Efficient Computation of Hedging Portfolios for Options with Discontinuous Payoffs

Jakša 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.

Key Words: Hedging options, Malliavin calculus, Monte Carlo methods.
JEL classification: C15, G11. **AMS Subject Classification:** 90A09, 90A46.

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

*Department of Mathematics, USC, 1042 W Downey Way, MC 1113, Los Angeles 90089-1113. E-mail: cvitanic@math.usc.edu. Research supported in part by NSF grant DMS-00-99549.

†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.

‡School of Mathematics, University of Minnesota, Minneapolis, MN 55455; email: jfzhang@math.umn.edu. This author is supported in part by NSF grant #9971720.

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 models in which the portfolio process might enter the drift of the wealth process in a nonlinear fashion.

The models considered in this paper have the following two main features: 1) the number of factors may be larger than the number of Brownian motions, and 2) the payoff is discontinuous. A typical model with feature 1) is one in which the underlying stock driven by one Brownian motion, but the interest rate and volatility are also diffusion processes driven by the same Brownian motion. The prototypical example of the feature 2) is the digital option, which will be the focus of our results.

We shall consider two numerical methods for computing the hedging portfolios. The first one is based on integration by parts in Malliavin calculus, a technique used in the aforementioned papers. This leads to a new representation of the hedging portfolio that is not covered by any existing result. Since such a representation by nature does not involve differentiating the payoff function, it can then 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 by perturbing the volatility matrix to a non-singular (square) one with appropriate number of additional columns, indexed by ε. We then use the results from the aforementioned papers to obtain a representation of the hedging portfolios in the "artificial markets". Finally, we show that as $\varepsilon \to 0$ these portfolios converge to the hedging portfolio in the original market model.

We then apply these methods to options with discontinuous payoffs, and confirm by numerical experiments that these procedures provide feasible algorithms for computing hedging portfolios. We compare our two methods, called M and M_ε methods, respectively, to other methods that are applicable to models with discontinuous payoffs and with the number of Brownian motions being smaller than that of factors. These include the standard finite difference "delta method", or Δ method, and the "Retrieval of Volatility Method" of Cvitanic, Goukasian and Zapatero (2001), which we call the RVM method. We show that, 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 same amount of processing time. 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. 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.

In the special framework of models for LIBOR rates Glasserman and Zhao (1999) address similar issues, but they focus on the computation of "greeks", and not on the hedging portfolio, using methods different from ours. The conclusions they derive from their numerical experiments seem to be consistent with ours.

The rest of the paper is organized as follows. Section 2 describes the model and the problem. Section 3 gives the new prepresentation formulae for hedging portfolios. Section 4 computes the Skorohod integral involved in a representation. Section 5 presents the second, approximation based method, and 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 $\mathcal{F} = \mathcal{F}_\infty$. 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 $\partial_x \psi \stackrel{\triangle}{=} (\partial_{x_i} \psi^j)^n_{i,j=1}$ 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 $\gamma_t = [\gamma_t^{ij}]$ is invertible and the discounted stock prices are martingales. 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_t^i the amount of money the investor holds in the stock i at time t, and we denote $\pi = (\pi^1, \dots, \pi^d)$. Typically, the payoff of the option is 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.

The *discounted price process* satisfies $Y_t = Y_0 + \int_0^t R_s \pi_s \gamma_s dW_s$, where $R_t = \exp\{-\int_0^t r_s ds\}$. 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:

$$Y_t = Y_0 + \int_0^t Z_s dW_s. \tag{2.2}$$

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

$$dX_t^i = b_i(t, X_t)dt + \sum_{j=1}^{d} \sigma_{ij}(t, X_t) dW_t^j, \qquad i = 1, \dots, n, \tag{2.3}$$

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

$$Y_t = E\{g(X_T)|\mathcal{F}_t\} = Y_0 + \int_0^t Z_s dW_s = g(X_T) - \int_t^T Z_s dW_s. \tag{2.4}$$

It is noted that the triplet (X, Y, Z) is now an $\{\mathcal{F}_t\}$-adapted solution to the *forward-backward SDE* (2.3) and (2.4). 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 SDE's 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, or, equivalently, 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**) $n \ge d$. 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

