

Monte Carlo estimation of a joint density using Malliavin Calculus, and application to American options

Moez Mrad	Nizar Touzi	Amina Zeghal
Société Générale Group	CREST	University Paris IX,
Fixed Income Quantitative Research	and CEREMADE	CEREMADE,
Paris, France	Paris, France	Paris, France
moez.mrad@socgen.com	touzi@ensae.fr	zeghal@ceremade.dauphine.fr

October 17, 2005

Abstract

We use the Malliavin integration by parts formula in order to provide a family of representations of the joint density (which does not involve Dirac measures) of $(X_\theta, X_{\theta+\delta})$, where X is a d -dimensional Markov diffusion ($d \geq 1$), $\theta > 0$ and $\delta > 0$. Following [5], the different representations are determined by a pair of *localizing functions*. We discuss the problem of variance reduction within the family of *separable localizing functions*: We characterize a pair of exponential functions as the unique integrated-variance minimizer among this class of separable localizing functions. We test our method on the d -dimensional Brownian motion and provide an application to the problem of American options valuation by the quantization tree method introduced by [2].

Key words: Monte Carlo, Malliavin calculus, quantization, American options.

MSC 1991 subject classifications: Primary 60H07 65C05 ; secondary 49-00.

1 Introduction

Let X be a Markov process. Given M simulated paths of X , we are interested in the Monte Carlo estimation of the joint density function of $(X_\theta, X_{\theta+\delta})$:

$$f(x, y) = \mathbb{E}(\varepsilon_x(X_\theta)\varepsilon_y(X_{\theta+\delta}))$$

where $\theta > 0$, $\delta > 0$, and ε_x is the Dirac measure at the point x . The purpose of this paper is to introduce an estimator based on the Malliavin integration by parts formula which allows to integrate the Dirac measure up to the Heaviside function.

This problem is motivated by the recent financial literature which has been interested in the Monte Carlo approach for the numerical approximation of the value of American options. The Monte Carlo technique is particularly adapted for high dimensional problems.

By passing to a discrete-time approximation, the American option valuation problem is reduced to a backward algorithm which requires to compare at each step the reward from exercising the option to the expected reward from continuing. The main difficulty from the numerical viewpoint lies in the computation of the expected reward conditional on the actual information. In order to approximate the required conditional expectations, one can use the classical tools from non-parametric regression methods in statistics. The Kernel regression method (see e.g. Bosq [3]) has been suggested in this context by Carriere [8]. The basis projection method consists in approximating the conditional expectation by the orthogonal projection on some finite truncation of an orthonormal basis of L^2 , and has been used in the context of American options by Longstaff and Schwartz [14], and Tsitsiklis and van Roy [16].

An alternative approach, based on the integration by parts formula from Malliavin calculus, has been suggested by Fournié et al. [10], and implemented for American options in the context of the multivariate Black and Scholes model. The rate of convergence of this method has been analyzed for general smooth processes by [5] and [6]. We also mention the approach of Braodie and Glassermann [7] which use a random grid at each time step associated with some conveniently chosen weights defined by some statistical ideas based on the importance sampling theorem.

In this paper, we focus on the quantization technique of Bally, Pagès and Printemps [2]. Loosely speaking, this method consists in "quantizing" the state variable of the problem at each time step, i.e. projecting X_{t_k} on a grid containing N_k discrete points \widehat{X}_k^i , $i = 1, \dots, N_k$. The state space is then partitioned into a family of cells $C(\widehat{X}_k^i)$ with $\widehat{X}_k^i \in C(\widehat{X}_k^i)$. Given m independent copies $(X_{t_k}^{(m)}, X_{t_{k+1}}^{(m)})$ of the pair of random variables $(X_{t_k}, X_{t_{k+1}})$, the

transition probability from the cell $C(\widehat{X}_k^i)$ to the cell $C(\widehat{X}_{k+1}^j)$ is estimated by :

$$\widehat{\pi}_k^{i,j} = \frac{\frac{1}{M} \sum_{m=1}^M \mathbf{1}_{X_{t_k}^{(m)} \in C(\widehat{X}_k^i)} \mathbf{1}_{X_{t_{k+1}}^{(m)} \in C(\widehat{X}_{k+1}^j)}}{\frac{1}{M} \sum_{m=1}^M \mathbf{1}_{X_{t_k}^{(m)} \in C(\widehat{X}_k^i)}}.$$

Observe that as the size of these cells gets smaller (which is the case in practice), the number of scenarios falling in each cell decreases dramatically, inducing a large variance of the above estimator.

The method of estimation of transition probabilities exposed above, is in fact an extension of the classical approach used to estimate density function of a random variable using M independent copies. In order to improve this method, one can think about Malliavin integration by parts formula that has been used recently used in [11], [10], [13], [12] and [5] in order to obtain an alternative representation of the conditional expectation :

$$r(x) = \mathbb{E}[g(X_{\theta+\delta}) | X_\theta = x] = \frac{\mathbb{E}[g(X_{\theta+\delta}) \varepsilon_x(X_\theta)]}{\mathbb{E}[\varepsilon_x(X_\theta)]}$$

where g is a mapping from \mathbb{R}^d into \mathbb{R} . The representation in [5] does not involve the Dirac measure ε_x (it uses instead H_x : the Heaviside measure at the point x) and depends only on the mapping g . It also allows to recover the \sqrt{M} rate of convergence.

In this paper, we extend the results of [5] by providing a family of alternative representations of the joint density which does not involve Dirac measures. Each representation is determined by the choice of a pair of localization functions. We then study the problem of minimizing the integrated mean square error within the class of all pairs of separable localization functions. Our main result provides an explicit characterization of such optimal localization functions.

The paper is organized as follows. Section 2. introduces the main notations and some preliminary results. Section 3 contains the proof of the family of alternative representations of the joint density. The variance reduction issues are discussed in Section 4. Section 5 specializes the discussion to the multivariate Brownian motion, and provides some numerical tests proving the performance of the Malliavin-based estimator. We also proceed to a comparison with the Kernel density estimator. Finally section 6 provides and application of this technique to the quantization tree method.

2 Notations

Throughout this paper we shall denote by \mathcal{J} the subset of \mathbb{N}^k whose elements $I = (i_1, \dots, i_k)$ satisfy $1 \leq i_1 < \dots < i_k \leq d$. The cardinality of I will be denoted by $|I|$. Let $I = (i_1, \dots, i_m)$ and $J = (j_1, \dots, j_n)$ be two arbitrary elements in \mathcal{J} . Then $\{i_1, \dots, i_m\} \cup \{j_1, \dots, j_n\} =$

$\{k_1, \dots, k_p\}$ for some $\max\{n, m\} \leq p \leq \min\{d, m+n\}$, and $1 \leq k_1 < \dots < k_p \leq d$. We then denote $I \vee J := (k_1, \dots, k_p) \in \mathcal{J}$. For all positive integer $i \leq d$, we denote by $\mathcal{J}^{(-i)}$ the subset of \mathcal{J} whose elements I do not contain i , i.e. $\mathcal{J}^{(-i)} := \{I \in \mathcal{J} : i \notin I\}$.

For a vector $x \in \mathbb{R}^d$, we shall denote x^i its i -th component, and $x^{(-i)}$ the vector of \mathbb{R}^{d-1} obtained by omitting the i -th component, i.e. $x^{(-i)} := (x^1, \dots, x^{i-1}, x^{i+1}, \dots, x^d)$. For $x, y \in \mathbb{R}^d$, we denote by $x \cdot y := \sum_{i=1}^d x^i y^i$ the Euclidean scalar product.

Let φ be a map from \mathbb{R}^d into \mathbb{R} . Following [5], we say that φ is a *smooth localizing function* if

$$\varphi(0) = 1 \quad \text{and} \quad \partial_I \varphi \text{ exists and lies in } C_b^0 \text{ for all } I \in \mathcal{J}.$$

Here, C_b^0 is the set of all bounded continuous functions from \mathbb{R}^d to \mathbb{R} , and for $I = (i_1, \dots, i_{|I|})$, the notation $\partial_I \varphi$ stands for the I -th cross derivative, i.e. $\partial_I \varphi = \partial^{|I|} \varphi / \partial x_{i_1} \cdots \partial x_{i_{|I|}}$. For $I = \emptyset$ we set $\partial_\emptyset \varphi = \varphi$. We denote by \mathcal{L} the set of all smooth localization functions.

For $I = (1, \dots, d)$ and $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, we denote by $\nabla \varphi := \partial_I \varphi$ the gradient of φ , if it does exist.

3 Alternative representation of joint density function

3.1 Malliavin derivatives and Skorohod integrals

Let (Ω, \mathcal{F}, P) be a complete probability space equipped with a d -dimensional standard Brownian motion $W = (W^1, \dots, W^d)$. We shall restrict the time interval to $\mathbf{T} := [0, \theta + \delta]$ for some $\theta, \delta > 0$. We denote by $\mathbb{F} := \{\mathcal{F}_t, t \in \mathbf{T}\}$ the P -completion of the filtration generated by W . Throughout this paper, we consider a Markov process X such that X_θ and $X_{\theta+\delta}$ belong to the Sobolev spaces $\mathbb{D}^{k,p}$ ($p, k \geq 1$) of k -times Malliavin differentiable random variables satisfying :

$$\|X\|_{\mathbb{D}^{k,p}} := \left[E(|X|^p) + \sum_{j=1}^k E\left(\|D^j X\|_{L^p(\mathbf{T}^j)}^p\right) \right]^{1/p} < \infty$$

where

$$\|D^j X\|_{L^p(\mathbf{T}^j)} = \left(\int_{\mathbf{T}^j} |D_{t_1} \cdots D_{t_j} X|^p dt_j \cdots dt_1 \right)^{1/p}.$$

Given a matrix-valued process h , with columns denoted by h^i , and a random variable F , we denote

$$S_i^h(F) := \int_{\mathbf{T}} F h_t^i \cdot dW_t \quad \text{for } i = 1, \dots, d, \text{ and } S_I^h(F) := S_{i_1}^h \circ \dots \circ S_{i_k}^h(F)$$

for $I = (i_1, \dots, i_k) \in \mathcal{J}$, whenever these stochastic integrals exist in the Skorohod sense. We extend this definition to the empty set by setting $S_\emptyset^h(F) := F$. For the maximal element $I = (1, \dots, d)$, we shall simply write

$$S^h(F) := S_{(1, \dots, d)}^h(F).$$

Similarly, for $I \in \mathcal{J}$, we set :

$$S_{-I}^h(F) := S_I^h(F) \quad \text{where} \quad \bar{I} \vee I = (1, \dots, d) \quad \text{and} \quad I \cap \bar{I} = \emptyset.$$

We next introduce the set $\mathbf{H}(X)$ as the collection of all pairs of matrix-valued $L^2(\mathcal{F})$ processes (g, h) satisfying

$$\int_{\mathbf{T}} D_t X_\theta g_t dt = 0, \quad \int_{\mathbf{T}} D_t X_{\theta+\delta} g_t dt = I_d, \quad (3.1)$$

$$\int_{\mathbf{T}} D_t X_\theta h_t dt = I_d, \quad \int_{\mathbf{T}} D_t X_{\theta+\delta} h_t dt = 0 \quad (3.2)$$

and such that

$$S_J^g \circ S_I^h(\varphi(X_\theta)\psi(X_{\theta+\delta})) \text{ is well-defined in } \mathbb{D}^{1,2} \quad (3.3)$$

for all $I, J \in \mathcal{J}$ and $\varphi, \psi \in \mathcal{L}$.

