Big\} \\
&\le Ce^{\beta\Delta t_i}|\Delta t_i|^2 E\Big\{1 + |x|^2 + |\xi|^2 + |Y_{t_i} - Y^\pi_{t_i}|^2 + |\widetilde{Z}^{\pi,1}_{t_{i-1}}|^2\Big\},
\end{aligned}
\tag{4.30}
$$

thanks to (4.28).

It remains to estimate I_3, which is a little involved. Note that (4.27) leads to

$$
|I_3| \le Ce^{\beta\Delta t_i}\int_{t_{i-1}}^{t_i}\Big|(Y_t - Y^\pi_t) - (Y_{t_i} - Y^\pi_{t_i})\Big|\Big|f(t_i, X^\pi_{t_i}, Y^\pi_{t_i}, \widetilde{Z}^{\pi,1}_{t_{i-1}})\Big|dt
$$

$$+Ce^{\beta\Delta t_i}\int_{t_{i-1}}^{t_i}\Big|Y_t-Y_t^\pi\Big|\Big|f(t,\Theta_t)-f(t_i,X_{t_i}^\pi,Y_{t_i}^\pi,\widetilde{Z}_{t_{i-1}}^{\pi,1})\Big|dt \qquad (4.31)$$
$$= I_3^1+I_3^2,$$

where I_3^1 and I_3^2 are defined in an obvious way. Now let us estimate I_3^2 first. Since $f\in C_L^{\frac{1}{2}}$, we have

$$\begin{aligned}
&E\Big\{\Big|f(t,\Theta_t)-f(t_i,X_{t_i}^\pi,Y_{t_i}^\pi,\widetilde{Z}_{t_{i-1}}^{\pi,1})\Big|^2\Big\}\\
&\le CE\Big\{\Delta t_i+|X_t-X_{t_i}^\pi|^2+|Y_t-Y_{t_i}^\pi|^2+|Z_t-\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\Big\}\\
&\le CE\Big\{\Delta t_i+|X_t-X_{t_i}|^2+|X_{t_i}-X_{t_i}^\pi|^2 \qquad (4.32)\\
&\quad +|Y_t-Y_{t_i}|^2+|Y_{t_i}-Y_{t_i}^\pi|^2+|Z_t-\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\Big\}\\
&\le CE\Big\{(1+|x|^2+|\xi|^2)|\pi|+\int_{t_{i-1}}^{t_i}Z_r^2dr+|Y_{t_i}-Y_{t_i}^\pi|^2+|Z_t-\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\Big\},
\end{aligned}$$

thanks to Lemmas 1.2.5, 1.2.6 and 4.2.1. Recall (3.77) and apply Lemma 3.4.2, we have

$$\begin{aligned}
&E\Big\{\int_{t_{i-1}}^{t_i}|Z_t-\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2dt\Big\}\le CE\Big\{\int_{t_{i-1}}^{t_i}\Big[|Z_t-\widetilde{Z}_t^\pi|^2+|\widetilde{Z}_t^\pi-\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\Big]dt\Big\}\\
&\le CE\Big\{\int_{t_{i-1}}^{t_i}\Big[|Z_t-Z_t^\pi|^2+|Z_t^\pi-\widetilde{Z}_t^\pi|^2+|\widetilde{Z}_t^\pi-Z_{t_{i-1}}^{\pi,1}|^2\Big]dt\Big\} \qquad (4.33)\\
&\le CE\Big\{\int_{t_{i-1}}^{t_i}\Big[|Z_t-Z_t^\pi|^2+|Z_t^\pi-\widetilde{Z}_t^\pi|^2+|\widetilde{Z}_t^\pi-Z_t|^2+|Z_t-Z_{t_{i-1}}^{\pi,1}|^2\Big]dt\Big\}\\
&\le CE\Big\{\int_{t_{i-1}}^{t_i}\Big[|Z_t-Z_t^\pi|^2+|Z_t^\pi-\widetilde{Z}_t^\pi|^2+|Z_t-Z_{t_{i-1}}^{\pi,1}|^2\Big]dt\Big\}.
\end{aligned}$$

By (4.23) one can easily get

$$E\Big\{\int_{t_{i-1}}^{t_i}|Z_t^\pi-\widetilde{Z}_t^\pi|^2dt\Big\}\le CE\{|\widehat{Z}_{t_i}^{\pi,1}-\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\}\Delta t_i^2\le CE\{|\widehat{Z}_{t_i}^{\pi,1}|^2+|\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\}\Delta t_i^2. \quad (4.34)$$

Plug (4.34) into (4.33), we have

$$\begin{aligned}
&E\Big\{\int_{t_{i-1}}^{t_i}|Z_t-\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2dt\Big\}\\
&\le CE\Big\{|\Delta t_i|^2\Big[|\widehat{Z}_{t_i}^{\pi,1}|^2+|\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\Big]+\int_{t_{i-1}}^{t_i}\Big[|Z_t-Z_t^\pi|^2+|Z_t-Z_{t_{i-1}}^{\pi,1}|^2\Big]dt\Big\}. (4.35)
\end{aligned}$$

Then plug (4.35) into (4.32), we get

$$E\Big\{\int_{t_{i-1}}^{t_i}\Big|f(t,\Theta_t)-f(t_i,X_{t_i}^\pi,Y_{t_i}^\pi,\widetilde{Z}_{t_{i-1}}^{\pi,1})\Big|^2dt\Big\}$$

$$\leq CE\Big\{\Delta t_i\Big[(1+|x|^2+|\xi|^2)|\pi|+\int_{t_{i-1}}^{t_i} Z_r^2dr\Big]+\Delta t_i|Y_{t_i}-Y_{t_i}^\pi|^2 \tag{4.36}$$
$$+|\Delta t_i|^2\Big[|\widehat{Z}_{t_i}^{\pi,1}|^2+|\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\Big]+\int_{t_{i-1}}^{t_i}\Big[|Z_t-Z_t^\pi|^2+|Z_t-Z_{t_{i-1}}^{\pi,1}|^2\Big]dt\Big\}.$$

Note that for $\forall 0<\varepsilon<1, ab\leq \frac{1}{4\varepsilon}a^2+\varepsilon b^2$, we deduce from (4.30) and (4.36) that

$$E\{|I_3^2|\}\leq CE\Big\{\int_{t_{i-1}}^{t_i}\Big[e^{2\beta\Delta t_i}\varepsilon^{-1}|Y_t-Y_t^\pi|^2+\varepsilon\Big|f(t,\Theta_t)-f(t_i,X_{t_i}^\pi,Y_{t_i}^\pi,\widetilde{Z}_{t_{i-1}}^{\pi,1})\Big|^2\Big]dt\Big\}$$
$$\leq Ce^{2\beta\Delta t_i}\varepsilon^{-1}E\Big\{\int_{t_{i-1}}^{t_i}|Y_t-Y_t^\pi|^2dt\Big\} \tag{4.37}$$
$$+CE\Big\{\Delta t_i\Big[(1+|x|^2+|\xi|^2)|\pi|+\int_{t_{i-1}}^{t_i}Z_r^2dr\Big]+\int_{t_{i-1}}^{t_i}|Z_t-\bar{Z}_{t_{i-1}}^{\pi,1}|^2dt\Big\}$$
$$+CE\Big\{\Delta t_i|Y_{t_i}-Y_{t_i}^\pi|^2+|\Delta t_i|^2\Big[|\widehat{Z}_{t_i}^{\pi,1}|^2+|\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\Big]+\varepsilon\int_{t_{i-1}}^{t_i}|Z_t-Z_t^\pi|^2dt\Big\}.$$

We now estimate I_3^1. By (1.2) and (4.23) we have

$$|(Y_{t_i}-Y_{t_i}^\pi)-(Y_t-Y_t^\pi)|$$
$$\leq |\int_t^{t_i}f(r,\Theta_r)dr|+|\int_t^{t_i}(Z_r-Z_r^\pi)dW_r|+|f(t_i,X_{t_i}^\pi,Y_{t_i}^\pi,\widetilde{Z}_{t_{i-1}}^{\pi,1})|\Delta t_i$$
$$\leq \int_{t_{i-1}}^{t_i}\Big|f(r,\Theta_r)-f(t_i,X_{t_i}^\pi,Y_{t_i}^\pi,\widetilde{Z}_{t_{i-1}}^{\pi,1})\Big|dr$$
$$+|\int_t^{t_i}(Z_r-Z_r^\pi)dW_r|+2|f(t_i,X_{t_i}^\pi,Y_{t_i}^\pi,\widetilde{Z}_{t_{i-1}}^{\pi,1})|\Delta t_i.$$

Thus, using (4.36) and (4.28) we obtain

$$E\Big\{|(Y_{t_i}-Y_{t_i}^\pi)-(Y_t-Y_t^\pi)|^2\Big\}$$
$$\leq CE\Big\{\Delta t_i\int_{t_{i-1}}^{t_i}\Big|f(r,\Theta_r)-f(t_i,X_{t_i}^\pi,Y_{t_i}^\pi,\widetilde{Z}_{t_{i-1}}^{\pi,1})\Big|^2dr$$
$$+\int_{t_{i-1}}^{t_i}|Z_r-Z_r^\pi|^2dr+|f(t_i,X_{t_i}^\pi,Y_{t_i}^\pi,\widetilde{Z}_{t_{i-1}}^{\pi,1})|^2|\Delta t_i|^2\Big\}$$
$$\leq CE\Big\{|\Delta t_i|^2\Big[(1+|x|^2+|\xi|^2)|\pi|+\int_{t_{i-1}}^{t_i}Z_r^2dr\Big]+|\Delta t_i|^2|Y_{t_i}-Y_{t_i}^\pi|^2 \tag{4.38}$$
$$+|\Delta t_i|^3\Big[|\widehat{Z}_{t_i}^{\pi,1}|^2+|\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\Big]+\Delta t_i\int_{t_{i-1}}^{t_i}\Big[|Z_t-Z_t^\pi|^2+|Z_t-Z_{t_{i-1}}^{\pi,1}|^2\Big]dt$$
$$+\int_{t_{i-1}}^{t_i}|Z_r-Z_r^\pi|^2dr+|\Delta t_i|^2\Big[1+|x|^2+|\xi|^2+|Y_{t_i}-Y_{t_i}^\pi|^2+|\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\Big]\Big\}$$
$$\leq CE\Big\{\Delta t_i\Big[(1+|x|^2+|\xi|^2)|\pi|+\int_{t_{i-1}}^{t_i}Z_r^2dr\Big]+\Delta t_i|Y_{t_i}-Y_{t_i}^\pi|^2$$
$$+|\Delta t_i|^2\Big[|\widehat{Z}_{t_i}^{\pi,1}|^2+|\widetilde{Z}_{t_{i-1}}^{\pi,1}|^2\Big]+\int_{t_{i-1}}^{t_i}\Big[|Z_t-Z_t^\pi|^2+|Z_t-\bar{Z}_{t_{i-1}}^{\pi,1}|^2\Big]dt.$$

Therefore, note that $ab \le \frac{\varepsilon}{\Delta t_i}a^2 + \frac{\Delta t_i}{4\varepsilon}b^2$, (4.31), (4.38) and (4.28) imply that

$$
\begin{aligned}
E\{|I_3^1|\} &\le C\frac{\varepsilon}{\Delta t_i}\int_{t_{i-1}}^{t_i} E\Big\{|(Y_{t_i}-Y_{t_i}^\pi)-(Y_t-Y_t^\pi)|^2\Big\}dt \\
&\qquad +Ce^{2\beta\Delta t_i}|\Delta t_i|^2\varepsilon^{-1}E\Big\{|f(t_i,X_{t_i}^\pi,Y_{t_i}^\pi,\widetilde Z_{t_{i-1}}^{\pi,1})|^2\Big\} \\
&\le C\varepsilon E\Big\{\Delta t_i\Big[(1+|x|^2+|\xi|)|\pi|+\int_{t_{i-1}}^{t_i} Z_r^2dr\Big]+\int_{t_{i-1}}^{t_i}|Z_t-Z_{t_{i-1}}^{\pi,1}|^2dt \\
&\quad +\Delta t_i|Y_{t_i}-Y_{t_i}^\pi|^2+|\Delta t_i|^2\Big[|\widehat Z_{t_i}^{\pi,1}|^2+|\widetilde Z_{t_{i-1}}^{\pi,1}|^2\Big]+\int_{t_{i-1}}^{t_i}|Z_t-Z_t^\pi|^2dt\Big\} \qquad (4.39)\\
&\quad +e^{2\beta\Delta t_i}|\Delta t_i|^2\varepsilon^{-1}E\Big\{1+|x|^2+|\xi|^2+|Y_{t_i}-Y_{t_i}^\pi|^2+|\widetilde Z_{t_{i-1}}^{\pi,1}|^2\Big\} \\
&\le C\varepsilon E\Big\{\int_{t_{i-1}}^{t_i}|Z_t-Z_t^\pi|^2dt\Big\}+Ce^{2\beta\Delta t_i}\varepsilon^{-1}E\Big\{\Delta t_i\Big[(1+|x|^2+|\xi|^2)|\pi|+\int_{t_{i-1}}^{t_i} Z_r^2dr\Big] \\
&\quad +\int_{t_{i-1}}^{t_i}|Z_t-Z_{t_{i-1}}^{\pi,1}|^2dt+\Delta t_i|Y_{t_i}-Y_{t_i}^\pi|^2+|\Delta t_i|^2\Big[|\widehat Z_{t_i}^{\pi,1}|^2+|\widetilde Z_{t_{i-1}}^{\pi,1}|^2\Big]\Big\}.
\end{aligned}
$$

Combined with (4.39) and (4.37), (4.31) leads to

$$
\begin{aligned}
E\{I_3\} &\le C\varepsilon E\Big\{\int_{t_{i-1}}^{t_i}|Z_t-Z_t^\pi|^2dt\Big\} \\
&\quad +Ce^{2\beta\Delta t_i}\varepsilon^{-1}E\Big\{\Delta t_i\Big[(1+|x|^2+|\xi|^2)|\pi|+\int_{t_{i-1}}^{t_i} Z_r^2dr\Big]+\int_{t_{i-1}}^{t_i}|Z_t-Z_{t_{i-1}}^{\pi,1}|^2dt \qquad (4.40)\\
&\quad +\Delta t_i|Y_{t_i}-Y_{t_i}^\pi|^2+|\Delta t_i|^2\Big[|\widehat Z_{t_i}^{\pi,1}|^2+|\widetilde Z_{t_{i-1}}^{\pi,1}|^2\Big]+\int_{t_{i-1}}^{t_i}|Y_t-Y_t^\pi|^2dt\Big\}.
\end{aligned}
$$

Now plug (4.29), (4.30) and (4.40) into (4.26), we obtain

$$
\begin{aligned}
&E\Big\{|Y_{t_{i-1}}-Y_{t_{i-1}}^\pi|^2+\beta\int_{t_{i-1}}^{t_i}|Y_t-Y_t^\pi|^2dt+\int_{t_{i-1}}^{t_i}|Z_t-Z_t^\pi|^2dt\Big\} \\
&\le E\Big\{|Y_{t_{i-1}}-Y_{t_{i-1}}^\pi|^2+\beta\int_{t_{i-1}}^{t_i}e^{\beta(t-t_{i-1})}(Y_t-Y_t^\pi)^2dt+\int_{t_{i-1}}^{t_i}e^{\beta(t-t_{i-1})}(Z_t-Z_t^\pi)^2dt\Big\} \\
&\le e^{2\beta\Delta t_i}E\Big\{(1+C\varepsilon^{-1}\Delta t_i)|Y_{t_i}-Y_{t_i}^\pi|^2+C\varepsilon\int_{t_{i-1}}^{t_i}|Z_t-Z_t^\pi|^2dt \qquad (4.41)\\
&\quad +C\varepsilon^{-1}\Big[\int_{t_{i-1}}^{t_i}|Y_t-Y_t^\pi|^2dt+\Delta t_i\Big[(1+|x|^2+|\xi|^2)|\pi|+\int_{t_{i-1}}^{t_i} Z_r^2dr\Big] \\
&\quad +\int_{t_{i-1}}^{t_i}|Z_t-Z_{t_{i-1}}^{\pi,1}|^2dt+|\Delta t_i|^2\Big[|\widehat Z_{t_i}^{\pi,1}|^2+|\widetilde Z_{t_{i-1}}^{\pi,1}|^2\Big]\Big]\Big\}.
\end{aligned}
$$

