Efficient Computation of Hedging Portfolios for Options with Discontinuous Payoffs

Jaksa Cvitanić, * Jin Ma, † and Jianfeng Zhang ‡

Abstract

We consider the problem of computing hedging portfolios for options that may have discontinuous payoffs, in the framework of diffusion models in which the number of factors may be larger than the number of Brownian motions driving the model. Extending the work of Fournie et al (1999), as well as Ma and Zhang (2000), using integration by parts of Malliavin calculus, we find two representations of the hedging portfolio in terms of expected values of random variables that do not involve differentiating the payoff function. Once this has been accomplished, the hedging portfolio can be computed by simple Monte Carlo. We find the theoretical bound for the error of the two methods. We also perform numerical experiments in order to compare these methods to two existing methods, and find that no method is clearly superior to others.

*Departments of Mathematics and Economics, USC, Los Angeles 90089-1113. E-mail: cvitanic@math.usc.edu. Research supported in part by NSF grant DMS-97-32810.

†Department of Mathematics, Purdue University, West Lafayette, IN 47907-1395; email: majin@math.purdue.edu. This author is supported in part by NSF grant #9971720.

‡Department of Mathematics, Purdue University, West Lafayette, IN 47907-1395; email: jfzhang@math.purdue.edu. This author is supported in part by NSF grant #9971720.

1 Introduction.

Until quite recently the method mostly used in practice for evaluating hedging portfolios of options in standard diffusion models has been based on the fact that the optimal number of shares to be held is typically obtained by differentiating the option price with respect to the underlying factors: namely, one would compute the price for some initial value of a factor X, increase it by a small amount Δx, find the price for the perturbed factor, compute the difference and divide by Δx. This division usually makes the method much more computationally expensive than finding the option price. Typically, if one uses Monte Carlo with n steps, the error goes from the order of $n^{-1/2}$ to the order of $n^{-1/4}$ or $n^{-1/3}$; see Boyle, Broadie and Glasserman (1997), henceforth [BBG97]. On the other hand, in special cases one can use so-called "direct estimates" that avoid re-simulation for the perturbed initial conditions; see [BBG97] for references. In the recent papers by Fournie et al.(1999,2001), this was generalized to finite-dimensional Markovian models, and shown that the hedging portfolio can be calculated as an expected value of a functional involving the gradient of the option payoff. This brings down the order of the error to $n^{1/2}$, if Monte Carlo is used. Moreover, using Malliavin calculus, they show that this expression can be transformed to avoid the need for computing the gradient of the payoff, which is quite useful, because typical option payoffs are not everywhere differentiable. Ma and Zhang (2000) extended these representations to the models in which the portfolio process enters the drift of the wealth process in a nonlinear fashion. Typical examples for these are markets with different rates for borrowing and lending, or modeling the trading of a large investor that can affect the return rates of traded assets. In the language of Backward Stochastic Differential Equations (BSDEs), they deal with BSDEs with nonlinear drift.

In the present paper, we extend those results to the case of models in which the number of factors may be larger than the number of Brownian motions. This is typically the case in many applications: for instance, in the case of pricing options on one underlying stock driven by one Brownian motion, but with the interest rate and volatility also being stochastic processes driven by the same Brownian motion. We consider two methods. The first one is based on integration by parts in Malliavin calculus, the technique used in above mentioned papers. This enables us to obtain a representation of the hedging portfolio as an expected value that does not involve differentiating the payoff function. Hence, it can be computed by direct Monte Carlo. The second method is as follows: we first artificially increase the number of Brownian motions to match the number of factors, transforming the volatility matrix to a non-singular square matrix by adding the appropriate number of columns with terms of size ε; we can then use the results from the above mentioned papers to get a representation of the hedging portfolios in these artificial markets; finally, we show that letting $\varepsilon \to 0$, these portfolios converge to the hedging portfolio in the original market model.

We show that these methods can be applied even for options with discontinuous payoffs, the prototypical example of which is the digital option, which is the focus of our results. Finally, we confirm by numerical experiments that these procedures provide feasible algorithms for computing hedging portfolios. We compare the two methods, called M and M_ε methods, to other methods applicable to discontinuous payoffs, that allow the number of Brownian motions to be smaller than the number of factors: the standard "delta method", or Δ method, and the "retrieval of volatility method" of Cvitanic, Goukasian and Zapatero (2000), called the RVM method. With the appropriate choice of numerical parameters, the M_ε method has the smallest standard error, but not much smaller than the Δ method and the RVM method. However, it seems that the M_ε method is not as sensitive with respect to the choice of ε as much as the Δ method is, with respect to the size of the perturbation, or as much as the RVM method is, with respect to its free parameters. Moreover, in the example we have only two-dimensional factor process. The computational time of the Δ method increases linearly with the number of factors, since it has to compute approximate derivatives of the option price with respect to all the factors. On the other hand, a disadvantage of the three methods other than the Δ method is that they do not provide the sensitivities of the option price to individual parameters – they only provide the value of the hedging portfolio.

Somewhat surprisingly, the M method has the largest error, for the smae amount of processing tiem. This is in contrast with the payoffs which are continuous and almost everywhere differentiable, in which case the Malliavin calculus based methods are very efficient. On the other hand, we do not have to worry about the choice of any small parameters for the M method, other than the time step size. However, there is a trade-off: the M method requires, in general, the most sophisticated preparation. In particular, it requires computing the value of a Skorohod integral. In fact, the M method requires the smallest number of time steps and simulation paths, but it seems to require a lot of time for computing the Skorohod integral.

Section 2 describes the model and the problem. Section 3 states the main results on the Malliavin calculus method and provides the proofs. Section 4 computes the needed Skorohod integral in terms of standard Ito and Riemann integrals. Section 5 presents the second, approximation based method. Section 6 reports results of numerical experiments.

2 Problem Formulation

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

Consider the following market model: there are d risky assets and 1 riskless asset, whose prices at time t are denoted by $S_t = (S^1_t, \cdots, S^d_t)^T$ and S^0_t, respectively. We assume that the prices follow the following SDE:

$$\begin{cases} dS^0_t = S^0_t r_t dt; \\ dS^i_t = S^i_t [r_t dt + \sum_{j=1}^d \gamma^{ij}_t dW^j_t], \quad i = 1, \cdots, d. \end{cases} \tag{2.1}$$

We suppose that the volatility matrix $\gamma_t = [\gamma^{ij}_t]$ is such that there exists a unique equivalent martingale measure (under which the discounted stock prices are martingales). This is why we have set the drift of the risky assets to be equal to r, i.e., we have assumed that we are in the risk-neutral model.

Let us recall the standard option pricing framework. Suppose that the seller (hereafter called the investor) of the option is trying to replicate the option payoff by investing in the market. We denote by π^i_t the amount of money the investor holds in the stock i at time t, and we denote $\pi = (\pi^1, \ldots, \pi^d)$. A standard argument shows that the wealth process V of the investor will follow the following SDE (under the equivalent martingale measure):

$$V_t = V_0 + \int_0^t r_s V_s ds + \int_0^t \pi_s \gamma_s dW_s. \tag{2.2}$$

Typically, the actual payoff of the option will be given by $\tilde{g}(S_T)$ for some function $\tilde{g}$, and by definition, the (option) price process is equal to the wealth process which replicates the option at the maturity time T, namely we must have $V_T = \tilde{g}(S_T)$.

To simplify these equations, it is common to introduce the *discounted wealth process*. That is, define

$$R_t = \exp\{-\int_0^t r_s ds\}, \qquad Y_t = R_t V_t, \qquad t \geq 0, \tag{2.3}$$

then, using Ito's rule we see that (2.2) becomes

$$Y_t = Y_0 + \int_0^t R_s \pi_s \gamma_s dW_s. \tag{2.4}$$

Since γ and R are both invertible, we can simply set $Z_t = R_t \pi_t \gamma_t$ so that the discounted wealth process Y is now described by a simple form:

$$Y_t = Y_0 + \int_0^t Z_s dW_s. \tag{2.5}$$

In this paper we shall assume that r and γ are components of a finitely-dimensional diffusion process. To be more precise, we shall consider a state process X of the form $X = (S, R, X_{d+2}, \dots, X_n)$, and assume that X is an n-dimensional diffusion which satisfies the following SDE

$$dX_t^i = b_i(t, X_t)dt + \sum_{j=1}^{d} \sigma_{ij}(t, X_t)dW_t^j, i = 1, \dots, n, \tag{2.6}$$

where $(X_1, \dots, X_d) = S$ and $X_{d+1} = R$. We can then set the discounted payoff function to be

$$g(X) = R\tilde{g}(S),$$

and thus the discounted price process Y of the option at each time is given by

$$Y_t = E\{g(X_T)|\mathcal{F}_t\} = Y_0 + \int_0^t Z_s dW_s = g(X_T) - \int_t^T Z_s dW_s. \tag{2.7}$$

It is noted that the triplet (X, Y, Z) is now an $\{\mathcal{F}_t\}$-adapted solution to the *forward-backward SDE* (2.6) and (2.7). We refer the readers to El Karoui, Peng and Quenez (1997) and Ma and Yong (1999) for a complete account regarding the theory of backward/forward-backward SDEs and their applications in finance.