Throughout this paper, we shall assume that

$$\mathbf{H}(X) \neq \emptyset.$$

Remark 3.1 For later use, we observe that for all $(g, h) \in \mathbf{H}(X)$, we have

$$S_I^h(\varphi(X_\theta)\psi(X_{\theta+\delta})) = \psi(X_{\theta+\delta})S_I^h(\varphi(X_\theta)),$$

for all $I \in \mathcal{J}$ and $\varphi, \psi \in \mathcal{L}$. To see this, observe that it is sufficient to prove these equalities for $|I| = 1$. For a positive integer $i \leq d$, it follows from the connection between the Itô and the Skorohod integral (see [15] p40) that :

$$\begin{aligned} S_i^h(\varphi(X_\theta)\psi(X_{\theta+\delta})) &= \psi(X_{\theta+\delta})S_i^h(\varphi(X_\theta)) - \int_{\mathbf{T}} D_t \psi(X_{\theta+\delta}) h_t^i dt \\ &= \psi(X_{\theta+\delta})S_i^h(\varphi(X_\theta)) - \nabla \psi(X_{\theta+\delta}) \int_{\mathbf{T}} D_t X_{\theta+\delta} h_t^i dt \\ &= \psi(X_{\theta+\delta})S_i^h(\varphi(X_\theta)) \end{aligned}$$

by (3.2).

3.2 The case of Markov diffusions

Before going any further, let us verify that the conditions listed in the above paragraph are satisfied in the case of *regular* Markov Diffusions. Let $X = \{X_t\}$ be the continuous-time process defined by the stochastic differential equation

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

together with an initial condition X_0 . Here $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ are respectively vector and matrix valued maps, the matrix $\sigma(x)$ is invertible for all $x \in \mathbb{R}^d$, b , σ and σ^{-1} are C_b^∞ . Then, for all $p, k \geq 1$, the process X lies in the set $\mathbb{L}_{\mathbf{T}}^{k,p}$, i.e. $X_t \in \mathbb{D}^{k,p}$ for all $t \in \mathbf{T}$ and :

$$\|X\|_{\mathbb{L}_{\mathbf{T}}^{k,p}} := \left[E\left(\int_{\mathbf{T}} |X_t|^p dt\right) + \sum_{j=1}^k E\left(\int_{\mathbf{T}} \|D^j X_t\|_{L^p(T^j)}^p dt\right) \right]^{\frac{1}{p}} < \infty.$$

We denote by $\mathbb{L}_{\mathbf{T}}^\infty := \bigcap_{p \geq 1} \bigcap_{k \geq 1} \mathbb{L}_{\mathbf{T}}^{k,p}$. We similarly define \mathbb{D}^∞ . Notice that $f(X) \in \mathbb{L}_{\mathbf{T}}^\infty$ whenever $f \in C_b^\infty$. In particular, $\sigma^{-1}(X) \in \mathbb{L}_{\mathbf{T}}^\infty$ (see [15] Proposition 1.5.1).

The first variation process of X is the matrix-valued process defined by :

$$Y_0 = I_d \quad \text{and} \quad dY_t = \nabla b(X_t)Y_t dt + \sum_{i=1}^d \nabla \sigma^i(X_t)Y_t dW_t^i, \quad (3.4)$$

where σ^i is the i -th column vector of σ . By [15] Lemma 2.2.2, the processes Y and Y^{-1} also belong to $\mathbb{L}_{\mathbf{T}}^\infty$.

The Malliavin derivative is related to the first variation process by :

$$D_s X_t = Y_t Y_s^{-1} \sigma(X_s) \mathbf{1}_{\{s \leq t\}}; \quad s \geq 0, \quad (3.5)$$

so that :

$$D_s X_\theta = Y_\theta Y_{\theta+\delta}^{-1} D_s X_{\theta+\delta} \mathbf{1}_{\{s \leq \theta\}}; \quad s \geq 0. \quad (3.6)$$

It follows that $\mathbf{H}(X)$ is not empty. Indeed, since X , Y , Y^{-1} and $\sigma^{-1}(X)$ are in $\mathbb{L}_{\mathbf{T}}^\infty$,

$$\hat{g}_t := (D_t X_{\theta+\delta})^{-1} \delta^{-1} \mathbf{1}_{t \in [\theta, \theta+\delta]} \quad (3.7)$$

$$\hat{h}_t := (D_t X_{\theta+\delta})^{-1} Y_{\theta+\delta} Y_\theta^{-1} (\theta^{-1} \mathbf{1}_{t \in [0, \theta]} - \delta^{-1} \mathbf{1}_{t \in [\theta, \theta+\delta]}) \quad (3.8)$$

defines a pair of processes in $\mathbb{L}_{\mathbf{T}}^\infty$ satisfying (3.1)-(3.2). Moreover, for each real-valued $F \in \mathbb{D}^\infty$, and $i = 1, \dots, d$, $S_i^{\hat{h}}(F)$ is well defined and belongs to \mathbb{D}^∞ (see [15] Property 2 p38 and Proposition 3.2.1 p158). By simple iteration of this argument, we also see that (\hat{g}, \hat{h}) satisfies (3.3).

3.3 Alternative representation of the joint density distribution

In this paragraph, we provide an extension of Theorem 3.1. of [5]. This result induces an expectation representation of the joint density of $(X_\theta, X_{\theta+\delta})$ which does not involve Dirac measures.

Theorem 3.1 *Let $\{A_i, i \leq d\}$, $\{B_i, i \leq d\}$ be two families of Borel subsets of \mathbb{R} , and set $A := A_1 \times \cdots \times A_d$, $B := B_1 \times \cdots \times B_d$. Then for all pair of localizing functions $\varphi, \psi \in \mathcal{L}$, we have:*

$$\begin{aligned} & E[\mathbf{1}_A(X_\theta)\mathbf{1}_B(X_{\theta+\delta})] \\ &= \iint_{A \times B} E \left[H_x(X_\theta) H_y(X_{\theta+\delta}) S^g \circ S^h (\varphi(X_\theta - x)\psi(X_{\theta+\delta} - y)) \right] dx dy \end{aligned}$$

where, for $\xi = (\xi_1, \dots, \xi_d)$ and $\zeta = (\zeta_1, \dots, \zeta_d) \in \mathbb{R}^d$, $H_\xi(\zeta) := \prod_{i=1}^d \mathbf{1}_{\{\zeta_i > \xi_i\}}$ is the so-called Heaviside function.

Proof. We shall prove the required representation result by using repeatedly a lemma to be derived in the first part of this proof. In comparison to the proof of Theorem 3.1 of [5], the main new ingredient is located in Step 3 below. Let us first introduce the following additional notations:

$$\pi_i(x) := (0, \dots, 0, x_{i+1}, \dots, x_d) \text{ for } i = 1, \dots, d-1 \quad \text{and} \quad \pi_d(x) := 0$$

for $x = (x_1, \dots, x_d) \in \mathbb{R}^d$,

$$\varphi_i := \varphi \circ \pi_i \quad \text{for all } \varphi \in \mathcal{L} \text{ and } i \leq d,$$

and

$$I_i := (i+1, \dots, d) \in \mathcal{J} \text{ for } i = 1, \dots, d-1, \quad \text{and} \quad I_d := \emptyset.$$

1. By a classical density argument, it is sufficient to prove the result for $A_i = [a_i, b_i]$ with $a_i < b_i$ and $B_i = [c_i, d_i]$ with $c_i < d_i$.
2. By the same argument as in [5], we have

$$\begin{aligned} & E \left[\mathbf{1}_{A_i}(X_\theta^i) \phi \left((X_\theta^{(-i)}) f(X_{\theta+\delta}) S_{I_i}^h (\varphi_i(X_\theta - x)F) \right) \right] \\ &= \int_{A_i} E \left[H_{x^i}(X_\theta^i) \phi \left((X_\theta^{(-i)}) f(X_{\theta+\delta}) S_{I_{i-1}}^h (\varphi_{i-1}(X_\theta - x)F) \right) \right] dx^i. \end{aligned} \tag{3.9}$$

Here, $i \leq d$, $f : \mathbb{R}^d \rightarrow \mathbb{R}$, $\phi : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ are two bounded maps, F is a real-valued random variable in $\mathbb{D}^{1,2}$, $(g, h) \in \mathbf{H}(X)$, $\varphi \in \mathcal{L}$, and A_i is a Borel subset of \mathbb{R} .

3. We now prove that

$$\begin{aligned} \alpha &:= E \left[f(X_\theta) \mathbf{1}_{B_i}(X_{\theta+\delta}^i) \phi \left(X_{\theta+\delta}^{(-i)} \right) S_{I_i}^g(\psi_i(X_{\theta+\delta} - y)F) \right] \\ &= \int_{B_i} E \left[f(X_\theta) H_{y^i}(X_{\theta+\delta}^i) \phi \left(X_{\theta+\delta}^{(-i)} \right) S_{I_{i-1}}^g(\psi_{i-1}(X_{\theta+\delta} - y)F) \right] dy^i, \end{aligned} \quad (3.10)$$

where $i, f, \phi, F, (g, h)$ are as above, and $\psi \in \mathcal{L}, B_i \in \mathcal{B}(\mathbb{R})$. To see this, we first observe that it is sufficient to consider the case where f and ϕ are smooth. The general case is then deduced by a classical density argument. Next, since $S_{I_i}^g(\psi_i(X_{\theta+\delta} - y)F) \in \mathbb{D}^{1,2}$, it follows from the smoothness of f and ϕ that the random variable

$$V := \int_{-\infty}^{X_{\theta+\delta}^i} f(X_\theta) \mathbf{1}_{B_i}(y^i) \phi \left(X_{\theta+\delta}^{(-i)} \right) S_{I_i}^g(\psi_{i-1}(X_{\theta+\delta} - y)F) dy^i$$

is Malliavin differentiable. By the usual rules of differentiation, we directly compute that

$$\begin{aligned} \int_{\mathbf{T}} D_t V \cdot g_t^i dt &= f(X_\theta) \mathbf{1}_{B_i}(X_{\theta+\delta}^i) \phi \left(X_{\theta+\delta}^{(-i)} \right) S_{I_i}^g(\psi_i(X_{\theta+\delta} - y)F) \int_{\mathbf{T}} D_t X_{\theta+\delta}^i \cdot g_t^i dt \\ &\quad + \int_{-\infty}^{X_{\theta+\delta}^i} \mathbf{1}_{B_i}(y^i) \int_{\mathbf{T}} D_t \left\{ f(X_\theta) \phi \left(X_{\theta+\delta}^{(-i)} \right) S_{I_i}^g(\psi_{i-1}(X_{\theta+\delta} - y)F) \right\} \cdot g_t^i dt dy^i \\ &= f(X_\theta) \mathbf{1}_{B_i}(X_{\theta+\delta}^i) \phi \left(X_{\theta+\delta}^{(-i)} \right) S_{I_i}^g(\psi_i(X_{\theta+\delta} - y)F) \\ &\quad + \int_{-\infty}^{X_{\theta+\delta}^i} \mathbf{1}_{B_i}(y^i) f(X_\theta) \phi \left(X_{\theta+\delta}^{(-i)} \right) \int_{\mathbf{T}} D_t \left\{ S_{I_i}^g(\psi_{i-1}(X_{\theta+\delta} - y)F) \right\} \cdot g_t^i dt dy^i, \end{aligned}$$

by (3.1). Recall the definition of the parameter α in (3.10), and set $S := S_{I_i}^g(\psi_{i-1}(X_{\theta+\delta} - y)F)$.

By the Malliavin integration by parts formula, this provides :

$$\begin{aligned} \alpha &= E \left[V \int_{\mathbf{T}} g_t^i \cdot dW_t \right] \\ &\quad - E \left[\int_{-\infty}^{X_{\theta+\delta}^i} \mathbf{1}_{B_i}(y^i) f(X_\theta) \phi \left(X_{\theta+\delta}^{(-i)} \right) \int_{\mathbf{T}} D_t S \cdot g_t^i dt dy^i \right] \\ &= \int_{B_i} E \left[H_{y^i}(X_{\theta+\delta}^i) f(X_\theta) \phi \left(X_{\theta+\delta}^{(-i)} \right) \left(S \int_{\mathbf{T}} g_t^i \cdot dW_t - \int_{\mathbf{T}} D_t S \cdot g_t^i dt \right) \right] dy^i \\ &= \int_{B_i} E \left[H_{y^i}(X_{\theta+\delta}^i) f(X_\theta) \phi \left(X_{\theta+\delta}^{(-i)} \right) \int_{\mathbf{T}} S g_t^i \cdot dW_t \right] dy^i, \end{aligned}$$

by the connection between the Itô and the Skorohod integrals, see [15] p40. We then deduce (3.10) by observing that $\int_{\mathbf{T}} S g_t^i \cdot dW_t = S_{I_{i-1}}^g(\psi_{i-1}(X_{\theta+\delta} - y)F)$.

