Applications of Malliavin calculus to the pricing and hedging of Bermudan options

James Newbury

Mathematical Institute
University of Oxford



A dissertation submitted for the degree of:
Master of Science in Mathematical and Computational Finance
Trinity 2011

Abstract

The pricing of Bermudan options, which give the holder the right to buy or sell an underlying asset at a predetermined price and at a discretely spaced number of times prior to maturity, can be based on a deterministic method or on a probabilistic one. Deterministic methods such as finite differences lose their efficiency as the dimension of the problem increases, and they are therefore known to suffer from the "curse of dimensionality". Probabilistic methods enable us to overcome this problem by using Monte Carlo simulations. One particular method is the Malliavin pricing and hedging algorithm, which uses representation formulas for conditional expectation and its derivative to approximate the price and delta of a Bermudan option. This paper specifically deals with how the powerful tools of Malliavin calculus are applied in the derivation of such representation formulas, and looks at how the latter are subsequently used in the pricing and hedging algorithm.

Key words: Bermudan option, dynamic programming principle, Malliavin derivative operator, Skorohod integral, first and second variational processes, representation formula, localizing function.

Acknowledgements

I would like to thank my supervisor, Dr Lajos Gergely Gyurkó, for his guidance as well as for all the helpful remarks he has given me throughout the duration of the dissertation.

Contents

Introduction **3**

1 Malliavin Calculus **8**
1.1 Context and general framework 8
1.2 The Malliavin derivative operator 9
1.3 The Skorohod integral and other important results 10

2 Representation formulas for conditional expectation and its derivative **13**
2.1 General framework 13
2.2 Application to SDEs 15
2.3 The notion of localization 22

3 Applications to the pricing and hedging of Bermudan options **28**
3.1 Overview of the problem and alternative probabilistic pricing methods 28
3.2 The Malliavin pricing and hedging algorithm 31
3.2.1 Applications of the representation formulas 31
3.2.2 Sketch of the algorithm 34
3.3 Numerical applications 37

Concluding considerations **41**

Bibliography **44**

Introduction

Bermudan options, which can be considered as intermediate contracts between American and European options, give the holder the right to buy or sell an underlying asset at a predetermined price and at a discretely spaced number of times prior to a given maturity. Intuitively speaking, the price of an American option is always greater or equal than the price of its Bermudan counterpart, since it gives its holder the right to exercise his right at any given time prior to maturity. In the same way, the price of a Bermudan option is always greater or equal than the price of a European option with the same characteristics, given that the latter can only be exercised at maturity. A direct consequence of this feature is that the price of a European option can explicitly be determined by computing the discounted expectation of the terminal payoff. However, this specific procedure cannot be extended to the case of Bermudan options, since the idea here is to determine the optimal exercise date. In the case of a non-dividend paying Bermudan call, it can be shown that the optimal exercise date is in effect at maturity. Apart from this very isolated case, the determination of the optimal exercise date is far from being a trivial affair. We now look at how the problem can be formalised.

Let T be a positive constant representing the maturity of the option and let $W = (W_t)_{0 \leq t \leq T}$ be a 1-dimensional Wiener process on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{Q})$, where $(\mathcal{F}_t)_{0 \leq t \leq T}$ represents the augmented filtration generated by W, and where $\mathbb{Q}$ is the risk-neutral probability. We make the assumption of a complete market, i.e all contingent claims are $\mathbb{Q}$-attainable, which guarantees the uniqueness of $\mathbb{Q}$, and enables us to determine the unique price of European contingent claims as the discounted expectation of their payoff at maturity. We denote by $\mathbb{E}(.)$ the risk-neutral expectation.

We then proceed to the discretization of the interval $[0, T]$ by introducing the time step $\delta t = \frac{T}{M}$ and by taking into consideration the discrete times $t_k = k\delta t$ for all $k \in \{0, ..., M\}$. A Bermudan option with maturity T on an underlying asset with price X is defined as a contract which gives the holder the right to buy or sell the underlying at a strike price K and at any exercise date within the set $\{t_1, ..., t_M\}$. We denote by $\Phi : \mathbb{R}_+ \to \mathbb{R}$ the payoff of the option, i.e $\Phi(X_{k\delta t})$ represents the amount received by the holder if he decides to exercise his right at time t_k, and we assume that the underlying asset's price process

$X = (X_t)_{0 \leq t \leq T}$ satisfies the following SDE on $[0, T]$:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x \tag{1}$$

where b and σ are two sufficiently smooth functions with bounded derivatives. We also assume that the following integrability condition is satisfied by the discounted payoff at time 0:

$$\mathbb{E}\left(\sup_{t\in[0,T]} |e^{-rt}\Phi(X_t)|\right) < +\infty$$

It can be shown that the price at time $t \in [0, T]$ of such an option is given by:

$$V(t, x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}_{t,x}\left(e^{-r(\tau-t)}\Phi(X_\tau)\right)$$

where $\mathbb{E}_{t,x}$ represents the risk-neutral expectation given that the price process starts at x at time t, where $\mathcal{T}_{t,T}$ denotes the set of all admissible stopping times in $[t, T]$, and where r denotes the constant risk-free interest rate.

We are now interested in determining two fundamental quantities which characterize the option. On the one hand, the premium of the option, defined as its price $V(0, x)$ at time 0, is the quantity paid upfront by the holder of the option to its seller. On the other hand, the delta of the option at time t, denoted by $\Delta(t, x)$, corresponds to the derivative of the price of the option at time t with respect to the initial value of the underlying asset:

$$\Delta(t, x) = \partial_x V(t, x) = \partial_x \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}_{t,x}\left(e^{-r(\tau-t)}\Phi(X_\tau)\right)$$

The above expressions for the price and delta of a Bermudan option precisely show that we are in fact confronted with an optimal stopping problem. We therefore have to consider the Bellman dynamic programming principle, starting with the following backward recursive system, valid for all $k \in \{M - 1, ..., 0\}$:

$$\begin{cases} \hat{V}(t_M, X_{M\delta t}) = \Phi(X_{M\delta t}) \\ \hat{V}(t_k, X_{k\delta t}) = \max\left(\Phi(X_{k\delta t}), e^{-r\delta t}\mathbb{E}\left(\hat{V}(t_{k+1}, X_{(k+1)\delta t})|X_{k\delta t}\right)\right) \end{cases} \tag{2}$$

where $\hat{V}(t_k, X_{k\delta t})$ corresponds to the approximation of the option's price at time t_k. This system shows that the value of the option at maturity is simply equal to the payoff at maturity, given the fact that the holder cannot exercise the option at a later date. It also states that the value of the option at time t_k is equal to the maximum of the intrinsic value at time t_k and the continuation value, i.e the discounted conditional expectation of the value of the option at time t_{k+1}. Intuitively speaking, the continuation value can be considered as the value of keeping the option alive rather than exercising it at a given discrete time. The price of the option $V(0, x)$ is then approximated by $\hat{V}(0, x)$.

As for the delta of the option, we heuristically take into account the derivative of the price at time t_1 with respect to the price of the underlying. More precisely, the approximation $\hat{\Delta}(0,x)$ of the delta at time 0 is given by:

$$\begin{cases} \hat{\Delta}(t_1, X_{\delta t}) = \partial_\alpha \Phi(\alpha) \mathbb{1}_{C^c} + e^{-r\delta t} \partial_\alpha \mathbb{E}\left(\hat{V}(t_2, X_{2\delta t}) | X_{\delta t} = \alpha\right) \mathbb{1}_C \\ \hat{\Delta}(0,x) = \mathbb{E}\left(\hat{\Delta}(t_1, X_{\delta t})\right) \end{cases} \tag{3}$$

where the continuation region C is defined by $C = \{\alpha \in \mathbb{R}_+ : \hat{V}(t_1, \alpha) > \Phi(\alpha)\}$.

Generally speaking, the existing methods used for pricing Bermudan options can either be deterministic or probabilistic. As far as the deterministic methods are concerned, a simple application of Itô's lemma to the price $V^A(t,x)$ of the associated American option leads to the following PDE:

$$\left(\frac{\partial V^A}{\partial t} + \mathcal{L}V^A - rV^A\right) \mathbb{1}_{\{V^A(t,x) > \Phi(x)\}} = 0$$

with the terminal condition $V^A(T,x) = \Phi(x)$ and where $\mathcal{L}$ denotes the infinitesimal generator of the process X defined in the following way:

$$\mathcal{L}.(x) = b(x)\frac{\partial .}{\partial x}(x) + \frac{1}{2}\sigma^2(x)\frac{\partial^2 .}{\partial x^2}(x)$$

A weak formulation of the previous problem with variational inequalities (in Bensoussan and Lions [2] for example) leads to a linear complementarity problem which can be solved by standard finite difference methods (for example, see Duffy [5]). This method subsequently gives us the price of the associated Bermudan option, and an approximation of its American counterpart. Despite its relatively high efficiency for low-dimensional problems, the finite difference approach suffers from the so-called "curse of dimensionality". There seems to be a consensus based on the idea that in three or more dimensions, the computing speed of the finite difference approach is uncompetitive in comparison with probabilistic techniques.

The problem of pricing Bermudan options using probabilistic methods is based on the use of Monte Carlo simulations. Such probabilistic approaches can roughly be divided into three distinct categories. The first one is based on the approximation of the conditional expectations which appear in the dynamic programming principle formulation (2) using a regression on a truncated basis of $\mathbb{L}^2$. This specific method is known as the Longstaff-Schwartz algorithm, and is introduced in [11]. The second category involves the presence of a tree which enables us to discretize the underlying price process on a given grid (for example, see Broadie and Glasserman [4]). Finally, the last category exploits the existence of representation formulas for conditional expectation and its derivative (which appear in systems (1) and (2)) via a Malliavin calculus approach, which is exactly what is dealt with throughout this paper. This method was first

presented by Lions and Reigner in [10], and then by Bally *et al.* in [1], where the algorithm is presented for a geometric Brownian motion underlying process. One of the aims of this paper is to provide a slightly more generalised version of the approach presented in [1], most notably by presenting the problem and the algorithm with an underlying price process following the SDE which appears in (1).

Malliavin calculus, also known as stochastic calculus of variations, provides us with the mechanics to compute derivatives of random variables. Powerful features such as the Malliavin derivative operator and its adjoint, the Skorohod integral, enable us to express conditional expectations of the form $\mathbb{E}(f(X_t)|X_s = \alpha)$ and its derivative $\partial_\alpha \mathbb{E}(f(X_t)|X_s = \alpha)$, where f is a function satisfying a certain number of regularity conditions, as a ratio of two non-conditioned expectations which involve quantities such as the Skorohod integral. In other words, the information "contained" in the conditioning is subsequently transmitted to the Skorohod integral. These formulas are obtained by Fournié *et al.* in [7].

This paper is organised as follows. In chapter 1, we introduce the necessary tools of Malliavin calculus which are used later on in the derivation of the representation formulas. Given that the formalism behind Malliavin calculus can rapidly become very complex, we try to follow a simple and logical approach and we therefore do not provide any proofs in this introductory chapter, many of which can be found in Nualart [12]. Once these tools have been introduced, we proceed to the derivation of representation formulas for conditional expectation and its derivative in chapter 2. Here, we give details of all the proofs as this section undeniably represents the core of the problem we are dealing with. We finally apply these representation formulas within the context of the pricing and hedging of Bermudan options in chapter 3, most prominently by introducing the resulting algorithm, and by extending our discussion to the analysis of various numerical results.

Let us finally mention that in this paper, the representation formulas as well as the Malliavin pricing and hedging algorithm are presented in a one-dimensional setting. We have deliberately chosen not to introduce these ideas in the multi-dimensional case for the sake of clarity. However, all the results presented here are easily extendable to higher dimensions (this is done in the geometric Brownian motion case by Bally *et al.* in [1]), and one must not forget that the main benefits of Monte Carlo based approaches over finite difference methods arise in a multi-dimensional setting.

1 Malliavin Calculus

This chapter aims to give an introduction to Malliavin calculus and presents the main theoretical results which will be needed later on in the derivation of the representation formulas for conditional expectation and its derivative. We follow the approach presented in [6], and a more complete description of the following results along with their proofs can be found in [8] and [12].

1.1 Context and general framework

We start by introducing the general framework used in this introductory chapter. We start by fixing a positive constant $T \in]0, +\infty]$. We then consider an n-dimensional Wiener process $W = (W_t)_{0 \leq t \leq T}$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$, where $(\mathcal{F}_t)_{0 \leq t \leq T}$ represents the augmented filtration generated by W. Let us mention that this multi-dimensional setting is only used in this chapter, and we later move on to a one-dimensional framework.

Let us now recall the basic spaces and norms which are commonly used in stochastic analysis.

Definition 1.1.1 *The Hilbert space $\mathbb{L}^2(\Omega)$ is the set of real-valued square integrable random variables. It is equipped with the following inner product:*

$$\forall X, Y \in \mathbb{L}^2(\Omega), \ < X, Y >_{\mathbb{L}^2(\Omega)} := \mathbb{E}(XY)$$

The norm induced by this inner product is:

$$\forall X \in \mathbb{L}^2(\Omega), \|X\|_{\mathbb{L}^2(\Omega)} = \sqrt{< X, X >_{\mathbb{L}^2(\Omega)}} = \mathbb{E}(X^2)^{1/2}$$

Definition 1.1.2 *The Hilbert space $\mathbb{L}^2(\Omega \times [0,T])$ is the set of random processes ϕ such that:*

$$\mathbb{E}\left(\int_0^T \phi^2(\omega, t)dt\right) < \infty$$

$\mathbb{L}^2(\Omega \times [0,T])$ is equipped with the inner product:

$$\forall \phi, \psi \in \mathbb{L}^2(\Omega \times [0,T]), < \phi, \psi >_{\mathbb{L}^2(\Omega \times [0,T]}} := \mathbb{E}\left(\int_0^T \phi(\omega, t)\psi(\omega, t)dt\right)$$

Here, the corresponding norm is naturally:

$$\forall \phi \in \mathbb{L}^2(\Omega \times [0,T]), \|\phi\|_{\mathbb{L}^2(\Omega\times[0,T])} = \sqrt{< \phi, \phi >_{\mathbb{L}^2(\Omega\times[0,T])}} = \mathbb{E}\left(\int_0^T \phi^2(\omega,t)dt\right)^{1/2}$$

Definition 1.1.3 *We denote by* $\mathbb{M}^2(\Omega \times [0,T]) \subset \mathbb{L}^2(\Omega \times [0,T])$ *the Hilbert space of progressively measurable random processes* ϕ *such that:*

$$\mathbb{E}\left(\int_0^T \phi^2(\omega,t)dt\right) < \infty$$

The corresponding inner product and norm are defined in a similar way as above.