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

(**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 the $(n \times d)$ matrix σ as

$$\sigma(t,x) = \begin{bmatrix} \sigma_1(t,x) \\ \sigma_2(t,x) \end{bmatrix}, \tag{2.5}$$

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 $(\mathbf{A1}) - (\mathbf{A2})$; and that the dimension of Y is 1. Our method of computing hedging portfolios is based heavily on the Feynman-Kac type representation of the process Z of a BSDE, as in Ma and Zhang (2000). However, our payoff function may not even be continuous, much less uniformly Lipschitz, as was assumed in Ma and Zhang (2000). For the sake of completeness, we begin by the following modified representation theorem for the process Z, which can be regarded as a special case of the Clark-Ocone formula. 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 $A = \{x \in \mathbb{R}^n :\ \partial_x g(x)$ does not exist$\}$. Assume further that $P\{X_T \in A\} = 0$ and that $\partial_x g$ is* *uniformly bounded outside A. Then, we have*

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

Here, ∇X is the solution to the variational equation

$$\nabla X_t = I_{n\times n} + \int_0^t \partial_x b(s, X_s)\nabla X_s ds + \int_0^t \partial_x \sigma(s, X_s)\nabla X_s dW_s, \tag{3.3}$$

where $I_{n\times n}$ denotes the $n \times n$ identity matrix, 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 SDE's (cf. e.g., Ma and Yong (1999)) one has

$$E\int_0^T |Z_t^\varepsilon - Z_t|^2 dt \to 0, \qquad \text{as } \varepsilon \to 0. \tag{3.4}$$

Furthermore, since g^ε is differentiable with bounded derivatives, we can apply the representation theorem of Ma and Zhang (2000) (or Clark-Ocone formula) 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

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

where $|v| \stackrel{\Delta}{=} [|v_1|, \cdots, |v_n|]^T$ 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 (3.4), 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, and 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) holds.*

Proof. Let $\hat{X} \triangleq (X^1, \cdots, X^d)^T$. 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,

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

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

$$D_t X_T = \nabla X_T [\nabla X_t]^{-1} \sigma(t, X_t). \tag{3.5}$$

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 (3.5) 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 $\tilde{\gamma}$ 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. ∎

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. The proof is motivated by Proposition 2.1.1 in Nualart (1995).

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 (3.3), $\delta_t(\cdot)$ is the indefinite Skorohod integral over $[t,T]$;

$$F_t \triangleq \frac{\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 u_t] \nabla X_T^1}{\|DX_T^1\|_{[t,T]}^2}, \qquad u_t \triangleq [\nabla X_t]^{-1}\sigma(t, X_t), \quad t \in [0,T]; \tag{3.6}$$

and

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

To prove the theorem, we need a technical lemma, whose proof we omit.

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

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 $A \triangleq \{x_1 \in \mathbb{R} : g \text{ is not differentiable at } x_1\}$. Define g^ε be a modification of g as follows.

$$g^\varepsilon(x_1, \mathbf{x}_2) \triangleq \begin{cases} g(x_1, \mathbf{x}_2), & |x_1 - x_1^*| > \varepsilon; \\ \frac{(x_1^*+\varepsilon) - x_1}{2\varepsilon} g(x_1^* - \varepsilon, \mathbf{x}_2) + \frac{x_1 - (x_1^* - \varepsilon)}{2\varepsilon} g(x_1^* + \varepsilon, \mathbf{x}_2), & \text{otherwise.} \end{cases} \tag{3.7}$$

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 $A^\varepsilon \triangleq (A \cap \{|x_1 - x_1^*| > \varepsilon\}) \cup \{x_1^* + \varepsilon, x_1^* - \varepsilon\}$. 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 BSDE's tells us that

$$\lim_{\epsilon \to 0} E\left\{\int_0^T |Z_t^\varepsilon - Z_t|^2 dt\right\} = 0. \tag{3.8}$$