In order to perform hedging in our model, we have to find an efficient numerical method for computing the portfolio π for which $V_T = \tilde{g}(S_T)$ holds. Recasting this in terms of Y and Z, we are trying to find an efficient numerical method to compute the process Z that makes $Y_T = R_T\tilde{g}(S_T) = g(X_T)$. We are particularly interested in the case where g is a discontinuous function. A typical example is the so-called *digital option* (or binary option), that is, $\tilde{g}(s) = 1_{\{s \geq K\}}$, for some $K > 0$.

The main difficulty in the numerical computation is that the discontinuity of g will cause many technical problems using standard arguments via the PDE theory. Secondly, since the dimension of X is n, this increases the dimension of the corresponding PDE to n state variables (plus one time variable), which makes the PDE methods very slow for $n > 2$.

A method which can circumvent these difficulties has been developed by Fournie et al. (1999); however, in that paper it is assumed that the number of factors n is equal to the number of Brownian motions (and the number of stocks) d. This is not the case in many models used in practice, such as the case of one Brownian motion, but with r or γ being random.

To conclude this section we give the *Standing Assumptions* that we will use throughout this paper.

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

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

(**A2**) The function $g : \mathbb{R}^n \mapsto \mathbb{R}$ is a measurable function; and there exists a constant $K > 0$ such that $|g(x)| \le K(1 + |x|)$.

We also introduce the following notation: we represent matrix σ as

$$\sigma(t,x) = \begin{bmatrix} \sigma_1(t,x) \\ \sigma_2(t,x) \end{bmatrix} , \tag{2.8}$$

where σ_1 is a $d \times d$ matrix.

3 Representations of Hedging Portfolios

Recall from the previous section that we are considering the following system of stochastic differential equations:

$$\begin{cases} X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \\ Y_t = g(X_T) - \int_t^T Z_t dW_t. \end{cases} \tag{3.1}$$

Let dimension of X be n, and that of W be d. We shall assume (**A1**) – (**A2**); and that the dimension of Y is 1.

Our method of computing hedging portfolios is based heavily on the new Feynman-Kac type representation of the process Z of a BSDE (see, for example, Ma and Zhang (2000)). However, as we pointed out before, the main difficulty in the present situation is that the dimensions d and n are not equal in general, and that the payoff function may not even be continuous, much less uniform Lipschitz, as was assumed in Ma and Zhang (2000). Therefore, some more careful considerations must be given here.

We begin by the following modified representation theorem for the process Z. Note that we do not require that the matrix σ be a square matrix.

Theorem 3.1 *Assume (**A1**); and assume that the function g is continuous. Denote $A = \{x \in \mathbb{R}^n : \ \partial_x g(x) \text{ does not exist}\}$. Assume further that $P\{X_T \in A\} = 0$ and that $\partial_x g$ is uniformly bounded outside A. Then, it holds that*

$$Z_t = E_t[\partial_x g(X_T)\nabla X_T \mathbf{1}_{\{X_T \notin A\}}][\nabla X_t]^{-1}\sigma(t, X_t). \tag{3.2}$$

In particular,

$$Z_0 = E[\partial_x g(X_T)\nabla X_T \mathbf{1}_{\{X_T \notin A\}}]\sigma(0, x). \tag{3.3}$$

Here, ∇X is the solution to the variational equation

$$\nabla X_t = I + \int_0^t \partial_x b(s, X_s)\nabla X_s ds + \int_0^t \partial_x \sigma(s, X_s)\nabla X_s dW_s, \tag{3.4}$$

and E_t is the expectation conditional on $\mathcal{F}_t$.

Proof. Let $\{g^\varepsilon\}$ be a sequence of molifiers of g. That is, g^ε's are smooth functions such that $\partial_x g^\varepsilon$ are uniformly bounded, $g^\varepsilon \to g$ uniformly, and $\partial_x g^\varepsilon(x) \to \partial_x g(x)$ for all $x \notin A$, as $\varepsilon \to 0$. Since $g^\varepsilon(X_T) \to g(X_T)$, by standard stability results for backward SDEs (cf. e.g., Ma and Yong (1999)) one has

$$E\int_0^T |Z_t^\varepsilon - Z_t|^2 dt \to 0, \qquad \text{as } \varepsilon \to 0. \tag{3.5}$$

Furthermore, since g^ε is differentiable with bounded derivatives, we can apply the representation theorem of Ma and Zhang (2000) to get

$$Z_t^\varepsilon = E_t[\partial_x g^\varepsilon(X_T)\nabla X_T][\nabla X_t]^{-1}\sigma(t, X_t).$$

On the other hand, let us denote the right hand side of (3.2) by $\tilde{Z}_t$. It is easy to see that

$$\begin{aligned} |Z_t^\varepsilon - \tilde{Z}_t| \quad \le \quad & E_t[|\partial_x g^\varepsilon(X_T) - \partial_x g(X_T)||\nabla X_T|\mathbf{1}_{\{X_T\notin A\}}]|[\nabla X_t]^{-1}||\sigma(t, X_t)| \\ & + E_t[|\partial_x g^\varepsilon(X_T)||\nabla X_T|\mathbf{1}_{\{X_T\in A\}}]|[\nabla X_t]^{-1}||\sigma(t, X_t)|, \end{aligned}$$

where $|v| \stackrel{\triangle}{=} [|v_1|, \cdots, |v_n|]^T$ whenever $v = [v_1, \cdots, v_n]^T$. Noting that $P\{X_T \in A\} = 0$, and that $\partial_x g^\varepsilon(X_T) \to \partial_x g(X_T)$ for $X_T \notin A$, applying the dominated convergence theorem we have

$$E\int_0^T |Z_t^\varepsilon - \tilde{Z}_t|^2 dt \to 0.$$

This, together with (3.5), implies that $\tilde{Z} = Z, dt \times dP$-a.s. Finally, note that being the product of a martingale and a continuous process $\tilde{Z}$ has a càdlàg version. Thus as a modification of $\tilde{Z}$, we conclude that Z has a càdlàg version as well, completing the proof of the theorem. ∎

We remark that in Theorem 3.1 the assumption $P\{X_T \in A\} = 0$ plays a crucial role. However, in practice such an assumption is not easy to verify, especially in the case when $d < n$. The following sufficient condition is therefore useful for our future discussion.

Theorem 3.2 *Assume (***A1***), and that g is uniformly Lipschitz in all variables but differentiable with respect to $(x_{d+1}, \cdots, x_n)$. Assume further that $det(\sigma_1(T, X_T)) \neq 0$. Then, $P\{X_T \in A\} = 0$. In particular, (3.2) and (3.3) hold.*

Proof. Let $\hat{X} \stackrel{\triangle}{=} (X^1, \cdots, X^d)^T$. We first show that the law of $\hat{X}_T$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, denoted by $|\cdot|_d$.

To this end, let $\hat{A} = \text{Proj}_{\mathbb{R}^d}(A)$ be the projection of set A on $\mathbb{R}^d$, where A is the set defined in Theorem 3.1. That is,

$$\hat{A} \stackrel{\triangle}{=} \{\hat{x} = (x_1, \cdots, x_d) : \exists (x_{d+1}, \cdots, x_n), \quad \text{such that } x \stackrel{\triangle}{=} (x_1, \cdots, x_n) \in A\}.$$

Since g is Lipschitz continuous on $(x_1, \cdots, x_d)$, and differentiable on $(x_{d+1}, \cdots, x_n)$, we see that $|\hat{A}|_d = 0$.

Next, note that by standard arguments one shows that X_T (whence $\hat{X}_T$) is "Malliavin differentiable", that is, $X_T \in \mathbb{D}^{1,2}$ and

$$D_t X_T = \nabla X_T [\nabla X_t]^{-1} \sigma(t, X_t). \tag{3.6}$$

In particular, we have

$$D_T X_T = \sigma(T, X_T), \quad \text{and} \quad D_T \hat{X}_T = \sigma_1(T, X_T).$$

Define $\tilde{\gamma} \triangleq \int_0^T D_t \hat{X}_T (D_t \hat{X}_T)^T dt$. From (3.6) it is readily seen that $(D_t X_T)$ is continuous in t, and that

$$\det(D_T \hat{X}_T) = \det(\sigma_1(T, X_T)) \neq 0, \quad a.s..$$

Therefore, for $\forall x \in \mathbb{R}^d$ such that $x \neq \mathbf{0}$, $x D_t \hat{X}_T (D_t \hat{X}_T)^T x^T$ is nonnegative for all $t \in [0, T]$ and is positive for t close to T. Thus we have $x \left\{ \int_0^T (D_t X_T)(D_t X_T)^T dt \right\} x^T > 0$, which implies that the symmetric matrix γ has positive determinant. Now we can apply Theorem 2.1.2. of Nualart (1995) to conclude that the law of $\hat{X}_T$ is absolutely continuous with respect to $|\cdot|_d$, which, combined with the fact that $|\hat{A}|_d = 0$, implies that $P(\hat{X}_T \in \hat{A}) = 0$. Since g is differentiable with respect to $x_{d+1}, \cdots, x_n$, we see that $P(X_T \in A) = 0$, and the result follows from Theorem 3.1. ■

So far we have assumed that g is continuous. Motivated by the digital option, we now consider the case where g is allowed to be discontinuous. To the best of our knowledge, the representation theorem in such a situation is new. We state the theorem specifically for an option with one discontinuity point, but the result can easily be extended to an option with finitely many discontinuities.