We also need to specify the general form of the random variables we shall be working with. More precisely:

Definition 1.1.4 *We define* $\mathcal{C} \subset \mathbb{L}^2(\Omega)$ *as the set of random variables* F *of the form:*

$$F = f\left(\int_0^T g_1(t)dW(t), ..., \int_0^T g_n(t)dW(t)\right)$$

where f *belongs to* $C_p^\infty(\mathbb{R}^n)$*, the set of infinitely continuously differentiable functions on* $\mathbb{R}^n$ *such that all partial derivatives have polynomial growth, and where the functions* $g_1, ..., g_n$ *are in* $\mathbb{L}^2(\Omega \times [0,T])$*.*

1.2 The Malliavin derivative operator

We start this section by giving a precise mathematical meaning to the derivative of an element $F \in \mathcal{C}$:

Definition 1.2.1 *If* $F \in \mathcal{C}$*, the Malliavin derivative* DF *of* F *is defined as the stochastic process* $DF = (D_tF)_{0\leq t\leq T}$ *belonging to* $\mathbb{L}^2(\Omega \times [0,T])$ *such that:*

$$D_tF = \sum_{i=1}^n \frac{\partial f}{\partial x_i}\left(\int_0^T g_1(t)dW(t), ..., \int_0^T g_n(t)dW(t)\right) g_i(t)$$

One interesting feature of $\mathcal{C}$ is that its completion with respect to a certain norm is a Banach space:

Definition 1.2.2 *If* $F \in \mathcal{C}$*, we define the norm* $\|.\|_{1,2}$ *by:*

$$\|F\|_{1,2} := \mathbb{E}(F^2)^{1/2} + \mathbb{E}\left(\int_0^T (D_tF)^2 dt\right)^{1/2}$$

Furthermore, we define the Banach space $\mathbb{D}^{1,2}$ *as the completion of* $\mathcal{C}$ *with respect to* $\|.\|_{1,2}$*.*

We are now able to give a first important result related to the Malliavin derivative operator D:

Proposition 1.2.1 *The Malliavin derivative operator D, defined on $\mathbb{D}^{1,2}$ and with values in $\mathbb{L}^2(\Omega \times [0,T])$, is a closed unbounded operator.*

As a matter of fact, the Malliavin derivative operator shares a certain number of properties which are satisfied by the derivative operator of ordinary differential calculus:

Proposition 1.2.2 *The Malliavin derivative operator D satisfies the following properties:*

1. Linearity:

$$\forall(\alpha,\beta) \in \mathbb{R}^2, \forall F, G \in \mathbb{D}^{1,2}, D_t(\alpha F + \beta G) = \alpha D_t(F) + \beta D_t(G)$$

2. Product rule:

$$\forall F, G \in \mathbb{D}^{1,2}, D_t(FG) = F D_t(G) + G D_t(F)$$

3. Chain rule: If $u : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function with bounded partial derivatives, if $F = (F_1, ..., F_n)$ is a random vector with components in $\mathbb{D}^{1,2}$, then $u(F) \in \mathbb{D}^{1,2}$ and:

$$D_t u(F) = \sum_{i=1}^{n} \frac{\partial u}{\partial x_i}(F) D_t F_i$$

1.3 The Skorohod integral and other important results

According to the Riesz representation theorem, the unboundedness of the Malliavin derivative operator D enables us to define its adjoint, namely the Skorohod integral, denoted by δ:

Definition 1.3.1 *Let u be a random process in $\mathbb{L}^2(\Omega \times [0,T])$. Then $u \in Dom(\delta)$ if for any $F \in \mathbb{D}^{1,2}$, we have:*

$$< DF, u >_{\mathbb{L}^2(\Omega\times[0,T])} = \mathbb{E}\left(\int_0^T (D_t F) u_t dt\right) \leq C(u)\|F\|_{1,2}$$

where $C(u)$ is a constant which only depends on u. Subsequently, if $u \in Dom(\delta)$, the Skorohod integral $\delta(u)$ is defined for any $F \in \mathbb{D}^{1,2}$ by:

$$\mathbb{E}(F\delta(u)) = < DF, u >_{\mathbb{L}^2(\Omega\times[0,T])} = \mathbb{E}\left(\int_0^T (D_t F) u_t dt\right)$$

When $u \in Dom(\delta)$, we say that u is Skorohod-integrable, and $\delta(u)$ defined above is an element of $\mathbb{L}^2(\Omega)$. Furthermore, the following property enables us to establish a connection between the Skorohod integral and the Itô integral:

Proposition 1.3.1 *$Dom(\delta)$ contains all the progressively measurable random processes in $\mathbb{L}^2(\Omega \times [0,T])$, i.e $Dom(\delta) \supset \mathbb{M}^2(\Omega \times [0,T])$, and the Skorohod integral restricted to $\mathbb{M}^2(\Omega \times [0,T])$ coincides with the Itô integral.*

We can now state a fundamental result related to the Skorohod integral, known as the integration by parts formula:

Theorem 1.3.1 *Let F be an $\mathcal{F}_T$-measurable random variable in $\mathbb{D}^{1,2}$. Then for any $u \in Dom(\delta)$:*

$$\delta(Fu) = F\delta(u) - \int_0^T (D_t F) u_t dt$$

We now look at how Malliavin calculus can be applied within the context of stochastic differential equations. This fundamental application will prove to be vital throughout the main steps of the derivation of the representation formulas for conditional expectation and its derivative.

Proposition 1.3.2 *Let $X = (X_t)_{t\in[0,T]}$ be an n-dimensional Itô process characterized by the following dynamics:*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

where b and σ are two continuously differentiable functions with bounded derivatives. Define $Y = (Y_t)_{t\in[0,T]}$ as the first variational process of X:

$$dY_t = b'(X_t)Y_t dt + \sum_{i=1}^{n} \sigma_i'(X_t)Y_t dW_t^i, \quad Y_0 = I_n$$

where I_n denotes the identity matrix of $\mathbb{R}^n$. Then $X \in \mathbb{D}^{1,2}$ and satisfies:

$$D_s X_t = Y_t Y_s^{-1} \sigma(X_s) \mathbb{1}_{\{s\leq t\}}, \quad s \geq 0 \ \ a.s$$

Moreover, if ψ is a continuously differentiable function with bounded partial derivatives, we have:

$$D_s \psi(X_T) = \nabla\psi(X_T) Y_T Y_s^{-1} \sigma(X_s) \mathbb{1}_{\{s\leq T\}}, \quad s \geq 0 \ \ a.s$$

We now have a look at some examples of applications of the previous proposition:

Example 1.3.1 *We suppose that the process $X = (X_t)_{t\in[0,T]}$ satisfies the following SDE:*

$$dX_t = rX_t dt + \sigma X_t dW_t, \ \ X_0 = x$$

In this case, the first variational process $Y = (Y_t)_{t\in[0,T]}$ satisfies:

$$dY_t = rY_t dt + \sigma Y_t dW_t, \;\; Y_0 = 1$$

and therefore $Y_t = \frac{X_t}{x}$ for all $t \geq 0$. Applying proposition 1.3.2, we obtain:

$$D_s X_t = \sigma X_t \mathbb{1}_{\{s\leq t\}}$$

Example 1.3.2 *We suppose that the process $X = (X_t)_{t\in[0,T]}$ satisfies the following SDE:*

$$dX_t = a(\theta - X_t)dt + \sigma dW_t, \;\; X_0 = x$$

Here, the dynamics of the first variational process $Y = (Y_t)_{t\in[0,T]}$ are given by:

$$dY_t = -aY_t dt, \;\; Y_0 = 1$$

We therefore have $Y_t = e^{-at}$ for all $t \geq 0$, and the Malliavin derivative of X_t is now given by:

$$D_s X_t = \sigma e^{-a(t-s)} \mathbb{1}_{\{s\leq t\}}$$

We end this chapter by introducing commutation formulas between the Malliavin derivative operator D and the Riemann-Stieltjes and Itô operators of integration. These results will prove to be very helpful within the context of the derivation of representation formulas for the derivative of conditional expectation.

Proposition 1.3.3 *We consider a process $u = (u_t)_{t\in[0,T]}$ in $\mathbb{M}^2(\Omega \times [0,T])$ satisfying $u_t \in \mathbb{D}^{1,2}$ for all $t \in [0,T]$ and $(D_s u_t)_{t\geq s} \in \mathbb{M}^2(\Omega \times [s,T))$ for all $s \in [0,T]$. Then $\int_0^T u_t dW_t$ and $\int_0^T u_t dt$ belong to $\mathbb{D}^{1,2}$, and for all $s \in [0,T]$, we have the two following equalities:*

$$\begin{cases} D_s\left(\int_0^T u_t dW_t\right) = u_s + \int_s^T (D_s u_t) dW_t \\ D_s\left(\int_0^T u_t dt\right) = \int_s^T (D_s u_t) dt \end{cases}$$

2 Representation formulas for conditional expectation and its derivative

In this chapter, we apply the previous results in order to derive representation formulas for conditional expectation and its derivative. These formulas will enable us to approximate the price and delta of a Bermudan option using Monte Carlo simulations later on. In order to derive such formulas, we follow the approach developed in [7] and [1].

2.1 General framework

We place ourselves in the probabilistic framework set in chapter 1, taking $n = 1$. Let us recall that our aim is to take into consideration conditional expectations of the form:

$$\mathbb{E}(f(F)|G = 0)$$

where F and G are two $\mathcal{F}_T$-measurable random variables in $\mathbb{D}^{1,2}$, and where f is a Borel-measurable function with polynomial growth. We also assume that F and G are smooth enough in a Malliavin sense, meaning that D_tF and D_tG belong to $\mathbb{L}^2(\Omega \times [0,T])$. We finally assume that D_tG is non-degenerate, i.e there exists a sufficiently smooth process $u = (u_t)_{0\leq t\leq T}$ satisfying:

$$\mathbb{E}\left(\int_0^T (D_tG)u_t dt|\sigma(F,G)\right) = 1 \tag{2.1}$$

where $\sigma(F,G)$ denotes the σ-algebra generated by F and G. As a result, we will choose $u_t = \frac{1}{TD_tG}$ whenever it is possible. We can now state the first representation formula:

Theorem 2.1.1 *Assuming F, G, u and f satisfy the previous assumptions, and if f is continuously differentiable, we have:*

$$\mathbb{E}(f(F)|G = 0) = \frac{\mathbb{E}\left(f(F)H(G)\delta(u) - f'(F)H(G)\int_0^T (D_tF)u_t dt\right)}{\mathbb{E}(H(G)\delta(u))}$$

where H denotes the standard Heaviside funtion.

Proof: Using the definition of conditional expectation, we can write:

$$\mathbb{E}(f(F)|G=0) = \frac{\mathbb{E}(f(F)\delta_0(G))}{\mathbb{E}(\delta_0(G))}$$

where δ_0 denotes the Dirac delta function at 0. On the one hand, taking derivatives in a distributional sense, we have:

$$\begin{aligned}
\mathbb{E}(\delta_0(G)) &= \mathbb{E}(H'(G)) \\
&= \mathbb{E}\left(H'(G)\mathbb{E}\left(\int_0^T (D_tG)u_t dt|\sigma(F,G)\right)\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(H'(G)\int_0^T (D_tG)u_t dt|\sigma(F,G)\right)\right) \\
&= \mathbb{E}\left(H'(G)\int_0^T (D_tG)u_t dt\right) \\
&= \mathbb{E}\left(\int_0^T (D_tH(G))u_t dt\right) \\
&= \mathbb{E}(H(G)\delta(u))
\end{aligned}$$

On the other hand, using the integration by parts formula, we have:

$$\begin{aligned}
\mathbb{E}(H(G)\delta(f(F)u)) &= \mathbb{E}\left(f(F)H(G)\delta(u) - H(G)\int_0^T (D_tf(F))u_t dt\right) \\
&= \mathbb{E}\left(f(F)H(G)\delta(u) - f'(F)H(G)\int_0^T (D_tF)u_t dt\right)
\end{aligned}$$

But by definition of the Skorohod integral, we see that:

$$\begin{aligned}
\mathbb{E}(H(G)\delta(f(F)u)) &= \mathbb{E}\left(\int_0^T (D_tH(G))f(F)u_t dt\right) \\
&= \mathbb{E}\left(f(F)\delta_0(G)\int_0^T (D_tG)u_t dt\right) \\
&= \mathbb{E}\left(f(F)\delta_0(G)\right)
\end{aligned}$$

□

Remark 2.1.1 *Rigorously speaking, the previous computation can be justified by the fact that:*

$$\mathbb{E}(f(F)|G=0) = \lim_{\epsilon\to 0^+} \frac{\mathbb{E}(f(F)\mathbb{1}_{[-\epsilon,\epsilon]}(G))}{\mathbb{E}(\mathbb{1}_{[-\epsilon,\epsilon]}(G))}$$

We then use the integration by parts formula to write:

$$\mathbb{E}(f(F)\mathbb{1}_{[-\epsilon,\epsilon]}(G)) = \mathbb{E}\left(f(F)H_\epsilon(G)\delta(u) - f'(F)H_\epsilon(G)\int_0^T (D_tF)u_t dt\right)$$

where H_ϵ denotes the Heaviside function at ϵ, i.e defined by $H_\epsilon(x) = \mathbb{1}_{\{x>\epsilon\}} + C$, C being an arbitrary constant. The conclusion then comes by letting ϵ go to 0^+.