Denote the constant C above as C_1. Choose $\varepsilon \stackrel{\triangle}{=} \frac{1}{2C_1e^2}$ (so that $C_1e^2\varepsilon=\frac12$), $\beta \stackrel{\triangle}{=} \frac{C_1e^2}{\varepsilon}$, and assume π fine enough so that $|\pi| \le \frac{1}{\beta} = \frac{1}{2C_1^2e^4}$, and note that $e^{2\beta\Delta t_i} \le 1 + C\Delta t_i$ for some constant C, then (4.41) leads to

$$
E\Big\{|Y_{t_{i-1}}-Y_{t_{i-1}}^\pi|^2+\frac12\int_{t_{i-1}}^{t_i}|Z_t-Z_t^\pi|^2dt\Big\}
$$

$$\leq E\Big\{(1+C\Delta t_i)|Y_{t_i}-Y^\pi_{t_i}|^2+C|\Delta t_i|^2\Big[|\widehat{Z}^{\pi,1}_{t_i}|^2+|\widetilde{Z}^{\pi,1}_{t_{i-1}}|^2\Big] \qquad (4.42)$$

$$C\Big[\Delta t_i\Big[(1+|x|^2+|\xi|^2)|\pi|+\int_{t_{i-1}}^{t_i} Z_r^2 dr\Big]+\int_{t_{i-1}}^{t_i}|Z_t-Z^{\pi,1}_{t_{i-1}}|^2 dt\Big]\Big\}.$$

Now we need estimate $|\Delta t_i|^2 E\{|\widehat{Z}^{\pi,1}_{t_i}|^2+|\widetilde{Z}^{\pi,1}_{t_{i-1}}|^2\}$. To this end we recall (4.24). It follows that

$$\begin{aligned}
&|\Delta t_i|^2 E\Big\{|\widehat{Z}^{\pi,1}_{t_i}|^2+|\widetilde{Z}^{\pi,1}_{t_{i-1}}|^2\Big\}\leq E\Big\{\Big|\int_{t_i}^{t_{i+1}} Z^\pi_t dt\Big|^2+\Big|\int_{t_{i-1}}^{t_i}\widetilde{Z}^\pi_t dt\Big|^2\Big\}\\
&\leq E\Big\{\Delta t_{i+1}\int_{t_i}^{t_{i+1}}|Z^\pi_t|^2 dt+\Delta t_i\int_{t_{i-1}}^{t_i}|\widetilde{Z}^\pi_t|^2 dt\Big\}\\
&\leq CE\Big\{\Delta t_{i+1}\int_{t_i}^{t_{i+1}}\Big[|Z_t-Z^\pi_t|^2+|Z_t|^2\Big]dt \qquad (4.43)\\
&\qquad +\Delta t_i\int_{t_{i-1}}^{t_i}\Big[|\widetilde{Z}^\pi_t-Z^\pi_t|^2+|Z^\pi_t-Z_t|^2+|Z_t|^2\Big]dt\Big\}\\
&\leq CE\Big\{(\Delta t_i\vee\Delta t_{i+1})\int_{t_{i-1}}^{t_{i+1}}\Big[|Z_t-Z^\pi_t|^2+|Z_t|^2\Big]dt+\Delta t_i\int_{t_{i-1}}^{t_i}|\widetilde{Z}^\pi_t-Z^\pi_t|^2\Big\}.
\end{aligned}$$

Denote the constant C in the above as C_2. Choose π fine enough so that $C_2|\pi|\leq\frac{1}{2}$. Then (4.34) and (4.43) imply that

$$E\Big\{\int_{t_{i-1}}^{t_i}|\widetilde{Z}^\pi_t-Z^\pi_t|^2\Big\}\leq C(\Delta t_i\vee\Delta t_{i+1})E\Big\{\int_{t_{i-1}}^{t_{i+1}}\Big[|Z_t-Z^\pi_t|^2+|Z_t|^2\Big]dt\Big\}.$$

Plugging this into (4.43) we have

$$|\Delta t_i|^2 E\Big\{|\widehat{Z}^{\pi,1}_{t_i}|^2+|\widetilde{Z}^{\pi,1}_{t_{i-1}}|^2\Big\}\leq C(\Delta t_i\vee\Delta t_{i+1})E\Big\{\int_{t_{i-1}}^{t_{i+1}}\Big[|Z_t-Z^\pi_t|^2+|Z_t|^2\Big]dt\Big\}.$$

Again, denote C_3 as the product of the constant C in the above and the C in (4.42), and assume π fine enough so that $C_3|\pi|\leq\frac{1}{4}$, (4.42) leads to

$$\begin{aligned}
&E\Big\{|Y_{t_{i-1}}-Y^\pi_{t_{i-1}}|^2+\frac{1}{4}\int_{t_{i-1}}^{t_i}|Z_t-Z^\pi_t|^2 dt\Big\}\\
&\leq E\Big\{(1+C\Delta t_i)|Y_{t_i}-Y^\pi_{t_i}|^2+C(\Delta t_i\vee\Delta t_{i+1})\int_{t_i}^{t_{i+1}}|Z_t-Z^\pi_t|^2 dt \qquad (4.44)\\
&+C|\pi|\Big[(1+|x|^2+|\xi|^2)\Delta t_i+\int_{t_{i-1}}^{t_{i+1}}|Z_r|^2 dr\Big]+C\int_{t_{i-1}}^{t_i}|Z_t-Z^{\pi,1}_{t_{i-1}}|^2 dt\Big\}.
\end{aligned}$$

At last we can apply the Gronwall inequality again. From (4.44) we conclude that, for $|\pi|\leq\min(\frac{1}{2C_1^2e^4},\frac{1}{2C_2},\frac{1}{4C_3})$,

$$E\Big\{|Y_{t_{i-1}}-Y^\pi_{t_{i-1}}|^2+\frac{1}{4}\int_{t_{i-1}}^{t_i}|Z_t-Z^\pi_t|^2 dt\Big\}$$

$$\leq CE\Big\{\Big[|\pi|(1+|x|^2+|\xi|^2+\int_0^T|Z_r|^2dr)+\sum_{i=1}^n\int_{t_{i-1}}^{t_i}|Z_t-Z_{t_{i-1}}^{\pi,1}|^2dt\Big]\Big\} \quad (4.45)$$
$$\leq C\Big[(1+|x|^2+E|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big]$$

thanks to Theorem 3.4.3 and (4.22).

Now using (4.45) and (4.44) we obtain that

$$E\Big\{|Y_{t_{i-1}}-Y_{t_{i-1}}^\pi|^2+\frac{1}{4}\int_{t_{i-1}}^{t_i}(Z_t-Z_t^\pi)^2dt\Big\}$$
$$\leq E\{|Y_{t_i}-Y_{t_i}^\pi|^2\}+C|\pi|\Big[(1+|x|^2+|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big] \quad (4.46)$$
$$+CE\Big\{|\pi|\Big[(1+|x|^2+|\xi|^2)\Delta t_i+\int_{t_{i-1}}^{t_{i+1}}Z_r^2dr\Big]+\int_{t_{i-1}}^{t_i}|Z_t-Z_{t_{i-1}}^{\pi,1}|^2dt\Big\}.$$

Summing (4.46) over i and apply Theorem 3.4.3 again, we can easily get

$$E\Big\{\int_0^T|Z_t-Z_t^\pi|^2dt\Big\}\leq C\Big[(1+|x|^2+|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big], \quad (4.47)$$

which, together with (4.45), proves the theorem. ■

Now let us define a step approximation of (Y,Z):

$$\widehat{Y}_{t_n}^\pi=Y_{t_n}^\pi;\quad \widehat{Z}_{t_n}^\pi=Z_{t_n}^\pi;\quad \widehat{Y}_t^\pi=Y_{t_{i-1}}^\pi;\quad \widehat{Z}_t^\pi=\widehat{Z}_{t_{i-1}}^{\pi,1}.$$

Then we have the following theorem.

Theorem 4.4.2 Assume that all the conditions in Theorem 4.4.1 hold; and that $\xi=\Phi(X)$ where Φ satisfies the L^∞ Lipschitz condition (1.3). Then

$$\sup_{0\leq t\leq T}E\{|Y_t-\widehat{Y}_t^\pi|^2\}+E\Big\{\int_0^T|Z_r-\widehat{Z}_r^\pi|^2dr\Big\}\leq C\Big[(1+|x|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big].$$

Proof. First, for $\forall t\in[t_{i-1},t_i)$, applying Theorem 4.4.1, Lemma 1.2.6, and Corollary 3.2.4, we have

$$E\{|Y_t-\widehat{Y}_t^\pi|^2\}=E\{|Y_t-Y_{t_{i-1}}^\pi|^2\}\leq 2E\{|Y_t-Y_{t_{i-1}}|^2+|Y_{t_{i-1}}-Y_{t_{i-1}}^\pi|^2\}$$
$$\leq C\Big[(1+|x|^2+|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big]. \quad (4.48)$$

On the other hand, using Lemma 3.4.2 we have

$$
\begin{aligned}
E\Big\{\int_0^T |Z_t-\widehat{Z}_t^\pi|^2 dt\Big\} &= \sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i} |Z_t-\widehat{Z}_{t_{i-1}}^{\pi,1}|^2 dt\Big\} \\
&\le C\sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i} \Big[|Z_t-Z_t^\pi|^2+|Z_t^\pi-\widehat{Z}_{t_{i-1}}^{\pi,1}|^2\Big]dt\Big\} \\
&\le C\sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i} \Big[|Z_t-Z_t^\pi|^2+|Z_t^\pi-Z_{t_{i-1}}^{\pi,1}|^2\Big]dt\Big\} \\
&\le C\sum_{i=1}^n E\Big\{\int_{t_{i-1}}^{t_i} \Big[|Z_t-Z_t^\pi|^2+|Z_t-Z_{t_{i-1}}^{\pi,1}|^2\Big]dt\Big\} \\
&\le C\Big[(1+|x|^2+|\xi|^2)|\pi|+E\{|\xi-\xi^\pi|^2\}\Big],
\end{aligned}
\tag{4.49}
$$

where the last inequality is due to Theorems 4.4.1 and 3.4.3. Now recall (4.22), (4.48) and (4.49) clearly prove the theorem. ■

4.5 Explicit Numerical Schemes

In this section we propose numerical schemes based on the results of previous sections. For $\xi = \Phi(X)$, naturally one would like to choose $\xi^\pi = \Phi(\hat{X}^\pi)$ in (4.11) and (4.23), where $\hat{X}^\pi$ is as defined in (4.3). Note that we have to compute the values $\Phi(\hat{X}^\pi)$ for all possible choices of the high dimensional vector $(X_{t_0}^\pi, \cdots, X_{t_n}^\pi)$. That is, if we partition the state space into M parts, then there are at least M^{n+1} values involved in the scheme, which is beyond our computational resources in most cases. Therefore, to make our scheme implementable, the following restriction on Φ is important.

Definition 4.5.1 A functional $\Phi : \mathbb{D}[0,T] \mapsto \mathbb{R}^k$, for some positive integer k, is called *constructible* with *construction* φ if there exist $\Phi_t : \mathbb{D}[0,t] \mapsto \mathbb{R}^k$ and $\varphi_{s,t} : \mathbb{R}^k \times \mathbb{D}[s,t] \mapsto \mathbb{R}^k$, for $0 \le s < t \le T$, such that:

(i) $\Phi_T = \Phi$, and $\Phi_t(\mathbf{x}) = \Phi_t(\mathbf{x}1_{[0,t)})$;

(ii) $\Phi_t(\mathbf{x}) = \varphi_{s,t}(\Phi_s(\mathbf{x}), \mathbf{x}1_{[s,t)})$;

(iii) All $\varphi_{s,t}$ satisfy the L^∞-Lipschitz condition (1.3) with uniformly Lipschitz condition.

Remark 4.5.2 Clearly Φ also satisfies the L^∞-Lipschitz condition (1.3).

Remark 4.5.3 In general $\{\Phi_t(X)\}_{0\le t\le T}$ is not Markovian. But since X is Markovian, by (ii) we can easily see that $\{(\Phi_t(X), X_t)\}_{0\le t\le T}$ is Markovian.

Remark 4.5.4 $\Phi(\mathbf{x}) = \displaystyle\int_0^T \mathbf{x}(t)dt$ is constructible. Actually we may define

$$\Phi_t(\mathbf{x}) \triangleq \int_0^t \mathbf{x}(r)dr; \quad \varphi_{s,t}(a,\mathbf{x}) \triangleq a + \int_s^t \mathbf{x}(r)dr.$$

$\Phi(\mathbf{x}) = \sup_{0\le t\le T} \mathbf{x}(t)$ is also constructible. We may define

$$\Phi_t(\mathbf{x}) = \sup_{0\le r\le t} \mathbf{x}(r); \quad \varphi_{s,t}(a,\mathbf{x}) \triangleq a \vee \sup_{s\le r\le t} \mathbf{x}(r).$$

Remark 4.5.5 In the sequel, we would abuse the notation φ in the sense that

$$\varphi_{s,t}(a,x) \triangleq \varphi_{s,t}(a, x1_{[s,t)}), \quad \forall (a,x) \in \mathbb{R}^k \times \mathbb{R}^{d_1}.$$

For $0 \le s < t \le T$, define

$$X_t^{s,x} \triangleq x + b(s,x)(t-s) + \sigma(s,x)(W_t - W_s). \tag{4.50}$$

Recalling (4.2) one can easily see that

$$X_{t_i}^\pi = X_{t_i}^{t_{i-1}, X_{t_{i-1}}^\pi}. \tag{4.51}$$

To facilitate our proof, we need the following lemma.

Lemma 4.5.6 Assume that g is a Lipschitz continuous function, and that $g(W_T) = E\{g(W_T)\} + \displaystyle\int_0^T \eta_t dW_t$. Then η_t is a martingale, and $\eta_0 = \frac{1}{T}E\{g(W_T)W_T\}$.

Proof. Since g is Lipschitz, by [42] we know there exists $\xi \in L^2(\mathcal{F}_T)$ such that $\eta_t = E\{\xi|\mathcal{F}_t\}$, thus η_t is a martingale. The formula for η_0 is due to Theorem 2.2.4 by considering X as W itself. ■

Now we are ready for the main theorems of this section.

Theorem 4.5.7 Let $\xi = g(\Phi(X), X_T)$ and $\xi^\pi = g(\Phi(\hat{X}^\pi), \hat{X}^\pi_T)$, and (Y, Z) and (Y^π, Z^π) are solutions to BSDEs (4.10) and (4.11), respectively. Assume that Assumption 3.4.1 holds, g is Lipschitz continuous and that Φ is constructible with construction φ. Define, for $\forall(a,x) \in \mathbb{R}^{k+d_1}$, where k is the dimension of the range of Φ,

$$\begin{cases} u^\pi_n(a,x) \triangleq g(a,x); \quad v^\pi_n(a,x) \triangleq 0; \\ U^\pi_i(a,x,\omega) \triangleq u^\pi_i(\varphi_{t_{i-1},t_i}(a,x), X^{t_{i-1},x}_{t_i}); \\ u^\pi_{i-1}(a,x) \triangleq E\Big\{U^\pi_i(a,x) + f(t_i, X^{t_{i-1},x}_{t_i}, U^\pi_i(a,x))\Delta t_i\Big\}; \\ v^\pi_{i-1}(a,x) \triangleq E\Big\{\frac{W_{t_i}-W_{t_{i-1}}}{\Delta t_i}\Big[U^\pi_i(a,x) + f(t_i, X^{t_{i-1},x}_{t_i}, U^\pi_i(a,x))\Delta t_i\Big]\Big\}. \end{cases} \tag{4.52}$$

Then we have

$$\widehat{Y}^\pi_{t_i} = Y^\pi_{t_i} = u^\pi_i(\Phi_{t_i}(\hat{X}^\pi), \hat{X}^\pi_{t_i}); \quad \widehat{Z}^\pi_{t_i} = Z^\pi_{t_i} = v^\pi_i(\Phi_{t_i}(\hat{X}^\pi), \hat{X}^\pi_{t_i}), \tag{4.53}$$

and

$$\sup_{0\le t\le T} E\{|Y_t - \widehat{Y}^\pi_t|^2\} + E\Big\{\int_0^T |Z_t - \widehat{Z}^\pi_t|^2 dt\Big\} \le C(1+|x|^2)|\pi| \log\frac{1}{|\pi|}. \tag{4.54}$$

Moreover, if Φ satisfies the L^1 Lipschitz condition (1.4), then we have

$$\sup_{0\le t\le T} E\{|Y_t - \widehat{Y}^\pi_t|^2\} + E\Big\{\int_0^T |Z_t - \widehat{Z}^\pi_t|^2 dt\Big\} \le C(1+|x|^2)|\pi|. \tag{4.55}$$

Remark 4.5.8 By (4.50) and (4.52), to calculate $(u^\pi_{i-1}, v^\pi_{i-1})$ from (u^π_i, v^π_i) we need only approximate d-dimensional integrals.