Now note that

$$\begin{aligned} & \partial_x g^\varepsilon(X_T) \nabla X_T \mathbf{1}_{\{X_T^1 \notin A^\varepsilon\}} = \left[\partial_1 g^\varepsilon(X_T) \nabla X_T^1 + \partial_{\mathbf{x}_2} g^\varepsilon(X_T) \nabla \hat{X}_T^2\right] \mathbf{1}_{\{X_T^1 \notin A^\varepsilon\}} \\ = \;& \frac{1}{2\varepsilon}[g(x_1^* + \varepsilon, \hat{X}_T^2) - g(x_1^* - \varepsilon, \hat{X}_T^2)] \nabla X_T^1 \mathbf{1}_{\{|X_T^1 - x_1^*| < \varepsilon\}} \\ & + \partial_1 g^\varepsilon(X_T) \nabla X_T^1 \mathbf{1}_{\{X_T^1 \in (\{|x_1 - x_1^*| > \varepsilon\} \setminus A)\}} + \partial_{\mathbf{x}_2} g^\varepsilon(X_T) \nabla \hat{X}_T^2 \mathbf{1}_{\{X_T^1 \notin A^\varepsilon\}} \\ = \;& \frac{1}{2\varepsilon} \Delta g(x_1^*, \hat{X}_T^2) \nabla X_T^1 \mathbf{1}_{\{|X_T^1 - x_1^*| < \varepsilon\}} + \frac{1}{2\varepsilon} \mathbf{1}_{\{|X_T^1 - x_1^*| < \varepsilon\}} \times \\ & \quad [(g(x_1^* + \varepsilon, \hat{X}_T^2) - g(x_1^*+, \hat{X}_T^2)) + (g(x_1^*-, X_T^2) - g(x_1^* - \varepsilon, \hat{X}_T^2))] \nabla X_T^1 \\ & + \partial_1 g^\varepsilon(X_T) \nabla X_T^1 \mathbf{1}_{\{X_T^1 \in (\{|x_1 - x_1^*| > \varepsilon\} \setminus A)\}} + \partial_{\mathbf{x}_2} g^\varepsilon(X_T) \nabla \hat{X}_T^2 \mathbf{1}_{\{X_T^1 \notin A^\varepsilon\}} \\ = \;& I_1^\varepsilon + I_2^\varepsilon + I_3^\varepsilon + I_4^\varepsilon, \end{aligned} \tag{3.9}$$

where $I_1^\varepsilon, \dots, I_4^\varepsilon$ 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 (3.8) it suffices to show that

$$\lim_{\epsilon\to 0} E_t\{I_1^\varepsilon u_t\} = E_t\left\{\mathbf{1}_{\{X_T^1 > x_1^*\}}\delta_t(F_t u.)\right\}. \tag{3.10}$$

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 $t \le s \le T$,

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

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

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

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

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

On the other hand, by the Fubini Theorem we have

$$E_t\left\{\psi(X_T^1)\Delta g(x_1^*, \hat{X}_T^2)\nabla X_T^1\right\} = \int_a^b E_t\left\{\mathbf{1}_{\{X_T^1 > y\}}\delta_t(F_t u.)\right\} dy. \tag{3.11}$$

Note again that the law of X_T^1 has a density, thus the integrand in the right hand side of (3.11) is continuous with respect to y. Thus, letting $[a,b] = [x_1^* - \varepsilon, x_1^* + \varepsilon]$, dividing both sides of (3.11) by ε, and then sending $\varepsilon \to 0$ we obtain (3.10), whence the theorem. ∎

If, in fact, the volatility matrix σ is squared, then the following theorem gives a simpler representation result. We omit the proof. The result was given under stronger conditions in Fournie et al. (1999), and extended in Ma and Zhang (2000).

Theorem 3.5 *Assume $d = n$ and $(\mathbf{A1}), (\mathbf{A2})$, 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*

$$Z_t = E_t\{g(X_T)N_T^t\}\sigma(t, X_t), \tag{3.12}$$

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

$$Z_0 = E\{g(X_T)N_T^0\}\sigma(0, x). \tag{3.13}$$

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.