This theorem provides us with a representation formula of conditional expectation in the case of a continuously differentiable function f. However, in many financial applications, we often have to consider cases of non-smooth functions. Therefore, the previous result happens to be quite restrictive as far as assumptions are concerned. In order to deal with the potential non-smoothness of f, we have to introduce an additional assumption on the random process u:

Corollary 2.1.1 *We assume that F, G, u and f satisfy the previous assumptions, and we also assume that u satisifes:*

$$\mathbb{E}\left(\int_0^T (D_tF)u_t dt|\sigma(F,G)\right) = 0 \tag{2.2}$$

Then we have the following representation:

$$\mathbb{E}(f(F)|G=0) = \frac{\mathbb{E}(f(F)H(G)\delta(u))}{\mathbb{E}(H(G)\delta(u))}$$

Proof: Note that using the previous theorem, it suffices to prove that

$$\mathbb{E}\left(f'(F)H(G)\int_0^T (D_tF)u_t dt\right) = 0$$

But this is not hard to see, as we can write:

$$\begin{aligned}
\mathbb{E}\left(f'(F)H(G)\int_0^T (D_tF)u_t dt\right) &= \mathbb{E}\left(\mathbb{E}\left(f'(F)H(G)\int_0^T (D_tF)u_t dt|\sigma(F,G)\right)\right) \\
&= \mathbb{E}\left(f'(F)H(G)\mathbb{E}\left(\int_0^T (D_tF)u_t dt|\sigma(F,G)\right)\right) \\
&= 0.
\end{aligned}$$

□

2.2 Application to SDEs

Having introduced the representation formulas of conditional expectations of the form $\mathbb{E}(f(F)|G=0)$, we now wish to specialize these results to the

case of stochastic differential equations. More precisely, we consider a process $X = (X_t)_{0\leq t\leq T}$ and we suppose that it satisfies the following SDE on $[0,T]$:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x$$

where b and σ are two sufficiently smooth functions with bounded derivatives.

We now choose $F = X_t$ and $G = X_s - \alpha$, where $\alpha \in \mathbb{R}$ and where s and t are such that $0 \leq s \leq t \leq T$. Our aim is therefore to provide representation formulas for the conditional expectation $\mathbb{E}(f(X_t)|X_s = \alpha)$.

The first step is to introduce the first variational process $Y = (Y_t)_{0\leq t\leq T}$ associated with X. Recall that Y follows the following SDE on $[0,T]$:

$$dY_t = b'(X_t)Y_t dt + \sigma'(X_t)Y_t dW_t, \quad Y_0 = 1$$

The next step is to ensure the non-degeneracy of the Malliavin derivative of $X_s - \alpha$. In other words, we need to determine a process $u = (u_t)_{0\leq t\leq T}$ satisfying:

$$\mathbb{E}\left(\int_0^T (D_r(X_s - \alpha))u_r dr | \sigma(X_t, X_s)\right) = 1$$

Given that $D_r(X_s - \alpha) = D_r(X_s) = \frac{Y_s}{Y_r}\sigma(X_r)\mathbb{1}_{\{r\leq s\}}$, we naturally choose $u_r = \frac{Y_r}{sY_s\sigma(X_r)}\mathbb{1}_{\{r\leq s\}}$. As a result, the representation formula becomes:

$$\begin{aligned}
\mathbb{E}(f(X_t)|X_s = \alpha) &= \frac{\mathbb{E}\left(f(X_t)H(X_s - \alpha)\delta(u) - f'(X_t)H(X_s - \alpha)\int_0^T (D_r X_t)u_r dr\right)}{\mathbb{E}(H(X_s - \alpha)\delta(u))} \\
&= \frac{\mathbb{E}\left(f(X_t)H(X_s - \alpha)\delta(u) - f'(X_t)H(X_s - \alpha)\frac{Y_t}{Y_s}\right)}{\mathbb{E}(H(X_s - \alpha)\delta(u))}
\end{aligned}$$

Before any Monte-Carlo simulation can be performed with this formula, we see that one difficulty arises, namely the presence of the Skorohod integral $\delta(u)$. Indeed, the random variable $u_r = \frac{Y_r}{sY_s\sigma(X_r)}\mathbb{1}_{\{r\leq s\}}$ is not $\mathcal{F}_r$-measurable and therefore the proces u is not $\mathcal{F}$-adapted. As a result, there is no explicit expression for u involving an Itô integral. However, if we notice that the random variable $Y_s u_r = \frac{Y_r}{s\sigma(X_r)}\mathbb{1}_{\{r\leq s\}}$ is $\mathcal{F}_r$-measurable, we can now consider the $\mathcal{F}$-adapted process $v = (v_r)_{0\leq r\leq T}$ defined by $v_r = \frac{Y_r}{s\sigma(X_r)}\mathbb{1}_{\{r\leq s\}}$ and use the integration by parts formula to determine $\delta(u)$. Assuming Y_s takes positive values a.s, we can write:

$$\begin{aligned}
\delta(u) &= \frac{\delta(v)}{Y_s} + \frac{1}{Y_s}\int_0^T (D_r Y_s)u_r dr \\
&= \frac{1}{sY_s}\int_0^s \frac{Y_r}{\sigma(X_r)}dW_r + \frac{1}{sY_s^2}\int_0^s \frac{(D_r Y_s)Y_r}{\sigma(X_r)}dr
\end{aligned}$$

We finally need to determine an explicit expression for $D_r Y_s$. This can be done using the following lemma:

Lemma 2.2.1 *Let $Z = (Z_t)_{0\leq t\leq T}$ be the second variational process associated with X, i.e the process which satisfies the following SDE on $[0,T]$:*

$$dZ_t = (b'(X_t)Z_t + b''(X_t)Y_t^2)dt + (\sigma'(X_t)Z_t + \sigma''(X_t)Y_t^2)dW_t, \;\; Z_0 = 0$$

Then D_rY_s can be expressed as:

$$D_rY_s = \left(\frac{\sigma(X_r)Z_s}{Y_r} + \sigma'(X_r)Y_s - \frac{\sigma(X_r)Z_rY_s}{Y_r^2}\right)\mathbb{1}_{\{r\leq s\}}$$

Proof: Recall that we have:

$$Y_s = \int_0^s b'(X_r)Y_r dr + \int_0^s \sigma'(X_r)Y_r dW_r$$

Applying the two commutation formulas introduced in Proposition 1.3.3 enables us to write, for $s \geq r$:

$$D_rY_s = \int_r^s (D_v b'(X_v)Y_v)dv + \sigma'(X_r)Y_r + \int_r^s (D_v\sigma'(X_v)Y_v)dW_v$$

Therefore, for $s \geq r$, we can write:

$$\begin{aligned} d(D_rY_s) &= Y_s b''(X_s)(D_rX_s)ds + b'(X_s)(D_rY_s)ds \\ &+ Y_s\sigma''(X_s)(D_rX_s)dW_s + \sigma'(X_s)(D_rY_s)dW_s \end{aligned}$$

Now using the fact that $D_rX_s = \frac{Y_s}{Y_r}\sigma(X_r)\mathbb{1}_{\{r\leq s\}}$, we obtain, for $s \geq r$:

$$\begin{aligned} d(D_rY_s) &= Y_s^2 b''(X_s)\frac{\sigma(X_r)}{Y_r}ds + b'(X_s)(D_rY_s)ds \\ &+ Y_s^2\sigma''(X_s)\frac{\sigma(X_r)}{Y_r}dW_s + \sigma'(X_s)(D_rY_s)dW_s \end{aligned}$$

Consequently, given the dynamics of the process Z, the process R defined by $R_s = D_rY_s - \frac{\sigma(X_r)Z_s}{Y_r}$ satisfies the following SDE, for all $s \in [r,t]$:

$$dR_s = b'(X_s)R_s ds + \sigma'(X_s)R_s dW_s$$

The solution of this SDE can be readily expressed as, for $s \geq r$:

$$R_s = R_r\frac{Y_s}{Y_r} = \sigma'(X_r)Y_s - \frac{\sigma(X_r)Z_rY_s}{Y_r^2}$$

We then deduce the required expression for D_rY_s. □

The representation formula is finally obtained in the following theorem:

Theorem 2.2.1 *Assuming f is a continuously differentiable function, we have for all $\alpha \in \mathbb{R}$:*

$$\mathbb{E}(f(X_t)|X_s = \alpha) = \frac{\mathbb{E}\left(f(X_t)H(X_s-\alpha)\delta(u) - f'(X_t)H(X_s-\alpha)\frac{Y_t}{Y_s}\right)}{\mathbb{E}(H(X_s-\alpha)\delta(u))}$$

where the Skorohod integral $\delta(u)$ is given by:

$$\delta(u) = \frac{1}{sY_s}\left(\int_0^s \frac{Y_r}{\sigma(X_r)}dW_r + \frac{sZ_s}{Y_s} + \int_0^s \left(\frac{\sigma'(X_r)Y_r}{\sigma(X_r)} - \frac{Z_r}{Y_r}\right)dr\right)$$

Once again, we notice that the previous theorem requires the smoothness of f. In order to take into account the case of non-smooth functions, recall that the process u must now also satisfy the following condition:

$$\mathbb{E}\left(\int_0^T (D_r X_t)u_r dr | \sigma(X_t, X_s)\right) = 0$$

Choosing $u_r = \frac{Y_r}{\sigma(X_r)Y_s}\left(\frac{1}{s}\mathbb{1}_{\{r\leq s\}} - \frac{1}{t-s}\mathbb{1}_{\{s\leq r\leq t\}}\right)$, u satisfies the two required conditions (2.1) and (2.2), and the representation formula becomes:

Theorem 2.2.2 *For all $\alpha \in \mathbb{R}$, we have:*

$$\mathbb{E}(f(X_t)|X_s = \alpha) = \frac{\mathbb{E}(f(X_t)H(X_s - \alpha)\delta(u))}{\mathbb{E}(H(X_s - \alpha)\delta(u))}$$

where the Skorohod integral $\delta(u)$ is now given by:

$$\delta(u) = \frac{1}{sY_s}\left(\int_0^s \frac{Y_r}{\sigma(X_r)}dW_r - \frac{s}{t-s}\int_s^t \frac{Y_r}{\sigma(X_r)}dW_r + \frac{sZ_s}{Y_s} + \int_0^s \left(\frac{\sigma'(X_r)Y_r}{\sigma(X_r)} - \frac{Z_r}{Y_r}\right)dr\right)$$

We now apply the previous representation formulas to specific stochastic differential equations. We start with the case of geometric Brownian motion:

Example 2.2.1 *Here, we suppose that the process $X = (X_t)_{0\leq t\leq T}$ satisfies the following SDE on $[0, T]$:*

$$dX_t = rX_t dt + \sigma X_t dW_t, \quad X_0 = x$$

In this particular case, the first and second variational processes are respectively given by $Y_t = \frac{X_t}{x}$ and $Z_t = 0$ for all $t \in [0, T]$. We can therefore explicitly determine the Skorohod integral involved in the representation formula of $\mathbb{E}(f(X_t)|X_s = \alpha)$. Using the expression of $\delta(u)$ given in theorem 2.2.2, we have:

$$\delta(u) = \frac{x}{X_s}\left(\frac{tW_s - sW_t}{\sigma s(t-s)} + 1\right)$$

Consequently, for all $\alpha \in \mathbb{R}$, we have:

$$\mathbb{E}(f(X_t)|X_s = \alpha) = \frac{\mathbb{E}\left(f(X_t)H(X_s - \alpha)\frac{1}{X_s}\left(\frac{tW_s - sW_t}{\sigma s(t-s)} + 1\right)\right)}{\mathbb{E}\left(H(X_s - \alpha)\frac{1}{X_s}\left(\frac{tW_s - sW_t}{\sigma s(t-s)} + 1\right)\right)}$$

We then apply the results to the case of an Ornstein-Uhlenbeck process:

Example 2.2.2 *We suppose that the process $X = (X_t)_{0\leq t\leq T}$ satisfies the following SDE on $[0,T]$:*

$$dX_t = a(\theta - X_t)dt + \sigma dW_t, \quad X_0 = x$$

Here, the first and second variational processes are respectively given by $Y_t = e^{-at}$ and $Z_t = 0$ for all $t \in [0,T]$. The expression of $\delta(u)$ in theorem 2.2.2 gives us:

$$\delta(u) = \frac{1}{\sigma s e^{-as}} \left(\int_0^s e^{-ar} dW_r - \frac{s}{t-s} \int_s^t e^{-ar} dW_r \right)$$

We therefore have, for all $\alpha \in \mathbb{R}$:

$$\mathbb{E}(f(X_t)|X_s = \alpha) = \frac{\mathbb{E}\left(f(X_t) H(X_s - \alpha) \left(\int_0^s e^{-ar} dW_r - \frac{s}{t-s} \int_s^t e^{-ar} dW_r \right) \right)}{\mathbb{E}\left(H(X_s - \alpha) \left(\int_0^s e^{-ar} dW_r - \frac{s}{t-s} \int_s^t e^{-ar} dW_r \right) \right)}$$

In order to be able to perform a Monte Carlo simulation using the above expression, we see that we need to determine the distribution of $\delta(u)$. More precisely, for all $s \in [0,T]$, we have to determine the distribution of the random variable A_s defined by $A_s = \sigma s e^{-as} \delta(u)$. We have:

$$\begin{aligned} A_s &= \int_0^s e^{-ar} dW_r - \frac{s}{t-s} \int_s^t e^{-ar} dW_r \\ &= \frac{s}{t-s} \int_0^t e^{-ar} \left(\frac{t-s}{s} \mathbb{1}_{\{r\leq s\}} - \mathbb{1}_{\{r\geq s\}} \right) dW_r \\ &= \frac{1}{t-s} \int_0^t e^{-ar} (t\mathbb{1}_{\{r\leq s\}} - s) dW_r \end{aligned}$$

A straightforward computation gives us $A_s \sim \mathcal{N}(0, a_s)$ where a_s is given by, for all $s \in [0,T]$:

$$a_s = \frac{1}{2a(t-s)^2} \left(\left(1 - e^{-2as}\right) t(t-2s) + s^2 \left(1 - e^{-2at}\right) \right)$$

Finally this gives us, for all $\alpha \in \mathbb{R}$:

$$\mathbb{E}(f(X_t)|X_s = \alpha) = \frac{\mathbb{E}(f(X_t) H(X_s - \alpha) N\sqrt{a_s})}{\mathbb{E}(H(X_s - \alpha) N\sqrt{a_s})}$$

where N is a standard Gaussian random variable.