Proof of Theorem. By Theorem 4.3.1 and Theorem 4.2.2, it suffices to prove (4.53). We would also show simultaneously that u^π_i is Lipschitz. We proceed both by induction. For $i = n$ obviously (4.53) holds true and u^π_n is Lipschitz.

Assume that for i both results hold true. By (ii) of Definition 4.5.1 and by recalling Remark 4.5.5 we have

$$\Phi_{t_i}(\hat{X}^\pi) = \varphi_{t_{i-1},t_i}(\Phi_{t_{i-1}}(\hat{X}^\pi), \hat{X}^\pi_{t_{i-1}}),$$

Recall further (4.51), we get $U_i^\pi(\Phi_{t_{i-1}}(\hat{X}^\pi), \hat{X}^\pi_{t_{i-1}}) = Y^\pi_{t_i}$. Thus we may rewrite (4.11) as:

$$Y_t^\pi = U_i^\pi(\Phi_{t_{i-1}}(\hat{X}^\pi), \hat{X}^\pi_{t_{i-1}}) + f(t_i, X_{t_i}^{t_{i-1},x}, U_i^\pi(a,x))\Delta t_i - \int_t^{t_i} Z_r^\pi dW_r, \tag{4.56}$$

which obviously implies that

$$Y^\pi_{t_{i-1}} = u^\pi_{i-1}(\Phi_{t_{i-1}}(\hat{X}^\pi), \hat{X}^\pi_{t_{i-1}}), \tag{4.57}$$

thanks to (4.52). Moreover, since u_i^π is Lipschitz, Lemma 4.5.6 leads to that the process $\{Z_r^\pi\}_{t_{i-1}\le r<t_i}$ is a martingale. Recalling (4.19) we have

$$\widehat{Z}^\pi_{t_{i-1}} = \widehat{Z}^{\pi,1}_{t_{i-1}} = Z^\pi_{t_{i-1}}.$$

Now applying Lemma 4.5.6 again we obtain

$$Z^\pi_{t_{i-1}} = v^\pi_{i-1}(\Phi_{t_{i-1}}(\hat{X}^\pi), \hat{X}^\pi_{t_{i-1}}).$$

Finally, by (iii) of Definition 4.5.1, we can easily show that u^π_{i-1} is also Lipschitz continuous. This finishes the induction and thus proves the theorem. ∎

In general case, we have a similar result:

Theorem 4.5.9 Let $\xi = g(\Phi(X), X_T)$ and $\xi^\pi = g(\Phi(\hat{X}^\pi), \hat{X}^\pi_T)$, and (Y, Z) and (Y^π, Z^π) are solutions to BSDEs (1.2) and (4.23), respectively. Assume that Assumption 3.4.1 holds, g is Lipschitz continuous and that Φ is constructible with construction φ. Define, for $\forall (a,x) \in \mathbb{R}^{k+d_1}$,

$$\left\{\begin{array}{l}
u_n^\pi(a,x) \triangleq g(a,x); \quad v_n^\pi(a,x) \triangleq 0; \quad \widetilde{v}_n^\pi(a,x) \triangleq 0;\\
\widetilde{v}^\pi_{i-1}(a,x) \triangleq E\Big\{\dfrac{W_{t_i} - W_{t_{i-1}}}{\Delta t_i}\Big[u_i^\pi(\varphi_{t_{i-1},t_i}(a,x), X_{t_i}^{t_{i-1},x})\\
\quad +f(t_i, X_{t_i}^{t_{i-1},x}, u_i^\pi(\varphi_{t_{i-1},t_i}(a,x), X_{t_i}^{t_{i-1},x}), v_i^\pi(\varphi_{t_{i-1},t_i}(a,x), X_{t_i}^{t_{i-1},x}))\Delta t_i\Big]\Big\};\\
U_i^\pi(a,x,\omega) \triangleq u_i^\pi(\varphi_{t_{i-1},t_i}(a,x), X_{t_i}^{t_{i-1},x});\\
u^\pi_{i-1}(a,x) \triangleq E\{U_i^\pi(a,x) + f(t_i, X_{t_i}^{t_{i-1},x}, U_i^\pi(a,x), \widetilde{v}^\pi_{i-1}(a,x))\Delta t_i\};\\
v^\pi_{i-1}(a,x) \triangleq E\{\dfrac{W_{t_i} - W_{t_{i-1}}}{\Delta t_i}\Big[U_i^\pi(a,x) + f(t_i, X_{t_i}^{t_{i-1},x}, U_i^\pi(a,x), \widetilde{v}^\pi_{i-1}(a,x))\Delta t_i\Big]\}.
\end{array}\right. \tag{4.58}$$

Then we have

$$\widehat{Y}^{\pi}_{t_i} = Y^{\pi}_{t_i} = u^{\pi}_i(\Phi_{t_i}(\hat{X}^{\pi}), \hat{X}^{\pi}_{t_i}); \quad \widehat{Z}^{\pi}_{t_i} = Z^{\pi}_{t_i} = v^{\pi}_i(\Phi_{t_i}(\hat{X}^{\pi}), \hat{X}^{\pi}_{t_i}), \tag{4.59}$$

and

$$\sup_{0\le t\le T} E\{|Y_t - \widehat{Y}^{\pi}_t|^2\} + E\Big\{\int_0^T |Z_t - \widehat{Z}^{\pi}_t|^2 dt\Big\} \le C(1+|x|^2)|\pi|\log\frac{1}{|\pi|}. \tag{4.60}$$

Moreover, if Φ satisfies the L^1 Lipschitz condition (1.4), then we have

$$\sup_{0\le t\le T} E\{|Y_t - \widehat{Y}^{\pi}_t|^2\} + E\Big\{\int_0^T |Z_t - \widehat{Z}^{\pi}_t|^2 dt\Big\} \le C(1+|x|^2)|\pi|. \tag{4.61}$$

Proof. This proof is a line by line analogy of that for Theorem 4.5.7, except that here our induction assumption is that (4.59) holds true and that u^{π}_i, v^{π}_i and $\widetilde{v}^{\pi}_i$ are all Lipschitz continuous. ■

Remark 4.5.10 In Theorems 4.5.7 and 4.5.9, if we assume that $\xi = g(X_T)$, then we may consider Φ as a constant functional. In this case, $\varphi_{s,t}$ are also constant functionals, thus (4.52) becomes

$$\begin{cases} u^{\pi}_n(x) \triangleq g(x); \quad v^{\pi}_n(x) \triangleq 0; \\ u^{\pi}_{i-1}(x) = E\{u^{\pi}_i(X^{t_{i-1},x}_{t_i}) + f(t_i, X^{t_{i-1},x}_{t_i}, u^{\pi}_i(X^{t_{i-1},x}_{t_i}))\Delta t_i\}; \\ v^{\pi}_{i-1}(x) \triangleq E\{\frac{W_{t_i}-W_{t_{i-1}}}{\Delta t_i}\Big[u^{\pi}_i(X^{t_{i-1},x}_{t_i}) + f(t_i, X^{t_{i-1},x}_{t_i}, u^{\pi}_i(X^{t_{i-1},x}_{t_i}))\Delta t_i\Big]\}, \end{cases} \tag{4.62}$$

and (4.58) becomes

$$\begin{cases} u^{\pi}_n(x) \triangleq g(x); \quad v^{\pi}_n(x) \triangleq 0; \quad \widetilde{v}^{\pi}_n(x) \triangleq 0; \\ \widetilde{v}^{\pi}_{i-1}(x) \triangleq E\Big\{\dfrac{W_{t_i}-W_{t_{i-1}}}{\Delta t_i}\Big[u^{\pi}_i(X^{t_{i-1},x}_{t_i}) \\ \quad +f(t_i, X^{t_{i-1},x}_{t_i}, u^{\pi}_i(X^{t_{i-1},x}_{t_i}), v^{\pi}_i(X^{t_{i-1},x}_{t_i}))\Delta t_i\Big]\Big\}; \\ u^{\pi}_{i-1}(x) = E\Big\{u^{\pi}_i(X^{t_{i-1},x}_{t_i}) + f(t_i, X^{t_{i-1},x}_{t_i}, u^{\pi}_i(X^{t_{i-1},x}_{t_i}), \widetilde{v}^{\pi}_{i-1}(x))\Delta t_i\Big\}; \\ v^{\pi}_{i-1}(x) \triangleq E\Big\{\dfrac{W_{t_i}-W_{t_{i-1}}}{\Delta t_i}\Big[u^{\pi}_i(X^{t_{i-1},x}_{t_i}) + f(t_i, X^{t_{i-1},x}_{t_i}, u^{\pi}_i(X^{t_{i-1},x}_{t_i}), \widetilde{v}^{\pi}_{i-1}(x))\Delta t_i\Big]\Big\}. \end{cases} \tag{4.63}$$

In both cases, we still have

$$\widehat{Y}^{\pi}_{t_i} = Y^{\pi}_{t_i} = u^{\pi}_i(\hat{X}^{\pi}_{t_i}); \quad \widehat{Z}^{\pi}_{t_i} = Z^{\pi}_{t_i} = v^{\pi}_i(\hat{X}^{\pi}_{t_i}),$$

and since $E\{|g(X_T) - g(\hat{X}_T^\pi)|^2\} \leq C(1+|x|^2)|\pi|$, we conclude that the rate of convergence in (4.54) and (4.60) becomes $C(1+|x|^2)|\pi|$, which coincides with the result of [37].

CHAPTER 5. EXTENSIONS TO INCOMPLETE MARKETS DRIVEN BY LEVY PROCESSES

5.1 Introduction

A. Background. The problems addressed in this chapter are motivated by questions arising in Financial Asset Pricing Theory where the market is not complete. The framework is as follows: let $X = (X_t)_{t \geq 0}$ be a semimartingale representing the price process of a risky asset. Under the standard assumption of the absence of arbitrage opportunities, there exists a probability measure P^*, equivalent to the original probability measure P (the "objective" probability), such that X is a P^*-local martingale (technically one requires X only to be a P^*-sigma-martingale ; see [16]). P^* is known as the risk neutral measure.

Let us assume that X is in fact a P^*-martingale in L^2, for $0 \leq t \leq T$, as is often the case. For a non-redundant contingent claim $H \in L^2(\mathcal{F}_T, dP^*)$ we have a unique decomposition:

$$H = \alpha + \int_0^T \xi_s^H dX_s + N_T, \tag{5.1}$$

where N is an $L^2(dP^*)$ martingale strongly orthogonal to X. (The decomposition (5.1) is called the Kunita-Watanabe L^2-martingale decomposition; see [15] or [46] for background.) Let $(\eta_t)_{0 \leq t \leq T}$ be an (optional) strategy devoted to the trading of a risk-free savings account, whose price is fixed at 1. The *value* of the portfolio at time t is then

$$V_t = \xi_t X_t + \eta_t,$$

and the *cost* up to time t is

$$C_t = V_t - \int_0^t \xi_s dX_s.$$

We require $V_T = H$. A strategy (ξ, η) is *self-financing* if $(C_t)_{t\geq 0}$ is constant, and it is *mean-self-financing* if $E\{C_T - C_t|\mathcal{F}_t\} = 0$; that is, if C is a martingale. If we wish to minimize the remaining risk after time t, we then wish to minimize the quantity

$$E\{(C_T - C_t)^2|\mathcal{F}_t\}, \tag{5.2}$$

interpreting risk in the L^2, or "squared error", sense. H. Föllmer and M. Schweizer [20] have shown that the strategy

$$(\xi_t, \eta_t) = (\xi_t^H, V_t - \int_0^t \xi_s^H dX_s)$$

is *optimal* in the sense that it minimizes the risk quantity (5.2).

Therefore, at least in this special case where X is an L^2-martingale under the risk-neutral measure and $H \in L^2(\mathcal{F}_T, dP^*)$, the hedging strategy ξ^H of (5.1), while of necessity not perfect replication, is nevertheless optimal under squared error loss.

Several issues arise immediately: (1) when are there formulae to describe ξ^H analogous to those available in the traditional Black-Scholes paradigm? (2) when formulae are *not* available, what can one infer about ξ^H? In particular, when can one be assured of path regularity of ξ^H?

Issue (1) above is addressed in [30], where it is shown that if there is an underlying quasi-left continuous strong Markov process $Y = (\Omega, \mathcal{F}, (\mathcal{F}_t), P_y^*, Y)$ and if X is an L^2-martingale under each P_y^* and if $H = g(Y_T)$ for an appropriate class of functions g, then there is an explicit formula for ξ. (Note that "explicit" in the preceding sentence can mean different things to different people.) The results from [30] are perhaps the most interesting when $Y = X$ and X is the solution of a stochastic differential equation driven by a Lévy martingale.

Issue (2) above is the topic of this paper. For most contingent claims it is not possible to obtain explicit formulas for ξ^H. Instead here we are concerned with when the processes ξ^H–which are *a priori* assumed to be only predictably measurable – have regular sample paths. In particular by "regular" we mean that it is at least *left continuous with right limits*, known by its French acronym *càglàd* . When ξ^H

can be shown to have càglàd paths it is useful for two reasons: (a) approximations of ξ^H will converge to it in a Skorohod-type topology; and (b) one can approximate $\int \xi^H_s dX_s$ with Riemann-type sums and have convergence uniformly in probability (or even almost surely if the partition size shrinks quickly). The importance of part (b) in Finance has been emphasized, for example, in [49].

B. New Results. We assume that under the risk neutral measure P^* we have a Wiener process W and a compensated Poisson random measure $\widetilde{\mu}(drdz) = \mu(drdz) - drF(dz)$, where F is a Lévy measure. We further assume that $\int_{\mathbb{R}} z^2 F(dz) < \infty$, so that the process

$$Z_t = W_t + \int_0^t \int_{\mathbb{R}} z\widetilde{\mu}(drdz)$$

is a Lévy process with $E\{Z_t\} = 0$ and $E\{Z_t^2\} < \infty$, $0 \le t \le T$. (Thus Z is an $L^2(dP^*)$ martingale. Our price process X satisfies

$$X_t = y + \int_0^t \sigma(r, X_r)dW_r + \int_0^t \int_{\mathbb{R}} b(r, X_{r-})z\widetilde{\mu}(drdz) \tag{5.3}$$

and thus is also an $L^2(dP^*)$ martingale with mild hypotheses on σ and b.

A contingent claim $H \in \mathcal{F}_T$ can be assumed to be of the form $H = \Phi(X_s; s \le T)$, where Φ is a functional mapping $\mathbb{D}$ to $\mathbb{R}$, where $\mathbb{D}$ is the Skorohod space of càdlàg (right continuous with left limits) functions. We find hypotheses on Φ such that ξ^H is càglàd . In Chapter 3 we used the L^∞-Lipschitz condition to study the path regularity problem for the solutions to backward SDEs driven by Brownian motions. This chapter in a sense extends the result there to the Lévy case. Some examples of path-dependent options covered by our results include, but are not limited to,

(i) $\Phi(X)_T = \dfrac{1}{T}\displaystyle\int_0^T X_s ds$ (Asian option);

(ii) $\Phi(X)_T = g(\sup_{0\le t\le T} h(t, X_t))$ (Lookback option);

(iii) $\Phi(X)_T = g(\int_0^T h(s, X_{s-})dX_s)$; or

(iv) $\Phi(X) = g(\Phi_1(X), \cdots, \Phi_n(X))$, where g is Lipschitz and Φ_i's are of any of the forms (i)–(iii). (For example, if $g(x) = (K - x)^+$, then g combined with (i) gives an Asian Option.)

We remark that to justify the price equation (5.3) one should note that in almost all of the existing theory of Financial Asset Pricing, the price process is assumed to be Markov under the risk neutral measure. (The price process need not be Markov under the objective measure however.) E. Çinlar and J. Jacod [11] have shown that all "reasonable" strong Markov martingale processes are solutions of equations of the form

$$X_t = y + \int_0^t \sigma(r, X_r)dW_r + \int_0^t b(r, X_{r-}, z)\widetilde{\mu}(dr dz). \tag{5.4}$$

Thus our assumption merely restricts the general case by assuming $b(r, x, z)$ is of the form $b(r, x)z$, as well as some restrictions on the integrability of the coefficients.