Theorem 3.3 *Assume $d = 1$, $\sigma_1(t, x) \geq c_0$ and $\sigma, b \in C^{0,2}$ with bounded first and second derivatives. Assume that g is differentiable with respect to $x_2, \cdots, x_n$ with bounded derivatives; and that g is uniformly Lipschitz continuous with respect to x_1, except for the point $x_1 = x_1^*$, and both $g(x_1^*+, x_2, \cdots, x_n)$ and $g(x_1^*-, x_2, \cdots, x_n)$ exist and are differentiable. Then for $t \in [0, T)$, we have*

$$Z_t = E_t \left\{ \partial_x g(X_T) \nabla X_T u_t \mathbf{1}_{\{X_T^1 \notin A\}} + \mathbf{1}_{\{X_T^1 > x_1^*\}} \delta_t(F_t u.) \right\},$$

where A is the same as in Theorem 3.1; $\hat{X}^2 \triangleq (X^2, \cdots, X^n)^T$; $\nabla X = \begin{pmatrix} \nabla X^1 \\ \nabla \hat{X}^2 \end{pmatrix}$ is the solution to (3.4), $\delta_t(\cdot)$ is the indefinite Skorohod integral over $[t, T]$;

$$F_t \triangleq \frac{\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 u_t] \nabla X_T^1}{\|DX_T^1\|_{[t,T]}^2}, \qquad u_t \triangleq [\nabla X_t]^{-1} \sigma(t, X_t), \quad t \in [0, T]; \tag{3.7}$$

and

$$\|DX_T^1\|^2_{[t,T]} \triangleq \int_t^T |D_sX_T^1|^2 ds, \quad \Delta g(x_1^*, \hat{x}^2) \triangleq g(x_1^*+, \hat{x}^2) - g(x_1^*-, \hat{x}^2).$$

To prove the theorem, we need a technical lemma.

Lemma 3.4 *Assume that all the assumptions of Theorem 3.3 are in force. Then for $\forall t < T$, the process $\{F_t u_s\}_{t\le s\le T}$ is Skorohod integrable over $[t,T]$.*

Proof. First note that

$$D_sX_T^1 = \nabla X_T^1 u_s, \quad 0 \le s \le T. \tag{3.8}$$

By the "product property" of Nualart (1995) (pp. 40, property (4)) and the integrability of X and its "derivatives" (see Ma and Zhang (2000), for example), it suffices to show that $\int_t^T (D_sF)u_s ds$ is in $L^2(\Omega)$. In order to show that, by (4.4) below it is sufficient to check that

$$E\left\{\frac{1}{\|DX_T^1\|^{16}_{[t,T]}}\right\} < \infty.$$

To this end, we first note that $D_sX_s^1 = \sigma_1(s, X_s) \ge c_0$, for $\forall s \in [t,T]$. Moreover, it holds that

$$\begin{aligned}
&|D_sX_s^1 - D_sX_T^1| = |[\nabla X_s^1 - \nabla X_T^1]u_s| \le |\nabla X_s^1 - \nabla X_T^1||u_s| \\
&\le C\Big[\int_s^T |\partial_x b \nabla X_r| dr + |\int_s^T \partial_x\sigma\nabla X_r dW_r|\Big]|u_s| \\
&\le C\Big[\Gamma_1(T-s) + \Gamma_2(T-s)^{\frac{1}{3}}\Big]\Gamma_3 \le C[\Gamma_1+\Gamma_2]\Gamma_3(T-s)^{\frac{1}{3}},
\end{aligned} \tag{3.9}$$

where

$$\Gamma_1 \triangleq \sup_{t\le r\le T} |\nabla X_r|; \quad \Gamma_2 \triangleq \sup_{t\le r\le T} \frac{|\int_r^T \partial_x\sigma\nabla X_\tau dW_\tau|}{(T-r)^{\frac{1}{3}}}; \quad \Gamma_3 \triangleq \sup_{t\le r\le T} |u_r|.$$

Therefore one has

$$|D_sX_T^1| \ge c_0 - C[\Gamma_1+\Gamma_2]\Gamma_3(T-s)^{\frac{1}{3}}.$$

Since the right hand side above is strictly increasing in s and has maximun value c_0 (when $s = T$), we can choose a random time (not stopping time!) $\tilde{T} < T$ so that

$$|D_sX_T^1| \ge \frac{c_0}{2}, \qquad \text{for all } s \in [\tilde{T}, T].$$

Consequently

$$\int_t^T |D_sX_T^1|^2 ds \ge \int_{\tilde{T}\vee t}^T |D_sX_T^1|^2 ds \ge \frac{c_0^2}{4}(T - \tilde{T}\vee t),$$

and it follows that

$$\frac{1}{\int_t^T |D_s X_T^1|^2 ds} \le \frac{4}{(T-\tilde{T}\vee t)c_0^2} \le \frac{4}{(T-t)c_0^2} + \frac{4}{(T-\tilde{T})c_0^2} = \frac{4}{(T-t)c_0^2} + C[\Gamma_1^3+\Gamma_2^3]\Gamma_3^3.$$

For $\forall p \ge 2$, by standard arguments one can easily prove that $\Gamma_1, \Gamma_3 \in L^p$, and that $\sup_{t\le r\le T} |\partial_x \sigma \nabla X_r| \in L^p$. Then by Theorem 2.1, Chap.1 of Revuz-Yor [17] we know that $\Gamma_2 \in L^p$, for $\forall p \ge 2$. Now the result follows easily. ∎

Proof of Theorem 3.3.

First we denote $\mathbf{x} = (x_1, x_2, \cdots, x_n) = (x_1, \mathbf{x}_2)$ and recall that $A \stackrel{\triangle}{=} \{x_1 \in \mathbb{R} : g$ is not differentiable at $x_1\}$. Define g^ε be a modification of g as follows.

$$g^\varepsilon(x_1, \mathbf{x}_2) \stackrel{\triangle}{=} \begin{cases} g(x_1, \mathbf{x}_2), & |x_1 - x_1^*| > \varepsilon; \\ \dfrac{(x_1^*+\varepsilon) - x_1}{2\varepsilon} g(x_1^* - \varepsilon, \mathbf{x}_2) + \dfrac{x_1 - (x_1^* - \varepsilon)}{2\varepsilon} g(x_1^* + \varepsilon, \mathbf{x}_2), & \text{otherwise.} \end{cases} \tag{3.10}$$

Then clearly $\lim_{\epsilon\to 0} |g^\varepsilon(\mathbf{x}) - g(\mathbf{x})| = 0$, for all $\mathbf{x}$ except for $\mathbf{x} = (x_1^*, \mathbf{x}_2)$. Now by Theorem 3.2, we have

$$Z_t^\varepsilon = E_t\left\{\partial_x g^\varepsilon(X_T)\nabla X_T \mathbf{1}_{\{X_T^1 \notin A^\varepsilon\}}\right\} u_t,$$

where $A^\varepsilon \stackrel{\triangle}{=} (A \cap \{|x_1 - x_1^*| > \varepsilon\}) \cup \{x_1^* + \varepsilon, x_1^* - \varepsilon\}$. Since Z^ε is the martingale integrand in the solution of BSDE (3.1) with g being replaced by g^ε, the stability result of BSDEs tells us that

$$\lim_{\epsilon\to 0} E\left\{\int_0^T |Z_t^\varepsilon - Z_t|^2 dt\right\} = 0. \tag{3.11}$$

Now note that

$$\begin{aligned} & \partial_x g^\varepsilon(X_T)\nabla X_T \mathbf{1}_{\{X_T^1\notin A^\varepsilon\}} = \left[\partial_1 g^\varepsilon(X_T)\nabla X_T^1 + \partial_{\mathbf{x}_2} g^\varepsilon(X_T)\nabla \hat{X}_T^2\right]\mathbf{1}_{\{X_T^1\notin A^\varepsilon\}} \\ = \; & \frac{1}{2\varepsilon}[g(x_1^*+\varepsilon, \hat{X}_T^2) - g(x_1^* - \varepsilon, \hat{X}_T^2)]\nabla X_T^1 \mathbf{1}_{\{|X_T^1 - x_1^*|<\varepsilon\}} \\ & + \partial_1 g^\varepsilon(X_T)\nabla X_T^1 \mathbf{1}_{\{X_T^1 \in (\{|x_1-x_1^*|>\varepsilon\}\setminus A)\}} + \partial_{\mathbf{x}_2} g^\varepsilon(X_T)\nabla\hat{X}_T^2 \mathbf{1}_{\{X_T^1\notin A^\varepsilon\}} \\ = \; & \frac{1}{2\varepsilon}\Delta g(x_1^*, \hat{X}_T^2)\nabla X_T^1 \mathbf{1}_{\{|X_T^1-x_1^*|<\varepsilon\}} + \frac{1}{2\varepsilon}\mathbf{1}_{\{|X_T^1-x_1^*|<\varepsilon\}} \times \\ & \quad [(g(x_1^*+\varepsilon, \hat{X}_T^2) - g(x_1^*+, \hat{X}_T^2)) + (g(x_1^*-, X_T^2) - g(x_1^*-\varepsilon, \hat{X}_T^2))]\nabla X_T^1 \\ & + \partial_1 g^\varepsilon(X_T)\nabla X_T^1 \mathbf{1}_{\{X_T^1 \in (\{|x_1-x_1^*|>\varepsilon\}\setminus A)\}} + \partial_{\mathbf{x}_2} g^\varepsilon(X_T)\nabla\hat{X}_T^2 \mathbf{1}_{\{X_T^1\notin A^\varepsilon\}} \\ = \; & I_1^\varepsilon + I_2^\varepsilon + I_3^\varepsilon + I_4^\varepsilon, \end{aligned} \tag{3.12}$$