We now wish to derive representation formulas for the derivative of conditional expectation, i.e for expressions of the form $\partial_\alpha \mathbb{E}(f(X_t)|X_s = \alpha)$. These formulas will subsequently enable us to compute the delta of Bermudan options. We only consider the more general case where the f is not necessarily smooth enough, and therefore our starting point is the second representation formula given in theorem 2.2.2:

Theorem 2.2.3 *For all $\alpha \in \mathbb{R}$, we have:*

$$\partial_\alpha \mathbb{E}(f(X_t)|X_s = \alpha) = \frac{\mathbb{E}(H(X_s - \alpha)\delta(u))\mathbb{L}_{s,t}[f](\alpha) - \mathbb{E}(f(X_t)H(X_s - \alpha)\delta(u))\mathbb{L}_{s,t}[1](\alpha)}{\mathbb{E}(H(X_s - \alpha)\delta(u))^2}$$

where the operator $\mathbb{L}_{s,t}[.](\alpha)$ is given by, for all $\alpha \in \mathbb{R}$:

$$\mathbb{L}_{s,t}[.](\alpha) = \mathbb{E}\left(.(X_t)H(X_s - \alpha)\delta^2(u) - .(X_t)H(X_s - \alpha)\int_0^t (D_r\delta(u))u_r dr\right)$$

and where the Skorohod integral $\delta(u)$ is given by:

$$\delta(u) = \frac{1}{sY_s}\left(\int_0^s \frac{Y_r}{\sigma(X_r)}dW_r - \frac{s}{t-s}\int_s^t \frac{Y_r}{\sigma(X_r)}dW_r + \frac{sZ_s}{Y_s} + \int_0^s \left(\frac{\sigma'(X_r)Y_r}{\sigma(X_r)} - \frac{Z_r}{Y_r}\right)dr\right)$$

Proof: We need to show that $\partial_\alpha \mathbb{E}(f(X_t)H(X_s - \alpha)\delta(u)) = \mathbb{L}_{s,t}[f](\alpha)$ and that $\partial_\alpha \mathbb{E}(H(X_s - \alpha)\delta(u)) = \mathbb{L}_{s,t}[1](\alpha)$. Assuming we can interchange the derivation and expectation operators, we have:

$$\begin{aligned}\partial_\alpha \mathbb{E}(f(X_t)H(X_s - \alpha)\delta(u)) &= \mathbb{E}(f(X_t)\partial_\alpha(H(X_s - \alpha))\delta(u)) \\ &= \mathbb{E}(f(X_t)\delta_\alpha(X_s)\delta(u))\end{aligned}$$

where δ_α denotes the Dirac delta function at α. Furthermore, using the specific expression of u and the fact that $\mathbb{E}\left(\int_0^t (D_r f(X_t))u_r dr|\sigma(f(X_t), X_s)\right) = 0$, we can write:

$$\begin{aligned}\partial_\alpha \mathbb{E}(f(X_t)H(X_s - \alpha)\delta(u)) &= \mathbb{E}(f(X_t)\delta_\alpha(X_s)\delta(u)) \\ &= \mathbb{E}\left(f(X_t)\delta_\alpha(X_s)\delta(u)\int_0^t (D_r X_s)u_r dr\right) \\ &= \mathbb{E}\left(\int_0^t (D_r H(X_s - \alpha)f(X_t))\delta(u)u_r dr\right) \\ &= \mathbb{E}(H(X_s - \alpha)f(X_t)\delta(\delta(u)u)) \\ &= \mathbb{E}\left(f(X_t)H(X_s - \alpha)\left(\delta^2(u) - \int_0^t (D_r\delta(u))u_r dr\right)\right) \\ &= \mathbb{L}_{s,t}[f](\alpha)\end{aligned}$$

where we have used the integration by parts formula between the last two lines. In order to show that $\partial_\alpha \mathbb{E}(H(X_s - \alpha)\delta(u)) = \mathbb{L}_{s,t}[1](\alpha)$, we repeat the same steps taking $f \equiv 1$. □

We now return to our two examples in order to derive expressions for the derivative of conditional expectations. Using theorem 2.2.3, we see that the only difficulty is to determine an expression for $D_r\delta(u)$.

Example 2.2.3 *In the case of geometric Brownian motion, recall that the Skorohod integral $\delta(u)$ is given by:*

$$\delta(u) = \frac{x}{X_s}\left(\frac{tW_s - sW_t}{\sigma s(t-s)} + 1\right)$$

We then have:

$$\begin{aligned} D_r\delta(u) &= -\frac{x}{X_s^2}\left(\frac{tW_s - sW_t}{\sigma s(t-s)} + 1\right)(D_r X_s) + \frac{x}{\sigma s(t-s)X_s}(D_r(tW_s - sW_t)) \\ &= -\frac{x}{X_s^2}\left(\frac{tW_s - sW_t}{\sigma s(t-s)} + 1\right)\sigma X_s \mathbb{1}_{\{r\leq s\}} + \frac{tx\mathbb{1}_{\{r\leq s\}} - sx\mathbb{1}_{\{r\leq t\}}}{\sigma s(t-s)X_s} \end{aligned}$$

A straightforward computation then gives us, after rearranging all the terms:

$$\int_0^t (D_r\delta(u))u_r dr = \frac{x^2}{X_s^2\sigma s(t-s)}\left(\frac{t}{\sigma} - \Delta W_{s,t}\right)$$

with $\Delta W_{s,t} = (tW_s - sW_t + \sigma s(t-s))$. *We then have:*

$$\delta^2(u) - \int_0^t (D_r\delta(u))u_r dr = \frac{x^2}{X_s^2\sigma s(t-s)}\left(\frac{\Delta W_{s,t}^2}{\sigma s(t-s)} - \frac{t}{\sigma} + \Delta W_{s,t}\right)$$

This finally yields, for all $\alpha \in \mathbb{R}$*:*

$$\mathbb{L}_{s,t}[.](\alpha) = \mathbb{E}\left(.(X_t)H(X_s - \alpha)\frac{x^2}{X_s^2\sigma s(t-s)}\left(\frac{\Delta W_{s,t}^2}{\sigma s(t-s)} - \frac{t}{\sigma} + \Delta W_{s,t}\right)\right)$$

The quantity $\partial_\alpha \mathbb{E}(f(X_t)|X_s = \alpha)$ *can then be readily computed by plugging the above expression of* $\mathbb{L}_{s,t}[.](\alpha)$ *into the formula given by theorem 2.2.3.*

Example 2.2.4 *In the case of the Ornstein-Uhlenbeck process, the Skorohod integral* $\delta(u)$ *is given by:*

$$\delta(u) = \frac{1}{\sigma s(t-s)e^{-as}}\int_0^t e^{-ar}(t\mathbb{1}_{\{r\leq s\}} - s)dW_r$$

The first step is to compute $D_r\delta(u)$*:*

$$\begin{aligned} D_r\delta(u) &= \frac{1}{\sigma s(t-s)e^{-as}} D_r\left(\int_0^t e^{-av}(t\mathbb{1}_{\{v\leq s\}} - s)dW_v\right) \\ &= \frac{t}{\sigma s(t-s)e^{-as}} D_r\left(\int_0^s e^{-av}dW_v\right) - \frac{s}{\sigma s(t-s)e^{-as}} D_r\left(\int_0^t e^{-av}dW_v\right) \\ &= \frac{te^{-ar}}{\sigma s(t-s)e^{-as}} - \frac{se^{-ar}}{\sigma s(t-s)e^{-as}} \\ &= \frac{e^{-a(r-s)}}{\sigma s} \end{aligned}$$

Given that $u_r = \frac{e^{-a(r-s)}}{\sigma}\left(\frac{1}{s}\mathbb{1}_{\{r\leq s\}} - \frac{1}{t-s}\mathbb{1}_{\{s\leq r\leq t\}}\right)$*, we now have:*

$$\begin{aligned} \int_0^t (D_r\delta(u))u_r dr &= \int_0^t \frac{e^{-2a(r-s)}}{s\sigma^2}\left(\frac{1}{s}\mathbb{1}_{\{r\leq s\}} - \frac{1}{t-s}\mathbb{1}_{\{s\leq r\leq t\}}\right)dr \\ &= \frac{e^{2as}}{s^2\sigma^2}\int_0^s e^{-2ar}dr - \frac{e^{2as}}{s(t-s)\sigma^2}\int_s^t e^{-2ar}dr \\ &= \frac{1}{2as^2\sigma^2}\left(e^{2as} - 1\right) + \frac{1}{2as(t-s)\sigma^2}\left(e^{-2a(t-s)} - 1\right) \end{aligned}$$

This finally gives us, for all $\alpha \in \mathbb{R}$:

$$\begin{aligned}\mathbb{L}_{s,t}[.](\alpha) &= \mathbb{E}\left(.(X_t)H(X_s-\alpha)\frac{1}{\sigma^2 s^2(t-s)^2 e^{-2as}}\left(\int_0^t e^{-ar}(t\mathbb{1}_{\{r\leq s\}}-s)dW_r\right)^2\right) \\ &- \mathbb{E}\left(.(X_t)H(X_s-\alpha)\left(\frac{1}{2as^2\sigma^2}\left(e^{2as}-1\right)+\frac{1}{2as(t-s)\sigma^2}\left(e^{-2a(t-s)}-1\right)\right)\right)\end{aligned}$$

2.3 The notion of localization

In the proof of theorem 2.1.1, the use of Bayes' formula within the context of conditional expectation introduces a Dirac delta function, which is very irregular. As a result, the technique involving the integration by parts formula enbables us to "eliminate" the Dirac delta function and we subsequently obtain an expression with a Heaviside function H. However, it has to be said that this procedure yields an increase in the variance of the resulting estimator, something which is specifically related to the presence of H. In order to overcome this potential drawback, the notion of localization has been developed in recent years (for example in [7] and [1]). Such a procedure involves the use of a certain localizing function $\psi : \mathbb{R} \to [0,\infty)$ satisfying $\int_{\mathbb{R}} \psi(t)dt = 1$. Let us mention that the use of the localized version of the representation formulas is absolutely necessary if one wishes to obtain satisfactory convergence results in the context of Monte Carlo simulation.

We now return to the case where the function f is not necessarily smooth, and therefore the random process u satisfies the two required conditions (2.1) and (2.2). The localized version of the representation theorem becomes, both for conditional expectation and its derivative:

Theorem 2.3.1 *For any function $\psi : \mathbb{R} \to [0,\infty)$ satisfying $\int_{\mathbb{R}} \psi(t)dt = 1$, and for all $\alpha \in \mathbb{R}$, we have:*

$$\mathbb{E}(f(X_t)|X_s = \alpha) = \frac{\mathbb{J}^{\psi}_{s,t}[f](\alpha)}{\mathbb{J}^{\psi}_{s,t}[1](\alpha)}$$

where the operator $\mathbb{J}^{\psi}_{s,t}[.](\alpha)$ is given by, for all $\alpha \in \mathbb{R}$:

$$\mathbb{J}^{\psi}_{s,t}[.](\alpha) = \mathbb{E}\left(.(X_t)\left(\psi(X_s-\alpha)+\delta(u)(H(X_s-\alpha)-\Psi(X_s-\alpha))\right)\right)$$

and where Ψ is the cumulative distribution function associated with ψ, i.e such that $\Psi(x) = \int_{-\infty}^{x} \psi(t)dt$. Furthermore, for all $\alpha \in \mathbb{R}$, we have:

$$\partial_\alpha \mathbb{E}(f(X_t)|X_s = \alpha) = \frac{\mathbb{J}^{\psi}_{s,t}[1](\alpha)\mathbb{L}^{\psi}_{s,t}[f](\alpha) - \mathbb{J}^{\psi}_{s,t}[f](\alpha)\mathbb{L}^{\psi}_{s,t}[1](\alpha)}{\mathbb{J}^{\psi}_{s,t}[1](\alpha)^2}$$

where the operator $\mathbb{L}^{\psi}_{s,t}[.](\alpha)$ *is given by, for all* $\alpha \in \mathbb{R}$*:*

$$\begin{aligned}\mathbb{L}^{\psi}_{s,t}[.](\alpha) &= \mathbb{E}\Big(.(X_t)\psi(X_s-\alpha)\delta(u)\Big) \\ &+ \mathbb{E}\left(.(X_t)(H(X_s-\alpha)-\Psi(X_s-\alpha))\Big(\delta^2(u)-\int_0^t (D_r\delta(u))u_r dr\Big)\right)\end{aligned}$$

Note that the Skorohod integral $\delta(u)$ *is still given by:*

$$\delta(u) = \frac{1}{sY_s}\left(\int_0^s \frac{Y_r}{\sigma(X_r)}dW_r - \frac{s}{t-s}\int_s^t \frac{Y_r}{\sigma(X_r)}dW_r + \frac{sZ_s}{Y_s} + \int_0^s \left(\frac{\sigma'(X_r)Y_r}{\sigma(X_r)} - \frac{Z_r}{Y_r}\right)dr\right)$$

Proof: According to theorem 2.2.2, we have for all $\alpha \in \mathbb{R}$:

$$\mathbb{E}(f(X_t)|X_s=\alpha) = \frac{\mathbb{E}(f(X_t)H(X_s-\alpha)\delta(u))}{\mathbb{E}(H(X_s-\alpha)\delta(u))}$$