4. Now, we are able to prove Theorem 3.1. We start by taking $F = 1$ and setting

$\phi(z^{(-d)}) := \prod_{i=1}^{d-1} \mathbf{1}_{A_i}(z^i)$. Since $\varphi(0) = 1$, we see that :

$$\begin{aligned} &E[\mathbf{1}_A(X_\theta) \mathbf{1}_B(X_{\theta+\delta})] \\ &= E \left[\mathbf{1}_A(X_\theta) \mathbf{1}_B(X_{\theta+\delta}) S_{I_d}^h(\varphi_d(X_\theta - x)) \right] \\ &= \int_{A_d} E \left[H_{x^d}(X_\theta^d) \phi \left(X_\theta^{(-d)} \right) \mathbf{1}_B(X_{\theta+\delta}) S_{I_{d-1}}^h(\varphi_{d-1}(X_\theta - x)) \right] dx^d \end{aligned}$$

by (3.9). Next, we set $\phi(z^{-(d-1)}) := H_{x^d}(z^d) \prod_{i=1}^{d-2} \mathbf{1}_{A_i}(z^i)$, we observe that :

$$\begin{aligned} & E[\mathbf{1}_A(X_\theta) \mathbf{1}_B(X_{\theta+\delta})] \\ &= \int_{A_d} \int_{A_{(d-1)}} E[H_{x^{d-1}}(X_\theta^{d-1}) \phi(X_\theta^{-(d-1)}) \mathbf{1}_B(X_{\theta+\delta}) S_{I_{d-2}}^h(\varphi_{d-2}(X_\theta - x))] dx^{d-1} dx^d \end{aligned}$$

Iterating this argument, we obtain that :

$$E[\mathbf{1}_A(X_\theta) \mathbf{1}_B(X_{\theta+\delta})] = \int_A \beta(x) dx ; \quad (3.11)$$

where

$$\beta(x) := E\left[H_x(X_\theta) \mathbf{1}_B(X_{\theta+\delta}) S^h(\varphi(X_\theta - x))\right].$$

We next apply (3.10) with $F = S^h(\varphi(X_\theta - x))$, $\phi(z^{(-d)}) := \prod_{i=1}^{d-1} \mathbf{1}_{B_i}(z^i)$. Since $\psi_d(x) = \psi(0) = 1$, we get

$$\begin{aligned} \beta(x) &= E\left[H_x(X_\theta) \mathbf{1}_B(X_{\theta+\delta}) S_{I_d}^g(\psi_d(X_{\theta+\delta} - y)F)\right] \\ &= \int_{B_d} E\left[H_x(X_\theta) H_{y^d}(X_{\theta+\delta}^d) \phi(X_{\theta+\delta}^{(-d)}) S_{I_{d-1}}^g(\psi_{d-1}(X_{\theta+\delta} - y)F)\right] dy. \end{aligned}$$

Iterating this argument, we obtain that :

$$\beta(x) = \int_{B_d} E\left[H_x(X_\theta) H_y(X_{\theta+\delta}^d) \phi(X_{\theta+\delta}^{(-d)}) S^g(\psi(X_{\theta+\delta} - y)F)\right] dy. \quad (3.12)$$

Now, the statement of the theorem follows by combining (3.11), (3.12) and by using Remark 3.1. \square

Remark 3.2 For later use, we observe that the above proof shows that

$$\begin{aligned} & E\left[\mathbf{1}_{A^{(-i)}}(X_\theta^{(-i)}) H_{x^i}(X_\theta^i) \mathbf{1}_B(X_{\theta+\delta})\right] \\ &= \int_{A^{(-i)} \times B} E\left[H_x(X_\theta) H_y(X_{\theta+\delta}) S^g \circ S_{(-i)}^h(\varphi(X_\theta - x)\psi(X_{\theta+\delta} - y))\right] dx^{(-i)} dy \end{aligned}$$

and

$$\begin{aligned} & E\left[\mathbf{1}_A(X_\theta) \mathbf{1}_{B^{(-j)}}(X_{\theta+\delta}^{(-j)}) H_{x^j}(X_{\theta+\delta}^j)\right] \\ &= \int_{A \times B^{(-j)}} E\left[H_x(X_\theta) H_y(X_{\theta+\delta}) S_{(-j)}^g \circ S^h(\varphi(X_\theta - x)\psi(X_{\theta+\delta} - y))\right] dx dy^{(-j)} \end{aligned}$$

where $A^{(-i)} := A_1 \times \dots \times A_{i-1} \times A_{i+1} \times \dots \times A_d$ and $B^{(-j)} := B_1 \times \dots \times B_{j-1} \times B_{j+1} \times \dots \times B_d$.

4 Variance reduction using localizing separable functions

Given a pair of localizing functions $(\varphi, \psi) \in \mathcal{L} \times \mathcal{L}$ and $(g, h) \in \mathbf{H}(X)$, Theorem 3.1 provides the following representation of the joint-density function:

$$f(x, y) = \mathbb{E} \left[H_x(X_\theta) H_y(X_{\theta+\delta}) S^g \circ S^h (\varphi(X_\theta - x) \psi(X_{\theta+\delta} - y)) \right].$$

This suggests the following Monte Carlo estimator of the joint density function :

$$\widehat{f}_n(x, y) := \frac{1}{n} \sum_{k=1}^n H_x \left(X_\theta^{(k)} \right) H_y \left(X_{\theta+\delta}^{(k)} \right) S^{g^{(k)}} \circ S^{h^{(k)}} \left(\varphi(X_\theta^{(k)} - x) \psi(X_{\theta+\delta}^{(k)} - y) \right),$$

where $(X_\theta^{(k)}, X_{\theta+\delta}^{(k)}, g^{(k)}, h^{(k)})$ are independent copies with the same distribution as $(X_\theta, X_{\theta+\delta}, g, h)$. Following [5], we consider the problem of minimizing the *integrated mean square error* (mse) :

$$I^{g,h}(\varphi, \psi) := \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E} \left[H_x(X_\theta) H_y(X_{\theta+\delta}) S^g \circ S^h (\varphi(X_\theta - x) \psi(X_{\theta+\delta} - y))^2 \right] dx dy$$

within a suitable class of localizing functions. We first observe that, for any pair (φ, ψ) of localizing functions, only the restrictions of φ and ψ to \mathbb{R}_+^d are involved in $I^{g,h}(\varphi, \psi)$. We then consider the set \mathcal{L}_+ of functions of the form $\varphi|_{\mathbb{R}_+^d}$.

The problem of minimizing the integrated mse has been introduced in [12] and completely solved in [5] in the case of *separable* localizing functions, i.e.

$$\varphi(x) = \prod_{i=1}^d \varphi_i(x^i),$$

where, for each $i = 1, \dots, d$, the function φ_i maps \mathbb{R}_+ into \mathbb{R} . We denote by \mathcal{L}_+^s the set of all separable localizing functions. We will then concentrate on the minimization problem

$$v^{g,h} := \inf_{(\varphi, \psi) \in \mathcal{L}_+^s \times \mathcal{L}_+^s} I^{g,h}(\varphi, \psi). \quad (4.1)$$

In order to ensure that $v^{g,h}$ is finite, we assume that:

$$\sum_{I, J \in \mathcal{J}} \mathbb{E} \left[S_{J,I}^{g,h}(1)^2 \right] < \infty. \quad (4.2)$$

The main result of this section is the following.

Theorem 4.1 *There exists a unique solution $(\widehat{\varphi}, \widehat{\psi}) \in \mathcal{L}_+^s \times \mathcal{L}_+^s$ to the integrated minimization problem (4.1) given by:*

$$\widehat{\varphi}(x) := e^{-\widehat{\eta} \cdot x} \text{ and } \widehat{\psi}(x) := e^{-\widehat{\rho} \cdot x} \text{ for } x \in \mathbb{R}_+^d,$$

where $(\hat{\eta}, \hat{\rho})$ is the unique solution in $(0, \infty)^{2d}$ of the system of polynomial equations:

$$\hat{\eta}_m^2 := \frac{\mathbb{E} \left[\left(\sum_{I \in \mathcal{J}^{(-m)}} \sum_{J \in \mathcal{J}} S_{-J}^g \circ S_{-I}^h(1) \prod_{i \in I} \hat{\eta}^i \prod_{j \in J} \hat{\rho}^j \right)^2 \right]}{\mathbb{E} \left[\left(\sum_{I \in \mathcal{J}^{(-m)}} \sum_{J \in \mathcal{J}} S_{-J}^g \circ S_{-(I \vee m)}^h(1) \prod_{i \in I} \hat{\eta}^i \prod_{j \in J} \hat{\rho}^j \right)^2 \right]} \quad (4.3)$$

$$\hat{\rho}_m^2 := \frac{\mathbb{E} \left[\left(\sum_{I \in \mathcal{J}} \sum_{J \in \mathcal{J}^{(-m)}} S_{-J}^g \circ S_{-I}^h(1) \prod_{i \in I} \hat{\eta}^i \prod_{j \in J} \hat{\rho}^j \right)^2 \right]}{\mathbb{E} \left[\left(\sum_{I \in \mathcal{J}} \sum_{J \in \mathcal{J}^{(-m)}} S_{-(J \vee m)}^g \circ S_{-I}^h(1) \prod_{i \in I} \hat{\eta}^i \prod_{j \in J} \hat{\rho}^j \right)^2 \right]}, \quad (4.4)$$

for all $m = 1, \dots, d$.

The proof of this result, is based on the classical calculus of variations technique, and follows essentially the line of arguments of the corresponding result in [5]. The details of this proof are reported in the Appendix section.

5 Application to the d -dimensional Brownian motion

In this section, we specialize the discussion to the d -dimensional Brownian motion. We first exploit the particular structure of the Brownian motion in order to push further the characterization of the optimal separable localization function. This leads to an explicit solution of the optimal separable localization problem, see (5.8) below. Then, we examine the numerical performance of our estimation method of the joint density of $(W_\theta, W_{\theta+\delta})$, and we provide a comparison with the corresponding kernel estimator.