Due to the Markovian nature of our setting, we shall consider only $t = 0$:

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

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

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

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. To begin our analysis, let us first use integration by parts formula for Skorohod integrals and noting that $\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1\sigma(0,x)]$ is a scalar, we have

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

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:

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

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

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

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

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

Next, since ∇X is the solution to the variational SDE (3.3), we have

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

Now applying Itô's formula we get

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

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 $D_t \nabla X_u^i = \nabla X_u^i \Gamma_u^t$. Moreover, since

$$\begin{cases} D_t(\partial_x b(s, X_s)) = \sum_{k=1}^n [\partial_k(\partial_x b)](s, X_s)[\nabla X_s^k u_t]; \\ D_t(\partial_x \sigma(s, X_s)) = \sum_{k=1}^n [\partial_k(\partial_x \sigma)](s, X_s)[\nabla X_s^k u_t], \end{cases} \qquad (4.8)$$

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

$$\Gamma_u^t = \gamma_t + \sum_{k=1}^n \int_t^u [\nabla X_s^k u_t]\left[\alpha_s^k ds + \beta_s^k dW_s\right], \qquad (4.9)$$

where

$$\begin{cases} \alpha_s^k \triangleq [\nabla X_s]^{-1}[\partial_k(\partial_x b) - \partial_x \sigma \partial_k(\partial_x \sigma)](s, X_s)\nabla X_s \\ \beta_s^k \triangleq [\nabla X_s]^{-1}\partial_k(\partial_x \sigma)(s, X_s)\nabla X_s; \\ \gamma_t \triangleq [\nabla X_t]^{-1}\partial_x \sigma(t, X_t)\nabla X_t. \end{cases} \qquad (4.10)$$

Combining (4.9) and $D_t \nabla X_u^i = \nabla X_u^i \Gamma_u^t$, from (4.4) we obtain

$$\begin{aligned}\int_0^T I_2(t)dt &= \frac{\Delta g(x_1^*, \hat{X}_T^2)\int_0^T [\nabla X_T^1 u_t] D_t[\nabla X_T^1]dt\sigma(0, x)}{\|DX_T^1\|_H^2} \\ &= \frac{\Delta g(x_1^*, \hat{X}_T^2)\sum_1^n \nabla_j X_T^1 I_2^j \sigma(0, x)}{\|DX_T^1\|_H^2}, \qquad (4.11)\end{aligned}$$

where

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

thanks to Fubini's Theorem. Using analogous arguments, we get

$$\int_0^T I_3(t)dt = \frac{\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1\sigma(0,x)]}{\|DX_T^1\|_H^2}\int_0^T D_t\nabla X_T^1 u_t dt \tag{4.12}$$

$$+\frac{\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1\sigma(0,x)]}{\|DX_T^1\|_H^2}\nabla X_T^1\left\{\int_0^T \gamma_t u_t dt + \sum_{k,j=1}^n \int_0^T \nabla_j X_s^k\Big[\int_0^s u_t^j u_t dt\Big]\Big[\alpha_s^k ds + \beta_s^k dW_s\Big]\right\}.$$

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

$$D_t\|DX_T^1\|_H^2 = 2D_t[\nabla X_T^1]\int_0^T u_r u_r^T dr[\nabla X_T^1]^T + 2\nabla X_T^1\int_t^T (D_t u_r)u_r^T dr[\nabla X_T^1]^T. \tag{4.13}$$

From $\nabla X_r u_r = \sigma(r, X_r)$, we have $(D_t[\nabla X_r])u_r + \nabla X_r(D_t u_r) = \partial_x\sigma(r, X_r)\nabla X_r u_t$, and

$$D_t u_r = [\nabla X_r]^{-1}\left[\partial_x\sigma(r, X_r)\nabla X_r u_t - (D_t\nabla X_r)u_r\right]. \tag{4.14}$$