We then write:

$$\begin{aligned}\mathbb{E}(f(X_t)H(X_s-\alpha)\delta(u)) &= \mathbb{E}(f(X_t)\psi(X_s-\alpha)) + \mathbb{E}(f(X_t)H(X_s-\alpha)\delta(u)) \\ &- \mathbb{E}(f(X_t)\psi(X_s-\alpha)) \\ &= \mathbb{E}(f(X_t)\psi(X_s-\alpha)) + \mathbb{E}(f(X_t)H(X_s-\alpha)\delta(u)) \\ &- \mathbb{E}(f(X_t)\Psi'(X_s-\alpha))\end{aligned}$$

But then we can apply exactly the same argument as in the proof of theorem 2.1.1 (i.e using the definition of the Skorohod integral and the integration by parts formula) to notice that $\mathbb{E}(f(X_t)\Psi'(X_s-\alpha)) = \mathbb{E}(f(X_t)\Psi(X_s-\alpha)\delta(u))$. This gives us the required result for the numerator. As far as the denominator is concerned, we apply the previous argument taking $f \equiv 1$. Moreover, let us notice that in order to obtain the representation formula of the derivative of conditional expectation, it now suffices to prove that $\mathbb{L}^{\psi}_{s,t}[.](\alpha) \equiv \mathbb{L}_{s,t}[.](\alpha)$ for all $\alpha \in \mathbb{R}$. We apply the same reasoning as above, where we introduce the quantity $\pi = \delta^2(u) - \int_0^t (D_r\delta(u))u_r dr$:

$$\begin{aligned}\mathbb{L}_{s,t}[f](\alpha) &= \mathbb{E}(f(X_t)H(X_s-\alpha)\pi) \\ &= \mathbb{E}(f(X_t)\psi(X_s-\alpha)\delta(u)) + \mathbb{E}(f(X_t)H(X_s-\alpha)\pi) \\ &- \mathbb{E}(f(X_t)\psi(X_s-\alpha)\delta(u)) \\ &= \mathbb{E}(f(X_t)\psi(X_s-\alpha)\delta(u)) + \mathbb{E}(f(X_t)H(X_s-\alpha)\pi) \\ &- \mathbb{E}(f(X_t)\Psi'(X_s-\alpha)\delta(u)) \\ &= \mathbb{E}(f(X_t)\psi(X_s-\alpha)\delta(u)) + \mathbb{E}(f(X_t)H(X_s-\alpha)\pi) \\ &- \mathbb{E}(f(X_t)\Psi(X_s-\alpha)\pi) \\ &= \mathbb{L}^{\psi}_{s,t}[f](\alpha)\end{aligned}$$

where we have used the definition of the Skorohod integral and the integration by parts formula between lines 3 and 4. The same reasoning can be applied to

the specific case $f \equiv 1$ and we get the required result. $\square$

Now that we have introduced the localized versions of the representation formulas for conditional expectation and its derivative, one natural question arises, namely the optimal choice of the localizing function ψ. Recall that in order to evaluate the quantity $\mathbb{E}(f(X_t)|X_s = \alpha)$, we need to compute $\mathbb{J}^{\psi}_{s,t}[.](\alpha)$, which is given by, for all $\alpha \in \mathbb{R}$:

$$\mathbb{J}^{\psi}_{s,t}[.](\alpha) = \mathbb{E}\left(.(X_t)\left(\psi(X_s - \alpha) + \delta(u)(H(X_s - \alpha) - \Psi(X_s - \alpha))\right)\right)$$

Practically speaking, this expectation is computed via Monte Carlo simulation, and the estimator which is used is simply the associated empirical mean. More precisely, if we denote by N the number of simulated paths, we have the following approximation, valid for all $\alpha \in \mathbb{R}$:

$$\mathbb{J}^{\psi}_{s,t}[.](\alpha) \approx \frac{1}{N}\sum_{q=1}^{N} .(X_t^{(q)})\left(\psi(X_s^{(q)} - \alpha) + \delta(u)(H(X_s^{(q)} - \alpha) - \Psi(X_s^{(q)} - \alpha))\right)$$

In order to reduce the variance as much as possible, the idea is to minimize the mean integrated variance with respect to the localizing function ψ. In other words, we have to solve the following program:

$$\inf_{\psi \in \mathcal{L}^1} I(\psi)$$

where $\mathcal{L}^1 = \{\psi : \mathbb{R} \to [0,\infty) : \psi \in \mathcal{C}^1(\mathbb{R}), \psi(+\infty) = 0, \int_{\mathbb{R}} \psi(t)dt = 1\}$ and where the mean integrated variance $I(\psi)$ is given by:

$$I(\psi) = \int_{\mathbb{R}} \mathbb{E}\left(.^2(X_t)\left(\psi(X_s - \alpha) + \delta(u)(H(X_s - \alpha) - \Psi(X_s - \alpha))\right)^2\right) d\alpha$$

The choice of the optimal localizing function ψ is given in the following proposition:

Proposition 2.3.1 *We have* $\inf_{\psi \in \mathcal{L}^1} I(\psi) = I(\psi^*)$*, where* ψ^* *is the probability density function of the Laplace distribution with parameter* $\lambda^*_.$*, i.e for all* $t \in \mathbb{R}$*,* $\psi^*(t) = \frac{\lambda^*_.}{2} e^{-\lambda^*_. |t|}$*, where* $\lambda^*_.$ *is given by:*

$$\lambda^*_. = \left(\frac{\mathbb{E}\left(.^2(X_t)\delta^2(u)\right)}{\mathbb{E}(.^2(X_t))}\right)^{1/2}$$

Remark 2.3.1 *The cumulative distribution function* Ψ *associated with the Laplace probability density function with parameter* $\lambda^*_.$ *is given by, for all* $t \in \mathbb{R}$*:*

$$\Psi(t) = \left(1 - \frac{1}{2}e^{-\lambda^*_. t}\right)\mathbb{1}_{\{t \geq 0\}} + \frac{1}{2}e^{\lambda^*_. t}\mathbb{1}_{\{t<0\}}$$

Proof: Using Fubini's theorem and the change of variables $a = X_s - \alpha$, we have:

$$I(\psi) = \mathbb{E}\left(\int_{\mathbb{R}} (.^2(X_t)\left(\psi(a) + \delta(u)(H(a) - \Psi(a))\right)^2 da\right)$$

We then fix $h \in \mathbb{R}$ and consider an integrable function $\hat{\psi}$ such that $\psi + h\hat{\psi}$ is in $\mathcal{L}^1$. The directional derivative of $I(\psi)$ in the direction $\hat{\psi}$ is then given by:

$$\begin{aligned}\nabla_{\hat{\psi}} I(\psi) &= \lim_{h\to 0} \frac{I(\psi + h\hat{\psi}) - I(\psi)}{h} \\ &= 2\mathbb{E}\left(\int_{\mathbb{R}} (.^2(X_t)\left(\hat{\psi}(a) - \delta(u)\hat{\Psi}(a)\right)\left(\psi(a) + \delta(u)(H(a) - \Psi(a))\right) da\right)\end{aligned}$$

The idea now is to prove that the only function in $\mathcal{L}^1$ such that $\nabla_{\hat{\psi}} I(\psi) = 0$ is the Laplace density function with parameter $\lambda_.^*$. Using the classical integration by parts formula, we can write:

$$\begin{aligned}\int_{\mathbb{R}} (.^2(X_t)\hat{\psi}(a)\left(\psi(a) + \delta(u)(H - \Psi)(a)\right) da &= \left[.^2(X_t)\hat{\Psi}(a)\left(\psi(a) + \delta(u)(H - \Psi)(a)\right)\right]_{-\infty}^{+\infty} \\ &- \int_{\mathbb{R}} (.^2(X_t)\hat{\Psi}(a)\left(\psi'(a) - \delta(u)\psi(a)\right) da \\ &= -\int_{\mathbb{R}} (.^2(X_t)\hat{\Psi}(a)\left(\psi'(a) - \delta(u)\psi(a)\right) da\end{aligned}$$

since the antiderivative $\hat{\Psi}$ of $\hat{\psi}$ satisfies $\lim_{a\to-\infty} \hat{\Psi}(a) = 0$ and given that $\lim_{a\to+\infty}(\psi(a) + (H - \Psi)(a)) = 0$. Using Fubini's theorem once again, this yields:

$$\nabla_{\hat{\psi}} I(\psi) = -2\int_{\mathbb{R}} \hat{\Psi}(a)\mathbb{E}\left((.^2(X_t)\left(\psi'(a) + \delta(u)(H - \Psi)(a)\right)\right) da$$

Given that $\hat{\psi}$ was chosen arbitrarily, $\nabla_{\hat{\psi}} I(\psi) = 0$ if and only if we have:

$$\mathbb{E}\left((.^2(X_t)\left(\psi'(a) + \delta(u)(H - \Psi)(a)\right)\right) = 0$$

This leads us to the following ODE:

$$\Psi''(a) - \lambda_.^2\Psi(a) + \lambda_.^2 H(a) = 0$$

subject to the boundary conditions $\Psi(+\infty) = 1$ and $\Psi(-\infty) = 0$ and where we have introduced the quantity:

$$\lambda_. = \left(\frac{\mathbb{E}\left(.^2(X_t)\delta^2(u)\right)}{\mathbb{E}(.^2(X_t))}\right)^{1/2}$$

Using the two boundary conditions and taking into consideration the fact that ψ is a probability density function, we obtain:

$$\Psi^*(a) = \left(1 - \frac{1}{2}e^{-\lambda_.^* a}\right)\mathbb{1}_{\{a\geq 0\}} + \frac{1}{2}e^{\lambda_.^* a}\mathbb{1}_{\{a<0\}}$$

and we subsequently deduce that:

$$\psi^*(t) = \frac{\lambda^*_\cdot}{2} e^{-\lambda^*_\cdot |t|}$$

The final step of the proof is to show that this choice of ψ^* effectively yields the minimum of I. If we choose another integrable function ϕ such that $\psi + \phi$ is in $\mathcal{L}^1$, and if we denote by Φ its antiderivative, we can write:

$$\begin{aligned} I(\psi^* + \phi) - I(\psi^*) &= \mathbb{E}\left(\int_{\mathbb{R}} .^2(X_t)(\phi(a) - \delta(u)\Phi(a))^2 da \right) + \nabla_\phi I(\psi^*) \\ &\geq 0 \end{aligned}$$

since $\nabla_\phi I(\psi^*) = 0$ for any integrable function ϕ. □

Now that we have introduced the notion of localization and given the general form of the optimal localizing functions, we go back to our two main examples, more specifically the geometric Brownian motion case and the Ornstein-Uhlenbeck case. Let us point out that we already have explicit expressions of $\delta(u)$, $D_r(\delta(u))$, ψ and Ψ, and that the only quantity we now need to determine is $\lambda^*_\cdot$. On the one hand, λ^*_f is approximated via a usual Monte Carlo estimator. On the other hand, we can find closed-form expressions for λ^*_1:

Example 2.3.1 *In the case of geometric Brownian motion, recall that the Skorohod integral $\delta(u)$ is given by:*

$$\delta(u) = \frac{x}{X_s}\left(\frac{tW_s - sW_t}{\sigma s(t-s)} + 1 \right)$$

Therefore we can write:

$$\begin{aligned} \mathbb{E}(\delta^2(u)) &= \mathbb{E}\left(\frac{x^2}{X_s^2}\left(\frac{tW_s - sW_t}{\sigma s(t-s)} + 1 \right)^2 \right) \\ &= \frac{x^2}{\sigma^2 s^2 (t-s)^2} \mathbb{E}\left(\frac{1}{X_s^2} \mathbb{E}\left(((t-s)(W_s + \sigma s) - s(W_t - W_s))^2 | \mathcal{F}_s \right) \right) \end{aligned}$$

By using the relations $\mathbb{E}\left(e^{-2\sigma W_s}\right) = e^{2s\sigma^2}$, $\mathbb{E}\left(W_s e^{-2\sigma W_s}\right) = -2\sigma s e^{2s\sigma^2}$ and $\mathbb{E}\left(W_s^2 e^{-2\sigma W_s}\right) = s(1 + 4s\sigma^2)e^{2s\sigma^2}$, a straightforward computation gives us:

$$\mathbb{E}(\delta^2(u)) = \frac{t + s(t-s)\sigma^2}{s(t-s)\sigma^2} e^{-2s(r - \frac{1}{2}\sigma^2) + 2s\sigma^2}$$

This finally yields:

$$\lambda^*_1 = \left(\frac{t + s(t-s)\sigma^2}{s(t-s)\sigma^2} \right)^{1/2} e^{-s(r - \frac{1}{2}\sigma^2) + s\sigma^2}$$

Example 2.3.2 *In the case of the Ornstein-Uhlenbeck process, the Skorohod integral $\delta(u)$ is given by:*

$$\delta(u) = \frac{1}{\sigma s(t-s)e^{-as}} \int_0^t e^{-ar}(t\mathbb{1}_{\{r\leq s\}} - s)dW_r$$

We then have:

$$\begin{aligned}
\mathbb{E}(\delta^2(u)) &= \frac{1}{\sigma^2 s^2(t-s)^2 e^{-2as}} \mathbb{E}\left(\left(\int_0^t e^{-ar}(t\mathbb{1}_{\{r\leq s\}} - s)dW_r\right)^2\right) \\
&= \frac{1}{\sigma^2 s^2(t-s)^2 e^{-2as}} \int_0^t e^{-2ar}(t\mathbb{1}_{\{r\leq s\}} - s)^2 dr \\
&= \frac{1}{2a\sigma^2 s^2(t-s)^2 e^{-2as}}(s^2(1-e^{-2at}) + t(t-2s)(1-e^{-2as}))
\end{aligned}$$

This finally gives us:

$$\lambda_1^* = \frac{1}{\sigma s(t-s)e^{-as}} \left(\frac{s^2(1-e^{-2at}) + t(t-2s)(1-e^{-2as})}{2a}\right)^{1/2}$$

3 Applications to the pricing and hedging of Bermudan options

The aim of this chapter is to apply the previous representation formulas for conditional expectation and its derivative to the context of pricing and hedging Bermudan options. We start with an overview of the problem and we briefly discuss existing alternative probabilistic pricing methods. We then introduce the Malliavin pricing and hedging algorithm before extending our analysis to some numerical applications.