5.1 Optimal separable localization for the Brownian motion

1. Preparation.

The Malliavin derivative of the d -dimensional Brownian motion is given by

$$D_s W_t = I_d \mathbf{1}_{\{s \leq t\}},$$

for any $t, s \geq 0$. It is then easily checked that the pair (\hat{g}, \hat{h}) defined by

$$\hat{g}_t := \delta^{-1} I_d \mathbf{1}_{[\theta, \theta+\delta]}(t) \quad \text{and} \quad \hat{h}_t := \theta^{-1} I_d \mathbf{1}_{[0, \theta]}(t) - \delta^{-1} I_d \mathbf{1}_{[\theta, \theta+\delta]}(t)$$

is in $\mathbf{H}(W)$. Given two separable localizing functions $\varphi, \psi \in \mathcal{L}^s$ and $I, J \in \mathcal{J}$, we easily calculate that :

$$\begin{aligned} S_I^{\hat{h}}(\varphi(W_\theta - x)) &= \left(\prod_{k \notin I} \varphi_k(W_\theta^k - x^k) \right) \prod_{i \in I} S_i^{\hat{h}}(\varphi_i(W_\theta^i - x^i)) \\ S_J^{\hat{g}}(\psi(W_{\theta+\delta} - y)) &= \left(\prod_{k \notin J} \psi_k(W_{\theta+\delta}^k - y^k) \right) \prod_{j \in J} S_j^{\hat{g}}(\psi_j(W_{\theta+\delta}^j - y^j)), \end{aligned}$$

where

$$S_i^{\hat{h}}(\varphi_i(W_\theta^i - x^i)) = \varphi_i(W_\theta^i - x^i)S_i^{\hat{h}}(1) - \varphi_i'(W_\theta^i - x^i), \quad (5.1)$$

$$S_j^{\hat{g}}(\psi_j(W_{\theta+\delta}^j - y^j)) = \psi_j(W_{\theta+\delta}^j - y^j)S_j^{\hat{g}}(1) - \psi_j'(W_{\theta+\delta}^j - y^j), \quad (5.2)$$

$$S_i^{\hat{h}}(1) = \frac{W_\theta^i}{\theta} - \frac{W_{\theta+\delta}^i - W_\theta^i}{\delta}, \quad S_j^{\hat{g}}(1) = \frac{W_{\theta+\delta}^j - W_\theta^j}{\delta}.$$

Observe that the random variables $S_i^{\hat{h}}(\varphi_i(W_\theta^i - x^i))$ and $S_j^{\hat{g}}(\psi_j(W_{\theta+\delta}^j - y^j))$ depend only on the one-dimensional Brownian motion W^i . We then directly calculate that

$$S^{\hat{g}} \circ S_I^{\hat{h}}(\varphi(W_\theta - x)\psi(W_{\theta+\delta} - y)) \quad (5.3)$$

$$= \prod_{j \notin I} S_j^{\hat{g}}(\psi_j(W_{\theta+\delta}^j - y^j)) \prod_{i \in I} S_i^{\hat{g}} \circ S_i^{\hat{h}}(\varphi_i(W_\theta^i - x^i)\psi_i(W_{\theta+\delta}^i - y^i)), \quad (5.4)$$

and

$$S_J^{\hat{g}} \circ S^{\hat{h}}(\varphi(W_\theta - x)\psi(W_{\theta+\delta} - y)) \quad (5.5)$$

$$= \prod_{i \notin J} S_i^{\hat{h}}(\varphi_i(W_\theta^i - x^i)) \prod_{j \in J} S_j^{\hat{g}} \circ S_j^{\hat{h}}(\varphi_j(W_\theta^j - x^j)\psi_j(W_{\theta+\delta}^j - y^j)). \quad (5.6)$$

We finally need the following direct computation :

$$\begin{aligned} A &= S_k^{\hat{g}} \circ S_k^{\hat{h}}(\varphi_k(W_\theta^k - x^k)\psi_k(W_{\theta+\delta}^k - y^k)) \\ &= S_k^{\hat{g}}(\psi_k(W_{\theta+\delta}^k - y^k)S_k^{\hat{h}}(\varphi_k(W_\theta^k - x^k))) - \int_0^{\theta+\delta} \psi_k'(W_{\theta+\delta}^k - y^k)\hat{h}_t^k dt \\ &= S_k^{\hat{g}}(\psi_k(W_{\theta+\delta}^k - y^k)S_k^{\hat{h}}(\varphi_k(W_\theta^k - x^k))) \\ &= S_k^{\hat{h}}(\varphi_k(W_\theta^k - x^k)) S_k^{\hat{g}}(\psi_k(W_{\theta+\delta}^k - y^k)) \\ &\quad - \int_0^{\theta+\delta} \psi_k(W_{\theta+\delta}^k - y^k)D_t S_k^{\hat{h}}(\varphi_k(W_\theta^k - x^k))\hat{g}_t^k dt \end{aligned}$$

which provides that

$$A = S_k^{\hat{h}}(\varphi_k(W_\theta^k - x^k)) S_k^{\hat{g}}(\psi_k(W_{\theta+\delta}^k - y^k)) + \frac{1}{\delta}\varphi_k(W_\theta^k - x^k)\psi_k(W_{\theta+\delta}^k - y^k) \quad (5.7)$$

2. Optimal separable localization function.

Now, we rewrite the characterization (4.3)-(4.4) of Theorem 4.1 in the context of the d -dimensional Brownian motion. In order to simplify the computations, we use the equivalent system (7.2)-(7.4). By (5.3), it follows that

$$(\hat{\eta}^i)^2 = \frac{E \left[\Gamma_i \int H_{y^i}(W_{\theta+\delta}^i) S_i^{\hat{g}} \circ S_i^{\hat{h}}(\psi_i(W_{\theta+\delta}^i - y^i))^2 dy^i \right]}{E \left[\Gamma_i \int H_{y^i}(W_{\theta+\delta}^i) S_i^{\hat{g}}(\psi_i(W_{\theta+\delta}^i - y^i))^2 dy^i \right]},$$

where

$$\Gamma_i := \prod_{k \neq i} \int \int H_{x^k} \left(W_{\theta}^k \right) H_{y^k} \left(W_{\theta+\delta}^k \right) S_k^{\hat{g}} \circ S_k^{\hat{h}} \left(\varphi_k(W_{\theta}^k - x^k) \psi_k(W_{\theta+\delta}^k - y^k) \right)^2 dx^k dy^k.$$

Since the random variables $S_i^{\hat{g}} \circ S_i^{\hat{h}} (\psi_i(W_{\theta+\delta}^i - y^i))$ and $S_i^{\hat{g}} (\psi_i(W_{\theta+\delta}^i - y^i))$ are both independent of Γ_i , the above expression reduces to

$$(\hat{\eta}^i)^2 = \frac{E \left[\int H_{y^i} (W_{\theta+\delta}^i) S_i^{\hat{g}} \circ S_i^{\hat{h}} (\psi_i(W_{\theta+\delta}^i - y^i))^2 dy^i \right]}{E \left[\int H_{y^i} (W_{\theta+\delta}^i) S_i^{\hat{g}} (\psi_i(W_{\theta+\delta}^i - y^i))^2 dy^i \right]}.$$

By (5.7) and (5.1)-(5.2), this provides :

$$\begin{aligned} (\hat{\eta}^i)^2 &= \frac{E \left[\int H_{y^i} (W_{\theta+\delta}^i) \left(S_i^{\hat{h}}(1) S_i^{\hat{g}} (\psi_i(W_{\theta+\delta}^i - y^i)) + \delta^{-1} \psi_i(W_{\theta+\delta}^i - y^i) \right)^2 dy^i \right]}{E \left[\int H_{y^i} (W_{\theta+\delta}^i) S_i^{\hat{g}} (\psi_i(W_{\theta+\delta}^i - y^i))^2 dy^i \right]} \\ &= \frac{E \left[\int H_{y^i} (W_{\theta+\delta}^i) \left(S_i^{\hat{h}}(1) S_i^{\hat{g}}(1) \psi_i - S_i^{\hat{h}}(1) \psi_i' + \delta^{-1} \psi_i \right) (W_{\theta+\delta}^i - y^i)^2 dy^i \right]}{E \left[\int H_{y^i} (W_{\theta+\delta}^i) \left(\psi_i S_i^{\hat{g}}(1) - \psi_i' \right) (W_{\theta+\delta}^i - y^i)^2 dy^i \right]} \\ &= \frac{E \left[\int_0^{\infty} \left(S_i^{\hat{h}}(1) S_i^{\hat{g}}(1) \psi_i(y^i) - S_i^{\hat{h}}(1) \psi_i'(y^i) + \delta^{-1} \psi_i(y^i) \right)^2 dy^i \right]}{E \left[\int_0^{\infty} \left(\psi_i(y^i) S_i^{\hat{g}}(1) - \psi_i'(y^i) \right)^2 dy^i \right]} \end{aligned}$$

Observing that $E \left[S_i^{\hat{h}}(1) S_i^{\hat{g}}(1) \right] = -\delta^{-1}$ and $E \left[S_i^{\hat{h}}(1)^2 S_i^{\hat{g}}(1) \right] = E \left[S_i^{\hat{h}}(1) \right] = 0$, we get :

$$\begin{aligned} (\hat{\eta}^i)^2 &= \frac{\int_0^{\infty} \left(E \left[S_i^{\hat{h}}(1)^2 S_i^{\hat{g}}(1)^2 \right] \psi_i(y^i)^2 + E \left[S_i^{\hat{h}}(1)^2 \right] \psi_i'(y^i)^2 - \delta^{-2} \psi_i(y^i)^2 \right) dy^i}{\int_0^{\infty} \left(E \left[S_i^{\hat{g}}(1)^2 \right] \psi_i(y^i)^2 + \psi_i'(y^i)^2 \right) dy^i} \\ &= \frac{\int_0^{\infty} \left(-\delta^{-2} + E \left[S_i^{\hat{h}}(1)^2 S_i^{\hat{g}}(1)^2 \right] + E \left[S_i^{\hat{h}}(1)^2 \right] (\hat{\rho}^i)^2 \right) \psi_i(y^i)^2 dy^i}{\int_0^{\infty} \left(E \left[S_i^{\hat{g}}(1)^2 \right] + (\hat{\rho}^i)^2 \right) \psi_i(y^i)^2 dy^i} \\ &= \frac{-\delta^{-2} + E \left[S_i^{\hat{h}}(1)^2 S_i^{\hat{g}}(1)^2 \right] + E \left[S_i^{\hat{h}}(1)^2 \right] (\hat{\rho}^i)^2}{E \left[S_i^{\hat{g}}(1)^2 \right] + (\hat{\rho}^i)^2}. \end{aligned}$$

By similar calculations, we also get

$$(\hat{\rho}^i)^2 = \frac{-\delta^{-2} + E \left[S_i^{\hat{h}}(1)^2 S_i^{\hat{g}}(1)^2 \right] + E \left[S_i^{\hat{g}}(1)^2 \right] (\hat{\eta}^i)^2}{E \left[S_i^{\hat{h}}(1)^2 \right] + (\hat{\eta}^i)^2}.$$

Recalling the definition of $S_i^{\hat{h}}(1)$ and $S_i^{\hat{g}}(1)$, we immediately calculate :

$$E \left[S_i^{\hat{h}}(1)^2 S_i^{\hat{g}}(1)^2 \right] = \frac{3}{\delta^2} + \frac{1}{\delta \theta} \quad \text{and} \quad E \left[S_i^{\hat{h}}(1)^2 \right] = \frac{1}{\theta} + \frac{1}{\delta}.$$

Therefore, we obtain the following system of equations :

$$\begin{aligned}(\hat{\eta}^i)^2 &= \frac{2\delta^{-2} + \theta^{-1}\delta^{-1} + (\theta^{-1} + \delta^{-1})(\hat{\rho}^i)^2}{\delta^{-1} + (\hat{\rho}^i)^2} \\(\hat{\rho}^i)^2 &= \frac{2\delta^{-2} + \theta^{-1}\delta^{-1} + \delta^{-1}(\hat{\eta}^i)^2}{\theta^{-1} + \delta^{-1} + (\hat{\eta}^i)^2},\end{aligned}$$

which can be solved explicitly and we get :

$$\hat{\eta}^i = (\delta^{-1} + \theta^{-1})^{1/4} (2\delta^{-1} + \theta^{-1})^{1/4} \quad \text{and} \quad \hat{\rho}^i = \delta^{-1/2} \left(\frac{2\delta^{-1} + \theta^{-1}}{\delta^{-1} + \theta^{-1}} \right)^{1/4}. \quad (5.8)$$

Remark 5.1 Let θ be a fixed positive parameter. Then, the above expression shows that

$$\hat{\eta}^i = O\left(\delta^{-1/2}\right) \quad \text{and} \quad \hat{\rho}^i = O\left(\delta^{-1/2}\right) \quad \text{for small } \delta > 0.$$

This is the same asymptotic behavior of the localization function that is in [6].

3. Optimal (non-separable) localization.

In this paper, we only focus on the problem of minimizing the integrated mean square error with the class of separable localization functions. The problem of minimizing the integrated mean square error, within the class of all localization functions, has been analyzed in [5] in the context of approximating $E[\varepsilon_x(X_\theta)g(X_{\theta+\delta})]$. An optimal (non-separable) localization function does exist in some convenient Sobolev space, but there was no explicit characterization of this solution in general. However, in the case of the Brownian motion, i.e. $X = W$, it was observed that the optimal separable localization function is also optimal for the problem of minimizing the integrates mean square error with the class of all localization functions. See Remark 4.3 in [5].