A simple way in which our model might arise is if the objective price process X is modeled as a geometric Lévy process:

$$dX_t = \sigma_t X_{t-} dZ_t + b_t X_t dt, \tag{5.5}$$

where Z is an L^2 Lévy martingale under P. Since Lévy processes give rise to incomplete markets, we have to choose an equivalent risk neutral measure P^* in a natural way. We can do this using the idea of Föllmer and Schweitzer [20]: an equivalent risk neutral probability P^* is *minimal* if any square-integrable P-martingale M orthogonal to Z is also a P^*-martingale. T. Chan [8] has shown, under some restrictive assumptions, that if U satisfies

$$dU_t = 1 + U_t \gamma_t dW_t,$$

where γ can be taken to be non-random if σ and b are non-random, then $dP^* = U_T dP$, and thus it follows that under this P^* the process X of (5.5) satisfies (5.3). That is, the drift is removed by the canonical risk neutral minimal martingale measure, and the price process has the desired form under P^*.

5.2 Preliminaries

First we note that in this chapter some notations may have different meanings from those in previous chapters.

Throughout this chapter we assume that $(\Omega, \mathcal{F}, P; \{\mathcal{F}_t\}_{0\le t\le T})$ is a complete filtered probability space satisfying the *usual hypotheses* (see, e.g., [46]), and $T > 0$ is a fixed time duration. We denote $\mathbf{F} \stackrel{\triangle}{=} \{\mathcal{F}_t\}_{t\ge 0}$ and assume $\mathbf{F}$ is quasi-left continuous. Let W be an $\mathbf{F}$-adapted Brownian motion, and μ be a random measure generated by an $\mathbf{F}$-adapted Lévy process (see, e.g., [31]). Let $\nu(dtdz) = dtF(dz)$ be the (non-random) compensator of μ, and denote $\widetilde{\mu} = \mu - \nu$. We assume that F integrates z^2, and by rescaling we may assume without loss of generality that $\int_{\mathbb{R}} z^2 F(dz) = 1$.

In what follows we denote $\mathbb{D} \stackrel{\triangle}{=} \mathbb{D}[0,T]$ to be the space of all càdlàg functions on $[0,T]$; and $C_b^{0,1}([0,T]\times\mathbb{R})$ to be the space of all continuous functions on $[0,T]\times\mathbb{R}$ that are continuously differentiable with bounded derivatives in the spatial variable x. To simplify the presentation, we assume all the processes are 1-dimensional, but all the results in this chapter can be extended to higher dimensional cases without substantial difficulties.

For $(s,y)\in[0,T)\times\mathbb{R}$, let us consider a local martingale, $\{X_t^{s,y}\}_{s\le t\le T}$, defined as the (unique) solution to the following stochastic differential equation:

$$X_t = y + \int_s^t \sigma(r, X_{r-})dW_r + \int_s^t\int_{\mathbb{R}} b(r, X_{r-}) z\widetilde{\mu}(drdz), \quad s\le t\le T. \tag{5.6}$$

We assume that the coefficients σ and b satisfy the following standing assumptions:

(**A1**) (i) The functions $\sigma, b \in C_b^{0,1}([0,T]\times\mathbb{R})$, such that $\sigma^2(t,x)\neq 0$, for all (t,x); and

(ii) There exists a constant $K>0$ such that

$$\begin{cases} \sup\limits_{0\le t\le T} [|\sigma(t,0)| + |b(t,0)|] \le K; \\ |\sigma_x'(t,x)| + |b_x'(t,x)| \le K. \end{cases} \tag{5.7}$$

Here and in the sequel we denote σ_x' and b_x' to be the partial derivatives of σ and b, respectively, with respect to the spatial variable x.

We first give two lemmas. Since the proofs are more or less standard (see, e.g., [30]), we omit them. The first lemma shows that the assumption (A1), together with the requirement that $\int z^2F(dz) < \infty$, renders X a martingale rather than a local martingale.

Lemma 5.2.1 Assume (A1). Then for any $0 \le s < t \le T$, there exists a constant $C > 0$ depending only on K and T, such that

$$E\{\sup_{s\le t\le T} |X_t^{s,y}|^2\} \le C(1+|y|^2), \tag{5.8}$$

and

$$E\{\sup_{s\le r\le t} |X_r^{s,y} - y|^2\} \le C(1+|y|^2)(t-s). \tag{5.9}$$

Consequently, for any $(s,y) \in [0,T] \times \mathbb{R}$, $X_\cdot^{s,y}$ is a true martingale defined on $[s,T]$. Furthermore, one has

$$d\langle X^{s,y}, X^{s,y}\rangle_t = (\sigma^2 + b^2)(t, X_{t-}^{s,y})dt, \quad t \in [s,T]. \tag{5.10}$$

Throughout the chapter, unless otherwise specified, we denote $C > 0$ to be a generic constant depending only on K and T, which may vary from line to line. The following *variational* process ∇X, defined in Lemma 5.2.2, is important in the chapter.

Lemma 5.2.2 Assume (A1), and let α and β be two $\mathbf{F}$-predictable processes that are bounded by K. For each $0 \le s < T$, let $Y_\cdot^s$ be the solution to the following (linear) SDE:

$$Y_t^s = 1 + \int_s^t \alpha_r Y_{r-}^s dW_r + \int_s^t \int_{\mathbb{R}} \beta_r Y_{r-}^s z \widetilde{\mu}(drdz). \tag{5.11}$$

Then it holds that

$$E\left\{\sup_{s\le t\le T} |Y_t^s|^2\right\} \le C. \tag{5.12}$$

In particular, if $\alpha_t = \sigma_x'(t, X_t^{s,y})$ and $\beta_t = b_x'(t, X_t^{s,y})$, then we denote the solution to (5.11) by $Y \triangleq \nabla X$, which satisfies the following relation:

$$E\left\{\Big|\frac{1}{h}(X_t^{s,x+h} - X_t^{s,x}) - \nabla X_t^{s,x}\Big|^2\right\} \to 0, \qquad \text{as } h \to 0.$$

We note that sometimes the notion of "*difference quotient*" of X will also be used: for any $y, h \in \mathbb{R}$ and $t \in [s,T]$,

$$\Delta_h X_t^{s,y} \triangleq \frac{1}{h}[X_t^{s,y+h} - X_t^{s,y}].$$

The following identity is then obvious:

$$\sup_{s\le t\le T}|X_t^{s,y_1}-X_t^{s,y_2}|=|y_1-y_2|\sup_{s\le t\le T}|\Delta_{y_1-y_2}X_t^{s,y_2}|. \tag{5.13}$$

Now let $\Phi:\mathbb{D}\mapsto\mathbb{R}$ be a functional such that $E|\Phi(X)|^2<\infty$. We consider the following $\mathbf{F}$-martingale:

$$M_t\stackrel{\triangle}{=}E\{\Phi(X)|\mathcal{F}_t\},\quad t\ge 0. \tag{5.14}$$

By martingale theory, there exists an $\mathbf{F}-$predictable process, ξ, such that

$$M_t=\alpha+\int_0^t\xi_s dX_s+N_t, \tag{5.15}$$

where $\alpha=M_0$ and N is an $\mathbf{F}$-martingale that is orthogonal to X (i.e., $[N,X]$ is an $\mathbf{F}$-martingale. For more on this theory, consult Dellacherie-Meyer [15] or Protter [46]). One of the main purposes of this paper is to find conditions on Φ, so that ξ has càglàd paths.

To this end, let us introduce a functional φ which plays an important role in the sequel. Define $\varphi:\mathbb{D}\times[0,T]\times\mathbb{R}\mapsto\mathbb{R}$ as follows

$$\varphi(\mathbf{x},t,y)=E\{\Phi(\mathbf{x}1_{[0,t)}+X^{t,y}1_{[t,T]})\},\quad(\mathbf{x},t,y)\in\mathbb{D}\times[0,T]\times\mathbb{R}. \tag{5.16}$$

We note that if $\Phi(X)=g(X_T)$, then $\varphi(\mathbf{x},t,y)=P_{T-t}g(y)$, where (P_t) is the transition semi-group of the strong Markov process X.

5.3 Discrete Case Revisited

In this section we look at the special case where Φ is a discrete functional, say

$$\Phi(\mathbf{x})=g(\mathbf{x}_{t_0},\cdots,\mathbf{x}_{t_n}),$$

where $\pi:0=t_0<t_1<\cdots<t_n=T$ is a partition of $[0,T]$ and $g\in C_b^1(\mathbb{R}^{n+1})$. We note that such a case was also studied by Jacod-Méléard-Protter [30], but we shall give a slightly different formula that is more useful for our future discussion. We assume that all the (first order) partial derivatives of g, denoted by $\partial_i g$ $(=\partial_{x_i}g)$,

$i = 0, 1, \cdots, n$, are bounded by a common constant $K > 0$. Recall the function φ defined by (5.16). Clearly, for $t \in (t_{i-1}, t_i]$ and $\mathbf{x} \in \mathbb{D}$,

$$\varphi(\mathbf{x}; t, y) = E\Big\{g\Big(\mathbf{x}(t_0), \cdots, \mathbf{x}(t_{i-1}), X_{t_i}^{t,y}, \cdots, X_{t_n}^{t,y}\Big)\Big\}. \tag{5.17}$$

Since X is Markovian, it can be shown that the martingale defined by (5.14) can be written as

$$M_t = \varphi(X; t, X_t) = \lim_{\substack{t_{i-1}<u<t \\ u \downarrow t_{i-1}}} \varphi(X1_{[0,u)}; t, X_t),$$

where the second equality holds true for $t \in (t_{i-1}, t_i]$. Note that by (5.15) we have

$$dM_t = \xi_t dX_t + dN_t. \tag{5.18}$$

We shall follow the idea of [30] to identify ξ. To this end, we define a function $\phi : [0, T] \times \mathbb{R}^i$ by

$$\phi(t, x_1, \cdots, x_{i-1}, y) = E\Big\{g(x_1, \cdots, x_{i-1}, X_{t_i}^{t,y}, \cdot, X_{t_n}^{t,y})\Big\}.$$

Clearly, we have $\varphi(X; t, X_t) = \phi(t, X_{t_0}, \cdots, X_{t_{i-1}}, X_t)$ and (suppressing all variables) $\partial_t \varphi = \partial_t \phi$, $\partial_y \varphi = \partial_y \phi$. We shall first assume that that $\phi \in C^{1,2}((t_{i-1}, t_i] \times \mathbb{R}^i)$. Also, when the context is clear, we shall simply write $\phi(t, y) = \phi(t, X_{t_0}, \cdots, X_{t_{i-1}}, y)$ for notational convenience.

Now, for $j = 1, \cdots, i-1$, we define $X_t^j \equiv X_{t_j}$, for all $t \geq t_{i-1} \geq t_j$. Then, applying Itô's formula over $[t_{i-1}, t_i]$, and noting that $dX_s^j \equiv 0$, $j = 1, \cdots, i-1$ and $d\langle X^j, Y \rangle \equiv 0$, for $Y = X, X^1, \cdots, X^{i-1}$, and $t \geq t_{i-1}$, we get:

$$\begin{aligned} dM_t &= d\phi(t, X_t^1, \cdots, X_t^i, X_t) \\ &= \partial_t \phi(t, X_{t-})dt + \partial_y \phi(t, X_{t-})dX_t + \frac{1}{2}\partial_{yy}\phi(t, X_{t-})\sigma^2(t, X_{t-})dt \\ &+ \int_{\mathbb{R}} \Big[\phi(t, X_{t-} + b(t, X_{t-})z) - \phi(t, X_{t-}) - \partial_y \phi(t, X_{t-})b(t, X_{t-})z\Big]\mu(dt, dz). \end{aligned} \tag{5.19}$$

Since M is a martingale, the finite variation terms on the right side of (5.19) must equal 0; and an argument analogous to that in [30] then shows that

$$\begin{aligned} dM_t &= \partial_y \phi(t, X_{t-})dX_t + \int_{\mathbb{R}} \Big[\phi(t, X_{t-} + b(t, X_{t-})z) - \phi(t, X_{t-}) \\ &\qquad - \partial_y \phi(t, X_{t-})b(t, X_{t-})z\Big]\widetilde{\mu}(dtdz) \\ &= (\partial_y \phi \sigma)(t, X_{t-})dW_t + \int_{\mathbb{R}} [\phi(t, X_{t-} + b(t, X_{t-})z) - \phi(t, X_{t-})]\widetilde{\mu}(dtdz). \end{aligned} \tag{5.20}$$

Since X and N are orthogonal, by combining (5.10), (5.18) and (5.20) we have

$$\begin{aligned} d\langle M, X\rangle_t &= \xi_t(\sigma^2+b^2)(t,X_{t-})dt \\ &= \Big[(\partial_y\phi\sigma^2)(t,X_{t-}) + b(t,X_{t-})\int_{\mathbb{R}} z\Lambda_t(z)F(dz)\Big]dt, \end{aligned}$$

where

$$\Lambda_t(z) \stackrel{\triangle}{=} \varphi(X,t,X_{t-}+b(t,X_{t-})z) - \varphi(X,t,X_{t-}). \tag{5.21}$$

Consequently, one has, for $t \in (t_{i-1}, t_i)$,

$$\xi_t = \frac{(\partial_y\phi\sigma^2)(t,X_{t-}) + b(t,X_{t-})\int_{\mathbb{R}} z\Lambda_t(z)F(dz)}{(\sigma^2+b^2)(t,X_{t-})}. \tag{5.22}$$

Now we are ready to prove the following theorem.

Theorem 5.3.1 Assume (A1), and assume that the function g is continuously differentiable with bounded derivatives. Then it holds that

$$\xi_t = \frac{\sigma^2(t,X_{t-})[\nabla X_{t-}]^{-1}\xi^1_{t-} + \xi^2_t}{(\sigma^2+b^2)(t,X_{t-})} \tag{5.23}$$

is càglàd , where

$$\begin{cases} \xi^1_t \stackrel{\triangle}{=} E\Big\{\sum_{t_j>t} \partial_j g(X_{t_0},\cdots,X_{t_n})\nabla X_{t_j}\Big|\mathcal{F}_t\Big\}; \\ \xi^2_t \stackrel{\triangle}{=} b(t,X_{t-})\int_{\mathbb{R}} z\Lambda_t(z)F(dz), \end{cases} \tag{5.24}$$

and Λ is defined by (5.21).

Proof. Recall that $\nu(dt,dx) = dtF(dx)$. Assume first that F has compact support, that σ, b are infinitely differentiable with bounded derivatives of all orders, and that g is bounded, and twice continuously differentiable with bounded derivatives. Following the argument of Lemma 5.1 in [30] we know that in each subinterval (t_{i-1}, t_i), $\phi \in C^{1,2}((t_{i-1},t_i)\times\mathbb{R})$. Then following the same argument as before we see that (5.22) must hold for $t \in (t_{i-1}, t_i)$. Now using (5.17) we derive that

$$\partial_y\phi(t,X_t) = [\nabla X_t]^{-1}E\Big\{\sum_{j\geq i} g_j(X_{t_0},\cdots,X_{t_n})\nabla X_{t_j}\Big|\mathcal{F}_t\Big\} = [\nabla X_t]^{-1}\xi^1_t.$$

Noting that $\partial_y \phi$ is continuous, and taking left limits on both sides above we obtain (5.23) on (t_{i-1}, t_i).

To show that ξ is càglàd we first observe from (5.24) that ξ^1 is obviously càdlàg , hence the mapping $t \mapsto \sigma^2(t, X_{t-})[\nabla X_{t-}]^{-1}\xi^1_{t-}$ is càglàd . Furthermore, from (5.21) we see that the process Λ is càglàd . Also, applying Lemma 5.2.2 one shows that $|\phi_x| \le C$ for some constant $C > 0$. Thus,

$$|\Lambda_t(z)| \le C|b(t, X_{t-})||z| \le C(1 + \sup_{0\le s\le T} |X_s|)|z|.$$

Now a simple application of the Dominated Convergence Theorem shows that ξ^2, defined by (5.24), is also càglàd . Therefore so is ξ.

It remains to remove the extra assumptions made on F, g, σ and b. Here we can follow closely the approximation techniques of [30]. We leave it to the interested readers. The proof is now complete. ■

For future applications, we now extend Theorem 5.3.1 slightly:

Theorem 5.3.2 If g is continuously differentiable, and further it is of linear growth and all its partial derivatives composed with X, $\partial_i g(X_{t_0}, \cdots, X_{t_n})$, are square integrable, then (5.23) holds and ξ is càglàd .