By (4.4) and (4.13), and applying Fubini's theorem again, we have

$$\int_0^T I_4(t)dt = \frac{2\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1\sigma(0,x)]}{\|DX_T^1\|_H^4}\sum_{j=1}^n \nabla_j X_T^1 \times \tag{4.15}$$

$$\left\{\int_0^T u_t^j D_t\nabla X_T^1 dt\cdot\int_0^T u_r u_r^T dr + \nabla X_T^1\int_0^T\Big[\int_0^r u_t^j D_t u_r dt\Big]u_r^T dr\right\}[\nabla X_T^1]^T \quad .$$

Using (4.14), we can rewrite (4.15) as

$$\begin{aligned}\int_0^T I_4(t)dt &= \frac{2\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1\sigma(0,x)]}{\|DX_T^1\|_H^4}\sum_{j=1}^n \nabla_j X_T^1 \times\\ &\left\{\int_0^T u_t^j D_t\nabla X_T^1 dt\cdot\int_0^T u_r u_r^T dr + \nabla X_T^1\int_0^T \gamma_r\left(\int_0^r u_t^j u_t dt\right)u_r^T dr\right.\\ &\left.-\nabla X_T^1\int_0^T [\nabla X_r]^{-1}\left(\int_0^r u_t^j D_t\nabla X_r dt\right)u_r u_r^T dr\right\}[\nabla X_t^1]^T \qquad (4.16)\\ &= \frac{2\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1\sigma(0,x)]}{\|DX_T^1\|_H^4}\sum_{j=1}^n \nabla_j X_T^1\left\{I_4^{j,1} + I_4^{j,2} - I_4^{j,3}\right\}[\nabla X_T^1]^T,\end{aligned}$$

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

$$I_4^{j,2} \triangleq \nabla X_T^1\int_0^T \gamma_r\left(\int_0^r u_t^j u_t dt\right)u_r^T dr. \tag{4.17}$$

Recalling (4.9), (4.7) and $D_t\nabla X_u^i = \nabla X_u^i\Gamma_u^t$, we have

$$\begin{aligned}I_4^{j,1} &= \nabla X_T^1\left\{\int_0^T u_t^j\gamma_t dt + \sum_{k=1}^n\int_0^T[\nabla X_s^k\int_0^s u_t^j u_t dt]\Big[\alpha_s^k ds + \beta_s^k dW_s\Big]\right\}\Big(\int_0^T u_r u_r^T dr\Big);\\ I_4^{j,3} &= \nabla X_T^1\int_0^T\left\{\int_0^r u_t^j\gamma_t dt + \sum_{k=1}^n\int_0^r[\nabla X_s^k\int_0^s u_t^j u_t dt]\Big[\alpha_s^k ds + \beta_s^k dW_s\Big]\right\}u_r u_r^T dr.\end{aligned} \tag{4.18}$$

Finally, since (4.16) can be calculated via (4.18) and (4.17), we can compute $\int_0^T (D_t F) u_t dt$ by computing (4.5), (4.11), (4.12) and (4.16). 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). Recall the processes X_t, ∇X_t and $[\nabla X_t]^{-1}$ and define:

$$u_t = [\nabla X_t]^{-1}\sigma(t, X_t); \quad A_t = \int_0^t u_s dW_s; \quad B_t = \int_0^t u_s u_s^T ds;$$
$$\alpha_t^k = [\nabla X_t]^{-1}[\partial_k\partial_x b - \partial_x\sigma\partial_k\partial_x\sigma](t, X_t)\nabla X_t; \quad \beta_t^k = [\nabla X_t]^{-1}\partial_k\partial_x\sigma(t, X_t)\nabla X_t;$$
$$\gamma_t = [\nabla X_t]^{-1}\partial_x\sigma(t, X_t)\nabla X_t; \quad C_t^j = \int_0^t u_s^j \gamma_s ds + \sum_{k=1}^n \int_0^t [\nabla X_s^k B_s^j]\Big[\alpha_s^k ds + \beta_s^k dW_s\Big];$$
$$H_t = \int_0^t \gamma_s u_s ds + \sum_{j,k=1}^n \int_0^t \nabla_j X_s^k B_s^j \Big[\alpha_s^k ds + \beta_s^k dW_s\Big]; \quad L_t^j = \int_0^t \Big[\gamma_s A_s^j - C_s^j u_s\Big] u_s^T ds;$$