A similar comment is valid for the context of this paper. Following the same proof as in [5], we see that *the optimal separable localization functions for the Brownian motion, determined in the previous paragraph, solve the problem of integrated mean square error minimization within the class of all localization functions.*

5.2 Numerical results

In this section, we apply our technique to estimate the joint density $f(x, y)$ of $(W_\theta, W_{\theta+\delta})$, and compare to the exact value

$$f(x, y) = (2\pi)^{-d}(\theta\delta)^{-d/2} \exp\left(-\frac{1}{2\theta\delta} \sum_{i=1}^d (\theta + \delta)x_i^2 - 2\theta x_i y_i + \theta y_i^2\right). \quad (5.9)$$

For a fixed number of scenarios $M = 200\,000$, we analyze the impacts of the dimension, the dispersion between x and y and the time step δ on the quality of the estimator.

5.2.1 Impact of the dimension

In order to study the impact of the dimension on our estimator, we fix $\theta = 1$, $\delta = 1$, estimate the joint density at the origin $(x, y) = (0, 0)$ for d ranging from 1 to 4. We collect the estimation results together with the comparison to the exact value given by (5.9) in the following table. The mean and the standard deviation are estimated by their empirical counterpart on an independent sample of 500 Malliavin based estimations of $f(0, 0)$.

d	Mean	Std. deviation	Th. value	$\frac{\text{Mean} - \text{Th. value}}{\text{Th. value}}$	$\frac{\text{Std. deviation}}{\text{Th. value}}$
Localization by exact $\hat{\eta}$ and $\hat{\rho}$, see (5.8)					
1	0.1592	$0.7982 \cdot 10^{-3}$	0.1592	0.001%	0.48%
2	0.0253	$0.3408 \cdot 10^{-3}$	0.0254	-0.03%	1.35%
3	0.040	$0.1377 \cdot 10^{-3}$	0.0040	0.07%	3.42%
4	0.0006	$0.0566 \cdot 10^{-3}$	0.0006	-0.13%	8.82%
Localization by $\eta^i = \rho^i = \delta^{-1/2}$, see Remark 5.1					
1	0.1592	$0.6534 \cdot 10^{-3}$	0.1592	0.01%	0.41%
2	0.0253	$0.2745 \cdot 10^{-3}$	0.0254	-0.06%	1.08%
3	0.0040	$0.1081 \cdot 10^{-3}$	0.0040	-0.25%	2.68%
4	0.0006	$0.0397 \cdot 10^{-3}$	0.0006	-0.20%	6.19%

Notice that the mean of the Malliavin based estimator is very close to the exact value. The above results also show that the quality of the estimator is worsened by increasing the dimension.

5.2.2 Impact of the dispersion

We now analyze the impact of the distance between x and y . We fix $\theta = \delta = 1$, and we estimate the joint density $f(x, y)$ for different pairs (x_i, y_i) . We perform numerical experiments for $d = 1$ and $d = 2$. The following table reports our estimation results together with the comparison to the exact valued obtained from (5.9).

Dimension 1

x	y	Mean	Std deviation	Th value	$\frac{\text{Mean} - \text{Th. value}}{\text{Th. value}}$	$\frac{\text{Std deviation}}{\text{Th value}}$
Localization by exact $\hat{\eta}$ and $\hat{\rho}$, see (5.8)						
0	0	0.1591	$0.7342 \cdot 10^{-3}$	0.1592	-0.018%	0.46%
0	1	0.0965	$0.5840 \cdot 10^{-3}$	0.0965	0.004%	0.60%
-1	1	0.0131	$0.2792 \cdot 10^{-3}$	0.0131	-0.096%	2.14%
Localization by $\eta^i = \rho^i = \delta^{-1/2}$, see Remark 5.1						
0	0	0.1591	$0.6628 \cdot 10^{-3}$	0.1592	-0.007%	0.42%
0	1	0.0966	$0.5424 \cdot 10^{-3}$	0.0965	0.04%	0.56%
-1	1	0.131	$0.5086 \cdot 10^{-3}$	0.0131	0.37%	3.89%

Dimension 2

x	y	Mean	Std. dev.	Th val.	$\frac{\text{Mean} - \text{Th. val.}}{\text{Th. val.}}$	$\frac{\text{Std dev.}}{\text{Th val.}}$
Localization by exact $\hat{\eta}$ and $\hat{\rho}$, see (5.8)						
(0, 0)	(0, 0)	0.0253	$0.3502 \cdot 10^{-3}$	0.0253	0.06%	1.38%
(0, 0)	$(1, 1)/\sqrt{2}$	0.0154	$0.2614 \cdot 10^{-3}$	0.0154	0.002%	1.70%
$-(1, 1)/\sqrt{2}$	$(1, 1)/\sqrt{2}$	0.0021	$0.0523 \cdot 10^{-3}$	0.0021	0.016%	2.52%
Localization by $\eta^i = \rho^i = \delta^{-1/2}$, see Remark 5.1						
(0, 0)	(0, 0)	0.0253	$0.2913 \cdot 10^{-3}$	0.0253	0.0091%	1.15%
(0, 0)	$(1, 1)/\sqrt{2}$	0.0154	$0.2240 \cdot 10^{-3}$	0.0154	-0.0029%	1.46%
$-(1, 1)/\sqrt{2}$	$(1, 1)/\sqrt{2}$	0.0021	$0.1052 \cdot 10^{-3}$	0.0021	-0.0609%	5.06%

The above result confirm the expected conclusion that the quality of the estimator is worsened by increasing the distance between x and y . Notice that the mean of the joint density estimator is very close to the exact values.

5.2.3 Impact of the time step

We now fix $\theta = 1$ and compute the Malliavin based estimator of the joint density at the origin $(x, y) = (0, 0)$ for different values of the time step δ . We again perform numerical experiments for the one-dimensional and the two-dimensional case, and we compare to the exact values given by (5.9).

Dimension 1 : $(x, y) = (0, 0)$

δ	Mean	Std deviation	Th value	$\frac{\text{Mean} - \text{Th. value}}{\text{Th. value}}$	$\frac{\text{Std deviation}}{\text{Th value}}$
Localization by exact $\hat{\eta}$ and $\hat{\rho}$, see (5.8)					
0.2	0.3559	0.0023	0.3559	0.006%	0.63%
0.5	0.2251	0.0012	0.2251	0.032%	0.53%
1	0.1592	0.0007	0.1592	0.022%	0.45%
Localization by $\eta^i = \rho^i = \delta^{-1/2}$, see Remark 5.1					
0.2	0.3558	0.0024	0.3559	-0.017%	0.66%
0.5	0.2251	0.0011	0.2251	0.012%	0.50%
1	0.1591	0.0007	0.1592	-0.026%	0.43%

Dimension 1 : $(x, y) = (-1/2, 1/2)$

δ	Mean	Std deviation	Th value	$\frac{\text{Mean} - \text{Th. value}}{\text{Th. value}}$	$\frac{\text{Std deviation}}{\text{Th value}}$
Localization by exact $\hat{\eta}$ and $\hat{\rho}$, see (5.8)					
0.2	0.0257	$0.87 \cdot 10^{-3}$	0.0258	-0.0016	3.37%
0.5	0.0731	$0.52 \cdot 10^{-3}$	0.0731	0.0009	0.71%
1	0.0852	$0.46 \cdot 10^{-3}$	0.0852	0.0004	0.54
Localization by $\eta^i = \rho^i = \delta^{-1/2}$, see Remark 5.1					
0.2	0.0257	0.0012	0.0258	-0.12%	4.83%
0.5	0.0731	0.0007	0.0731	0.03%	0.90%
1	0.0852	0.0005	0.0852	-0.03%	0.55%

Dimension 2 : $(x, y) = (0, 0)$

δ	Mean	Std deviation	Th value	$\frac{\text{Mean} - \text{Th. value}}{\text{Th. value}}$	$\frac{\text{Std deviation}}{\text{Th value}}$
Localization by exact $\hat{\eta}$ and $\hat{\rho}$, see (5.8)					
0.2	0.1266	0.0028	0.1267	-0.0081%	2.18%
0.5	0.0506	0.0007	0.0507	-0.0356%	1.47%
1	0.0253	0.0004	0.0253	-0.0933%	1.39%
Localization by $\eta^i = \rho^i = \delta^{-1/2}$, see Remark 5.1					
0.2	0.1266	0.0029	0.1267	-0.069%	2.26%
0.5	0.0507	0.0007	0.0507	0.055%	1.47%
1	0.0253	0.0003	0.0253	0.023%	1.13%

Dimension 2 : $(x^1, x^2) = -(y^1, y^2) = \frac{-1}{2\sqrt{2}}(1, 1)$

δ	Mean	Std deviation	Th value	$\frac{\text{Mean} - \text{Th. value}}{\text{Th. value}}$	$\frac{\text{Std deviation}}{\text{Th value}}$
Localization by exact $\hat{\eta}$ and $\hat{\rho}$, see (5.8)					
0.2	0.0092	$0.44 \cdot 10^{-3}$	0.0092	-0.14%	4.82%
0.5	0.0165	$0.33 \cdot 10^{-3}$	0.0164	0.15%	2.02%
1	0.0136	$0.20 \cdot 10^{-3}$	0.0136	-0.01%	1.50%
Localization by $\eta^i = \rho^i = \delta^{-1/2}$, see Remark 5.1					
0.2	0.0092	$0.66 \cdot 10^{-3}$	0.0092	-0.089%	7.20%
05	0.0165	$0.31 \cdot 10^{-3}$	0.0164	0.037%	1.89%
1	0.0136	$0.17 \cdot 10^{-3}$	0.0136	0.012%	1.29%

The above results show that the quality of the estimator is worsened by shrinking the time step. This is in agreement with the variance explosion observation of the Malliavin based estimator observed in [6]. We also observe, as in the previous experiments, that the mean of the Malliavin based estimator is very close to the exact value.

Remark 5.2 From the examples above, we notice that the standard deviation of our estimations using $\eta^i = \rho^i = \delta^{-1/2}$ is very close to (and even sometimes better than) the standard of our estimations using the optimal values of η and ρ . This is due to the fact that these optimal values minimize the integrated-variance and not the variance of the joint density for a particular choice of x and y . Notice however that, for a large dispersion between x and y , or for a small time step δ , the optimal values $\hat{\eta}$ and $\hat{\rho}$ produce better estimators of the joint density.

5.3 Comparison with the kernel method

We next compare the malliavin based joint density estimator to the so-called kernel estimation method. Before reporting the numerical experiments, We start by recalling briefly the regression methods using convolution kernels.

5.3.1 Kernel convolution regression method

Let Z be a random variable valued in \mathbb{R}^k , and let $(Z_i)_{i \geq 1}$ be a sequence of independent copies of Z . In order to estimate the density function $\ell(z) := \mathbb{E}[\delta_z(Z)]$, we approximate the Dirac measure ε_z by the function $\zeta \mapsto b_n^{-k} \mathbf{1}_{|\zeta - z| \leq b_n}$, where $(b_n)_{n \geq 1}$ is a sequence of positive numbers converging to zero. This suggest to approximate $\ell(z)$ by

$$\frac{1}{nb_n^k} \sum_{i=1}^n \mathbf{1}_{|Z_i - z| \leq b_n}.$$

More generally, given a bounded function $K : \mathbb{R}^k \rightarrow \mathbb{R}_+$ satisfying

$$\int_{\mathbb{R}^d} K(r) dr = 1 \quad \text{and} \quad \lim_{|r| \rightarrow \infty} |r|^{2d} K(r) = 0,$$

we define the approximation of the Dirac measure ε_z :

$$K_n(\zeta) = \frac{1}{b_n^k} K\left(\frac{\zeta}{h}\right).$$

In the statistics literature, K is called a kernel function, and induces the following kernel estimator of ℓ :

$$\hat{\ell}_n(z) := \frac{1}{n} \sum_{i=1}^n K_n(Z_i - z).$$

Under some regularity conditions on ℓ and by judicious choice of the rate of convergence of $(b_n)_n$ towards zero, it is well-known that $\hat{\ell}_n(z)$ is a consistent estimator of $\ell(z)$. The rate of convergence decreases dramatically with the dimension, and is always slower than the parametric rate of convergence \sqrt{n} . We refer to Bosq [3] for an overview of the asymptotic properties of $\hat{\ell}_n$.