3.1 Overview of the problem and alternative probabilistic pricing methods

We place ourselves within the same framework as the one considered in the introduction. We briefly recall the general aspects of this setting. We first discretize the interval $[0,T]$ using the time step $\delta t = \frac{T}{M}$ and we then take into consideration the discrete times $t_k = k\delta t$ for all $k \in \{0,...,M\}$. We consider a Bermudan option with maturity T with a payoff function Φ on an underlying asset whose price process $X = (X_t)_{0\leq t\leq T}$ satisfies the following SDE on $[0,T]$:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x$$

We now recall the backward recursive system, valid for all $k \in \{M-1,...,0\}$, which gives us the approximation $\hat{V}(0,x)$ of the option's price at time 0:

$$\begin{cases} \hat{V}(t_M, X_{M\delta t}) = \Phi(X_{M\delta t}) \\ \hat{V}(t_k, X_{k\delta t}) = \max\left(\Phi(X_{k\delta t}), e^{-r\delta t}\mathbb{E}\left(\hat{V}(t_{k+1}, X_{(k+1)\delta t})|X_{k\delta t}\right)\right) \end{cases}$$

As for the delta of the option, its approximation $\hat{\Delta}(0,x)$ at time 0 is given by:

$$\begin{cases} \hat{\Delta}(t_1, X_{\delta t}) = \partial_\alpha\Phi(\alpha)\mathbb{1}_{C^c} + e^{-r\delta t}\partial_\alpha\mathbb{E}\left(\hat{V}(t_2, X_{2\delta t})|X_{\delta t} = \alpha\right)\mathbb{1}_C \\ \hat{\Delta}(0,x) = \mathbb{E}\left(\hat{\Delta}(t_1, X_{\delta t})\right) \end{cases}$$

where the continuation region C is defined by $C = \{\alpha \in \mathbb{R}_+ : \hat{V}(t_1,\alpha) > \Phi(\alpha)\}$.

We now briefly introduce the method proposed by Longstaff and Schwartz in [11], which can be seen as the "closest" alternative pricing method to the Malliavin approach, as it involves the approximation of conditional expectations. From a practical viewpoint, the aim is to determine a stopping rule enabling the holder to know when exactly to exercise the option. We therefore need to derive an approximation of the conditional expectation of the value of the option at time t_{k+1} given the value of the underlying at time t_k. More precisely, for all $\alpha \in \mathbb{R}$, we consider the quantity $A_k(\alpha)$ defined by:

$$A_k(\alpha) = \mathbb{E}\left(\hat{V}(t_{k+1}, X_{(k+1)\delta t})|X_{k\delta t} = \alpha\right)$$

The major idea behind applying least square methods in this specific case is to approximate such a conditional expectation via a set of R basis functions $\psi_1, ..., \psi_R$ (which vary from Fourier series to specific polynomials):

$$A_k(\alpha) \approx \sum_{r=1}^{R} \beta_{kr}\psi_r(\alpha) = \beta_k^T \psi(\alpha)$$

where $\beta_k = (\beta_{k1}, ..., \beta_{kR})^T$ and $\psi(\alpha) = (\psi_1(\alpha), ..., \psi_R(\alpha))^T$. The aim is therefore to solve the following minimization program:

$$\min_{\beta} \mathbb{E}\left((A_k(\alpha) - \beta_k^T\psi(\alpha))^2\right)$$

Setting the derivative of the above quantity with respect to β equal to zero and using the tower property for conditional expectation, the optimal vector β_k^* is given by:

$$\beta_k^* = \mathbb{M}_\psi^{-1}\mathbb{V}_{(\psi,\hat{V})}$$

where $\mathbb{M}_\psi$ is assumed to be a nonsingular $R \times R$ matrix given by:

$$\mathbb{M}_\psi = \mathbb{E}\left(\psi(X_{k\delta t})\psi(X_{k\delta t})^T\right)$$

and where the vector $\mathbb{V}_{(\psi,\hat{V})}$ of size R is defined by:

$$\mathbb{V}_{(\psi,\hat{V})} = \mathbb{E}\left(\psi(X_{k\delta t})\hat{V}(t_{k+1}, X_{(k+1)\delta t})\right)$$

The components of the vector β_k^* are then estimated by observing the pairs $(X_{k\delta t}^{(q)}, \hat{V}(t_{k+1}, X_{(k+1)\delta t}^{(q)}))$, for all $q \in \{1, ..., N\}$, N being the number of samples. The least square estimate of the vector β_k^* is then given by:

$$\hat{\beta}_k^* = \hat{\mathbb{M}}_\psi^{-1}\hat{\mathbb{V}}_{(\psi,\hat{V})}$$

where entry (i, j) of the R-dimensional square matrix $\hat{\mathbb{M}}_\psi$ is given by:

$$(\hat{\mathbb{M}}_\psi)_{(i,j)} = \frac{1}{N}\sum_{q=1}^{N} \psi_i(X_{k\delta t}^{(q)})\psi_j(X_{k\delta t}^{(q)})$$

and where the r-th component of the R-dimensional vector $\hat{\mathbb{V}}_{(\psi,\hat{V})}$ is given by:

$$(\hat{\mathbb{V}}_{(\psi,\hat{V})})_r = \frac{1}{N}\sum_{q=1}^{N} \psi_r(X_{k\delta t}^{(q)})\hat{V}(t_{k+1}, X_{(k+1)\delta t}^{(q)})$$

The resulting vector $\hat{\beta}_k^*$ enables us to deduce an approximation of the continuation value for all $\alpha \in \mathbb{R}$:

$$A_k(\alpha) \approx (\hat{\beta}_k^*)^T \psi(\alpha)$$

Once all the continuation values have been estimated, the idea is to go backwards in time to finally deduce the price of the option at time 0. More precisely, the main steps of the Longstaff-Schwartz algorithm can be summed up as follows:

1. Simulate N paths of M time steps.

2. At maturity, set $\hat{V}(t_M, X_{M\delta t}) = \Phi(X_{M\delta t})$.

3. Use the following backward recursion, for all $k = M-1, ..., 1$:

   (a) Compute $\hat{\beta}_k^* = \hat{\mathbb{M}}_\psi^{-1}\hat{\mathbb{V}}_{(\psi,\hat{V})}$.

   (b) Evaluate the continuation value at time t_k for path q.

   (c) Compute the value of the option at time t_k for path q:

   $$\hat{V}(t_k, X_{k\delta t}^{(q)}) = \max\left(\Phi(X_{k\delta t}^{(q)}), e^{-r\delta t}\mathbb{E}\left(\hat{V}(t_{k+1}, X_{(k+1)\delta t}^{(q)})|X_{k\delta t}^{(q)}\right)\right)$$

4. Approximate the price at time 0 by averaging over the N paths:

$$V(0,x) \approx \frac{1}{N}\sum_{q=1}^{N} \hat{V}(\delta t, X_{\delta t}^{(q)})$$

As we shall see in the next section, the fundamental difference between the Longstaff-Schwartz approach and the Malliavin pricing and hedging algorithm is based on the way the conditional expectations which are involved in the dynamic programming principle formulation are approximated. Furthermore, the Longstaff-Schwartz method is unable to provide an *ad hoc* method for the computation of the delta. It has to be combined with other methods such as the likelihood ratio method or pathwise sensitivities based approaches in order to obtain an approximation of the delta. However, with the representation formulas for the derivative of conditional expectation, the algorithm presented here is able to provide an estimation of the delta using its own specific method.

3.2 The Malliavin pricing and hedging algorithm

We now present the Malliavin pricing and hedging algorithm by putting into practice the representation forumulas for conditional expectation and its derivative. The first part of this section is devoted to an explanation of how these formulas have to be used from a practical point of view. The second part provides a sketch of the pricing and hedging algorithm. Here, we follow the approach developed in [1], albeit in a somewhat more generalised setting, as the algorithm we present is applicable to any underlying process following the SDE in (1). Let us mention that the basic data involved throughout this algorithm are represented by the following set:

$$\mathcal{Z} = \{Z_k^{(q)} \sim \mathcal{N}(0,1), k \in \{M, ..., 1\}, q \in \{1, ..., N\}\}$$

where N is the number of samples and where M is the number of subintervals for the discretization of time (and hence the number of discrete exercise dates).

3.2.1 Applications of the representation formulas

The backward recursive system which enables us to approximate the price of a Bermudan option at time 0 is given by, for all $k \in \{M-1, ..., 0\}$:

$$\begin{cases} \hat{V}(t_M, X_{M\delta t}) = \Phi(X_{M\delta t}) \\ \hat{V}(t_k, X_{k\delta t}) = \max\left(\Phi(X_{k\delta t}), e^{-r\delta t}\mathbb{E}\left(\hat{V}(t_{k+1}, X_{(k+1)\delta t})|X_{k\delta t}\right)\right) \end{cases}$$

In order to be able to go back recursively in time, we see that we need to determine conditional expectations $\mathbb{E}\left(\hat{V}(t_{k+1}, X_{(k+1)\delta t})|X_{k\delta t}\right)$ for all $k \in \{M-1, ..., 0\}$. Using the localized version of the representation theorem, i.e theorem 2.3.1, we can write, for all $\alpha \in \mathbb{R}$:

$$\mathbb{E}\left(\hat{V}(t_{k+1}, X_{(k+1)\delta t})|X_{k\delta t} = \alpha\right) = \frac{\mathbb{J}_{k,k+1}^{\psi}[\hat{V}](\alpha)}{\mathbb{J}_{k,k+1}^{\psi}[1](\alpha)}$$

where the operator $\mathbb{J}_{k,k+1}^{\psi}[.](\alpha)$ is given by, for all $\alpha \in \mathbb{R}$:

$$\mathbb{J}_{k,k+1}^{\psi}[.](\alpha) = \mathbb{E}\left(.(X_{(k+1)\delta t})\left(\psi(X_{k\delta t} - \alpha) + \delta_k(u)(H(X_{k\delta t} - \alpha) - \Psi(X_{k\delta t} - \alpha))\right)\right)$$

where we have introduced the subscript k on the Skorohod integral in order to emphasize its dependence on the time step.

In order to compute such conditional expectations, we naturally take into consideration the associated empirical means. To this purpose, for a given sample $q \in \{1, ..., N\}$, we introduce the operator $\mathcal{E}_k[.](X_{k\delta t}^{(q)})$ defined by:

$$\mathcal{E}_k[.](X_{k\delta t}^{(q)}) = \mathbb{E}\left(.(X_{(k+1)\delta t})|X_{k\delta t} = \alpha\right)\Big|_{\alpha = X_{k\delta t}^{(q)}}$$

One crucial point of the algorithm is to understand that for a given time t_k, we have to take into account a stochastic space component $X_{k\delta t}^{(q)}$, and therefore the quantity $\mathcal{E}_k[.](X_{k\delta t}^{(q)})$ has to be computed for each sample $q \in \{1, ..., N\}$. Introducing the empirical means, the approximation is given by:

$$\mathcal{E}_k[.](X_{k\delta t}^{(q)}) \approx \frac{\mathcal{J}_{k,k+1}^{\psi}[.](X_{k\delta t}^{(q)})}{\mathcal{J}_{k,k+1}^{\psi}[1](X_{k\delta t}^{(q)})} \tag{3.1}$$

where the operator $\mathcal{J}_{k,k+1}^{\psi}[.](\alpha)$ is given by:

$$\begin{aligned}\mathcal{J}_{k,k+1}^{\psi}[.](X_{k\delta t}^{(q)}) &= \frac{1}{N}\sum_{q'=1}^{N} .(X_{(k+1)\delta t}^{(q')})\psi(X_{k\delta t}^{(q')} - X_{k\delta t}^{(q)}) \\ &+ \frac{1}{N}\sum_{q'=1}^{N} .(X_{(k+1)\delta t}^{(q')})\delta_k^{(q')}(u)(H(X_{k\delta t}^{(q')} - X_{k\delta t}^{(q)}) - \Psi(X_{k\delta t}^{(q')} - X_{k\delta t}^{(q)}))\end{aligned}$$

and where the Skorohod integral also depends on the sample.

The Laplace probability density function ψ and its associated cumulative distribution function Ψ also have to be computed at each time t_k, since they depend on the parameter $\lambda_{\cdot}^{*}$, which depends on k. From now on, we shall drop the subscript $*$ and replace it by k so as to underline the dependence of this parameter on the time step:

$$\lambda_{\cdot}^{k} = \left(\frac{\mathbb{E}\left(.^2(X_{k\delta t})(\delta_k(u))^2\right)}{\mathbb{E}(.^2(X_{k\delta t}))}\right)^{1/2}$$

In order to compute $\lambda_{\cdot}^{*}$, we take into consideration the associated empirical means:

$$\lambda_{\cdot}^{k} = \left(\frac{\sum_{q=1}^{N} .^2(t_k, X_{k\delta t}^{(q)})(\delta_k^{(q)}(u))^2}{\sum_{q=1}^{N} .^2(t_k, X_{k\delta t}^{(q)})}\right)^{1/2}$$

In the particular case of λ_1^k, we generally have closed-form expressions, which we shall give explicitly in the case of geometric Brownian motion and the Ornstein-Uhlenbeck process.