Then we have

$$\begin{aligned} I \;=\; & \frac{1}{\nabla X_T^1 B_T [\nabla X_T^1]^T}\Big\{\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1 \sigma(0,x)][\nabla X_T^1 A_T] \\ & -[\nabla X_T^1\sigma(0,x)]\partial_2\Delta g(x_1^*, \hat{X}_T^2)\nabla \hat{X}_T^2 B_T [\nabla X_T^1]^T \\ & -\Delta g(x_1^*, \hat{X}_T^2)\sum_{j=1}^n \nabla_j X_T^1 \nabla X_T^1 C_T^j \sigma(0,x) - \Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1\sigma(0,x)]\nabla X_T^1 H_T\Big\} \\ & + \frac{2\Delta g(x_1^*, \hat{X}_T^2)[\nabla X_T^1\sigma(0,x)]}{\Big(\nabla X_T^1 B_T [\nabla X_T^1]^T\Big)^2} \sum_{j=1}^n \nabla_j X_T^1 \nabla X_T^1 \Big\{C_T^j B_T + L_T^j\Big\}[\nabla X_T^1]^T. \end{aligned} \tag{4.19}$$

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.

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 $\widetilde{W}_t \triangleq (W_t^T, (W_t^0)^T)^T$, and denote $\widetilde{\mathbf{F}}$ to be the filtration generated by $\widetilde{W}$. For each $\varepsilon > 0$ we define

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

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):

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

By the stability results for both forward and backward SDE's 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:

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

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

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*

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

The proof of the above inequality is standard in BSDE theory, using Gronwall's inequality. 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

$$Z_t^{\varepsilon,1} = E_t\left\{g(X_T^\varepsilon)N_T^{\varepsilon,t}\right\}\sigma(t, X_t^\varepsilon), \tag{5.7}$$

where

$$N_r^{\varepsilon,t} = \frac{1}{r-t}\left[\int_t^r \left([\sigma^\varepsilon(s,X_s^\varepsilon)]^{-1}\nabla X_s^\varepsilon\right)^T d\widetilde{W}_s\right]^T (\nabla X_t)^{-1}; \tag{5.8}$$

and ∇X^ε satisfies the linear SDE

$$\nabla_i X_t^\varepsilon = e_i + \int_0^t \partial_x b(s,X_s^\varepsilon)\nabla X_s^\varepsilon ds + \sum_{j=1}^d \int_0^t [\partial_x \sigma^j(s,X_s^\varepsilon)]\nabla_i X_s^\varepsilon dW^j, \quad i=1,\cdots,n, \tag{5.9}$$

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:

$$\tilde{g}(x) = \sum_{i=1}^K \tilde{g}^i(x)1_{[a_{i-1},a_i)}(x), \tag{5.10}$$

where $\tilde{g}^i(x) = A^i x + B^i$, $i=1,\cdots,K$. We first give the following approximation lemma, whose proof we omit.

Lemma 5.2 *Under the above assumption there exists a sequence of smooth functions $\tilde{g}_k(x)$ such that*

(i) $|\tilde{g}_k(x)| \le C(1+|x|)$, $\forall x$, for some $C>0$;

(ii) for each k, $\sup_x |\tilde{g}_k'(x)| \le C_k$, for some $C_k > 0$; and

(iii) for all $x \in \mathbb{R}\setminus \cup_{i=1}^K \{a_i\}$, $\tilde{g}_k'(x) \to \tilde{g}'(x)$ and $\tilde{g}_k(x)\to\tilde{g}(x)$, as $k\to\infty$.