5.3.2 Malliavin versus Kernel joint density estimation

We fix $\theta = \delta = 1$, and estimate the joint density by both methods, for d ranging from 1 to 3, at the point $(x, y) = (0, u/|u|)$, where u is a vector of ones. The numerical results are collected in the following table.

d	Th value	Malliavin Calculus based approach			kernel based approach		
		Mean	Std. dev.	$\frac{\text{Std. dev.}}{\text{Th. value}}$	Mean	Std. dev.	$\frac{\text{Std. dev.}}{\text{Th. value}}$
1	0.0965	0.0965	$0.57 \cdot 10^{-3}$	0.59%	0.0958	0.0015	1.54%
2	0.0154	0.0154	$0.28 \cdot 10^{-3}$	1.81%	0.0140	0.0005	2.95%
3	0.0024	0.0024	$0.12 \cdot 10^{-3}$	4.96%	0.0018	0.0001	3.72%

Notice that, unlike the Malliavin-based estimator, the mean of the kernel based estimator is far from the exact value. This is due to the fact that the Malliavin-based estimator is unbiased.

6 Solving the American option valuation problem by quantization

In this section, we focus on the problem of approximating the value of an optimal stopping problem by the *quantization tree method* introduced by Bally, Pagès and Printemps [1].

The quantization tree methods consists in approximating a continuous-time process with values in a continuous state space by a discrete-time process with finite state space. A key step in this quantization method is to estimate the transition matrix of the approximating discrete-time process. This is usually obtained by classical Monte Carlo technique. The main objective of this section is to use instead our Malliavin-based Monte Carlo estimator of the joint density, and to analyze the performance of this procedure.

6.1 The American option problem

Let $T > 0$ be a finite horizon, and $(\Omega, \mathcal{F}, \mathbb{P})$ a complete probability endowed with a filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ satisfying the usual conditions. The filtration \mathcal{F}_t represents the information available up to time t . We also assume that \mathcal{F}_0 is trivial and that $\mathcal{F}_T = \mathcal{F}$. Let Y be an \mathbb{F} -adapted process valued in \mathbb{R}_+ . For each $t \in [0, T]$ and $\omega \in \Omega$, $Y_t(\omega)$ represents the reward for stopping at time t in the state ω . Our main concern in this section is the valuation of American options which reduces to the following optimal stopping problem

$$U_0 := \sup_{\tau \in \mathcal{T}_0} \mathbb{E}[Y_\tau], \quad (6.1)$$

where \mathcal{T}_0 is the collection of all \mathbb{F} -stopping times valued in $[0, T]$.

6.2 Discrete-time approximation

Let $0 = t_0 < t_1 < \dots < t_n = T$ be a partition of the interval $[0, T]$ with mesh $\max_{1 \leq i \leq n} |t_i - t_{i-1}| \rightarrow 0$ as $n \rightarrow \infty$. One may choose for instance a constant mesh $t_i = iT/n$. A natural approximation of the optimal stopping problem (6.1) is obtained by restricting the maximization to those stopping times with values in $\{t_0, \dots, t_n\}$. We shall call \mathcal{T}_0^n the collection of all such stopping times, and

$$U_0^n := \sup_{\tau \in \mathcal{T}_0^n} \mathbb{E}[Y_\tau]. \quad (6.2)$$

The convergence of U_0^n towards U_0 holds under mild conditions. Also, it is well-known that the value of the discrete-time stopping problem (6.2) can be computed by the following backward induction :

$$U_n^n := Y_{t_n} \quad \text{and} \quad U_i^n = \max \{Y_{t_i}, \mathbb{E}[U_{i+1}^n | \mathcal{F}_{t_i}]\}$$

for $i = 0, \dots, n-1$. Assume in addition that the process Y is Markov. Then we can rewrite the above backward algorithm in :

$$U_n^n := Y_{t_n} \quad \text{and} \quad U_i^n = \max \{Y_{t_i}, \mathbb{E}[U_{i+1}^n | Y_{t_i}]\}.$$

6.3 Estimation of the discrete-time American option value by quantization

We next approximate, at each time t_i , the random variable Y_{t_i} by some random variable \hat{Y}_{t_i} which can only take a finite number N_i of values $y_i := (y_{i,j})_{1 \leq j \leq N_i}$. This is the quantization step. A *Voronoi tessellation* of y_i is a Borel partition $(C_j(y_i))_{1 \leq j \leq N_i}$ of \mathbb{R}^k satisfying

$$C_j(y_i) \subset \left\{ y \in \mathbb{R}^k : |y_{i,j} - y| = \min_{1 \leq k \leq N_i} |y_{i,k} - y| \right\}.$$

The quantizer of Y_{t_i} is defined by :

$$\hat{Y}_{t_i} := \sum_{j=1}^{N_i} y_{i,j} \mathbf{1}_{C_j(y_i)}(Y_{t_i}).$$

We now turn to the problem of choosing y_i :

1. We say that y_i is an L^p optimal quantizer of Y_{t_i} if y_i solves the problem of minimizing the L^p mean of the approximation error

$$D_i := \min_{y_i \in \mathbb{R}^{N_i}} E |\hat{Y}_{t_i} - Y_{t_i}|^p.$$

This approach is presented by [2], and the above minimization problem is solved by means of a (time consuming) stochastic gradient descent method. The quantity D_i is called the L^p -distorsion of the quantizer y_i .

2. A random quantization of Y_{t_i} consists in choosing N_i independent realizations of a random variable whose support contains the support of Y_{t_i} . Given the density function of Y_{t_i} , Cohort [9] provided the density function whose associated random quantization has the smallest rate of convergence of the distorsion when $N_i \rightarrow \infty$, and proved that this procedure achieves the same rate of convergence than the optimal quantization method.

We next turn to the problem of approximation of the value of the discrete-time optimal stopping problem. Set

$$\pi_k^{i,j} := \mathbb{P} \left[\hat{Y}_{t_{k+1}} = y_{k+1,j} | \hat{Y}_{t_k} = y_{k,i} \right],$$

and define the approximation

$$\hat{U}_n^n := \hat{Y}_{t_n} \quad \text{and} \quad \hat{U}_k^n = \max \left\{ \hat{Y}_{t_k}, \pi_k \hat{U}_{k+1}^n \right\},$$

where $U_k^n \in R^{N_k}$, and π_k is a matrix with entries $\pi_k^{i,j}$, and the maximum is taken component-wise.

6.4 Monte Carlo approximation of the transition matrices

We now concentrate on the approximation of the transition probabilities $\pi_k^{i,j}$. Let $(Y^{(m)})_{1 \leq m \leq M}$ be M copies of the process Y . The classical Monte Carlo estimator of the transition probabilities $\pi_k^{i,j}$ is given by :

$$\tilde{\pi}_k^{i,j} = \frac{\sum_{m=1}^M \mathbf{1}_{Y_{t_k}^{(m)} \in C_i(y_k)} \mathbf{1}_{Y_{t_{k+1}}^{(m)} \in C_j(y_{k+1})}}{\sum_{m=1}^M \mathbf{1}_{Y_{t_k}^{(m)} \in C_i(y_k)}}$$

As proved in [2], the bigger the N_k are, the smaller the error bounds are. So, in practice, in order to have a satisfying estimation of the value of the stopping-problem one has to take a big number of cells for each time step (usually N_k lies between 500 and 2000). When, we take a big number of cells N_k , the size of these cells become small and the variance of the estimators of the transition probabilities $\tilde{\pi}_k^{i,j}$ becomes large, since the number of scenarios falling in each cell is small. In order to overcome this problem, we suggest an approach using Malliavin calculus estimation of the joint density function. recall that the transition probability from the cell $C_i(y_k)$ to the cell $C_j(y_{k+1})$ is given by:

$$\pi_k^{i,j} = \frac{\iint_{C_i(y_k) \times C_j(y_{k+1})} f^{t_k, t_{k+1}}(x, y) dx dy}{\int_{C_i(y_k)} f^{t_k}(x) dx}$$

where $f^{t_k, t_{k+1}}$ is the joint density function of Y_{t_k} and $Y_{t_{k+1}}$ and f^{t_k} is the density function of Y_{t_k} . We will then approximate $\pi_k^{i,j}$ by the following expression:

$$\hat{\pi}_k^{i,j} := \frac{\hat{f}^{t_k, t_{k+1}}(y_{k,i}, y_{k+1,j}) \mathcal{V}(C_j(y_{k+1}))}{\sum_{l=1}^{N_{k+1}} \hat{f}^{t_k, t_{k+1}}(y_{k,i}, y_{k+1,l}) \mathcal{V}(C_l(y_{k+1}))},$$

where $\mathcal{V}(C_j(y_{k+1})) = \int_{C_j(y_{k+1})} dy$ is the hyper-volume¹ of the cell $C_j(y_{k+1})$, and $\hat{f}^{t_k, t_{k+1}}(x_1, x_2)$ is the Malliavin based estimator of the joint density $\mathbb{E}[\varepsilon_{x_1}(X_{t_k}) \varepsilon_{x_2}(X_{t_{k+1}})]$ as developed in the previous sections of this paper.

¹This hyper-volume is calculated using the function “convhulln” of MATLAB (www.mathworks.com). This function uses essentially a free software named “Qhull” that was created by the *National Science and Technology Research Center for Computation and Visualization of Geometric Structures of university of Minnesota* (www.geom.umn.edu/software/qhull/).

6.5 Numerical Implementation

6.5.1 The model

Let S be a process defined by the multi-dimensional Black-Scholes model :

$$\frac{dS_t^i}{S_t^i} = rdt + \sum_{j=1}^d \sigma_{ij} dW_t^j, i = 1, \dots, d$$

Where $W = (W^1, \dots, W^d)$ is a d-dimensional standard Brownian motion and r is a strictly positive constant and the matrix $[\sigma_{ij}]_{(i,j) \in \llbracket 1, d \rrbracket^2}$ is invertible. The reward process Y is defined by

$$Y_t := e^{-rt} f(S_t),$$

where f is a non-negative mapping from \mathbb{R}_+^d into \mathbb{R}_+ .

The most important feature of the Black-Scholes model is that for any time $t \in [0, T]$, the asset price S_t is a function of t and the Brownian motion W_t :

$$S_t^i = S_0^i \exp \left(\left(r - \sum_{j=1}^d \frac{\sigma_{ij}^2}{2} \right) t + \sum_{j=1}^d \sigma_{ij} W_t^j \right).$$

Therefore, instead of quantizing S_t , one can quantize W_t . This remark, will be very helpful in practice.

6.5.2 The numerical scheme

Applying the quantization algorithm to the optimal stopping problem with reward process Y , we are reduced to the following backward scheme :

$$\begin{cases} \hat{u}_n^i = e^{-rt_n} f(s_i^n), & i = 1, \dots, N_n \\ \hat{u}_k^i = \max(e^{-rt_k} f(s_i^k), e^{-r(t_{k+1}-t_k)} \sum_{j=1}^{N_{k+1}} \hat{\pi}_k^{i,j} \hat{u}_{k+1}^j), & k = 0, \dots, n-1, i = 1, \dots, N_k \end{cases}$$

Where $s_i^k \in \mathbb{R}^d$ is given by:

$$\left(s_i^k \right)_l = \left(s_1^0 \right)_l \exp \left(\left(r - \sum_{j=1}^d \frac{\sigma_{lj}^2}{2} \right) t_k + \sum_{j=1}^d \sigma_{lj} \left(x_i^k \right)_l \right)$$

and $x_i^k \in \mathbb{R}^d$ the i -th scenario ($i = 1 \dots N_k$) of the multi-dimensional Brownian motion $W_{t_k}, k = 0, \dots, n$.