Proof. Let $\{\phi^m\} \subset C^1_0(\mathbb{R}^{n+1})$ be a sequence of *truncation* functions satisfying $0 \le \phi^m \le 1$; $|\partial_i \phi^m| \le 1$; and

$$\phi^m(x_0, \cdots, x_n) = \begin{cases} 1, & |(x_0, \cdots, x_n)| \le m; \\ 0, & |(x_0, \cdots, x_n)| \ge m+1. \end{cases}$$

Define $\Phi_m : \mathbb{D} \mapsto \mathbb{R}$ by

$$\Phi_m(\mathbf{x}) = (g\phi^m)(\mathbf{x}(t_0), \cdots, \mathbf{x}(t_n)). \tag{5.25}$$

Then clearly Φ^m has compact support with all derivatives bounded. Applying Theorem 5.3.1 we have

$$\Phi_m(X) = \alpha_m + \int_0^T \xi^m_t dX_t + N^m_T,$$

where ξ^m is càglàd and satisfies, for $t \in (t_{i-1}, t_i]$,

$$\xi_t^m(\sigma^2 + b^2)(t, X_{t-}) = \sigma^2(t, X_{t-})(\nabla X_{t-})^{-1}\xi_{t-}^{m,1} + \xi_t^{m,2}, \quad \text{a.s.}$$

Here $\xi^{m,1}$, $\xi^{m,2}$ are defined in the same way as those in (5.24), as well as (5.21), with g being replaced by $g\phi^m$. Since g is of linear growth, and $\partial_j g(X_{t_0}, \cdots X_{t_n})$'s are all square-integrable, letting $m \to 0$ on both sides above and applying Lebesgue's dominated convergence theorem, one concludes that (5.23) holds and ξ is càglàd . ∎

5.4 L^∞-Lipschitz Functional Case

5.4.1 General Result

In this subsection we present our first path regularity result, under a rather general condition on the functional Φ, say Φ satisfies the L^∞-Lipschitz condition (1.3). Two important cases under such an assumption will be studied separately in the next subsection.

We first give a lemma that shows the implication of (1.3) on the function φ defined by (5.16).

Lemma 5.4.1 Suppose that Φ satisfies (1.3), and let φ be defined by (5.16). Then there exists a constant $C > 0$, depending only on the time duration T and the constants K in (A1) and L in (1.3), such that for any $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{D}$ and $y, y_1, y_2 \in \mathbb{R}$,

$$\begin{cases} |\varphi(\mathbf{x}_1, t, y) - \varphi(\mathbf{x}_2, t, y)| \le C \sup\limits_{0 \le s < t} |\mathbf{x}_1(s) - \mathbf{x}_2(s)|; \\ |\varphi(\mathbf{x}, t, y_1) - \varphi(\mathbf{x}, t, y_2)| \le C|y_1 - y_2|. \end{cases} \tag{5.26}$$

Proof. First note that, by (5.16) and (1.3),

$$\begin{aligned} &|\varphi(\mathbf{x}_1, t, y) - \varphi(\mathbf{x}_2, t, y)| \le E\Big\{|\Phi(\mathbf{x}_1 1_{[0,t)} + X^{t,y}1_{[t,T]}) - \Phi(\mathbf{x}_2 1_{[0,t)} + X^{t,y}1_{[t,T]})|\Big\} \\ \le\ & L\|[\mathbf{x}_1 1_{[0,t)} + X^{t,y}1_{[t,T]}] - [\mathbf{x}_2 1_{[0,t)} + X^{t,y}1_{[t,T]}]\|_\infty = L \sup_{0 \le s < t} |\mathbf{x}_1(s) - \mathbf{x}_2(s)|. \end{aligned}$$

Similarly, for any $\mathbf{x} \in \mathbb{D}$ and $y_1, y_2 \in \mathbb{R}$,

$$\begin{aligned} &|\varphi(\mathbf{x}, t, y_1) - \varphi(\mathbf{x}, t, y_2)| \le E\Big\{|\Phi(\mathbf{x}1_{[0,t)} + X^{t,y_1}1_{[t,T]}) - \Phi(\mathbf{x}1_{[0,t)} + X^{t,y_2}1_{[t,T]})|\Big\} \\ \le\ & LE\Big\{\sup_{t \le s \le T} |X_s^{t,y_1} - X_s^{t,y_2}|\Big\} = L|y_1 - y_2|E\Big\{\sup_{t \le s \le T} |\Delta_{y_1 - y_2} X_s^{t,y_2}|\Big\} \le C|y_1 - y_2|, \end{aligned}$$

thanks to (5.13) and Lemma 5.2.2. This completes the proof. ■

In light of the idea in Chapter 3 for the Brownian case, we shall approximate the functional Φ by a sequence of discrete functionals, which we now describe. For any partition $\pi : 0 = t_0 < t_1 < ... < t_n = T$, we define a mapping $\pi : \mathbb{D} \mapsto \mathbb{D}$ by $\mathbf{x} \mapsto \pi(\mathbf{x}) \stackrel{\triangle}{=} \mathbf{x}_\pi$, where

$$\mathbf{x}_\pi(t) \stackrel{\triangle}{=} \sum_{i=1}^{n} \mathbf{x}(t_i) 1_{[t_{i-1},t_i)}(t) + \mathbf{x}(T) 1_{\{T\}}(t). \tag{5.27}$$

Denote $|\pi| = \max_i |t_{i+1} - t_i|$ to be the mesh size of the partition π. Then, using the right continuity of $\mathbf{x}$ it is easily seen that

$$\lim_{|\pi|\to 0} |\mathbf{x}_\pi(t) - \mathbf{x}(t)| = 0. \tag{5.28}$$

The following lemma is a slight variation of Lemma 3.2.1. We shall state it without proof.

Lemma 5.4.2 Suppose that Φ satisfies (1.3). Let $\Pi = \{\pi\}$ be a family of partitions of $[0, T]$. Then there exists a family of discrete functionals $\{g^\pi : \pi \in \Pi\}$ such that

(i) for each $\pi \in \Pi$, $g^\pi \in C_b^1(\mathbb{R}^{n+1})$, and satisfies

$$\sum_{i=0}^{n} |g_i^\pi(x_0, \cdots, x_n)| \le K. \tag{5.29}$$

with constant K being the same as that in (1.3).

(ii) for any $\mathbf{x} \in \mathbb{D}$, it holds that

$$|g^\pi(\mathbf{x}(t_0), \cdots, \mathbf{x}(t_n)) - \Phi(\mathbf{x}_\pi)| \le |\pi|. \tag{5.30}$$

Our main theorem of this section is the following.

Theorem 5.4.3 Assume (A1), and that Φ satisfies (1.3). Assume further that the function φ defined by (5.16) is càglàd with respect to t, for each fixed $(\mathbf{x}, y) \in \mathbb{D} \times \mathbb{R}$, and that

$$\lim_{|\pi|\to 0} |\Phi(\mathbf{x}_\pi) - \Phi(\mathbf{x})| = 0; \qquad \forall \mathbf{x} \in \mathbb{D}, \tag{5.31}$$

where $\mathbf{x}_\pi$ is defined by (5.27). Then ξ in (5.15) admits a càglàd version.

Proof. Let the partition $\pi : 0 = t_0 < t_1 < \cdots < t_n = T$ be given. For $(\mathbf{x}, t, y) \in \mathbb{D} \times [0, T) \times \mathbb{R}$, define

$$\Phi_\pi(\mathbf{x}) \triangleq g^\pi(\mathbf{x}(t_0), \cdots, \mathbf{x}(t_n)); \qquad \varphi^\pi(\mathbf{x}, t, y) \triangleq E\{\Phi_\pi(\mathbf{x}1_{[0,t)} + X^{t,y}1_{[t,T]})\}. \tag{5.32}$$

where g^π is the approximation of g given by Lemma 5.4.2. By (5.30) and (1.3),

$$|\Phi_\pi(\mathbf{x})| \le |\Phi(\mathbf{x}_\pi)| + |\pi| \le C\Big(1 + \sup_{0\le t\le T} |\mathbf{x}(t)|\Big). \tag{5.33}$$

Applying Lemma 5.2.1 and the Dominated Convergence Theorem we derive from (5.30) and (5.31) that

$$\lim_{|\pi|\to 0} E|\Phi_\pi(X) - \Phi(X)|^2 = 0. \tag{5.34}$$

Now Theorem 5.3.1 tells us that in the following representation

$$\Phi_\pi(X) = \alpha_\pi + \int_0^T \xi_t^\pi dX_t + N_T^\pi. \tag{5.35}$$

the process ξ^π is càglàd ; and it has an explicit form:

$$\xi_t^\pi(\sigma^2 + b^2)(t, X_{t-}) = \sigma^2(t, X_{t-})(\nabla X_{t-})^{-1}\xi_{t-}^{\pi,1} + \xi_t^{\pi,2}, \tag{5.36}$$

where

$$\begin{cases} \xi_t^{\pi,1} \triangleq E\Big\{ \sum_{t_j > t}^{n} g_j^\pi(X_{t_0}, \cdots, X_{t_n})\nabla X_{t_j} \Big| \mathcal{F}_t \Big\}; \qquad \forall t \in (t_{i-1}, t_i]; \\ \xi_t^{\pi,2} \triangleq b(t, X_{t-}) \int_{\mathbb{R}} (\varphi^\pi(X, t, X_{t-} + b(t, X_{t-})z) - \varphi^\pi(X, t, X_{t-}))zF(dz). \end{cases} \tag{5.37}$$

Further, by virtue of (5.34), we see that as $|\pi| \to 0$ we must have $\alpha_\pi \to \alpha$ and

$$E\Big\{ \int_0^T |\xi_t^\pi - \xi_t|^2(\sigma^2 + b^2)(t, X_{t-})dt \Big\} \to 0. \tag{5.38}$$

To show that ξ has a càglàd version, it suffices to show that there exist càglàd processes ξ^1 and ξ^2 such that ξ has the explicit form:

$$\xi_t(\sigma^2 + b^2)(t, X_{t-}) = \sigma^2(t, X_{t-})(\nabla X_{t-})^{-1}\xi_t^1 + \xi_t^2, \quad \text{a.s.} \tag{5.39}$$

To this end, we note that from (5.32) and (5.29) one has

$$\begin{aligned} & |\varphi^\pi(X, t, X_{t-} + b(t, X_{t-})z) - \varphi^\pi(X, t, X_{t-})| \\ \le\ & CE\Big\{ \sup_{t\le s\le T} |X_s^{t,X_{t-}+b(t,X_{t-})z} - X_s^{t,X_{t-}}| \Big| \mathcal{F}_t \Big\} \\ =\ & CE\Big\{ |b(t, X_{t-})z| \sup_{t\le s\le T} |\Delta_{b(t,X_{t-})z} X_s| \Big| \mathcal{F}_t \Big\} \le C|b(t, X_{t-})||z|. \end{aligned}$$

where the last inequality is due to Lemma 5.2.1. Thus if we define

$$\xi_t^2 \triangleq b(t, X_{t-}) \int_{\mathbb{R}} (\varphi(X, t, X_{t-} + b(t, X_{t-})z) - \varphi(X, t, X_{t-}))zF(dz), \tag{5.40}$$

and note that $\int_{\mathbb{R}} F(dz)z^2 = 1$, then by applying the Dominated Convergence Theorem, we derive from (5.31) that

$$\lim_{|\pi|\to 0} E\{\int_0^T |\xi_t^{\pi,2} - \xi_t^2|dt\} = 0. \tag{5.41}$$

Therefore, possibly along a subsequence, $\xi^{\pi,2} \to \xi^2$ as $|\pi| \to 0$, in measure, a.s. Now since φ is càglàd in t, we see that so is ξ^2.

It remains to identify ξ^1 and show it has a càglàd version. We shall make use of Meyer-Zheng's tightness criterion again. First, combining (5.36), (5.38), and (5.41) we see that, possibly along a subsequence, one has

$$\xi_{t-}^{\pi,1} \to \xi_t^1 \triangleq \frac{\xi_t(\sigma^2 + b^2)(t, X_{t-}) - \xi_t^2}{\sigma^2(t, X_{t-})[\nabla X_{t-}]^{-1}}, \quad \text{as}|\pi| \to 0, \tag{5.42}$$

and the convergence is in measure, a.s.

On the other hand, let $\pi' : 0 = s_0 < \cdots < s_m = T$ be any partition of $[0, T]$ that is finer than π. We assume that $t_i = s_{l_i}$, $i = 1, 2, \cdots, n$. Then,

$$\begin{aligned} &\sum_{j=1}^m E\{|E\{\xi_{s_j}^{\pi,1} - \xi_{s_{j-1}}^{\pi,1}|\mathcal{F}_{s_{j-1}}\}|\} = \sum_{i=1}^n \sum_{j=l_{i-1}+1}^{l_i} E\{|E\{\xi_{s_j}^{\pi,1} - \xi_{s_{j-1}}^{\pi,1}|\mathcal{F}_{s_{j-1}}\}|\} \\ &= \sum_{i=1}^n E\Big\{|E\{g_i^\pi \nabla X_{t_i}|\mathcal{F}_{t_{i-1}}\}|\Big\} \le E\Big\{\sum_{i=1}^n |g_i^\pi \nabla X_{t_i}|\Big\} \\ &\le CE\Big\{\sup_{0\le t\le T} |\nabla X_t|\Big\} \le C. \end{aligned} \tag{5.43}$$

thanks to (5.29). Therefore, the processes $\xi^{\pi,1}$'s all have bounded conditional variation, and hence they are all quasimartingales as defined in §1.3.4. Furthermore, we note that the uniform bound of these conditional variations are indeed independent of the choice of π. Consequently, applying the Meyer-Zheng theorem (Lemma 1.2.11), there exists a càdlàg process $\widetilde{\xi}^1$ such that the càdlàg version of $\xi^{\pi,1}$ converges to $\widetilde{\xi}^1$ weakly under the Meyer-Zheng topology, as $|\pi| \to 0$. Note that the Meyer-Zheng (pseudo-path) topology is equivalent to convergence in measure (see, e.g., Lemma

1.2.10 or [41]), so if by a slight abuse of notation we denote the càglàd version of $\widetilde{\xi}^1$ by itself, then the uniqueness of the limit shows that $\widetilde{\xi}^1_t = \xi^1_t$, a.s., $\forall t$. In other words, ξ^1, whence ξ, has a càglàd version. This completes the proof. ■

5.4.2 Two Sufficient Conditions

One of the main differences between the Lévy case and the Brownian case is that the L^∞-Lipschitz condition alone *does not* guarantee the path regularity of the process ξ. In fact, in Theorem 5.4.3 we required an extra assumption on the mapping $t \mapsto \varphi(\mathbf{x}, t, y)$, which is not easy to verify in general. In this section we consider two cases where the functional Φ satisfies some stronger "Lipschitz" conditions so that the extra assumption can either be removed or be replaced by a more easily verifiable one. These two cases are mainly motivated by connections to finance. The first one corresponds to the "Asian Option", while the second one corresponds to the "Lookback Option". At the end of the section we shall give an example to show that there exists a non-regular ξ even though Φ satisfies the L^∞-Lipschitz condition.

Theorem 5.4.4 Suppose that Φ satisfies the L^1-Lipschitz condition (1.4), then the process ξ in the representation (5.15) admits a càglàd version.

Proof. It is clear that the L^1-Lipschitz condition (1.4) implies the L^∞-Lipschitz condition (1.3). Thus, by Theorem 5.4.3, it suffices to prove that φ is càglàd and that (5.31) holds.

Note that for any $\mathbf{x} \in \mathbb{D}$, it holds that $\|\mathbf{x}_\pi\|_\infty \le \|\mathbf{x}\|_\infty < \infty$. Then, by (1.4) and the Dominated Convergence Theorem, we have

$$|\Phi(\mathbf{x}_\pi) - \Phi(\mathbf{x})| \le L \int_0^T |\mathbf{x}_\pi(t) - \mathbf{x}(t)|dt \to 0; \qquad as \quad |\pi| \to 0, \tag{5.44}$$

thanks to (5.28). Furthermore, by Lemma 5.2.1 one has, for $0 \le t_1 < t_2 \le T$,

$$\begin{aligned} &|\varphi(\mathbf{x}, t_1, y) - \varphi(\mathbf{x}, t_2, y)| \\ &\le E\Big\{\Big|\Phi(\mathbf{x}1_{[0,t_1)} + X^{t_1,y}1_{[t_1,T]}) - \Phi(\mathbf{x}1_{[0,t_2)} + X^{t_2,y}1_{[t_2,T]})\Big|\Big\} \end{aligned}$$

$$\le CE\Big\{\int_{t_1}^{t_2}|X_s^{t_1,y}-\mathbf{x}(s)|ds+\int_{t_2}^{T}|X_s^{t_1,y}-X_s^{t_2,y}|ds\Big\} \tag{5.45}$$
$$\le CE\Big\{\sup_{t_1\le s\le t_2}(|X_s^{t_1,y}|+|\mathbf{x}(s)|)(t_2-t_1)+|X_{t_2}^{t_1,y}-y|\sup_{t_2\le s\le T}|\Delta_{X_{t_2}^{t,y}-y}X_s^{t_2,y}|\Big\}$$
$$\le C(1+|y|+\|\mathbf{x}\|_\infty)(t_2-t_1)^{\frac{1}{2}}.$$

Therefore, φ is continuous, which, together with (5.44), enables us to apply Theorem 5.4.3 to conclude that ξ has a càglàd version. ■

The second case, motivated by the Lookback option, is a little more involved.