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:*

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

Proof. Let $\{\tilde{g}_k\}$ be the smooth sequence that approximates $\tilde{g}(p)$, as in Lemma 5.2. 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 (5.7) 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 $\tilde{g}_k$'s have uniform linear growth (Lemma 5.2-(i)), one can easily check that, for any $x \in \mathbb{R}^n$, it holds that

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

Recalling (5.2) and noting that b and σ^ε are uniformly Lipschitz, by standard SDE arguments one can easily show that $E_{t,x}|X^\varepsilon_T|^4 < \infty$, 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 $g_k(X^\varepsilon_T) \to g(X^\varepsilon_T)$ in $L^2(\Omega)$ for each $\varepsilon > 0$. Thus, by the stability results for backward SDE's, 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

$$Z^\varepsilon_t = \lim_{k\to\infty} Z^{\varepsilon,k}_t = \lim_{k\to\infty} E_t\{g_k(X^\varepsilon_T)N^{\varepsilon,t}_T\}\sigma^\varepsilon(t, X^\varepsilon_t). \qquad \blacksquare$$

6 Numerical Experiments.

We compare here four methods for computing the hedging portfolio of a digital option:

- 1. The Malliavin calculus method, called M method, for Malliavin. We use Theorem 3.3 and (4.19) to do the computations.

- 2. The Malliavin calculus method using the approximation of σ. Let us call it M_ε method. We use the formulas (5.7)–(5.9).

- 3. The "Retrieval of Volatility Method" of Cvitanic, Goukasian and Zapatero (2001), 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 finite difference Δ method, called here Δ. In this method one computes (by central differences approximation and simulation) the derivatives of the price process Y at initial time with respect to the initial conditions $(X^1_t, \dots, X^n_t) = (x_1, \dots, x_n)$, 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 a stochastic volatility extension of the Black-Scholes model:

$$S_t = S_0 + \int_0^t \sigma_r S_r dW_r; \quad \sigma_t = \sigma_0 + \int_0^t k_\sigma(\bar{\sigma} - \sigma_r)dr + \int_0^t \rho_\sigma \sigma_r dW_r.$$

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 $N_M = 45, N_{RVM} = 80$, $K_{RVM} = 2, N_\Delta = 100$, $N_{M_\varepsilon} = 80, \varepsilon = .05$, $\Delta S = .01S_0$, $\Delta\sigma = .01\sigma_0, dt = .05$. 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 $N_M = 80$, $N_{RVM} = 100$, $K_{RVM} = 2$, $N_\Delta = 125$, $N_{M_\varepsilon} = 90$, $\varepsilon = .05$, $\Delta S = .01S_0$, $\Delta\sigma = .01\sigma_0$, $dt = .05$. 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. In this example the Δ method is not inferior if its parameters are carefully chosen to avoid high bias and high standard error.

Table 1: 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 2: 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

Acknowledgements: We are grateful to P. Glasserman for useful comments.

REFERENCES

- Boyle, P., M. Broadie and P. Glasserman (1997), "Monte Carlo Methods for Security Pricing," *Journal of Economic Dynamics and Control*, 21, 1267–1321.
- Cvitanić, J., Goukasian, L. and F. Zapatero (2001), "Hedging with Monte Carlo simulation." In E. Kontoghiorghes, B. Rustem and S. Siokos (eds.), *Computational Methods in Decision-Making, Economics and Finance*, Kluwer Academic Publishers, forthcoming.
- Duffie, D. and Glynn, P. (1995), "Efficient Monte Carlo Simulation of Security Prices," *The Annals of Applied Probability* 5, 897-905.
- El Karoui, N., Peng, S., and Quenez, M.C. (1997), *Backward stochastic differential equations in finance*, *Mathematical Finance*, **7**, 1-72.
- Fournié, E., Lasry, J-M., Lebuchoux, J., Lions, P-L., and Touzi, N. (1999) "Application of Malliavin calculus to Monte Carlo methods in Finance", *Fin. & Stoch.*, **391**, 391-412.
- 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.
- Glasserman, P. and Zhao, X. (1999) "Fast greeks by simulation in forward LIBOR models", *J. of Computational Finance* **3**, 5-39.
- Ma, J. and Yong, J. (1999), *Forward-Backward Stochastic Differential Equations and Their Applications*, Lecture Notes in Math., **1702**, Springer.
- Ma, J. and Zhang, J. (2000) "Representation theorems for backward stochastic differential equations", *Ann. Appl. Probab.*, to appear.
- Nualart, D. (1995) *The Malliavin Calculus and Related Topics*, Springer.