where $I_1^\varepsilon, \ldots, I_4^\varepsilon$ are defined in the obvious way. Clearly

$$\lim_{\epsilon\to 0} I_3^\varepsilon = \partial_1 g(X_T)\nabla X_T^1 \mathbf{1}_{\{X_T^1\notin A\}}; \quad \lim_{\epsilon\to 0} I_4^\varepsilon = \partial_{\mathbf{x}_2} g^\varepsilon(X_T)\nabla\hat{X}_T^2 \mathbf{1}_{\{X_T^1\notin A\}}.$$

Now similar to Theorem 3.2 we can show that the law of X_T^1 has a density, thus $\lim_{\epsilon\to 0} I_2^\varepsilon = 0$. So by (3.11) it suffices to show that

$$\lim_{\epsilon\to 0} E_t\{I_1^\varepsilon u_t\} = E_t\left\{\mathbf{1}_{\{X_T^1 > x_1^*\}}\delta_t(F_t u.)\right\}. \tag{3.13}$$

To do this, for $\forall a < b$, we define $\psi(x) \triangleq \mathbf{1}_{[a,b]}(x)$ and $\varphi(x) \triangleq \int_{-\infty}^x \psi(y)dy$. Then for $t \le s \le T$,

$$D_s\varphi(X_T^1) = \psi(X_T^1)D_sX_T^1.$$

Multiplying both sides above by $D_sX_T^1$ and then integrating over $[t,T]$, we have

$$\int_t^T D_s\varphi(X_T^1)D_sX_T^1 ds = \psi(X_T^1)\int_t^T |D_sX_T^1|^2 ds.$$

Since $F_t u.$ is Skorohod integrable over $[t,T]$, thanks to Lemma 3.4, applying integration by parts formula we get

$$E_t\left\{\psi(X_T^1)\Delta g(x_1^*,\hat{X}_T^2)\nabla X_T^1 u_t\right\} = E_t\left\{\int_t^T D_s\varphi(X_T^1)\frac{D_sX_T^1\Delta g(x_1^*,\hat{X}_T^2)\nabla X_T^1 u_t}{\|DX_T^1\|_{[t,T]}^2}ds\right\}$$

$$= E_t\left\{\int_t^T D_s\varphi(X_T^1)F_t u_s ds\right\} = E_t\left\{\varphi(X_T^1)\delta_t(F_t u.)\right\}.$$

On the other hand, by the Fubini Theorem we have

$$E_t\left\{\psi(X_T^1)\Delta g(x_1^*,\hat{X}_T^2)\nabla X_T^1\right\} = \int_a^b E_t\left\{\mathbf{1}_{\{X_T^1 > y\}}\delta_t(F_t u.)\right\}dy. \tag{3.14}$$

Noting again that the law of X_T^1 has a density, thus the integrand in the right hand side of (3.14) is continuous with respect to y. Thus, letting $[a,b] = [x_1^* - \varepsilon, x_1^* + \varepsilon]$, dividing both sides of (3.14) by ε, and then sending $\varepsilon \to 0$ we obtain (3.13), whence the theorem. ∎

We observe that if in fact the volatility matrix σ is squared, then the degree of difficulty should reduce. The following theorem gives an alternative representation result in such cases. The first version of such a result was given under stronger conditions in Fournie et al. (1999), and extended in Ma and Zhang (1999).

Theorem 3.5 *Assume $d = n$, that* $(\mathbf{A1}), (\mathbf{A2})$ *hold and that the matrix σ is non-degenerate. Assume further that $|A|_d = 0$, where $A \subset \mathbb{R}^d$ is the set of all discontinuity points of g. Then*

$$Z_t = E_t\{g(X_T)N_T^t\}\sigma(t,X_t), \tag{3.15}$$

where $N_T^t \triangleq \dfrac{1}{T-t}\left[\int_t^T (\sigma^{-1}(r,X_r)\nabla X_r)^T dW_r\right]^T [\nabla X_t]^{-1}$. *In particular,*

$$Z_0 = E\{g(X_T)N_T^0\}\sigma(0,x). \tag{3.16}$$

Proof. Let g^ε be smooth molifiers of g, and $(Y^\varepsilon, Z^\varepsilon)$ be solutions to the following BSDE:

$$Y_t^\varepsilon = g^\varepsilon(X_T) - \int_t^T Z_r^\varepsilon dW_r.$$

By Ma and Zhang (2000) we know

$$Z_t^\varepsilon = E_t\{g^\varepsilon(X_T)N_T^t\}\sigma(t, X_t). \tag{3.17}$$

For $\forall x \in A^c$, we have $g^\varepsilon(x) \to g(x)$. Since $d = n$, it is known (see, for example, Nualart (1995), or the proof of Theorem 3.2) that when σ is non-degenerate X_T has a density. Therefore, $P\{X_T \in A\} = 0$ whenever $|A|_d = 0$. Consequently, we have $\lim_{\epsilon\to 0} g^\varepsilon(X_T) = g(X_T)$ a.s. The assumption (**A2**) and the Dominated Convergence Theorem then lead to

$$\lim_{\varepsilon\to 0} E\{|g^\varepsilon(X_T) - g(X_T)|^2\} = 0. \tag{3.18}$$

By the standard stability result for BSDE one has

$$\lim_{\epsilon\to 0} E\left\{\int_0^T |Z_t^\varepsilon - Z_t|^2\right\} = 0. \tag{3.19}$$

On the other hand, applying the Dominated Convergence Theorem we also deduce from (3.17) that

$$\lim_{\epsilon\to 0} E\left\{\int_0^T |Z_t^\varepsilon - E_t\{g(X_T)N_T^t\}\sigma(t, X_t)|^2\right\} = 0,$$

which, combined with (3.19), leads to (3.15), whence (3.16). ∎

4 Computation of the Skorohod Integral

In the case of a digital option, we see from Theorem 3.3 that the representation of the hedging portfolio involves a Skorohod integral, which needs to be dealt with numerically so that the representation is useful in practice. In this section we propose a scheme to compute this Skorohod integral explicitly.

Let us first recall the representation of the hedging strategy Z from Theorem 3.3. Due to the Markovian nature of our setting, we shall consider only $t = 0$. This corresponds to the case of interest in practice: we need to compute the hedging strategy only at the present time for the current trading period, and not for the future periods. In other words, we will be interested in computing the following quantity:

$$Z_0 = E\left\{\partial_x g(X_T)\nabla X_T \sigma(0, x)\mathbf{1}_{\{X_T^1 \notin A\}} + \mathbf{1}_{\{X_T^1 > x_1^*\}}\delta(Fu.)\right\}, \tag{4.1}$$

where $\delta = \delta_0$, $\|DX_T^1\|_H^2 = \|DX_T^1\|_{[0,T]}^2$, and

$$F \triangleq F_0 = \frac{\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)] \nabla X_T^1}{\|DX_T^1\|_H^2}.$$

We remark here that although the Skorohod integral can be approximated by Riemann sums (see Nualart (1995)), in our case the Riemann summand will still contain the Malliavin Derivative DX_T^1, which is quite undesirable in practice. We now try to derive a scheme that involves only computations of Itô integrals and Lebesgue integrals, which can be simulated simultaneously with the underlying assets. Hence this is expected to be much more efficient. To begin our analysis, let us first use integration by parts formula for Skorohod integrals and note that $\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)]$ is a scalar, we have

$$I \triangleq \delta(Fu.) = F \int_0^T u_t dW_t - \int_0^T (D_t F) u_t dt. \tag{4.2}$$

Since u is an adapted process, the integral $\int_0^T u_t dW_t$ is indeed an Itô integral and there is no substantial difficulty in its computation. We now look only at the random variable F and the integral $\int_0^T (D_t F) u_t dt$. But it is easy to see that the only "unusual" part in F, which involves the Malliavin derivative, is $\|DX_T^1\|_H^2$. However, it can be calculated as follows:

$$\begin{aligned} \|DX_T^1\|_H^2 &= \int_0^T |D_t X_T^1|^2 dt = \int_0^T |\nabla X_T^1 [\nabla X_t]^{-1} \sigma(t, X_t)|^2 dt \\ &= \int_0^T |\nabla X_T^1 u_t|^2 dt = \nabla X_T^1 \left\{ \int_0^T u_t u_t^T dt \right\} [\nabla X_T^1]^T. \end{aligned} \tag{4.3}$$