Theorem 5.4.5 If $\Phi(\mathbf{x})=g(\sup_{0\le t\le T}h(t,\mathbf{x}(t)))$, where g and $h(t,\cdot)$ are uniformly Lipschitz with a common constant K, and $h(\cdot,x)$ is continuous for all x. Then, ξ in (5.15) admits a càglàd version.

Proof. That Φ satisfies (1.3) is obvious. So again we need to show only that φ is càglàd in t, and (5.31) holds.

First fix $\mathbf{x}\in\mathbb{D}$ and let $\pi:0=t_0<\cdots<t_n=T$ be any partition. Since g is Lipschitz, we have

$$\begin{aligned}|\Phi(\mathbf{x}_\pi)-\Phi(\mathbf{x})| &\le C\Big|\sup_{0\le t\le T}h(t,\mathbf{x}_\pi(t))-\sup_{0\le t\le T}h(t,\mathbf{x}(t))\Big|\\ &\le C\Big(\sup_{0\le t\le T}h(t,\mathbf{x}(t))-\max_i h(t_i,\mathbf{x}(t_i))\Big)\\ &\quad +C\max_i\sup_{t_{i-1}\le t<t_i}|h(t,\mathbf{x}(t_i))-h(t_i,\mathbf{x}(t_i))|.\end{aligned} \tag{5.46}$$

For any $\varepsilon>0$, choose $t^\varepsilon\in[0,T]$ such that

$$\sup_{0\le t\le T}h(t,\mathbf{x}(t))\le h(t^\varepsilon,\mathbf{x}(t^\varepsilon))+\varepsilon.$$

Now for each i we have

$$|h(t^\varepsilon,\mathbf{x}(t^\varepsilon))-h(t_i,\mathbf{x}(t_i))|\le|h(t_i,\mathbf{x}(t^\varepsilon))-h(t^\varepsilon,\mathbf{x}(t^\varepsilon))|+C|\mathbf{x}(t_i)-\mathbf{x}(t^\varepsilon)|,$$

and, thanks to the continuity of h,

$$h(t^\varepsilon,\mathbf{x}(t^\varepsilon))\le\varliminf_{|\pi|\to0}\max_i h(t_i,\mathbf{x}(t_i)).$$

Since ε is arbitrary, we deduce that

$$\sup_{0\le t\le T} h(t,\mathbf{x}(t)) \le \varliminf_{|\pi|\to 0} \max_i h(t_i,\mathbf{x}(t_i)). \tag{5.47}$$

Furthermore, since $\mathbf{x}\in\mathbb{D}$, it is bounded on $[0,T]$. Let $|\mathbf{x}(t)|\le K$ for some constant $K>0$. Also, since h is continuous, it is uniformly continuous on $[0,T]\times[-K,K]$. Thus we have

$$\lim_{|\pi|\to 0}\max_i \sup_{t_{i-1}\le t<t_i} |h(t,\mathbf{x}(t_i)) - h(t_i,\mathbf{x}(t_i))| = 0. \tag{5.48}$$

Combining (5.47) and (5.48) we derive from (5.46) that

$$\lim_{|\pi|\to 0} |\Phi(\mathbf{x}_\pi) - \Phi(\mathbf{x})| = 0. \tag{5.49}$$

It remains to show φ is càglàd in t. To do this we observe that, for fixed $\mathbf{x}$, y, and any $0\le t_1<t_2\le T$,

$$\begin{aligned}
&|\varphi(\mathbf{x},t_1,y)-\varphi(\mathbf{x},t_2,y)| \\
= \;& \Big|E\Big\{\Phi(\mathbf{x}1_{[0,t_1)} + X^{t_1,y}1_{[t_1,T]}) - \Phi(\mathbf{x}1_{[0,t_2)} + X^{t_2,y}1_{[t_2,T]})\Big\}\Big| \\
\le \;& CE\Big\{\Big|\sup_{0\le s<t_1} h(s,\mathbf{x}(s)) \vee \sup_{t_1\le s\le T} h(s,X^{t_1,y}_s) \\
&\qquad - \sup_{0\le s<t_2} h(s,\mathbf{x}(s)) \vee \sup_{t_2\le s\le T} h(s,X^{t_2,y}_s)\Big|\Big\} \\
\le \;& CE\Big\{|\sup_{0\le s<t_1} h(s,\mathbf{x}(s)) - \sup_{0\le s<t_2} h(s,\mathbf{x}(s))| \\
&\qquad + |\sup_{t_1\le s\le T} h(s,X^{t_1,y}_s) - \sup_{t_2\le s\le T} h(s,X^{t_2,y}_s)|\Big\} \\
\le \;& CE\Big\{\sup_{t_1\le s<t_2} |h(s,\mathbf{x}(s)) - h(t_1,\mathbf{x}(t_1-))| \\
&\qquad + \sup_{t_1\le s<t_2} |h(s,X^{t_1,y}_s) - h(t_2,y)| + \sup_{t_2\le s\le T} |h(s,X^{t_1,y}_s) - h(s,X^{t_2,y}_s)|\Big\} \\
\le \;& CE\Big\{\sup_{t_1\le s<t_2} |h(s,\mathbf{x}(t_1-)) - h(t_1,\mathbf{x}(t_1-))| \\
&\qquad + \sup_{t_1\le s<t_2} |\mathbf{x}(s) - \mathbf{x}(t_1-)| + \sup_{t_1\le s<t_2} |h(s,y) - h(t_2,y)| \\
&\qquad + \sup_{t_1\le s<t_2} |X^{t_1,y}_s - y| + |X^{t_1,y}_{t_2} - y| \sup_{t_2\le s\le T} |\Delta_{X^{t,y}_{t_2}-y} X^{t_2,y}_s|\Big\} \\
\le \;& C\Big[\sup_{t_1\le s<t_2} |h(s,\mathbf{x}(t_1-)) - h(t_1,\mathbf{x}(t_1-))| + \sup_{t_1\le s<t_2} |\mathbf{x}(s) - \mathbf{x}(t_1-)| \\
&\qquad + \sup_{t_1\le s<t_2} |h(s,y) - h(t_2,y)| + (t_2-t_1)^{\frac12}\Big].
\end{aligned} \tag{5.50}$$

thanks to Lemma 5.2.1 and Lemma 5.2.2. Now fix $t_0 \in (0,T)$. For $\forall \varepsilon > 0$, since $\mathbf{x}$ is càdlàg , there exists $\delta > 0$ such that,

$$\begin{cases} |\mathbf{x}(t) - \mathbf{x}(t_0-)| \le \varepsilon; & \forall t \in (t_0 - \delta, t_0); \\ |\mathbf{x}(t) - \mathbf{x}(t_0)| \le \varepsilon; & \forall t \in (t_0, t_0 + \delta). \end{cases} \tag{5.51}$$

Thus

$$\begin{cases} \sup_{t \le s < t_0} |\mathbf{x}(s) - \mathbf{x}(t-)| \le 2\varepsilon; & \forall t \in (t_0 - \delta, t_0); \\ \sup_{t_1 \le s < t_2} |\mathbf{x}(s) - \mathbf{x}(t_1-)| \le 2\varepsilon; & \forall t_0 < t_1 < t_2 < t_0 + \delta. \end{cases} \tag{5.52}$$

which, combined with (5.50) and the fact that h is locally uniformly continuous, clearly implies that φ is càglàd . ■

5.4.3 A Counterexample

To conclude this section we give an example which shows that in general the L^∞-Lipschitz condition (1.3) alone does not guarantee that ξ is càglàd .

Example 5.4.6 Let $X_t = N_t - t$, where N is the standard Poisson process. (That is, $\sigma = 0, b = 1$ and $F(dz) = \delta_{\{1\}}(dz)$.) Let $A \in \mathcal{B}([0,T])$ such that A is dense in $[0,T]$ and $0 < |A| < T$, where $|A|$ denotes the Lebesgue measure of A. Define $\Phi : \mathbb{D} \mapsto \mathbb{R}$ by $\Phi(\mathbf{x}) = \sup_{t\in A} |\Delta \mathbf{x}(t)|$. Now consider the equation:

$$M_t = 1 - \int_0^t M_{s-} 1_A(s) dX_s.$$

By the *Doléans-Dade Exponential Formula* (cf. Protter [46]) we have

$$\begin{aligned} M_t &= \exp\Big\{ - \int_0^t 1_A(s) dX_s \Big\} \prod_{s\le t} [(1 - 1_A(s)\Delta X_s) e^{1_A(s)\Delta X_s}] \qquad (5.53) \\ &= \exp\Big\{ - \sum_{s\le t} 1_A(s)\Delta N_s + |A \cap [0,t]| \Big\} [\prod_{s\le t}(1 - 1_A(s)\Delta N_s)] \exp\{\sum_{s\le t} 1_A(s)\Delta N_s\} \\ &= e^{|A\cap[0,t]|} \prod_{s \le t} (1 - 1_A(s)\Delta N_s). \end{aligned}$$

Since $\Phi(N)$ only takes values 0 or 1, and if $\Phi(N) = 0$, then for all $s \in A$, $\Delta N_s = 0$, thus $1_A(s)\Delta N_s = 0$ for all $s \in [0,T]$, hence $M_T = e^{|A|}$, thanks to (5.53). On the other hand, if $\Phi(N) = 1$, then there exists $s \in A$ such that $\Delta N_s = 1$, again by (5.53), we have $M_T = 0$. So

$$M_T = e^{|A|} 1_{\{\Phi(N)=0\}} = e^{|A|}(1 - \Phi(N)).$$

then

$$\Phi(X) = \Phi(N) = 1 - e^{-|A|} M_T = 1 - e^{-|A|} + \int_0^T e^{-|A|} M_{t-} 1_A(t) dX_t. \tag{5.54}$$

Hence

$$\xi_t = e^{-|A|} M_{t-} 1_A(t)$$

which clearly is not càglàd . ■

5.5 Stochastic Integral Case

In this section we consider the case where $\Phi(X) = g\Big(\int_0^T h(t, X_{t-}) dX_t\Big)$. Note that in this case $\Phi(X)$ no longer depends on X in a path by path manner, and none of the "Lipschitz conditions" studied in the previous sections is satisfied. Let us first modify the function φ. Observe that in this case

$$\begin{aligned}
&\Phi(X 1_{[0,t)} + X^{t,y} 1_{[t,T]}) \\
&= g\Big(\int_0^{t-} h(s, X_{s-}) dX_s + h(t, X_{t-})(y - X_{t-}) + \int_t^T h(s, X_{s-}^{t,y}) dX_s^{t,y}\Big),
\end{aligned}$$

we shall introduce the following two new functions $\psi : \Omega \times \mathbb{R}^2 \times [0, T] \times \mathbb{R} \mapsto \mathbb{R}$, and $\varphi : \mathbb{R}^2 \times [0, T] \times \mathbb{R} \mapsto \mathbb{R}$.

$$\begin{cases} \psi(a, x, t, y) \triangleq a + h(t, x)(y - x) + \displaystyle\int_t^T h(s, , X_{s-}^{t,y}) dX_s^{t,y}; \\ \varphi(a, x, t, y) \triangleq E\{g(\psi(a, x, t, y))\}. \end{cases} \tag{5.55}$$

The following theorem is an extension of Theorem 5.4.5.

Theorem 5.5.1 Suppose that $\Phi(X) = g\Big(\int_0^T h(t, X_{t-}) dX_t\Big)$, where g and h satisfy

(i) h is bounded;

(ii) for fixed x, $h(\cdot, x)$ is càglàd ;

(iii) for fixed t, g and $h(t, \cdot)$ are uniformly Lipschitz continuous with Lipschitz constant K, Then, ξ admits a càglàd version.

Proof. We follow the similar line of the proof of Theorem 5.4.5, but make necessary adjustments. First let us assume that g and $h(t,\cdot)$ are continuously differentiable for fixed t, and that $h'_x(\cdot,x)$ is càglàd for fixed x. We define an approximating discrete functional as follows. For $\pi : 0 = t_0 < \cdots < t_n = T$, define

$$\Phi_\pi(X) = \widetilde{g}(X_{t_0},\cdots,X_{t_n}) \stackrel{\triangle}{=} g\Big(\sum_{i=1}^n h(t_{i-1},X_{t_{i-1}})(X_{t_i}-X_{t_{i-1}})\Big); \tag{5.56}$$

Note that X is a martingale, one has

$$\begin{aligned}
&E\{|\Phi_\pi(X)-\Phi(X)|^2\}\\
&= E\Big\{\Big|g\Big(\sum_{i=1}^n h(t_{i-1},X_{t_{i-1}})(X_{t_i}-X_{t_{i-1}})\Big) - g\Big(\sum_{i=1}^n \int_{t_{i-1}}^{t_i} h(s,X_{s-})dX_s\Big)\Big|^2\Big\}\\
&\le CE\Big\{\Big|\sum_{i=1}^n \Big(h(t_{i-1},X_{t_{i-1}})(X_{t_i}-X_{t_{i-1}}) - \int_{t_{i-1}}^{t_i} h(s,X_{s-})dX_s\Big)\Big|^2\Big\}\\
&= CE\Big\{\sum_{i=1}^n \Big|h(t_{i-1},X_{t_{i-1}})(X_{t_i}-X_{t_{i-1}}) - \int_{t_{i-1}}^{t_i} h(s,X_{s-})dX_s\Big|^2\Big\} \qquad (5.57)\\
&\le CE\Big\{\sum_{i=1}^n \int_{t_{i-1}}^{t_i} |h(s,X_{s-})-h(t_{i-1},X_{t_{i-1}})|^2(\sigma^2+b^2)(s,X_{s-})ds\Big\}
\end{aligned}$$

A simple application of the Dominated Convergence Theorem then gives that

$$\lim_{|\pi|\to 0} E\{|\Phi_\pi(X)-\Phi(X)|^2\} = 0. \tag{5.58}$$

Next, note that by the martingale representation theorem we have

$$\begin{cases} \Phi_\pi(X) = \alpha_\pi + \displaystyle\int_0^T \xi_t^\pi dt + N_T^\pi;\\ \Phi(X) = \alpha + \displaystyle\int_0^T \xi_t dt + N_T, \end{cases}$$

and by (5.58) we have

$$\lim_{|\pi|\to 0} E\Big\{\int_0^T |\xi_t^\pi - \xi_t|^2(\sigma^2+b^2)(t,X_{t-})dt\Big\} = 0. \tag{5.59}$$

Furthermore, by Theorem 5.3.2 we know that ξ^π admits a càglàd version and has an explicit formula. We want to show that ξ also has a càglàd version and to identify its explicit form. To this end, let us introduce two functions corresponding to those in (5.55): for $t \in (t_{i-1}, t_i]$,

$$\begin{cases} \psi^\pi(a,x,t,y) \triangleq a + h(t_{i-1},x)(X^{t,y}_{t_i} - x) + \sum_{j>i} h(t_{j-1}, X^{t,y}_{t_{j-1}})(X^{t,y}_{t_j} - X^{t,y}_{t_{j-1}}); \\ \varphi^\pi(a,x,t,y) \triangleq E\{g(\psi^\pi(a,x,t,y))\}. \end{cases} \tag{5.60}$$

Then, it is easily seen that

$$\Phi_\pi(X1_{[0,t)} + X^{t,y}1_{[t,T]}) = g\Big(\psi^\pi\Big(\sum_{j<i} h(t_{j-1}, X_{t_{j-1}}), X_{t_{j-1}}, t, y\Big)\Big), \tag{5.61}$$

and by Theorem 5.3.2 ξ^π can be written explicitly as:

$$\xi^\pi_t = \frac{\sigma^2(t,X_{t-})(\nabla X_{t-})^{-1}\xi^{\pi,1}_t + \xi^{\pi,2}_t}{(\sigma^2+b^2)(t,X_{t-})}, \tag{5.62}$$

where, for $t \in (t_{i-1}, t_i]$,

$$\begin{cases} \xi^{\pi,1}_t \triangleq \partial_y \varphi^\pi\Big(\sum_{j<i} h(t_{j-1}, X_{t_{j-1}})(X_{t_j} - X_{t_{j-1}}), X_{t_{i-1}}, t, X_{t-}\Big); \\ \xi^{\pi,2}_t \triangleq b(t,X_{t-}) \int_{\mathbb{R}} z\Lambda^\pi_t(z)F(dz). \end{cases} \tag{5.63}$$

and

$$\begin{aligned} \Lambda^\pi_t(z) \triangleq\; & \varphi^\pi\Big(\sum_{j<i} h(t_{j-1}, X_{t_{j-1}})(X_{t_j} - X_{t_{j-1}}), X_{t_{i-1}}, t, X_{t-} + b(t,X_{t-})z\Big) \\ & -\varphi^\pi\Big(\sum_{j<i} h(t_{j-1}, X_{t_{j-1}})(X_{t_j} - X_{t_{j-1}}), X_{t_{i-1}}, t, X_{t-}\Big). \end{aligned} \tag{5.64}$$