For each $k = 0, \dots, n-1$, $i = 1, \dots, N_k$, and $j = 1, \dots, N_{k+1}$, we approximate the transition probability $\pi_k^{i,j}$ by :

$$\hat{\pi}_k^{i,j} = \frac{\hat{f}^{t_k, t_{k+1}}(x_{k,i}, x_{k+1,j}) \mathcal{V}(C_j(x_{k+1}))}{\sum_{l=1}^N \hat{f}^{t_k, t_{k+1}}(x_{k,i}, x_{k+1,l}) \mathcal{V}(C_l(x_{k+1}))}$$

where $\hat{f}^{t_k, t_{k+1}}(x_1, x_2)$ is the Malliavin based estimator of the joint density $\mathbb{E}[\varepsilon_{x_1}(W_{t_k}) \varepsilon_{x_2}(W_{t_{k+1}})]$ as developed in the previous sections of this paper.

Remark 6.1 In the estimator given above, we omitted deliberately to deal with case where we have unbounded cells (i.e. the periferal cells in the quantization tree methods), because several solutions can be proposed.

One can for example (like in classical lattice methods) bound his workspace and say that the probability to find a realization of the underlying process outside a certain domain is null. Doing so, we transform all our cells into bounded cells.

An other possibility is to use the classical estimation of $\mathbb{P}(C_i(x_k), C_j(x_{k+1}))$ when the 2 cells are unbounded. The case where only one of the two cells is unbounded can be treated in the following way (Suppose for example that $C_i(x_k)$ is bounded and $C_j(x_{k+1})$ is unbounded): We make the following approximation

$$\iint_{C_i(x_k) \times C_j(x_{k+1})} \hat{f}^{t_k, t_{k+1}}(x, y) dx dy = \left(\int_{C_j(x_{k+1})} \hat{f}^{t_k, t_{k+1}}(x_{k,i}, y) dy \right) \mathcal{V}(C_j(x_k))$$

and we estimate $\hat{f}^{t_k, t_{k+1}}(x_{k,i}, y)$ by making only one Mallivin-integration by part on “the first variable” (by doing this, we retrieve the framework introduced by Bouchard, Ekeland and Touzi in [5]).

We tested both methods. The results given by the second one seems to be slightly more relevant. That’s why we decided to adopt it in the numerical tests below.

6.5.3 Numerical Results

Impact of the dimension In order to test the impact of the dimension, we consider a European option and a Bermudan one, on the geometric mean of d assets :

$$f(s) := \left[K - \left(\prod_{i=1}^d s^i \right)^{\frac{1}{d}} \right]^+ .$$

In the multidimensional Black and Scholes framework, the geometric mean of d non-dividend lognormal processes is equivalent to a particular lognormal process with a divi-

dend yield. We give below the features of this process:

Initial value X_0	$\left(\prod_{i=1}^d S_0^i\right)^{\frac{1}{d}}$
Equivalent volatility σ_{eq}	$\frac{1}{d^2} \sum_{j=1}^d \left(\sum_{i=1}^d \sigma_{ij}\right)^2$
Dividend yield q	$\frac{1}{2} \left(\frac{1}{d} \sum_{i,j=1}^d \sigma_{ij}^2 - \sigma_{eq}^2\right)$

This allows to obtain the exact value of the price of the European option using the usual Black-Scholes explicit formula. We also use a 1-dimensional finite difference scheme and we consider the obtained approximation as the “exact value” of the price of the Bermudan option.

On the other hand, we price these two options with quantization tree method. We give below the backward induction allowing for that:

$$\begin{cases} \hat{u}_n^i = e^{-rt_n} f(s_i^n), & i \in \langle 1, N_n \rangle \\ \hat{u}_k^i = \max \left(I_k^i, e^{-r(t_{k+1}-t_k)} \sum_{j=1}^{N_{k+1}} \hat{\pi}_k^{i,j} \hat{u}_{k+1}^j \right), & k \in \langle 0, n-1 \rangle, i \in \langle 1, N_k \rangle \end{cases}$$

where for any $k \in \langle 0, n-1 \rangle$ and $i \in \langle 1, N_k \rangle$, the quantity I_k^i is equal to $-\infty$ for the European option and to $f(s_i^k)$ for the Bermudan one.

The following are the parameters used for the simulations :

instantaneous interest rate r	0.06
volatility	$\sigma_i = 0.3$ and $\rho_{ij} = 0.7$
maturity T	1
initial values	$S_0^i = 36$
strike K	40
time step	1/5

We give here, the prices obtained with 6000 scenarios and a stochastic tree of size 300. Then, we compare them to the results obtained when, we estimate the transition probabilities by the classical approach (We also consider the same 6000 scenarios and the same 300 Voronoi cells at each time step):

European option

		Malliavin Calculus based approach			Classical estimation based approach		
d	Exact price	Mean	Std. dev.	$\frac{\text{Std. dev.}}{\text{Exact price}}$	Mean	Std. dev.	$\frac{\text{Std. dev.}}{\text{Exact price}}$
1	5.277	5.272	0.024	0.45%	5.283	0.007	0.13%
2	5.064	5.076	0.026	0.51%	5.090	0.009	0.18%
3	4.988	4.991	0.025	0.50%	5.057	0.009	0.18%
4	4.949	4.937	0.023	0.46%	5.082	0.012	0.24%

Bermudan option

		Malliavin Calculus based approach			Classical estimation based approach		
d	Exact price	Mean	Std. dev.	$\frac{\text{Std. dev.}}{\text{Exact price}}$	Mean	Std. dev.	$\frac{\text{Std. dev.}}{\text{Exact price}}$
1	5.657	5.665	0.042	0.74%	5.683	0.0061	0.11%
2	5.426	5.412	0.053	0.97%	5.464	0.0056	0.10%
3	5.347	5.337	0.052	0.97%	5.416	0.0068	0.13%
4	5.306	5.313	0.063	1.19%	5.467	0.0072	0.14%

The above two tables show that the numerical results obtained by the Malliavin based approach are closer to the exact value. This suggests that the Malliavin based approach produces a smaller bias than the Bally and Pages [1] one. However, the Malliavin based estimators typically exhibit a larger estimation error.

More exotic examples We consider three more examples of exotic American (or more exactly Bermudan) options from finance. The first one is a put option on the arithmetic mean of 4 assets

$$f_1(s^1, s^2, s^3, s^4) := [K - \frac{1}{4}(s^1 + s^2 + s^3 + s^4)]^+,$$

the second one is a put option on the minimum of same assets

$$f_2(s^1, s^2, s^3, s^4) := [K - \min(s^1, s^2, s^3, s^4)]^+,$$

and the third one is a put option on the maximum of the same assets

$$f_3(s^1, s^2, s^3, s^4) := [K - \max(s^1, s^2, s^3, s^4)]^+,$$

The parameters used for this simulation are the following :

instantaneous interest rate r	0.06
volatility	$\sigma_i = 0.3$ and $\rho_{ij} = 0.8$
maturity T	1
initial values	$S_0^i = 36$
strike K	40
time step	1/10

We next provide the values obtained with 6000 scenarios and a stochastic tree of size 300. Then we compare the results obtained when we estimate the transition probabilities by the classical approach (We consider two cases: (a) 6000 scenarios and the same 300 Voronoi cells at each time step or (b) 600000 scenarios and the same 300 Voronoi cells at each time step):

Put on	Mean	Std. dev.	$\frac{\text{Std. dev.}}{\text{Mean}}$
Malliavin Calculus based approach			
Arith. mean	5.40	0.031	0.57%
Min	8.44	0.058	0.69%
Max	4.07	0.028	0.76%
Classical estimation based approach (a)			
Arith. mean	5.52	0.020	0.36%
Min	8.59	0.029	0.34%
Max	4.21	0.026	0.62%
Classical estimation based approach (b)			
Arith. mean	5.40	0.0030	0.06%
Min	8.49	0.0034	0.04%
Max	4.06	0.0024	0.06%

7 Appendix : Proof of Theorem 4.1

We shall split the proof in various Lemmas. We will use classical calculus of variations technique. We start by defining "admissible" perturbations of the assumed solution. Let φ be an arbitrary function in \mathcal{L}_+^s , and α a map from \mathbb{R}_+ into \mathbb{R} . Set

$$\Phi_i[\varphi, \alpha](x) := \alpha(x^i) \prod_{j \neq i} \varphi_j(x^j).$$

It follows that if α is C^1 , bounded and satisfies $\alpha(0) = 1$, we have that $\Phi_i[\varphi, \alpha] \in \mathcal{L}_+^s$. In particular, for $\alpha(x) \equiv \mathbf{1}$, $x \in \mathbb{R}_+^d$, we simplify the notation by setting :

$$\varphi_{(-i)} := \Phi_i[\varphi, \mathbf{1}] .$$

Lemma 7.1 *For $\varphi, \psi \in \mathcal{L}_+^s$, we define :*

$$\begin{aligned} F_i &:= \varphi_i(X_\theta^i), F'_i := \varphi'_i(X_\theta^i), F_{(-i)} := \varphi_{(-i)}(X_\theta), \\ G_i &:= \psi_i(X_{\theta+\delta}^i), G'_i := \psi'_i(X_{\theta+\delta}^i), G_{(-i)} := \psi_{(-i)}(X_{\theta+\delta}), \end{aligned}$$

Notice that $F = F_i F_{(-i)}$ and $G = G_i G_{(-i)}$. Then, for all $(g, h) \in \mathbf{H}(X)$, we have that :

$$\begin{aligned} S^g \circ S^h(FG) &= F_i S^g \circ S^h(F_{(-i)}G) - F'_i S^g \circ S_{(-i)}^h(F_{(-i)}G) \\ &= G_i S^g \circ S^h(FG_{(-i)}) - G'_i S_{(-i)}^g \circ S_{(-i)}^h(FG_{(-i)}) . \end{aligned}$$

Proof. We use the same argument as the proof of Lemma 4.1 in [5]. From (3.2) we have that :

$$S^h(FG) = F_i S^h(F_{(-i)}G) - F'_i S_{(-i)}^h(F_{(-i)}G) .$$

Then, we deduce the first equality of the lemma from Remark 3.1. By the same Remark, we also have that :

$$\begin{aligned} S^g \circ S^h(FG) &= S^g(GS^h(F)) \\ &= G_i S^g(G_{(-i)}S^h(F)) - G'_i S_{(-i)}^g(G_{(-i)}S^h(F)) \\ &= G_i S^g \circ S^h(FG_{(-i)}) - G'_i S_{(-i)}^g \circ S_{(-i)}^h(FG_{(-i)}) , \end{aligned}$$

where the second equality follows from (3.1). □

The next step consists in considering the problem of partial minimization.

Lemma 7.2 *Let (φ, ψ) be an arbitrary pair of localizing functions in $\mathcal{L}_+^s \times \mathcal{L}_+^s$, and fix a positive integer $i \leq d$. Then the minimization problem*

$$\min_{\alpha \in C_b^1} I^{g,h}(\Phi_i(\varphi, \alpha), \psi) \tag{7.1}$$

has a unique solution:

$$\hat{\alpha}_i(\xi) := e^{-\eta^i \xi}, \quad \text{for all } \xi \in \mathbb{R}_+ ,$$

where

$$\eta^i := \sqrt{\frac{\mathbb{E} \int_{\mathbb{R}^{2d-1}} H_{x^{(-i)}} \left(X_\theta^{(-i)} \right) H_y \left(X_{\theta+\delta} \right) S^g \circ S^h \left(F_{(-i)}(x) G(y) \right)^2 dx^{(-i)} dy}{\mathbb{E} \int_{\mathbb{R}^{2d-1}} H_{x^{(-i)}} \left(X_\theta^{(-i)} \right) H_y \left(X_{\theta+\delta} \right) S^g \circ S_{(-i)}^h \left(F_{(-i)}(x) G(y) \right)^2 dx^{(-i)} dy}} \quad (7.2)$$

$$F_{(-i)}(x) := \varphi_{(-i)}(X_\theta - x) \quad \text{and} \quad G(y) := \psi(X_{\theta+\delta} - y)$$

Similarly, the partial minimization problem

$$\min_{\beta \in C_b^1} I^{g,h}(\varphi, \Phi_i(\psi, \beta)) \quad (7.3)$$

has a unique solution:

$$\widehat{\beta}_i(\xi) := e^{-\rho^i \xi}, \quad \text{for all } \xi \in \mathbb{R}_+,$$

where

$$\rho^i := \sqrt{\frac{\mathbb{E} \int_{\mathbb{R}^{2d-1}} H_x(X_\theta) H_{y^{(-i)}} \left(X_{\theta+\delta}^{(-i)} \right) S^g \circ S^h \left(F(x) G_{(-i)}(y) \right)^2 dx dy^{(-i)}}{\mathbb{E} \int_{\mathbb{R}^{2d-1}} H_x(X_\theta) H_{y^{(-i)}} \left(X_{\theta+\delta}^{(-i)} \right) S_{(-i)}^g \circ S^h \left(F(x) G_{(-i)}(y) \right)^2 dx dy^{(-i)}}}, \quad (7.4)$$

$$F(x) := \varphi(X_\theta - x) \quad \text{and} \quad G_{(-i)}(y) := \psi_{(-i)}(X_{\theta+\delta} - y).$$

Proof. We only report the proof for the partial minimization problem (7.1). The same line of argument works perfectly for the problem (7.3).