We now examine the application of the representation formula of the derivative of conditional expectation in order to be able to compute an approximation of the delta of the option. Recall that the approximation $\hat{\Delta}(0, x)$ of the delta at time 0 is given by:

$$\begin{cases} \hat{\Delta}(t_1, X_{\delta t}) = \partial_\alpha \Phi(\alpha) \mathbb{1}_{C^c} + e^{-r\delta t} \partial_\alpha \mathbb{E}\left(\hat{V}(t_2, X_{2\delta t}) | X_{\delta t} = \alpha\right) \mathbb{1}_C \\ \hat{\Delta}(0, x) = \mathbb{E}\left(\hat{\Delta}(t_1, X_{\delta t})\right) \end{cases}$$

Using theorem 2.3.1, we can write, for all $\alpha \in \mathbb{R}$:

$$\partial_\alpha \mathbb{E}(.(X_{2\delta t})|X_{\delta t} = \alpha) = \frac{\mathbb{J}^\psi_{1,2}[1](\alpha)\mathbb{L}^\psi_{1,2}[.](\alpha) - \mathbb{J}^\psi_{1,2}[.](\alpha)\mathbb{L}^\psi_{1,2}[1](\alpha)}{\mathbb{J}^\psi_{1,2}[1](\alpha)^2}$$

where the operator $\mathbb{L}^\psi_{1,2}[.](\alpha)$ is given by, for all $\alpha \in \mathbb{R}$:

$$\begin{aligned}\mathbb{L}^\psi_{1,2}[.](\alpha) &= \mathbb{E}\Big(.(X_{2\delta t})\psi(X_{\delta t} - \alpha)\delta_1(u)\Big) \\ &+ \mathbb{E}\Big(.(X_{2\delta t})(H(X_{\delta t} - \alpha) - \Psi(X_{\delta t} - \alpha))\left(\delta_1^2(u) - \int_0^t (D_r\delta_1(u))u_r dr\right)\Big)\end{aligned}$$

For a given sample $q \in \{1, ..., N\}$, we introduce the operator $\mathcal{D}[.](X^{(q)}_{\delta t})$ defined by:

$$\mathcal{D}[.](X^{(q)}_{\delta t}) = \partial_\alpha \mathbb{E}\left(.(X_{2\delta t})|X_{\delta t} = \alpha\right)|_{\alpha = X^{(q)}_{\delta t}}$$

Once again we use the associated empirical means in order to approximate the above quantity:

$$\mathcal{D}[.](X^{(q)}_{\delta t}) \approx \frac{\mathcal{J}^\psi_{1,2}[1](X^{(q)}_{\delta t})\mathcal{L}^\psi_{1,2}[.](X^{(q)}_{\delta t}) - \mathcal{J}^\psi_{1,2}[.](X^{(q)}_{\delta t})\mathcal{L}^\psi_{1,2}[1](X^{(q)}_{\delta t})}{\mathcal{J}^\psi_{1,2}[1](X^{(q)}_{\delta t})^2} \tag{3.2}$$

where the operator $\mathcal{L}^\psi_{1,2}[.](X^{(q)}_{\delta t})$ is given by:

$$\begin{aligned}\mathcal{L}^\psi_{1,2}[.](X^{(q)}_{\delta t}) &= \frac{1}{N}\sum_{q'=1}^N .(X^{(q')}_{2\delta t})\psi(X^{(q')}_{\delta t} - X^{(q)}_{\delta t})\delta_1^{(q')}(u) \\ &+ \frac{1}{N}\sum_{q'=1}^N .(X^{(q')}_{2\delta t})(H(X^{(q')}_{\delta t} - X^{(q)}_{\delta t}) - \Psi(X^{(q')}_{\delta t} - X^{(q)}_{\delta t}))\pi_1^{(q')}\end{aligned}$$

and where the quantity $\pi_1^{(q')}$ is given by:

$$\pi_1^{(q')} = (\delta_1^{(q')}(u))^2 - \int_0^t (D_r\delta_1^{(q')}(u))u_r dr$$

Before going into the detailed steps of the algorithm, we return to our two examples - geometric Brownian motion and the Ornstein-Uhlenbeck process - and give more specific expressions of the above quantities.

Example 3.2.1 *In the case of geometric Brownian motion, the Skorohod integral $\delta_k^{(q)}$ at time t_k and for sample q is given by:*

$$\delta_k^{(q)}(u) = \frac{x}{X^{(q)}_{k\delta t}}\left(\frac{(k+1)\delta t W^{(q)}_{k\delta t} - k\delta t W^{(q)}_{(k+1)\delta t}}{\sigma k \delta t^2} + 1\right)$$

Furthermore, the parameter λ_1^k is in this case given by:

$$\lambda_1^k = \left(\frac{(k+1)\delta t + k\delta t^2\sigma^2}{k\delta t^2\sigma^2} \right)^{1/2} e^{-k\delta t(r-\frac{1}{2}\sigma^2)+k\delta t\sigma^2}$$

Example 3.2.2 *In the case of the Ornstein-Uhlenbeck process, the Skorohod integral $\delta_k^{(q)}$ at time t_k and for sample q is simulated using the following distribution equality:*

$$\delta_k^{(q')}(u) = \frac{1}{\sigma k\delta t^2 e^{-ak\delta t}} \left(\frac{k^2\delta t^2(1-e^{-2a(k+1)\delta t}) + \delta t^2(1-k^2)(1-e^{-2ak\delta t})}{2a} \right)^{1/2} Z_k^{(q')}$$

where $Z_k^{(q')}$ is a standard Gaussian random variable.

As for the paramter λ_1^k, we have the follwing expression:

$$\lambda_1^k = \frac{1}{\sigma k\delta t^2 e^{-ak\delta t}} \left(\frac{k^2\delta t^2(1-e^{-2a(k+1)\delta t}) + \delta t^2(1-k^2)(1-e^{-2ak\delta t})}{2a} \right)^{1/2}$$

3.2.2 Sketch of the algorithm

Having introduced the applications of the representation formulas for conditional expectation and its derivative within the context of the pricing and hedging of Bermudan options, we are now in a position to provide a detailed sketch of how the algorithm actually works. This pricing and hedging algorithm is comprised of the five main following steps:

1. Set the parameters related to the underlying price process, set the payoff of the option Φ, its derivative $\partial\Phi$, and the maturity of the option T. Set the number of samples N and the number of discretized subintervals for time M. Define the time step $\delta t = \frac{T}{M}$.

2. At time step M, i.e at maturity, simulate the underlying price process $X_{M\delta t}^{(q)}$ for all $q \in \{1, ..., N\}$. Introduce the sequence of functions $v = (v_m)_{m\in\{M,...,0\}}$ and initialize it by setting $v_M(X_{M\delta t}^{(q)}) = \Phi(X_{M\delta t}^{(q)})$ for all $q \in \{1, ..., N\}$.

3. For all $k \in \{M-1, ..., 2\}$, do the following:

   (a) Simulate the underlying price process $X_{k\delta t}^{(q)}$ for all $q \in \{1, ..., N\}$.

   (b) Compute the Skorohod integral $\delta_k^{(q)}(u)$ for all $q \in \{1, ..., N\}$.

   (c) Compute $\mathcal{E}_k[v_{k+1}](X_{k\delta t}^{(q)})$ for all $q \in \{1, ..., N\}$ using (3.1).

   (d) Set $v_k(X_{k\delta t}^{(q)}) = \max(\Phi(X_{k\delta t}^{(q)}), e^{-r\delta t}\mathcal{E}_k[v_{k+1}](X_{k\delta t}^{(q)}))$ for all $q \in \{1, ..., N\}$.

4. Repeat step 3 with $k = 1$. Then do the following:

   (a) Compute $\mathcal{D}[v_2](X_{\delta t}^{(q)})$ for all $q \in \{1, ..., N\}$ using (3.2).

   (b) For all $q \in \{1, ..., N\}$, compute the quantity $w_1(X_{\delta t}^{(q)})$ defined by:

$$w_1(X_{\delta t}^{(q)}) = \partial\Phi(X_{\delta t}^{(q)})\mathbb{1}_{\{v_1(X_{\delta t}^{(q)}) < \Phi(X_{\delta t}^{(q)})\}} + e^{-r\delta t}\mathcal{D}[v_2](X_{\delta t}^{(q)})\mathbb{1}_{\{v_1(X_{\delta t}^{(q)}) \geq \Phi(X_{\delta t}^{(q)})\}}$$

5. Do the following:

   (a) Compute $\mathcal{E}_0[v_1](x) = \frac{1}{N}\sum_{q=1}^{N} v_1(X_{\delta t}^{(q)})$ and set $v_0(x) = \max(\Phi(x), e^{-r\delta t}\mathcal{E}_0[v_1](x))$.

   (b) Compute $w_0(x) = \frac{1}{N}\sum_{q=1}^{N} w_1(X_{\delta t}^{(q)})$.

   The approximation of the price and the delta are then respectively given by the quantities $v_0(x)$ and $w_0(x)$.

We now examine how the underlying price processes are simulated in our two specific examples:

Example 3.2.3 *We first look at the case of geometric Brownian motion. For a given sample $q \in \{1, ..., N\}$, recall the general expression of $X_{k\delta t}^{(q)}$ as a function of its initial value x, for all $k \in \{1, ..., M\}$:*

$$X_{k\delta t}^{(q)} = x \exp\left((r - \frac{1}{2}\sigma^2)k\delta t + \sigma W_{k\delta t}^{(q)}\right)$$

Given the fact that the algorithm is of backward type in time, we start by simulating $W_{M\delta t}^{(q)} = \sqrt{M\delta t}Z_M^{(q)}$, where $Z_M^{(q)}$ is a standard Gaussian random variable. We can then deduce $X_{M\delta t}^{(q)}$ using the above expression. If we go back in time by one time step, we see that we first need to determine $W_{(M-1)\delta t}^{(q)}$ in order to obtain $X_{(M-1)\delta t}^{(q)}$. To this purpose, we use the Brownian bridge method, i.e we use the fact that for any $t > s$, the distribution of W_s given $W_t = w$ is Gaussian with mean $\frac{sw}{t}$ and variance $\frac{s(t-s)}{t}$. This gives us the following distribution equality, valid for all $k \in \{M-1, ..., 1\}$:

$$W_{k\delta t}^{(q)} = \frac{k}{k+1}W_{(k+1)\delta t}^{(q)} + \sqrt{\frac{k\delta t}{k+1}}Z_k^{(q)}$$

where $Z_k^{(q)}$ is a standard Gaussian random variable.

Example 3.2.4 *We now examine the case of the Ornstein-Uhlenbeck process. Here, we use the Ornstein-Uhlenbeck bridge method introduced by Bladt and Sorensen in [3]. Recall that for any $t > s$, we have the following distribution equality:*

$$X_t = X_s e^{-a(t-s)} + \theta(1 - e^{-a(t-s)}) + \sigma\sqrt{\frac{1 - e^{-2a(t-s)}}{2a}}Z$$

where Z is a standard Gaussian random variable. In order to obtain the Ornstein-Uhlenbeck bridge, for a given sample $q \in \{1, ..., N\}$, we start by generating $\hat{X}^{(q)}_{\delta t}, ..., \hat{X}^{(q)}_{M\delta t}$ defined by:

$$\hat{X}^{(q)}_{k\delta t} = \hat{X}^{(q)}_{(k-1)\delta t} e^{-a\delta t} + \theta(1 - e^{-a\delta t}) + \sigma \sqrt{\frac{1 - e^{-2a\delta t}}{2a}} Z^{(q)}_k$$

where the random variables $Z^{(q)}_k$ are independent. We then set $X^{(q)}_{M\delta t}$ in the following way:

$$X^{(q)}_{M\delta t} = x e^{-aM\delta t} + \theta(1 - e^{-aM\delta t}) + \sigma \sqrt{\frac{1 - e^{-2aM\delta t}}{2a}} Z^{(q)}_M$$

We finally go back in time using the following distribution equality, valid for all $k \in \{M - 1, ..., 1\}$:

$$X^{(q)}_{k\delta t} = \hat{X}^{(q)}_{k\delta t} + (X^{(q)}_{M\delta t} - \hat{X}^{(q)}_{M\delta t}) \frac{e^{ak\delta t} - e^{-ak\delta t}}{e^{aM\delta t} - e^{-aM\delta t}}$$

Remark 3.2.1 *Apart from the use of localizing functions to reduce the variance of the two previous estimators, Bally et al. [1] explain how the introduction of a control variate can be used in order to gain more precision. Let us first recall the main idea behind this specific variance reduction technique. Heuristically speaking, if we want to approximate the expectation $\mathbb{E}(f)$ of a given payoff f, and if there exists another payoff g such that its expectation is explicitly known, it is possible to reduce the variance of the estimator $\frac{1}{N}\sum_{q=1}^{N} f^{(q)}$ by replacing it by $\frac{1}{N}\sum_{q=1}^{N} f^{(q)} - \lambda\left(\frac{1}{N}\sum_{q=1}^{N} g^{(q)} - \mathbb{E}(g)\right)$, where λ is a real constant to be determined. It can be shown that the variance of the new estimator is equal to $\frac{1}{N}(Var(f) - 2\lambda Cov(f, g) + \lambda^2 Var(g))$, and this quantity is minimal for $\lambda^* = \frac{Cov(f,g)}{Var(g)}$. The resulting variance, equal to $\frac{1}{N}Var(f)(1 - Corr(f, g)^2)$, precisely shows that the idea is to choose a payoff g which is well correlated with f in absolute value. In the particular case of Bermudan options, Bally et al. [1] suggest the use of the associated European option as a control variate. If we denote by $V^B(t, x)$ and $V^E(t, x)$ the respective prices of a Bermudan and European call with the same maturity and strike, we define the quantity $V(t, x) = V^B(t, x) - V^E(t, x)$. We then introduce the new time-dependent obstacle $\hat{\Phi}(t, x) = \Phi(x) - V^E(t, x)$, which naturally satisfies $\hat{\Phi}(T, x) = 0$, and by construction we have:*

$$V(t, x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}_{t,x}\left(e^{-r(\tau - t)} \hat{\Phi}(\tau, X_\tau)\right)$$

In order to obtain the approximations of the price and delta of the Bermudan option, we first carry out all the previous steps with the new obstacle $\hat{\Phi}(t, x)$, and then we simply add the associated price and delta of the European call at time 0 to the estimators $v_0(x)$ and $w_0(x)$.