To identify the limit of $\{\xi^\pi\}$, we first note that (5.55) and (5.60) yield that

$$\begin{cases} \partial_y \psi(a,x,t,y) = h(t,x) + \int_t^T h'_x(s, X^{t,y}_{s-})\nabla X^{t,y}_{s-} dX^{t,y}_s \\ \qquad\qquad + \int_t^T h(s, X^{t,y}_{s-}) d\nabla X^{t,y}_s; \\ \partial_y \varphi(a,x,t,y) = E\Big\{g'(\psi(a,x,t,y))\partial_y\psi(a,x,t,y)\Big\}. \end{cases} \tag{5.65}$$

and

$$\begin{cases} \partial_y \psi^\pi(a,x,t,y) = h(t_{i-1},x)\nabla X^{t,y}_{t_i} + \sum_{j>i} h'_x(t_{j-1}, X^{t,y}_{t_{j-1}})\nabla X^{t,y}_{t_{j-1}}(X^{t,y}_{t_j} - X^{t,y}_{t_{j-1}}) \\ \qquad\qquad + \sum_{j>i} h(t_{j-1}, X^{t,y}_{t_{j-1}})(\nabla X^{t,y}_{t_j} - \nabla X^{t,y}_{t_{j-1}}); \\ \partial_y \varphi^\pi(a,x,t,y) = E\{g'(\psi^\pi(a,x,t,y))\partial_y\psi^\pi(a,x,t,y))\}. \end{cases} \tag{5.66}$$

Since $h(\cdot,x)$ and $h'_x(\cdot,x)$ are both bounded and left continuous, applying the Burkholder-Davis-Gundy inequality and the Dominated Convergence Theorem we obtain that

$$\begin{cases} \lim\limits_{|\pi|\to 0} E\Big\{\Big|\sum\limits_{t_j\ge t} h(t_j, X^{t,y}_{t_j})(X^{t,y}_{t_{j+1}} - X^{t,y}_{t_j}) - \int_t^T h(s, X^{t,y}_{s-})dX^{t,y}_s\Big|\Big\} = 0; \\ \lim\limits_{|\pi|\to 0} E\Big\{\Big|\sum\limits_{t_j\ge t} h(t_j, X^{t,y}_{t_j})(\nabla X^{t,y}_{t_{j+1}} - \nabla X^{t,y}_{t_j}) - \int_t^T h(s, X^{t,y}_{s-})d\nabla X^{t,y}_s\Big|\Big\} = 0; \\ \lim\limits_{|\pi|\to 0} E\Big\{\Big|\sum\limits_{t_j\ge t} h'_x(t_j, X^{t,y}_{t_j})\nabla X^{t,y}_{t_i}(X^{t,y}_{t_{j+1}} - \nabla X^{t,y}_{t_j}) \\ \qquad\qquad - \int_t^T h'_x(s, X^{t,y}_{s-})\nabla X^{t,y}_{s-}dX^{t,y}_s\Big|\Big\} = 0; \end{cases} \tag{5.67}$$

Now if $a_\pi \to a$ and $x_\pi \to x$ as $|\pi| \to 0$, then (5.67) implies that

$$\begin{cases} \lim\limits_{|\pi|\to 0} E\{|\psi^\pi(a_\pi, x_\pi, t, y) - \psi(a, x, t, y)|\} = 0; \\ \lim\limits_{|\pi|\to 0} E\{|\partial_y\psi^\pi(a_\pi, x_\pi, t, y) - \partial_y\psi(a, x, t, y)|\} = 0. \end{cases} \tag{5.68}$$

Further, note that

$$\begin{aligned} & |g'(\psi^\pi(a,x,t,y))\psi^\pi_y(a,x,t,y) - g'(\psi(a,x,t,y))\partial_y\psi(a,x,t,y)| \\ \le\ & |g'(\psi^\pi(a,x,t,y)) - g'(\psi(a,x,t,y))||\partial_y\psi(a,x,t,y)| \\ & +C|\partial_y\psi^\pi(a,x,t,y) - \partial_y\psi(a,x,t,y)|, \end{aligned}$$

applying the Dominated Convergence Theorem we then conclude that

$$\begin{cases} \lim\limits_{|\pi|\to 0} |\varphi^\pi(a_\pi, x_\pi, t, y) - \varphi(a, x, t, y)| = 0; \\ \lim\limits_{|\pi|\to 0} |\partial_y\varphi^\pi(a_\pi, x_\pi, t, y) - \partial_y\varphi(a, x, t, y)| = 0. \end{cases} \tag{5.69}$$

Now note that, possibly along a subsequence we must have that, for each $t \in [0,T]$,

$$\sum_{j<i} h(t_{j-1}, X_{t_{j-1}})(X_{t_j} - X_{t_{j-1}}) \to \int_0^{t-} h(s, X_{s-})dX_s, \quad \text{a.s.},$$

and $X_{t_{i-1}} \to X_{t-}$, a.s., as $|\pi| \to 0$. Thus, applying the Dominated Convergence Theorem if necessary, we derive from (5.63) that, for all $t \in [0,T]$

$$\xi^{\pi,1}_t \to \xi^1; \qquad \xi^{\pi,2}_t \to \xi^2_t, \qquad \text{as}|\pi| \to 0,$$

thanks to (5.69), where

$$\begin{cases} \xi^1_t \triangleq \partial_y\varphi\Big(\int_0^{t-} h(s,X_{s-})dX_s, X_{t-}, t, X_{t-}); \\ \xi^2_t \triangleq b(t,X_{t-})\int_{\mathbb{R}} \Lambda_t(z) z F(dz), \end{cases} \tag{5.70}$$

and

$$\begin{aligned} \Lambda_t(z) \triangleq \varphi\Big(\int_0^{t-} h(s,X_{s-})dX_s, X_{t-}, t, X_{t-}+b(t,X_{t-})z\Big) \\ -\varphi\Big(\int_0^{t-} h(s,X_{s-}dX_s, X_{t-}, t, X_{t-}). \end{aligned}$$

Furthermore, (5.59) implies that

$$\xi_t = \frac{\sigma^2(t,X_{t-})\xi^1_t + \xi^2_t}{(\sigma^2+b^2)(t,X_{t-})}, \quad dP\times dt\text{-a.e..} \tag{5.71}$$

It remains to show that both ξ^1 and ξ^2 have càglàd version. To see this, note that X is driven by a Lévy process, which has no fixed jump time. Namely, for every $t\in[0,T]$, one has $X_t = X_{t-}$, a.s. and $\nabla X_t = \nabla X_{t-}$, a.s. Recalling (5.55) and (5.65), we see that, for each $t\in[0,T]$, it holds almost surely that

$$\begin{aligned} \xi^1_t &= \partial_y\varphi\Big(\int_0^{t-} h(s,X_{s-})dX_s, X_{t-}, t, X_t\Big) \\ &= E\Big\{g'(\int_0^T h(s,X_{s-})dX_s)[h(t,X_{t-}) + [\nabla X_{t-}]^{-1}\int_t^T h'_x(s,X_{s-})\nabla X_{s-}dX_s \\ &\quad +[\nabla X_{t-}]^{-1}\int_t^T h(s,X_{s-})d\nabla X_s]\Big|\mathcal{F}_t\Big\} \\ &= \Big[h(t,X_{t-}) - [\nabla X_{t-}]^{-1}\int_0^{t-} h'_x(s,X_{s-})\nabla X_{s-}dX_s \\ &\quad -[\nabla X_{t-}]^{-1}\int_0^{t-} h(s,X_{s-})d\nabla X_s\Big]\cdot E\Big\{g'\Big(\int_0^T h(s,X_{s-})dX_s\Big)\Big|\mathcal{F}_t\Big\} \\ &\quad +[\nabla X_{t-}]^{-1}E\Big\{g'\Big(\int_0^T h(s,X_{s-})dX_s\Big)\cdot\Big[\int_0^T h'_x(s,X_{s-})\nabla X_{s-}dX_s \\ &\quad +\int_0^T h(s,X_{s-})d\nabla X_s\Big]\Big|\mathcal{F}_t\Big\}, \end{aligned} \tag{5.72}$$

which clearly has a càglàd version, thanks to the assumption that h is càglàd with respect to t.

Furthermore, using (5.55) and (5.65) again we have

$$\begin{cases} |\partial_a\varphi(a,x,t,y)| \le C; \\ |\partial_x\varphi(a,x,t,y)| \le C(1+|y-x|); \\ |\partial_y\varphi(a,x,t,y)| \le C(1+|y|), \end{cases} \tag{5.73}$$

and for $0 \le t_1 < t_2 \le T$, denoting $X^i = X^{t_i,y}$, $i = 1, 2$, we have

$$\begin{aligned}
&|\varphi(a,x,t_1,y) - \varphi(a,x,t_2,y)| \le CE\{|\psi(a,x,t_1,y) - \psi(a,x,t_2,y)|\} \\
\le\ & CE\Big\{|h(t_1,x) - h(t_2,x)||y-x| + |\int_{t_1}^{t_2} h(s,X^1_{s-})dX^1_s| \\
&+\Big|\int_{t_2}^{T}[h(s,X^1_{s-})\sigma(s,X^1_{s-}) - h(s,X^2_{s-})\sigma(s,X^2_{s-})]dW_s\Big| \\
&+\Big|\int_{t_2}^{T}\int_{\mathbb{R}}[h(s,X^1_{s-})b(s,X^1_{s-}) - h(s,X^2_{s-})b(s,X^2_{s-})]z\widetilde{\mu}(ds,dz)\Big|\Big\} \\
\le\ & C\Big[|h(t_1,x) - h(t_2,x)||y-x| + (1+|y|)(t_2-t_1)^{1/2} \qquad (5.74)\\
&+E\Big\{\Big(\int_{t_2}^{T}|h(s,X^1_{s-})\sigma(s,X^1_{s-}) - h(s,X^2_{s-})\sigma(s,X^2_{s-})|^2 ds\Big)^{\frac{1}{2}}\Big\} \\
&+E\Big\{\int_{t_2}^{T}\int_{\mathbb{R}}|h(s,X^1_{s-})b(s,X^1_{s-}) - h(s,X^2_{s-})b(s,X^2_{s-})||z|F(dz)ds\Big\}\Big] \\
\le\ & C\Big[|h(t_1,x) - h(t_2,x)||y-x| + (1+|y|)(t_2-t_1)^{\frac{1}{2}} \\
&+E\Big(\int_{t_2}^{T}|X^1_{s-} - X^2_{s-}|^2 ds\Big)^{\frac{1}{2}} + E\Big\{\Big(\int_{t_2}^{T}|X^1_{s-} - X^2_{s-}|^2 ds\Big)^{\frac{1}{2}}\Big(1 + \sup_{t_2\le s\le T}|X^2_s|\Big)\Big\} \\
&+E\Big\{\int_{t_2}^{T}|X^1_{s-} - X^2_{s-}|ds + E\Big\{\Big(\int_{t_2}^{T}|X^1_{s-} - X^2_{s-}|ds\Big)\Big(1 + \sup_{t_2\le s\le T}|X^2_s|\Big)\Big\}\Big] \\
\le\ & C\Big[|h(t_1,x) - h(t_2,x)||y-x| + (1+|y|^2)(t_2-t_1)^{\frac{1}{2}}\Big],
\end{aligned}$$

where $C > 0$ is again some generic constant that is allowed to vary from line to line. This shows that φ is càglàd with respect to t. Combining this with (5.73) and (5.63) we see that ξ^2, whence ξ, is càglàd .

In the general case that g and $h(t,\cdot)$ are only Lipschitz continuous, we can again choose g^ε and h^ε to be the smooth molifiers with respect to the spatial variable x, and follow the standard arguments to show that

$$\lim_{\varepsilon\to 0} E\left\{\int_0^T |\xi^\varepsilon_t - \xi_t|^2(\sigma^2+b^2)(t,X_{t-})dt\right\} = 0. \qquad (5.75)$$

Next, since $(h^\varepsilon)'_x(\cdot,x)$ is càglàd for each x, using an argument similar to our previous one, we see that there exists a càglàd process ξ^ε of the form:

$$\xi^\varepsilon_t = \frac{\sigma^2(t,X_{t-})\xi^{\varepsilon,1}_t + \xi^{\varepsilon,2}_t}{(\sigma^2+b^2)(t,X_{t-})}, \qquad (5.76)$$

where $\xi^{\varepsilon,1}$ and $\xi^{\varepsilon,2}$ are defined in a by now obvious way, and such that

$$\lim_{\varepsilon\to 0} E\Big\{\int_0^T |\xi^{\varepsilon,2}_t - \xi^2_t|^2(\sigma^2+b^2)(t,X_{t-})dt\Big\} = 0; \qquad (5.77)$$

where ξ^2 is defined by (5.70). It is not too hard to show, as we did before, that ξ^2 is again càglàd . Define, in light of (5.76),

$$\xi^1_t \triangleq \frac{\xi_t(\sigma^2+b^2)(t,X_{t-})-\xi^2_t}{\sigma^2(t,X_{t-})}.$$

It suffices to show that ξ^1 is càglàd . To this end we denote, by (5.72),

$$\begin{aligned} A^\varepsilon_t &\triangleq \xi^{\varepsilon,1}_{t+}\nabla X_t - h^\varepsilon(t+,X_t)E\Big\{(g^\varepsilon)'(\int_0^T h^\varepsilon(s,X_{s-})dX_s)\Big|\mathcal{F}_t\Big\} \\ &= E\Big\{(g^\varepsilon)'(\int_0^T h^\varepsilon(s,X_{s-})dX_s)\Big[\int_t^T (h^\varepsilon)'_x(s,X_{s-})\nabla X_{s-}dX_s \\ &\qquad + \int_t^T h^\varepsilon(s,X_{s-})d\nabla X_s\Big]\Big|\mathcal{F}_t\Big\} \end{aligned} \tag{5.78}$$

For any $\lambda > 0$, define $\Omega_\lambda \triangleq \{\omega : \sup_t\{|\nabla X_t| + \dfrac{1}{\sigma^2(t,X_t)}\} \le \lambda\}$. From (5.75) and (5.77) we know that

$$\begin{aligned} & E\Big\{1_{\Omega_\lambda}\int_0^T|\xi^{\varepsilon,1}_t\nabla X_{t-}-\xi^1_t\nabla X_{t-}|dt\Big\} \\ \le\ & E\Big\{1_{\Omega_\lambda}\int_0^T\Big[\frac{\sigma^2+b^2}{\sigma^2}|\xi^\varepsilon_t-\xi_t||\nabla X_{t-}|+\frac{|\nabla X_{t-}|}{\sigma^2}|\xi^{\varepsilon,2}_t-\xi^2_t|\Big]dt\Big\} \\ \le\ & C(\lambda)\Big(E\Big\{1_{\Omega_\lambda}\int_0^T\Big[|\xi^\varepsilon_t-\xi_t|^2+|\xi^{\varepsilon,2}_t-\xi^2_t|^2\Big](\sigma^2+b^2)(t,X_{t-})dt\Big\}\Big)^{\frac12}\to 0,\ \text{as } \varepsilon\to 0. \end{aligned}$$

That is,

$$\lim_{\varepsilon\to 0}\xi^{\varepsilon,1}_\cdot\nabla X_{\cdot-} = \xi^1_\cdot\nabla X_{\cdot-}, \quad \text{strongly in} \quad L^1(\Omega_\lambda\times[0,T]). \tag{5.79}$$

Now denote $G^\varepsilon \triangleq (g_\varepsilon)'(\int_0^T h^\varepsilon(s,X_{s-})dX_s)$. Since it is uniformly bounded, there exists $G\in L^2(\Omega)$ such that, possibly along a subsequence, G^ε converges to G weakly in $L^2(\Omega)$. Noting that h^ε's are uniformly bounded and converge to h uniformly, we can easily derive

$$\lim_{\varepsilon\to 0}h^\varepsilon(\cdot+,X_\cdot)E\{G^\varepsilon|\mathcal{F}_\cdot\} = h(\cdot+,X_\cdot)E\{G|\mathcal{F}_\cdot\}, \quad \text{weakly in} \quad L^2(\Omega\times[0,T]),$$

which, together with (5.79), implies that

$$\lim_{\varepsilon\to 0}A^\varepsilon_{\cdot-} = A^0_\cdot, \quad \text{weakly in} \quad L^1(\Omega_\lambda\times[0,T]), \tag{5.80}$$

where $A^0_t \triangleq \xi^1_t\nabla X_{t-} - h(t,X_{t-})E\{G|\mathcal{F}_t\}_-$, $t\in[0,T]$. Now by Mazur's theorem (cf. e.g., [25]), there exists a sequence $\{B^{n,\lambda}\}$, where each $B^{n,\lambda}$ is a convex combinations of

A^ε, such that $B^{n,\lambda}$ converges to A^0 strongly in $L^1(\Omega_\lambda \times [0,T])$. Since $\Omega_\lambda \uparrow \Omega$, a simple diagonalization procedure shows that there exists a sequence $\{B^n\}$, where each B^n is a convex combination of $B^{n,\lambda}$'s (hence still a convex combination of A^ε's!), such that B^n converges to A^0, strongly in $L^1(\Omega_\lambda \times [0,T])$ for all λ, as $n \to \infty$. Consequently, for a.s. $\omega \in \Omega$, $B^n(\omega)$ converges to $A^0(\omega)$ in measure. On the other hand, by the definition (5.78) it is not hard to check that $\{A^\varepsilon\}$, whence $\{B^n\}$, is tight under the Meyer-Zheng topology. An argument similar to that in Chapter 3 shows that A^0 has a càglàd version. The proof is now complete. ■

Remark 5.5.2 If we let $h(t,x) = 1_{[0,t_0]}(t)$ for some $t_0 \in [0,T]$, then the assumptions (i)–(iii) of the theorem are all satisfied. Therefore our result covers the special case when $\Phi(X) = g(X_{t_0})$; that is, the case considered in [30].