It remains to calculate the integral $\int_0^T (D_t F) u_t dt$. To this end, let us denote $\partial_2 \Delta g = (\partial_{x_2}, \cdots, \partial_{x_n})(\Delta g)$. A direct computation shows that

$$\begin{aligned} (D_t F) u_t &= \frac{1}{\|DX_T^1\|_H^2} \Big\{ [\partial_2 \Delta g(x_1^*, \hat{X}_T^2)) \nabla \hat{X}_T^2 u_t][\nabla X_T^1 \sigma(0,x)][\nabla X_T^1 u_t] \\ &\quad + \Delta g(x_1^*, \hat{X}_T^2)[[D_t \nabla X_T^1] \sigma(0,x)][\nabla X_T^1 u_t] \\ &\quad + \Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)][D_t \nabla X_T^1 u_t] \Big\} \\ &\quad - \Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)][\nabla X_T^1 u_t] \frac{D_t \|DX_T^1\|_H^2}{\|DX_T^1\|_H^4} \\ &= I_1(t) + I_2(t) + I_3(t) - I_4(t), \end{aligned} \tag{4.4}$$

where I_i's are defined in the obvious way. We now analyze $\int_0^T I_i(t) dt$, $i = 1, \cdots, 4$ separately. First,

$$\int_0^T I_1(t) dt = \frac{\nabla X_T^1 \sigma(0,x)}{\|DX_T^1\|_H^2} \int_0^T \partial_2 \Delta g(x_1^*, \hat{X}_T^2) \nabla X_T^2 u_t u_t^T [\nabla X_T^1]^T dt \tag{4.5}$$

$$= \frac{\nabla X_T^1 \sigma(0,x)\partial_2 \Delta g(x_1^*, \hat{X}_T^2)\nabla X_T^2}{\|DX_T^1\|_H^2} \left\{ \int_0^T u_t u_t^T dt \right\} [\nabla X_T^1]^T.$$

Next, since ∇X is the solution to the variational SDE (3.4), we have

$$\begin{aligned} [\nabla X_t]^{-1} &= 1 - \int_0^t [\nabla X_s]^{-1}(\partial_x b - (\partial_x \sigma)^2)(s, X_s)ds + \int_0^t [\nabla X_s]^{-1}\partial_x \sigma(s, X_s)dW_s; \\ D_t \nabla X_u &= \partial_x \sigma(t, X_t)\nabla X_t + \int_t^u [D_t \partial_x b(s, X_s)\nabla X_s + \partial_x b(s, X_s) D_t \nabla X_s]ds \\ &\quad + \int_t^u [D_t \partial_x \sigma(s, X_s)\nabla X_s + \partial_x \sigma(s, X_s) D_t \nabla X_s]dW_s, \quad t \le u \le T. \end{aligned} \tag{4.6}$$

Now applying Itô's formula we get

$$\begin{aligned} \Gamma_u^t &\triangleq [\nabla X_u]^{-1} D_t(\nabla X_u) = [\nabla X_t]^{-1}\partial_x \sigma(t, X_t)\nabla X_t \\ &\quad + \int_t^u [\nabla X_s]^{-1}[D_t \partial_x b(s, X_s) - \partial_x \sigma(s, X_s) D_t \partial_x \sigma(s, X_s)]\nabla X_s ds \\ &\quad + \int_t^u [\nabla X_s]^{-1} D_t \partial_x \sigma(s, X_s)\nabla X_s dW_s. \end{aligned} \tag{4.7}$$

Note that the (i,j)-th entry of the $n \times n$ matrix ∇X is $\nabla_j X^i$. Thus for $i = 1, \cdots, n$ we deduce from (4.7) that

$$D_t \nabla X_u^i = \nabla X_u^i \Gamma_u^t. \tag{4.8}$$

Moreover, since

$$\left\{ \begin{array}{l} D_t(\partial_x b(s, X_s)) = \sum_{k=1}^n [\partial_k(\partial_x b)](s, X_s)[\nabla X_s^k u_t]; \\ D_t(\partial_x \sigma(s, X_s)) = \sum_{k=1}^n [\partial_k(\partial_x \sigma)](s, X_s)[\nabla X_s^k u_t], \end{array} \right. \tag{4.9}$$

where, for $\psi = b, \sigma$, $\partial_k(\partial_x \psi)$ is an $n \times n$ matrix whose (i,j)-th entry is $\partial_{x_k}\partial_{x_j}\psi^i$, we may rewrite (4.7) as

$$\Gamma_u^t = \gamma_t + \sum_{k=1}^n \int_t^u [\nabla X_s^k u_t] \left[\alpha_s^k ds + \beta_s^k dW_s \right], \tag{4.10}$$

where

$$\left\{ \begin{array}{l} \alpha_s^k \triangleq [\nabla X_s]^{-1}[\partial_k(\partial_x b) - \partial_x \sigma \partial_k(\partial_x \sigma)](s, X_s)\nabla X_s \\ \beta_s^k \triangleq [\nabla X_s]^{-1}\partial_k(\partial_x \sigma)(s, X_s)\nabla X_s; \\ \gamma_t \triangleq [\nabla X_t]^{-1}\partial_x \sigma(t, X_t)\nabla X_t. \end{array} \right. \tag{4.11}$$

Combining (4.10) and (4.8), from (4.4) we obtain

$$\begin{aligned} \int_0^T I_2(t)dt &= \frac{\Delta g(x_1^*, \hat{X}_T^2)}{\|DX_T^1\|_H^2} \int_0^T [\nabla X_T^1 u_t] D_t [\nabla X_T^1] dt \sigma(0, x) \\ &= \frac{\Delta g(x_1^*, \hat{X}_T^2)}{\|DX_T^1\|_H^2} \int_0^T \sum_{j=1}^n \nabla_j X_T^1 u_t^j D_t [\nabla X_T^1] dt \sigma(0, x) \\ &= \frac{\Delta g(x_1^*, \hat{X}_T^2)}{\|DX_T^1\|_H^2} \sum_{j=1}^n \nabla_j X_T^1 I_2^j \sigma(0, x), \end{aligned} \tag{4.12}$$

where

$$
\begin{aligned}
I_2^j &\triangleq \int_0^T u_t^j D_t \nabla X_T^1 dt = \int_0^T u_t^j \nabla X_T^1 \Gamma_T^t dt \\
&= \nabla X_T^1 \Big\{ \int_0^T u_t^j \gamma_t dt + \sum_{k=1}^n \int_0^T u_t^j \int_t^T [\nabla X_s^k u_t][\alpha_s^k ds + \beta_s^k dW_s] dt \Big\} \\
&= \nabla X_T^1 \Big\{ \int_0^T u_t^j \gamma_t dt + \sum_{k=1}^n \int_0^T [\nabla X_s^k \int_0^s u_t^j u_t dt] \Big[\alpha_s^k ds + \beta_s^k dW_s\Big] \Big\},
\end{aligned}
$$

thanks to Fubini's Theorem. Using analogous arguments we have

$$
\begin{aligned}
&\int_0^T I_3(t)dt = \frac{\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)]}{\|DX_T^1\|_H^2} \int_0^T D_t \nabla X_T^1 u_t dt \\
&= \frac{\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)]}{\|DX_T^1\|_H^2} \nabla X_T^1 \times \\
&\quad \Big\{ \int_0^T \gamma_t u_t dt + \sum_{k=1}^n \int_0^T \Big[\int_0^s [\nabla X_s^k u_t] u_t dt \Big] \Big[\alpha_s^k ds + \beta_s^k dW_s\Big] \Big\} \\
&= \frac{\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)]}{\|DX_T^1\|_H^2} \nabla X_T^1 \times \\
&\quad \Big\{ \int_0^T \gamma_t u_t dt + \sum_{k,j=1}^n \int_0^T \nabla_j X_s^k \Big[\int_0^s u_t^j u_t dt \Big] \Big[\alpha_s^k ds + \beta_s^k dW_s\Big] \Big\}.
\end{aligned} \tag{4.13}
$$

It remains to analyze $\int_0^T I_4(t)dt$. First, by (4.3) we have

$$
D_t \|DX_T^1\|_H^2 = 2D_t[\nabla X_T^1] \int_0^T u_r u_r^T dr [\nabla X_T^1]^T + 2\nabla X_T^1 \int_t^T (D_t u_r) u_r^T dr [\nabla X_T^1]^T. \tag{4.14}
$$

Note that $\nabla X_r u_r = \sigma(r, X_r)$, we have

$$
(D_t[\nabla X_r]) u_r + \nabla X_r (D_t u_r) = \partial_x \sigma(r, X_r) \nabla X_r u_t.
$$

Thus

$$
D_t u_r = [\nabla X_r]^{-1} \left[\partial_x \sigma(r, X_r) \nabla X_r u_t - (D_t \nabla X_r) u_r \right]. \tag{4.15}
$$