3.3 Numerical applications

In this final section, we discuss numerical results obtained after having implemented the Malliavin pricing and hedging algorithm in C++. As we have previously mentioned, the algorithm does not numerically work if the localization method is not taken into account. The following results are therefore obtained with the use of localizing functions. We shall base ourselves on the two main examples we have used throughout the dissertation, namely the case of geometric Brownian motion and the Ornstein-Uhlenbeck process.

We start with the case of geometric Brownian motion. To this purpose, we consider a non-dividend paying Bermudan put option with strike $K = 100$ and maturity $T = 1$ year on an underlying stock with the following characteristics:

- Annual risk-free interest rate $r = 9.53\%$.
- Annual volatility of returns $\sigma = 20\%$.
- Initial stock value $x = 100$.

According to online pricers (for example, the one which can be found at http://www.intrepid.com/robertl/option-pricer1.html) the "reference" price and delta of an American option with the same characteristics are 4.891 and -0.403. The following results are given with the number of paths N ranging from 100 to 20000, and the number of time steps M ranging from 10 to 30. Given that an American option can be exercised at any time before maturity, its price should in principle be greater than the price of the associated Bermudan option, which can only be exercised at a finite number of dates. We should therefore expect a convergence of the approximated prices towards the "upper bound" of 4.891:

N	$M = 10$	$M = 20$	$M = 30$
100	7.058	15.289	10.808
500	5.436	5.367	5.665
1000	5.318	5.298	5.154
2000	5.255	5.187	4.739
3000	5.109	4.769	4.803
10000	4.765	4.792	4.821
20000	4.777	4.814	4.829

Table 3.1: Approximation of the option's price with localization

The results given in table 3.1 illustrate several interesting points. To start with, the algorithm is very unstable for a very low number of paths, and only starts to provide reasonable prices from $N = 1000$ onwards. Furthermore, the condition that the Bermudan option's price should always be less than the "reference" value of the associated American option is not always satisfied. It only happens to be verified when the number of paths is sufficiently large. Finally,

taking a minimal number of paths equal to 10000, we do observe the reasonable increasing order of prices as the number of time steps (i.e the exercise dates) increases.

We now have a look at the standard deviation of the Malliavin price estimator as a function of the number of paths, with different time steps taken into account:

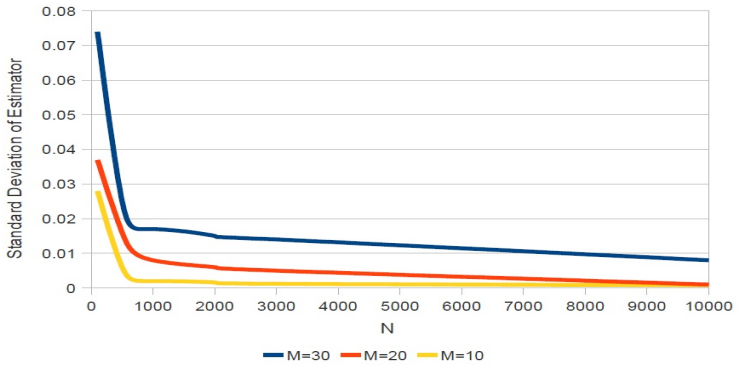


Figure 3.1: Standard Deviation of the Malliavin price estimator as a function of the number of paths

Figure 3.1 above underscores two fundamental aspects. On the one hand, regardless of the number of chosen time steps, we observe a decrease in the estimator's standard deviation as the number of paths increases. This is a standard property related to Monte Carlo estimators. On the other hand, the more interesting result is that the three curves plot above one another in a very particular order: for a given sample size, as the number of exercise dates increases, the standard deviation of the estimator also increases. This precisely shows us that if we want to increase the number of exercise dates, we also need to increase the number of paths in order to obtain the same precision. This particular trade-off between the number of exercise dates and standard deviation is underlined within the context of Bermudan options pricing by Hepperger [9]. Intuitively speaking, one possible explanation is that a higher number of exercise dates enables the estimator to capture the effects of more extreme points on the numerical grid, thus entailing an increase in its variance and standard deviation.

We now analyse the results of the approximation of the option's delta. Table 3.2 below shows that the approximations of the delta happen to be quite unstable for a number of paths less than 2000. The convergence towards the "reference " value of -0.403 is observable for higher values of N and M, although we do sense that additional variance reduction techniques such as the introduction of a control variate (Remark 3.2.1) would enable us to achieve a greater level of convergence. Unfortunately, our own implementation involving

the control variate method did not yield any satisfactory results, and they have therefore not been included here.

N	$M = 10$	$M = 20$	$M = 30$
100	-0.901	-0.256	-0.144
500	-0.159	-0.133	-0.278
1000	-0.303	-0.187	-0.532
2000	-0.567	-0.544	-0.512
3000	-0.511	-0.515	-0.485
10000	-0.502	-0.497	-0.469
20000	-0.491	-0.475	-0.442

Table 3.2: Approximation of the option's delta with localization

For illustrative purposes only, we also show the approximations of the price of the same Bermudan option when the localization technique is not taken into consideration:

N	Price	Standard Deviation
100	7011	1081
500	23701	3375
1000	19970	2748
2000	343588	28426
3000	2424299	202969

Table 3.3: Approximation of the option's price without localization, taking M=10 time steps

Table 3.3 precisely shows to what extent the non-localized version of the algorithm does not numerically work. We observe a "blow-up" of the prices and standard deviations as the number of paths increases.

We now consider a non-dividend paying Bermudan put option with strike $K = 3\%$ and maturity $T = 1$ year on an arbitrary underlying asset (which could be an interest rate in this case) represented by an Ornstein-Uhlenbeck price process. Let us mention that as "true" prices of an option on an underlying with such dynamics cannot be found online or in existing literature, the following analysis is purely indicative. We are therefore more interested in the behaviour of the estimators as N and M vary than in the actual value of the option. We choose an annual risk-free interest rate of $r = 9.53\%$ and the underlying asset has the following characteristics:

- Rate of mean-reversion $a = 10\%$.
- Long-term average $\theta = 4\%$.

- Annual volatility of returns $\sigma = 20\%$.
- Initial asset value $x = 3\%$.

N	$M = 10$	$M = 20$	$M = 30$
100	0.164	0.291	0.679
500	1.256	1.187	0.821
1000	1.071	1.501	2.218
2000	0.618	0.656	0.663
3000	0.740	0.783	0.794
10000	0.789	0.813	0.819
20000	0.799	0.826	0.831

Table 3.4: Approximation of the option's price (in %) with localization

The results are given in table 3.3 below. Once again, the approximation procedure seems to be rather unstable for a number of paths less than 2000. Even though we do not have any "reference" price in this particular case, we do observe a convergence towards a certain value greater than 0.831%. We also notice the increasing sequence of prices as the number of exercise dates increases for a given number of paths, which is consistent with our previous comments regarding this matter.

We end this numerical section by analysing the required CPU time of the Malliavin pricing algorithm, which is represented in figure 3.2 below:

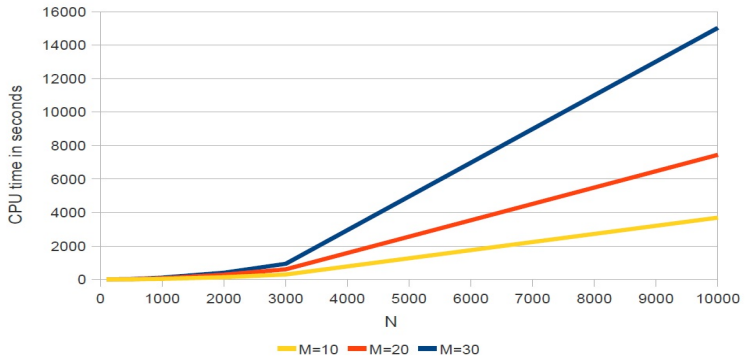


Figure 3.2: CPU time as a function of the number of paths N for the computation of the option's price

Having observed the required CPU time in figure 3.2, three essential remarks have to be made. To start with, there seems to be a quadratic relationship

between the CPU time and the number of simulated paths. As a matter of fact, this particular feature is simply explained by what actually has to be computed in step 3.(c) of the algorithm. More precisely, for all $q \in \{1, ..., N\}$ the following quantity is determined:

$$\mathcal{E}_k[v_{k+1}](X_{k\delta t}^{(q)}) \approx \frac{\mathcal{J}_{k,k+1}^{\psi}[v_{k+1}](X_{k\delta t}^{(q)})}{\mathcal{J}_{k,k+1}^{\psi}[1](X_{k\delta t}^{(q)})}$$

where the numerator and the denominator involve the computation of empirical means with another N sample paths. This naturally results in the CPU time being "squared" with respect to the number of paths. The second remark which can be made is related to the effect of the number of time steps. The required CPU time seems to be directly proportional to the number of time steps being used, which is something to be expected given the general body of the algorithm. Indeed, as the algorithm is of backward type, if we place ourselves at a given position in time, the different steps are carried out for each path. As a result, taking 20 time steps roughly doubles the CPU time in comparison with 10 time steps. Finally, the most important remark from a practical viewpoint is undeniably related to the overall computational cost in absolute terms. As we observed earlier on in the numerical examples, a very high number of paths is usually necessary to obtain satisfactory results. However, this is done at a considerably high computational cost. For example, choosing 10000 paths along with only 10 timsesteps yields a required CPU time of just over one hour. As a result, it has to be said that the this algorithm is rather uncompetitive with other algorithms such as Longstaff-Schwartz in terms of CPU time.

Concluding considerations

Let us first summarize the approach which has been followed here in order to be able to present the Malliavin pricing and hedging algorithm. In chapter 1, we started by introducing the main Malliavin calculus tools needed in the derivation of the representation formulas for conditional expectation and its derivative. We then introduced these formulas in chapter 2, starting with non-localized versions, before moving on to their localized counterparts. In chapter 3, starting from the dynamic programming principle formulation of the problem, we specifically showed how these formulas had to be applied within the context of the pricing and hedging of Bermudan options. We then provided a sketch of the algorithm itself, before finally discussing several numerical results.

As far as numerical aspects are concerned, our own implementation underlined the fact that the algorithm cannot numerically work without the use of localizing functions, as prices and standard deviations of estimators tend to "blow up". Even though our implementation of the control variate method presented in remark 3.2.1 did not prove to yield any satisfactory results, we believe that it most certainly would have enabled us to obtain more stable and more accurate prices and deltas with fewer sample paths. Indeed, the results presented here seem to be reasonably in line with "reference" values when the number of paths is relatively high, but this mechanically increases (and "squares") the required CPU time for the computations. This naturally leads us to question of the efficiency of the algorithm in terms of computational cost, especially in comparison with other methods such as the Longstaff-Schwartz algorithm. Bally *et al.* [1] state that the latter is undenaiably more competitive with regard to the computational speed, although it does not necessarily yield more accurate results. Furthermore, one fundamental advantage of the Malliavin approach is that it remains after all a pricing *and* hedging algorithm, unlike other probabilistic based methods, which do not provide their own specific procedure to compute the delta of the option.

Taking into consideration the sequence followed in this paper, the first natural extension which comes to mind is based on the introduction of the representation formulas for conditional expectation and its derivative in a multi-dimensional setting, as the main benefits of probabilistic methods over deterministic ones arise in higher dimensions. This can be done in a fairly straightforward

way when the underlying price process is characterized by geometric Brownian motion (see Bally *et al.* [1]), but whether or not such an extension can readily be developed for other processes remains to be seen. Finally, given the high required CPU time for the Malliavin pricing and hedging algorithm, another possible extension would be to look at how this particular method could be combined with other probabilistic based approaches in order to achieve more competitive time costs. The core problem we have to deal with here is related to the presence of step 3.(c) of the algorithm, which involves a very costly computation with the use of equation (3.1), and subsequently leads to the required CPU time being "squared". Other ways of evaluating (3.1) could potentially yield a more efficient computational cost, although it currently remains unclear as to how one could go about determining such alternative techniques.

Bibliography

[1] V. Bally, L. Caramellino, and A. Zanette, *Pricing and hedging American options by Monte Carlo methods using a Malliavin calculus approach*, Institut National de Recherche en Informatique et en Automatique, Research paper no 4804, 2003.

[2] A. Bensoussan and J-L. Lions, *Impulse control and quasivariational inequalities*, Gauthiers-Villars, 1984.

[3] M. Bladt and M. Sorensen, *Simple simulation of diffusion bridges with application to likelihood inference for diffusions*, Centre for Analytical Finance, University of Aarhus, Working paper no 225, 2007.

[4] M. Broadie and P. Glasserman, *Pricing American-style securities using simulation*, Journal of Economic Dynamics and Control, pp. 1323-1352, 1997.

[5] D. Duffy, *Finite difference methods in financial engineering: A partial differential equation approach*, Wiley Series, 2006.

[6] E. Fournié, J-M. Lasry, J. Lebuchoux, P-L. Lions, N. Touzi, *Applications of Malliavin calculus to Monte Carlo methods in finance*, Finance and Stochastics, Springer-Verlag, pp. 391-412, 1999.

[7] E. Fournié, J-M. Lasry, J. Lebuchoux, P-L. Lions, *Applications of Malliavin calculus to Monte Carlo methods in finance II*, Finance and Stochastics, Springer-Verlag, pp. 201-236, 2001.

[8] P. Friz, *An introduction to Malliavin calculus*, Courant Institute of Mathematical Sciences, New York University, 2002.

[9] P. Hepperger, *Pricing high-dimensional Bermudan options using variance-reduced Monte Carlo methods*, Zentrum Mathematik, Technische Universität München.

[10] P-L. Lions and H. Regnier, *Calcul du prix et des sensibilités d'une option américaine par une méthode de Monte Carlo*, Preprint, 2001.

[11] F. Longstaff and E. Schwartz, *Valuing American options by simulations: a simple least squares approach*, The Review of Financial Studies, no 14, pp. 113-148, 2001.

[12] D. Nualart, *The Malliavin calulus and related topics*, Springer-Verlag, 1995.