Remark 5.5.3 In the theorem we assumed that h is bounded so that the random variables involved are all square integrable. An alternative assumption (of (i)) could be that (i') $\int_{\mathbb{R}} z^4 F(dz) < \infty$. We leave the details to the interested readers.

5.6 General Case

In this section we shall summarize the results from previous sections to study some more general situations. We present them in two theorems.

Theorem 5.6.1 Assume that $\Phi(X) = g(\int_0^T \mathbf{h}_1(t,X_t)dt, \int_0^T \mathbf{h}_2(t,X_{t-})dX_t)$, where

$$\mathbf{h}_1(t,x) \triangleq (h_1^1, \cdots, h_m^1)(t,x), \quad \mathbf{h}_2(t,x) \triangleq (h_1^2, \cdots, h_n^2)(t,x).$$

Assume further that $\mathbf{h}_2$ is bounded and càglàd with respect to t. Define

$$\varphi : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \times [0,T] \times \mathbb{R} \mapsto \mathbb{R}$$

as follows:

$$\begin{aligned} &\varphi(A_1, A_2, x, t, y) \\ &\triangleq E\left\{ g(A_1 + \int_t^T \mathbf{h}_1(s, X_s^{t,y})ds, A_2 + \mathbf{h}_2(t,x)(y-x) + \int_t^T \mathbf{h}_2(s, X_{s-}^{t,y})dX_s^{t,y}) \right\}. \end{aligned}$$

i) If g, $\mathbf{h}_1$ and $\mathbf{h}_2$ are continuously differentiable with uniformly bounded derivatives with respect to the spatial variables, then ξ is càglàd . To be more precise, one has

$$\begin{aligned}
& \xi_t(\sigma^2+b^2)(t,X_{t-}) \\
= \;& \sigma^2(t,X_{t-})\varphi_y\Big(\int_0^{t-}\mathbf{h}_1(s,X_s)ds, \int_0^{t-}\mathbf{h}_2(s,X_{s-})dX_s, X_{t-}, t, X_{t-}\Big) \\
& +b^2(t,X_{t-})\int_{\mathbb{R}}\int_0^1 \varphi_y\Big(\int_0^{t-}\mathbf{h}_1(s,X_s)ds, \\
& \quad \int_0^{t-}\mathbf{h}_2(s,X_{s-})dX_s, X_{t-}, t, X_{t-}+b(t,X_{t-})zu\Big)z^2duF(dz),
\end{aligned} \tag{5.81}$$

where

$$\begin{aligned}
\varphi_y(A_1,A_2,x,t,y) &= E\Big\{\langle\, g_1, \int_t^T(\mathbf{h}_1)'_x(s,X_s^{t,y})\nabla X_s^{t,y}ds\,\rangle \\
&+\langle\, g_2, \mathbf{h}_2(t,x)+\int_t^T(\mathbf{h}_2)'_x(s,X_{s-}^{t,y})\nabla X_{s-}^{t,y}dX_s^{t,y}+\int_t^T\mathbf{h}_2(s,X_{s-}^{t,y})d\nabla X_s^{t,y}\,\rangle\Big\}.
\end{aligned} \tag{5.82}$$

ii) The same holds when g, $\mathbf{h}_1$ and $\mathbf{h}_2$ are differences of two functions convex with respect to the spatial variables, with right derivatives bounded and all the derivatives appearing in i) are replaced by the corresponding right derivative, provided we have $X_s^{t,y}$, $\int_0^T h_i^1(t,X_t)dt$ and $\int_0^T h_j^2(t,X_t)dX_t$ have no atoms.

Proof. (i) If $(\mathbf{h}_2)'_x$ is also càglàd with respect to t, then similar to (5.24) one can show that (5.81) also holds. In general we can again approximate $\mathbf{h}_2$ by the molifiers to conclude the same result.

(ii) This is a direct consequence of the arguments of Theorem 2.6-b) of [30]. ∎

Theorem 5.6.2 Assume that $\Phi(X)=g(\Phi_1(X),\cdots,\Phi_n(X))$, where g is uniformly Lipschitz, and $\Phi_i(X)$'s are of the form as those in Theorem 5.4.4, Theorem 5.4.5 or Theorem 5.5.1. Then, ξ is càglàd .

Proof. To simplify the presentation, let us assume that

$$\Phi(X)=g\Big(\Phi_1(X), \sup_{0\le t\le T}h(t,X_t), \int_0^T k(t,X_{t-})dX_t\Big),$$

where Φ_1 satisfies L^1-Lipschitz condition (1.3), and h, k satisfy the conditions in Theorem 5.4.4 and Theorem 5.5.1, respectively. Define $\varphi : \mathbb{D} \times \mathbb{R}^2 \times [0,T] \times \mathbb{R} \mapsto \mathbb{R}$ to be such that

$$\begin{aligned}\varphi(\mathbf{x}, a, x, t, y) \stackrel{\triangle}{=} E\Big\{g\Big(\Phi_1(\mathbf{x}1_{[0,t)} + X^{t,y}1_{[t,T]}), \sup_{0\le s<t} h(s, \mathbf{x}(s)) \vee \sup_{t\le s\le T} h(s, X^{t,y}_s), \\ a + k(t,x)(y-x) + \int_t^T k(s, X^{t,y}_{s-}) dX^{t,y}_s\Big)\Big\}.\end{aligned}$$

For any partition $\pi : 0 = t_0 < \cdots < t_n = T$, define φ^π similar to (5.37). Then, combining the arguments in the previous sections we see that ξ^π will converge to ξ in measure, as $|\pi| \to 0$, and that ξ is càglàd . ■

LIST OF REFERENCES

LIST OF REFERENCES

[1] Antonelli, F. (1993), *Backward-Forward Stochastic Differential Equations*, Ann. *Appl. Prob.*, **3**, 777-793.

[2] Asmussen, S., Glynn, P. and Pitman, J. (1995), *Discretization Error in Simulation of One-Dimensional Reflecting Brownian Motion*, Ann. Appl. Probab.., **5**, 875-896.

[3] Bally, V. (1995), *An Approximation Scheme for BSDEs and Applications to Control and Nonlinear PDEs*, prepublication **95-15** du Laboratoire de Statistique et *Processus de l'Universite du Maine.*

[4] Bensoussan, A. (1983), *Stochastic Maximum Principle for Distributed Parameter Systems*, *J. Franklin Inst.*, **315**, 387-406.

[5] Bermin, H-P. (1998), *Hedging Options: the Malliavin Calculus Approach Versus the* Δ*-hedging Approach*, preprint.

[6] Bismut, J. M. (1973), *Théorie Probabiliste du Contrôle des Diffusions*, Mem. *Amer. Math. Soc.* 176, Providence, Rhode Island.

[7] Björk, T. (1998), *Arbitrage Theory in Continuous Time*, Oxford.

[8] Chan, T. (1999), *Pricing Contingent Claims on Stocks Driven by Lévy Processes*, *Ann. Applied Proba.*, **9**, 504–528.

[9] Chevance, D. (1996), *Ph.D. thesis, Universite de Provence.*

[10] Chevance, D.(1997), *Numerical Methods for Backward Stochastic Differential Equations*, *Numerical methods in finance*, 232–244.

[11] Çinlar, E. and Jacod, J. (1981) *Representation of Semimartingale Markov Processes in Terms of Wiener Processes and Poisson Random Measures*, In *Seminar on Stochastic Processes*, Birkhauser, Boston, 159–242,

[12] Çinlar, E., Jacod, J., Protter, P., and Sharpe, M. (1980) *Semimartingales and Markov Processes*, *Z. Wahrsch. Verw. Geb.* **52**, 161–220.

[13] Cvitanić, J., Ma, J. and Zhang J. (2000), *Efficient Computation of Δ-Hedges for Options with Discontinuous Payoffs*, in preparation.

[14] Dai, M. and Jiang, L. (2000) *On Path-dependent Options*, *Mathematical Finance - Theory and Applications*, Ed. Yong, J. and Cont, R., Higher Education Press, China, 290-316.

[15] Dellacherie, C. and Meyer, P. (1978), *Probabilities and Potential*, North Holland.

[16] Delbaen, F. and Schachermayer, W. (1988) *The Fundamental Theorem for Unbounded Stochastic Processes*, *Math. Ann.*, **312**, 215–250.

[17] Douglas, J. Jr., Ma, J. and Protter, P. (1996), *Numerical Methods for Forward-Backward Stochastic Differential Equations*, *Ann. Appl. Probab..*, **6**, 940-968.

[18] El Karoui, N. and Mazliak, L. (1997), *Backward Stochastic Differential Equations*, Pitman research notes in mathematics series, **364**.

[19] El Karoui, N., Peng, S., and Quenez, M.C. (1997), *Backward Stochastic Differential Equations in Finance*, *Mathmatical Finance*, **7**, 1-72.

[20] Föllmer, H and Schweitzer, M. (1991) *Hedging of Contingent Claims under Incomplete Information*, *Applied Stochastic Analysis*, M.H.A. Davis and R.J. Elliott, eds., Gordon and Breach, London, 389–414.

[21] Fournié, E., Lasry, J-M., Lebuchoux, J., Lions, P-L., and Touzi, N. (1999) *An Applicaiton of Malliavin Calculus to Monte Carlo Methods in Finance*, *Finance & Stochastics*, **3**, 391-412.

[22] Fournié, E., Lasry, J-M., Lebuchoux, J., and Lions, P-L. (2001) *Applicaitons of Malliavin Calculus to Monte Carlo Methods in Finance. II*, *Finance & Stochastics*, **5**, 201-236.

[23] Gikhman, I.I. and Skorohod, A.V. (1972), *Stochastic Differential Equations*, Springer.

[24] Hamadene, S. (1996), *Équations Différentielles Stochastiques Rétrogrades: le Cas Localment Lipschitzien*, *Ann. Inst. Henri poincaré*, v. 32, **5**, 645–659.

[25] Hille, E. and Phillips, R.S. (1957) *Functional Analysis and Semi-Groups*, American Mathematical Society.

[26] Hu, Y. and Peng, S. (1995) *Solution of Forward-Backward Stochastic Differential Equations*, *Probab. Theory Relat. Fields.*, **103**, 273-283.

[27] Hu, Y. (2000), *On the Existence of Solutions to One-Dimensional Forward-Backward SDEs*, *Stochastic Anal. Appl.*, **18**, 101-111.

[28] Hu, Y. (2000), *On the Solutions of Forward-Backward SDEs with Monotone and Continuous Coefficients*, Nonlinear Anal., **42**, 1-12.

[29] Hu, Y. and Yong, J. (2000), *Forward-backward Stochastic Differential Equations with Nonsmooth Coefficients*, Stochastic Process. Appl., **87**, 93-106.

[30] Jacod, J., Méléard, S., and Protter, P. (2000), *Explicit Form and Robustness of Martingale Representations*, Ann. Probab., **28**, 1747–1780.

[31] Jacod, J. and Shiryaev (1981) *Limit Theorems for Stochastic Processes*, Springer-Verlag, Heidelberg.

[32] Karatzas, I. and Shreve, S.E. (1987), *Brownian Motion and Stochastic Calculus*, Springer.

[33] Lepeltier, M. and San Martin, J. (1997), *Backward Stochastic Differential Equations with NonLipschitz Coefficients*, Stat. and Prob. Letters, v.32, **4**, 425–430.

[34] Ma, J., Protter, P., San Martin, J., and Torres, S. (2000) *Numerical Method for Backward SDEs*, Preprint.

[35] Ma, J., Protter, P. and Yong, J. (1994), *Solving Forward-Backward Stochastic Differential Equations Explicitly - a Four Step Scheme*, Probab. Theory Relat. Fields., **98**, 339-359.

[36] Ma, J., Protter P. and Zhang, J. (2001), *Explicit Form and Path Regularity of Martingale Representations*, Lévy Processes - Theory and Applications, O.E. Barndorff-Nielsen, T. Mikosch and S.I. Resnick (Eds.), Birkhauser Boston, 337-360.

[37] Ma, J. and Yong, J. (1999), *Forward-Backward Stochastic Differential Equations and Their Applications*, Lecture Notes in Math., **1702**, Springer.

[38] Ma, J. and Zhang, J. (1999), *Representation Theorems for Backward Stochastic Differential Equations*, to appear in Annals of Applied Probability.

[39] Ma, J. and Zhang, J. (1999), *Path Regularity of Solutions to Backward Stochastic Differential Equations*, to appear in Probability Theory and Related Fields.

[40] Ma, J. and Zhang, J. (2000), *Regularity of Viscosity Solution to Quasilinear Parabolic Partial Differential Equations – a Probabilistic Approach*, in preparation.

[41] Meyer, P. and Zheng, W. (1984) *Tightness Criteria for Laws of Semimartingales*, Ann. Inst. Henri Poincaré, Vol. 20 **4**, 353-372.

[42] Nualart, D. (1995) *The Malliavin Calculus and Related Topics*, Springer.

[43] Pardoux, E. and Peng S. (1990), *Adapted Solutions of Backward Stochastic Equations*, System and Control Letters, **14**, 55-61.

[44] Pardoux, E. and Peng S. (1992), *Backward Stochastic Differential Equations and Quasilinear Parabolic Partial Differential Equations*, Lecture Notes in CIS, Springer, **176**, 200-217.

[45] Pardoux, E, and Peng, S. (1995), *Some Backward Stochastic Differential Equations with Non-Lipschitz Coefficients*, preprint.

[46] Protter, P. (1990), Integration and Stochastic Differential Equations, Springer.

[47] Revuz, D. and Yor, M. (1991), Brownian Motion and Continuous Martingales, Springer.

[48] Rogers, L. C. G. and Williams, D. (1987), Diffusions, Markov Processes, and Martingales, Volume 2: Itô Calculus, John Wiley & Sons.

[49] Willinger, W. and Taqqu, M.S. (1989) *Pathwise Stochastic Integration and Applications to the Theory of Continuous Trading*, Stoch. Proc. Appl., **32**, 253–280.

[50] Yong, J. (1997), *Finding Adapted Solutions of Forward-Backward Stochastic Differential Equations–Method of Continuation*, Prob. Theory and Rel. Fields, **107**, 537-572.

[51] Yong, J. and Zhou, X. (1999), Stochastic Controls: Hamiltonian Systems and HJB Equations, Springer.

[52] Zhang, J. (2000), *A Numerical Scheme for Backward Stochastic Differential Equations*, in preparation.

VITA

VITA

Jianfeng Zhang was born on August 4th, 1973 in Dongyang, Zhejiang, P.R. China. He got his Bachelor's degree in Mathematics at Fudan University in July of 1995. He entered the Mathematics Department of Purdue University in August of 1997 to pursue his doctoral studies.