By (4.4) and (4.14), and applying Fubini's theorem again, we have

$$
\begin{aligned}
&\int_0^T I_4(t)dt = \frac{\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)]}{\|DX_T^1\|_H^4} \sum_{j=1}^n \nabla_j X_T^1 \int_0^T u_t^j D_t \|DX_T^1\|_H^2 dt \\
&= \frac{2\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)]}{\|DX_T^1\|_H^4} \sum_{j=1}^n \nabla_j X_T^1 \times \\
&\quad \Big\{ (\int_0^T u_t^j D_t \nabla X_T^1 dt)(\int_0^T u_r u_r^T dr) + \nabla X_T^1 \int_0^T (\int_0^r u_t^j D_t u_r dt) u_r^T dr \Big\} [\nabla X_T^1]^T
\end{aligned} \tag{4.16}
$$

Using (4.15), we can rewrite (4.16) as

$$
\begin{aligned}
&\int_0^T I_4(t)dt = \frac{2\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)]}{\|DX_T^1\|_H^4} \sum_{j=1}^n \nabla_j X_T^1 \times \\
&\left\{ \left(\int_0^T u_t^j D_t \nabla X_T^1 dt \right) \left(\int_0^T u_r u_r^T dr \right) + \nabla X_T^1 \int_0^T \gamma_r \left(\int_0^r u_t^j u_t dt \right) u_r^T dr \right. \qquad (4.17) \\
&\left. -\nabla X_T^1 \int_0^T [\nabla X_r]^{-1} \left(\int_0^r u_t^j D_t \nabla X_r dt \right) u_r u_r^T dr \right\} [\nabla X_t^1]^T \\
&= \frac{2\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)]}{\|DX_T^1\|_H^4} \sum_{j=1}^n \nabla_j X_T^1 \left\{ I_4^{j,1} + I_4^{j,2} - I_4^{j,3} \right\} [\nabla X_T^1]^T,
\end{aligned}
$$

where $I_4^{j,l}, l = 1,2,3$, are defined in an obvious way. In particular,

$$
I_4^{j,2} \triangleq \nabla X_T^1 \int_0^T \gamma_r \left(\int_0^r u_t^j u_t dt \right) u_r^T dr. \tag{4.18}
$$

Recalling (4.10), (4.7) and (4.8), we have

$$
\begin{aligned}
I_4^{j,1} &= \nabla X_T^1 \left\{ \int_0^T u_t^j \gamma_t dt + \sum_{k=1}^n \int_0^T [\nabla X_s^k \int_0^s u_t^j u_t dt] \Big[\alpha_s^k ds + \beta_s^k dW_s \Big] \right\} \Big(\int_0^T u_r u_r^T dr \Big); \\
I_4^{j,3} &= \nabla X_T^1 \int_0^T \left\{ \int_0^r u_t^j \gamma_t dt + \sum_{k=1}^n \int_0^r [\nabla X_s^k \int_0^s u_t^j u_t dt] \Big[\alpha_s^k ds + \beta_s^k dW_s \Big] \right\} u_r u_r^T dr.
\end{aligned} \tag{4.19}
$$

Finally, since (4.17) can be calculated via (4.19) and (4.18), we can compute $\int_0^T (D_t F) u_t dt$ by computing (4.5), (4.12), (4.13) and (4.17). Consequently, combined with (4.3) we can compute the Skorohod integral I in (4.2). ∎

Summary of the algorithm.

We have obtained the following explicit scheme for computing the Skorohod integral in (4.2). First we define some (adapted) processes:

$$
\begin{aligned}
X_t &= x + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s; \\
\nabla X_t &= I + \int_0^t \partial_x b(s, X_s) \nabla X_s ds + \int_0^t \partial_x \sigma(s, X_s) \nabla X_s dW_s; \\
[\nabla X_t]^{-1} &= I - \int_0^t [\nabla X_s]^{-1} \Big[\partial_x b - (\partial_x \sigma)^2 \Big](s, X_s) ds - \int_0^t [\nabla X_s]^{-1} \partial_x \sigma(s, X_s) dW_s; \\
u_t &= [\nabla X_t]^{-1} \sigma(t, X_t); \quad A_t = \int_0^t u_s dW_s; \quad B_t = \int_0^t u_s u_s^T ds; \\
\alpha_t^k &= [\nabla X_t]^{-1} [\partial_k \partial_x b - \partial_x \sigma \partial_k \partial_x \sigma](t, X_t) \nabla X_t; \\
\beta_t^k &= [\nabla X_t]^{-1} \partial_k \partial_x \sigma(t, X_t) \nabla X_t;
\end{aligned}
$$

$$\begin{aligned}
\gamma_t &= [\nabla X_t]^{-1}\partial_x\sigma(t,X_t)\nabla X_t;\\
C_t^j &= \int_0^t u_s^j\gamma_s ds + \sum_{k=1}^n \int_0^t [\nabla X_s^k B_s^j]\Big[\alpha_s^k ds + \beta_s^k dW_s\Big];\\
H_t &= \int_0^t \gamma_s u_s ds + \sum_{j,k=1}^n \int_0^t \nabla_j X_s^k B_s^j\Big[\alpha_s^k ds + \beta_s^k dW_s\Big];\\
L_t^j &= \int_0^t \Big[\gamma_s A_s^j - C_s^j u_s\Big]u_s^T ds;
\end{aligned}$$

Then we have

$$\begin{aligned}
I \;=\; & \frac{1}{\nabla X_T^1 B_T[\nabla X_T^1]^T}\Big\{\Delta g(x_1^*,\hat{X}_T^2)[\nabla X_T^1\sigma(0,x)][\nabla X_T^1 A_T]\\
& -[\nabla X_T^1\sigma(0,x)]\partial_2\Delta g(x_1^*,\hat{X}_T^2)\nabla\hat{X}_T^2 B_T[\nabla X_T^1]^T \\
& -\Delta g(x_1^*,\hat{X}_T^2)\sum_{j=1}^n \nabla_j X_T^1\nabla X_T^1 C_T^j\sigma(0,x) - \Delta g(x_1^*,\hat{X}_T^2)[\nabla X_T^1\sigma(0,x)]\nabla X_T^1 H_T\Big\}\\
& +\frac{2\Delta g(x_1^*,\hat{X}_T^2)[\nabla X_T^1\sigma(0,x)]}{\Big(\nabla X_T^1 B_T[\nabla X_T^1]^T\Big)^2}\sum_{j=1}^n \nabla_j X_T^1\nabla X_T^1\Big\{C_T^j B_T + L_T^j\Big\}[\nabla X_T^1]^T.
\end{aligned} \tag{4.20}$$

5 A Perturbation Method

In this section we propose another method that can be applied when $d \neq n$. Our numerical experiments show that this method may be more efficient than the one described in previous sections. It is also conceptually easier to understand and to program. However, it is sensitive to the choice of the perturbation size.

The idea is to perturb the system a little bit so that the representation theorem of Ma and Zhang (2000) can be applied, and then using the idea similar to that of "vanishing viscosity" to obtain a numerical approximation for the desired hedging portfolio. More precisely, let W^0 be an $(n-d)$-dimensional standard Brownian motion defined on $(\Omega,\mathcal{F},P)$, such that it is independent of W. Let $\widetilde{W}_t \stackrel{\triangle}{=} (W_t^T,(W_t^0)^T)^T$, and denote $\widetilde{\mathbf{F}}$ to be the filtration generated by $\widetilde{W}$. For each $\varepsilon > 0$ we define

$$\sigma^\varepsilon(t,x) \stackrel{\triangle}{=} \begin{bmatrix} \sigma_1(t,x) & 0 \\ \sigma_2(t,x) & \varepsilon I_{(n-d)\times(n-d)} \end{bmatrix}. \tag{5.1}$$

Then it is clear that $\sigma^\varepsilon(t,x) > 0$ for all $(t,x)\in[0,T]\times\mathbb{R}^n$, and $\sigma^\varepsilon(t,x)\to\sigma^o(t,x)$, as $\varepsilon\to 0$, where $\sigma^0(t,x) \stackrel{\triangle}{=} [\sigma(t,x) \vdots 0]$.

Now consider the following perturbed version of (3.1):

$$\begin{cases} X_t^\varepsilon = x + \displaystyle\int_0^t b(s,X_s^\varepsilon)ds + \int_0^t \sigma^\varepsilon(s,X_s^\varepsilon)d\widetilde{W}_s, \\ Y_t^\varepsilon = g(X_T^\varepsilon) - \displaystyle\int_t^T Z_t^\varepsilon d\widetilde{W}_t. \end{cases} \tag{5.2}$$

By the stability results for both forward and backward SDEs we know that, as $\varepsilon \to 0$, the following limit must hold (see Ma and Yong (1999), for example):

$$E\left\{\sup_{0\le t\le T}|X_t^\varepsilon - X_t^0|^2 + \sup_{0\le t\le T}|Y_t^\varepsilon - Y_t^0|^2\right\} + E\int_0^T |Z_t^\varepsilon - Z_t^0|^2 dt \longrightarrow 0, \tag{5.3}$$

where (X^0, Y^0, Z^0) satisfies the following SDE:

$$\begin{cases} X_t^0 = x + \displaystyle\int_0^t b(s, X_s^0)ds + \int_0^t \sigma^0(s, X_s^0)d\widetilde{W}_s, \\ Y_t^0 = g(X_T^0) - \displaystyle\int_t^T Z_t^0 d\widetilde{W}_t. \end{cases} \tag{5.4}$$

Since $\sigma^0(t, X_t^0)d\widetilde{W}_t = \sigma(t, X_t^0)dW_t$, by uniqueness of the solution to the SDE we have $X_t \equiv X_t^0$, $\forall t \ge 0$, a.s. Thus by the uniqueness of the backward SDE (or the martingale representation theorem) we conclude that Z^0 must be of the form $Z_t^0 = (Z_t, 0)$, where Z is the solution of (3.1). Hence (5.4) is indeed (3.1), and consequently we must have

$$E\int_0^T |Z_t^{\varepsilon,1} - Z_t|^2 dt \longrightarrow 0, \qquad \text{as } \varepsilon \to 0, \tag{5.5}$$

where $Z^\varepsilon = (Z^{\varepsilon,1}, Z^{\varepsilon,2})$ is the solution to (5.2).

We have thus proved the first part of the following

Theorem 5.1 *Assume (**A1**) and that g is bounded and piecewise continuous. Then the hedging portfolio process Z can be approximated by $\{Z^{\varepsilon,1}\}$, in the sense of (5.5). Furthermore, if g is Lipschitz continuous (not necessarily bounded), then there exist a constant $K > 0$, independent of ε, such that*

$$E\int_0^T |Z_t^{\varepsilon,1} - Z_t|^2 dt \le K\varepsilon^2 T e^{KT}. \tag{5.6}$$

Proof. Let $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon)$ and (X^0, Y^0, Z^0) be the solutions to (5.2) and (5.4), respectively. Define $\delta\xi \stackrel{\triangle}{=} \xi^\varepsilon - \xi^0$, for $\xi = X, Y, Z$. By Itô's formula we have, for each $t \in [0, T]$,

$$|\delta Y_T^\varepsilon|^2 = |\delta Y_t^\varepsilon|^2 + 2\int_t^T \delta Y_s^\varepsilon \delta Z_s^\varepsilon d\widetilde{W}_s + \int_t^T |\delta Z_s^\varepsilon|^2 ds.$$

Setting $t = 0$ and taking expectation in the above we have

$$E\int_0^T |\delta Z_s^\varepsilon|^2 + E|\delta Y_0^\varepsilon|^2 = E|g(X_T^\varepsilon) - g(X_T^0)|^2. \tag{5.7}$$

Now note that δX_t^ε satisfies the following SDE:

$$\begin{aligned} \delta X_t^\varepsilon &= \int_0^t [b(s, X_s^\varepsilon) - b(s, X_s^0)]ds + \int [\sigma^\varepsilon(s, X_s^\varepsilon) - \sigma^0(s, X_s^0)]d\widetilde{W}_s \\ &= \int_0^t \delta b_s^\varepsilon X_s^\varepsilon ds + \int_0^t \delta\sigma_s^\varepsilon X_s^\varepsilon dW_s + \varepsilon W_t, \end{aligned} \tag{5.8}$$

where, with $\eta = b, \sigma$,

$$\delta\eta_s^\varepsilon \triangleq \begin{cases} \dfrac{\eta(s, X_s^\varepsilon) - \eta(s, X_s^0)}{X_s^\varepsilon - X_s^0}, & X_s^e - X_s^0 \neq 0 \\ 0 & X_s^\varepsilon - X_s^0 = 0. \end{cases}$$

Clearly, both δb^ε and $\delta\sigma^\varepsilon$ are uniformly bounded processes, thanks to (A1). Thus it is readily seen that

$$E|\delta X_t^\varepsilon|^2 \leq K \left\{ \int_0^t E|\delta X_s^\varepsilon|^2 ds + \varepsilon^2 t \right\}.$$

A simple application of Gronwall's inequality yields that

$$E|\delta X_t^\varepsilon|^2 \leq K\varepsilon^2 t e^{Kt}, \qquad \forall t \in [0, T]. \tag{5.9}$$

Consequently, if g is piecewise continuous and uniformly bounded, then applying the Bounded Convergence Theorem we obtain (5.5). If further g is uniformly Lipschitz, then combining (5.7) and (5.9) we obtain (5.6). ∎

The problem can now be reduced to computing $Z^{\varepsilon,1}$. Hence, we can use the results for $d = n$ from Fournie et al. (1999) and Ma and Zhang (2000). In particular, assuming that σ_1 is non-degenerate and that g is Lipshitz continuous, we have

$$Z_t^{\varepsilon,1} = E_t \left\{ g(X_T^\varepsilon) N_T^{\varepsilon,t} \right\} \sigma(t, X_t^\varepsilon), \tag{5.10}$$

where

$$N_r^{\varepsilon,t} = \frac{1}{r-t} \left[\int_t^r [\sigma^\varepsilon(s, X_s^\varepsilon)]^{-1} \nabla X_s^\varepsilon d\widetilde{W}_s \right] (\nabla X_t)^{-1}; \tag{5.11}$$

and ∇X^ε satisfies the linear SDE

$$\nabla_i X_t^\varepsilon = e_i + \int_0^t \partial_x b(s, X_s^\varepsilon) \nabla X_s^\varepsilon ds + \sum_{j=1}^d \int_0^t [\partial_x \sigma^j(s, X_s^\varepsilon)] \nabla_i X_s^\varepsilon dW^j, \quad i = 1, \cdots, n, \tag{5.12}$$

where $e_i = (0, \cdots, \overset{i}{1}, \cdots, 0)^T$ is the i-th coordinate vector of $\mathbb{R}^n$ and $\sigma^j(\cdot)$ is the j-th column of the matrix $\sigma(\cdot)$.

When g is not Lipshitz continuous, we consider the case $d = 1$, and assume that the discounted payoff function is of the form $g(X) = R\tilde{g}(S)$, where $\tilde{g}$ is a piecewise linear function:

$$\tilde{g}(x) = \sum_{i=1}^K \tilde{g}^i(x) 1_{[a_{i-1}, a_i)}(x), \tag{5.13}$$

where $\tilde{g}^i(x) = A^i x + B^i$, $i = 1, \cdots, K$.

We first give the following approximation lemma.

Lemma 5.2 *Under the above assumption there exists a sequence of smooth functions $\tilde{g}_k(x)$ such that for each k, $\tilde{g}_k'$ is uniformly bounded, and that $\tilde{g}_k'(x) \to \tilde{g}'(x)$, and $\tilde{g}_k(x) \to \tilde{g}(x)$ for all $x \in \mathbb{R} \setminus \cup_{i=1}^K \{a_i\}$, as $k \to \infty$.*

Proof . We only sketch the proof. Note that we can formally write the derivative of $\tilde{g}$ as

$$\tilde{g}'(x) = \sum_{i=1}^{K} \Big\{ A^i 1_{[a_{i-1},a_i)}(x) + [\tilde{g}^i(a_{i-1} - \tilde{g}^{i-1}(a_{i-1})]\delta_{\{a_{i-1}\}}(x) \Big\}. \tag{5.14}$$

For each i, let $\{\varphi_i^k(\cdot)\}$ be a smooth δ-sequence at a_{i-1}, for example,

$$\varphi_i^k(x) = \begin{cases} c_k \exp\left\{ -\dfrac{1}{1-k^2|x-a_{i-1}|^2} \right\}, & |x-a_{i-1}| < \frac{1}{k} \\ 0 & |x-a_{i-1}| \geq \frac{1}{k}. \end{cases}$$

and define

$$\tilde{g}_k'(x) = \sum_{i=1}^{K} \Big\{ A^i 1_{[a_{i-1},a_i)}(x) + [\tilde{g}^i(a_{i-1} - \tilde{g}^{i-1}(a_{i-1})]\varphi_i^k(x) \Big\}, \tag{5.15}$$

and $\tilde{g}_k(x) = \int_{-\infty}^x \tilde{g}_k'(y)dy$. Some additional straightforward calculations show that $\tilde{g}_k$ satisfy the stated properties.

Corollary 5.3 *Assume (**A1**), (**A2**), that σ_1 is non-degenerate and that the FBSDE (5.2) has a unique adapted solution $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon)$. Then the following relation holds:*

$$Z_t^\varepsilon = E_t\{g(X_T^\varepsilon)N_T^{\varepsilon,t}\}\sigma(t, X_t^\varepsilon).$$

Proof. Let $\{\tilde{g}_k\}$ be the smooth sequence as in Lemma 5.2 that approximates $\tilde{g}(p)$. Let $(X^{\varepsilon,k}, Y^{\varepsilon,k}, Z^{\varepsilon,k})$ be the solution to (5.2) with $g(X_T^\varepsilon)$ being replaced by $g_k(X_T^\varepsilon) = R_T^\varepsilon \tilde{g}_k(P_T^\varepsilon)$. Since the forward equation is not changed, we have $X^\varepsilon = X^{\varepsilon,k}$. Thus, applying (5.10) we have

$$Z_t^{\varepsilon,k} = E_t\{g_k(X_T^\varepsilon)N_T^{\varepsilon,t}\}\sigma^\varepsilon(t, X_t^\varepsilon).$$

We claim that $E|g_k(X_T^\varepsilon) - g(X_T^\varepsilon)|^2 \to 0$, as $k \to \infty$. Indeed, since X^ε is a (time-homogeneous) diffusion, we let $p^\varepsilon(x, y, t)$ be its transition density. Then

$$E_{t,x}|g_k(X_T^\varepsilon) - g(X_T^\varepsilon)|^2 = \int_{\mathbb{R}^n} |g_k(y) - g(y)|^2 p^\varepsilon(x, y, T-t)dy.$$

Since $g_k \to g$ holds except for only finitely many points, and by construction (5.15) $\tilde{g}_k'$ is bounded, uniformly for all k, one can easily check that, for any $x \in \mathbb{R}^n$, it holds that

$$|g_k(x)| + |g(x)| \leq K(1 + |x|^2).$$

Recalling (5.2) and noting that b and σ^ε are uniformly Lipschitz, by standard arguments on SDE one can easily show that $E_{t,x}|X_T^\varepsilon|^4 < \infty$, which implies that $\int_{\mathbb{R}^n} |y|^4 p^\varepsilon(s, y, T-t)dy < \infty$. Now we can apply the Dominated convergence theorem to conclude that $g_k(X_T^\varepsilon) \to g(X_T^\varepsilon)$ in $L^2(\Omega)$ for each $\varepsilon > 0$. Thus, by the stability results for backward SDEs, we know that $(Y^{\varepsilon,k}, Z^{\varepsilon,k}) \to (Y^\varepsilon, Z^\varepsilon)$ in the sense of (5.3). Therefore, P-almost surely, one has

$$Z_t^\varepsilon = \lim_{k\to\infty} Z_t^{\varepsilon,k} = \lim_{k\to\infty} E_t\{g_k(X_T^\varepsilon)N_T^{\varepsilon,t}\}\sigma^\varepsilon(t, X_t^\varepsilon).$$

■

6 Numerical Experiments.

In this section we compare four different methods for computing the hedging portfolio of a digital option:

- 1. The Malliavin calculus method using the Skorohod integral representation from the first part of this paper. Let us call it M method, for Malliavin.

- 2. The Malliavin calculus method using the approximation of σ. Let us call it M_ε method.

- 3. The "Retrieval of Volatility Method" of Cvitanic, Goukasian and Zapatero (2000), called RVM method. This method is based on the fact that the hedging portfolio process Z can be retrieved from the quadratic variation process of the process Y.

- 4. The standard Δ method, called here Δ. In this method one computes (by finite differences approximation and simulation) the derivatives of the price process Y at initial time with respect to the initial conditions $(X_t^1, \ldots, X_t^n) = (x_1, \ldots, x_n)$, and uses the fact (from Ito's lemma) that the value Z_t is determined from these derivatives and the matrix σ (See Boyle, Broadie and Glasserman (1997) for a survey).

We consider the following simple stochastic volatility extension of the Black-Scholes model:

$$S_t = S_0 + \int_0^t \sigma_r S_r dW_r;$$

$$\sigma_t = \sigma_0 + \int_0^t k_\sigma(\bar{\sigma} - \sigma_r)dr + \int_0^t \rho_\sigma \sigma_r dW_r.$$

The chosen parameter values are:

$$S_0 = 100, \sigma_0 = 0.1, k_\sigma = 0.695, \rho_\sigma = 0.21, \bar{\sigma} = 0.1$$

Note that we assume, for simplicity, that the interest rate is zero. We consider the digital option with payoff

$$g(S_T) = \mathbf{1}_{\{S_T > K\}}$$

with

$$T = .2, \quad K = 100.$$

The simulation of the paths of the underlying processes is done using the first order, Euler scheme. Denoting the number of time steps by N, the number of simulated paths is set to N^2 (this is of the optimal order for Euler scheme; see Duffie and Glynn (1995)). The mean portfolio value and the standard error are obtained by repeating the procedure 1000 times. The processing time is the total time for all the repetitions. The results are reported in Tables 1 and 2. The column labeled "portfolio" gives the number of shares to be held by the hedging portfolio.

Table 1 is computed so that all the processing times are similar. The RVM method involves computing an additional conditional expectation, conditioned on the first time period dt, using K_{RVM} simulated paths. We set

$$N_M = 45, \quad N_{RVM} = 80, \quad K_{RVM} = 2, \ N_\Delta = 100, \ N_{M_\varepsilon} = 80, \ \varepsilon = .05,$$

$$\Delta S = .01 S_0, \quad \Delta\sigma = .01\sigma_0, \; dt = .05.$$

Numerical experiments show that the standard error of the Δ method is very much sensitive to the choice of ΔS, $\Delta\sigma$. In particular, if these are not chosen carefully, we may achieve very small error, but with high bias.

With the above choice of parameters, the Δ method and the M_ε method have somewhat smaller standard error than the other two methods, for a similar processing time. However, it seems that the M_ε method is not that sensitive to the choice of ε. Moreover, in this example we have only two-dimensional factor process (S, σ). The computational time of the Δ method increases linearly with the number of factors, since it has to compute approximate derivatives of the option price with respect to all the factors. The RVM method has somewhat larger error than the Δ and M_ε methods, but not much. It is, however, also very sensitive to the choice of parameters dt and K_{RVM}. Surprisingly, the M method does worst here, even though not significantly worse. On the other hand, we do not have to worry about the choice of any small parameters for the M method. In fact, the M method uses the smallest number of time steps and simulated paths, but, apparently, takes a lot of time computing additional quantities such as the Skorohod integral.

Table 2 was computed so as to achieve a similar standard error for all methods. This was done with

$$N_M = 80, \quad N_{RVM} = 100, \quad K_{RVM} = 2, \; N_\Delta = 125, \; N_{M_\varepsilon} = 90, \; \varepsilon = .05,$$

$$\Delta S = .01 S_0, \quad \Delta\sigma = .01\sigma_0, \quad dt = .05.$$

The results are similar as in Table 1: The M method requires the longest time, even though it uses a smaller number of steps and simulated paths. The other three methods are comparable – they all need the amount of processing time of the same order.

The results are somewhat surprising considering that the direct method of Theorem 3.1 is typically more efficient than the Δ method. We see in this example that the Δ method is not inferior if its parameters are carefully chosen to avoid high bias and high standard error.

Table 2: Similar processing time			
Method	Error	Processing Time	Portfolio
M	0.003019	3414	0.088906
RVM	0.002057	3138	0.088750
M_ε	0.001647	3447	0.088843
Δ	0.001884	3379	0.088061

Table 1: Similar standard error			
Method	Error	Processing Time	Portfolio
M	0.001623	19095	0.088832
RVM	0.001580	6103	0.088755
M_ε	0.001495	4888	0.088832
Δ	0.001530	6509	0.088372

References

[1] Bermin, H-P. (1998), *Hedging options: the Malliavin Calculus approach versus the Δ-hedging approach*, preprint.

[2] Boyle, P., M. Broadie and P. Glasserman, 1997, "Monte Carlo Methods for Security Pricing," *Journal of Economic Dynamics and Control,* 21, 1267–1321.

[3] Cvitanić, J., Goukasian, L. and F. Zapatero, 1999, "Hedging with Monte Carlo simulation," working paper, USC.

[4] Dellacherie, C. and Meyer, P. (1978), *Probabilities and Potential*, North Holland.

[5] Duffie, D. and P. Glynn, 1995, "Efficient Monte Carlo Simulation of Security Prices," *The Annals of Applied Probability* 5, 897-905.

[6] El Karoui, N., Peng, S., and Quenez, M.C. (1997), *Backward stochastic differential equations in finance*, *Mathematical Finance*, **7**, 1-72.

[7] Fournié, E., Lasry, J-M., Lebuchoux, J., Lions, P-L., and Touzi, N. (1999) *An application of Malliavin calculus to Monte Carlo methods in Finance*, *Finance & Stochastics*, **391**, 391-412.

[8] Fournié, E., Lasry, J-M., Lebuchoux, J. and Lions, P-L. (2001) *Applications of Malliavin calculus to Monte Carlo methods in Finance II*, *Finance & Stochastics*, to appear.

[9] Karatzas, I. and Shreve, S.E. (1987), *Brownian Motion and Stochastic Calculus*, Springer.

[10] 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.

[11] Ma, J. and Yong, J. (1999), *Forward-Backward Stochastic Differential Equations and Their Applications*, Lecture Notes in Math., **1702**, Springer.

[12] Ma, J. and Zhang, J. (2000) *Representation theorems for backward stochastic differential equations*, Preprint.

[13] Nualart, D. (1995) *The Malliavin Calculus and Related Topics*, Springer.

[14] Pardoux, E. and Peng S. (1990), *Adapted solutions of backward stochastic equations*, *System and Control Letters*, **14**, 55-61.

[15] Pardoux, E. and Peng S. (1992), *Backward stochastic differential equations and quasilinear parabolic partial differential equations*, *Lecture Notes in CIS*, Springer, **176**, 200-217.

[16] Protter, P. (1990), *Integration and Stochastic Differential Equations*, Springer.

[17] Revuz, D. and Yor, M. (1991), *Brownian Motion and Continuous Martingales*, Springer.