1. We first derive the first order condition which must be satisfied by some potential solution $\hat{\alpha}$ of the minimization problem (7.1). Without loss of generality, we assume that $\varphi_i = \hat{\alpha}$ so that $\Phi_i(\varphi, \hat{\alpha}) = \varphi$. By the optimality of $\hat{\alpha}$, we have

$$I^{g,h}(\varphi, \psi) \leq I^{g,h}(\Phi_i(\varphi, \hat{\alpha} + \varepsilon \alpha), \psi)$$

for all $\alpha \in C_b^1$ and $\varepsilon > 0$. By the linearity of the Skorohod integral together with Remark 3.1, this provides :

$$\begin{aligned} I^{g,h}(\varphi, \psi) &\leq I^{g,h}(\Phi_i(\varphi, \psi) + \varepsilon^2 I^{g,h}(\Phi_i(\varphi, \alpha), \psi) \\ &\quad + 2\varepsilon \mathbb{E} \int_{\mathbb{R}^{2d}} H_x(X_\theta) H_y(X_{\theta+\delta}) S^g \circ S^h(\varphi(X_\theta - x) \psi(X_{\theta+\delta} - y)) \\ &\quad S^g \circ S^h(\Phi_i(\varphi, \alpha)(X_\theta - x), \psi(X_{\theta+\delta} - y)) dx dy. \end{aligned}$$

We next divide by ε , send ε to zero, and observe that the same calculation can be performed with the function $-\alpha$. This leads to the first order condition :

$$\begin{aligned} 0 &= \mathbb{E} \int_{\mathbb{R}^{2d}} H_x(X_\theta) H_y(X_{\theta+\delta}) S^g \circ S^h(\varphi(X_\theta - x) \psi(X_{\theta+\delta} - y)) \\ &\quad S^g \circ S^h(\Phi_i(\varphi, \alpha)(X_\theta - x) \psi(X_{\theta+\delta} - y)) dx dy \end{aligned} \quad (7.5)$$

for all $\alpha \in C_b^1$.

2. In this step, we consider an arbitrary $\alpha \in C_b^1$ with compact support, and we work out the integral with respect to the variable x^i in (7.5). Using the notations introduced in the statement of the lemma, we deduce from Lemma 7.1 that :

$$\begin{aligned}
& \int H_{x^i}(X_\theta^i) S^g \circ S^h (\varphi(X_\theta - x) \psi(X_{\theta+\delta} - y)) S^g \circ S^h (\Phi_i(\varphi, \alpha)(X_\theta - x), \psi(X_{\theta+\delta} - y)) dx^i \\
&= \int_{-\infty}^{X_\theta^i} \left\{ \hat{\alpha}(X_\theta^i - x^i) S^g \circ S^h (F_{(-i)}(x)G(y)) - \hat{\alpha}'(X_\theta^i - x^i) S^g \circ S_{(-i)}^h (F_{(-i)}(x)G(y)) \right\} \\
&\quad \left\{ \alpha(X_\theta^i - x^i) S^g \circ S^h (F_{(-i)}(x)G(y)) - \alpha'(X_\theta^i - x^i) S^g \circ S_{(-i)}^h (F_{(-i)}(x)G(y)) \right\} dx^i \\
&= \int_0^\infty \left\{ \hat{\alpha}(\xi) S^g \circ S^h (F_{(-i)}(x)G(y)) - \hat{\alpha}'(\xi) S^g \circ S_{(-i)}^h (F_{(-i)}(x)G(y)) \right\} \\
&\quad \left\{ \alpha(\xi) S^g \circ S^h (F_{(-i)}(x)G(y)) - \alpha'(\xi) S^g \circ S_{(-i)}^h (F_{(-i)}(x)G(y)) \right\} d\xi,
\end{aligned}$$

where the last equality follows by the trivial change of variable $\xi = X_\theta^i - x^i$. By Fubini's theorem, (7.5) provides :

$$\begin{aligned}
0 &= \int_0^\infty \left\{ a_i \hat{\alpha}(\xi) \alpha(\xi) + b_i \hat{\alpha}'(\xi) \alpha'(\xi) + c_i [\hat{\alpha}(\xi) \alpha'(\xi) + \hat{\alpha}'(\xi) \alpha(\xi)] \right\} d\xi, \\
&= \int_0^\infty \left\{ a_i \hat{\alpha}(\xi) \alpha(\xi) + b_i \hat{\alpha}'(\xi) \alpha'(\xi) \right\} d\xi
\end{aligned}$$

where we used the fact that $\alpha(0) = 0$, α has compact support, and the coefficients a_i, b_i, c_i are given by :

$$\begin{aligned}
a_i &:= \mathbb{E} \int_{\mathbb{R}^{2d-1}} H_{x^{(-i)}} \left(X_\theta^{(-i)} \right) H_y(X_{\theta+\delta}) S^g \circ S^h (F_{(-i)}(x)G(y))^2 dx^{(-i)} dy \\
b_i &:= \mathbb{E} \int_{\mathbb{R}^{2d-1}} H_{x^{(-i)}} \left(X_\theta^{(-i)} \right) H_y(X_{\theta+\delta}) S^g \circ S^h (F_{(-i)}(x)G(y))^2 dx^{(-i)} dy \\
c_i &:= \mathbb{E} \int_{\mathbb{R}^{2d-1}} H_{x^{(-i)}} \left(X_\theta^{(-i)} \right) H_y(X_{\theta+\delta}) S^g \circ S^h (F_{(-i)}(x)G(y)) \\
&\quad S^g \circ S_{(-i)}^h (F_{(-i)}(x)G(y)) dx^{(-i)} dy dx^{(-i)} dy.
\end{aligned}$$

Integrating by parts, and using again the fact that α has compact support and $\alpha(0) = 1$, this provides :

$$\int_0^\infty \{ a_i \hat{\alpha}(\xi) - b_i \hat{\alpha}'(\xi) \} \alpha(\xi) d\xi = 0.$$

Since α is an arbitrary C^1 function with compact support, this implies that $a_i \hat{\alpha} - b_i \hat{\alpha}' = 0$ on \mathbb{R}_+ . Since $\hat{\alpha}(0) = 1$ and $\hat{\alpha}$ is bounded, this leads to a unique solution to the first order conditions :

$$\hat{\alpha}(\xi) = e^{-\eta_i \xi} \quad \text{for all } \xi \geq 0 \text{ with } \eta_i := \sqrt{\frac{a_i}{b_i}}.$$

3. Problem (7.1) consists in minimizing a convex function on a convex domain. Then the above (unique) solution of the first order condition (7.5) is indeed the unique solution of the problem (7.1). □

By using repeatedly the decomposition of Lemma 7.1, we obtain the following result which be useful for the next Lemma.

Corollary 7.1 *Let (φ, ψ) be a pair of localizing functions in $\mathcal{L}_+^s \times \mathcal{L}_+^s$. Then :*

$$S^g \circ S^h (\varphi(X_\theta)\psi(X_{\theta+\delta})) = \sum_{I \in \mathcal{J}} \sum_{J \in \mathcal{J}} (-1)^{|I|+|J|} \partial_I \varphi(X_\theta) \partial_J \psi(X_{\theta+\delta}) S_{-J}^g S_{-I}^h (1).$$

In the following Lemma, we reduce the integrated mse minimization to the class of exponential separable localizing functions

$$\varphi_\eta(x) := e^{-\eta \cdot x}, \quad \text{for } \eta \in (0, \infty)^d.$$

Let us define the integrated mse problem within the family of exponential localizing function :

$$J^{g,h}(\eta, \rho) := I^{g,h}(\varphi_\eta, \varphi_\rho) \quad \text{and} \quad w^{g,h} := \inf_{(\eta, \rho) \in (0, \infty)^{2d}} J^{g,h}(\eta, \rho). \quad (7.6)$$

Lemma 7.3 *There is a unique solution $(\hat{\eta}, \hat{\rho}) \in (0, \infty)^{2d}$ to the optimization problem (7.6), i.e.*

$$w^{g,h} := J^{g,h}(\hat{\eta}, \hat{\rho}) < J^{g,h}(\eta, \rho) \quad \text{for all } (\eta, \rho) \in \mathbb{R}_+^{2d} \text{ with } (\eta, \rho) \neq (\hat{\eta}, \hat{\rho}).$$

Proof. The map $(\eta, \rho) \mapsto J^{g,h}(\hat{\eta}, \hat{\rho})$ is strictly convex and lower semi-continuous. Then, the statement of the lemma holds under the following claim :

$$\text{cl} \left(\left\{ (\eta, \rho) \in (0, \infty)^{2d} : J^{g,h}(\eta, \rho) \leq C \right\} \right) \quad \text{is a compact subset of } (0, \infty)^{2d} \quad (7.7)$$

for all constant $C > w^{g,h}$.

It remains to prove (7.7). We fix some positive integer $i \leq d$ and we denote :

$$\begin{aligned} U_i(x, y) &:= H_{x^{(-i)}}(X_\theta^{(-i)}) H_y(X_{\theta+\delta}) S^g \circ S^h (\varphi_{(-i)}(X_\theta - x) \psi(X_{\theta+\delta} - y)), \\ V_i(x, y) &:= H_{x^{(-i)}}(X_\theta^{(-i)}) H_y(X_{\theta+\delta}) S^g \circ S_{(-i)}^h (\varphi_{(-i)}(X_\theta - x) \psi(X_{\theta+\delta} - y)). \end{aligned}$$

Observe that $U_i(x, y)$ and $V_i(x, y)$ do not depend on the x^i variable. Proceeding as in Step 2 of the proof of Lemma 7.2, we see that

$$\begin{aligned} J^{g,h}(\eta, \rho) &= \mathbb{E} \int_{\mathbb{R}^{2d-1}} \int_{-\infty}^{X_\theta^i} \{ \varphi_i(X_\theta^i - x^i) U_i(x, y) - \varphi_i'(X_\theta^i - x^i) V_i(x, y) \}^2 dx^i dy \\ &= \mathbb{E} \int_{\mathbb{R}^{2d-1}} \int_0^\infty \{ \varphi_i(\xi) U_i(x, y) - \varphi_i'(\xi) V_i(x, y) \}^2 d\xi dx^{(-i)} dy . \end{aligned}$$

Now since $\varphi(x) = e^{-\eta x}$, we have $\varphi_i'(x^i) = -\eta^i \varphi_i(x^i)$, and therefore

$$J^{g,h}(\eta, \rho) = \frac{1}{2\eta^i} \mathbb{E} \int_{\mathbb{R}^{2d-1}} \{ U_i(x, y) + \eta_i V_i(x, y) \}^2 dx^{(-i)} dy .$$

Now, set $B_n := [-n, n]$, and use the trivial inequality

$$1 = (2n)^{-1} \int_{B_n} dx^i \geq (2n)^{-1} \int_{B_n} H_{x^i}(X_\theta^i) dx^i .$$

This provides that :

$$\begin{aligned} J^{g,h}(\eta, \rho) &\geq \frac{1}{2\eta^i} \mathbb{E} \int_{B_n^{2d-1}} \{ U_i(x, y) + \eta_i(\xi) V_i(x, y) \}^2 dx^{(-i)} dy \\ &\geq \frac{1}{4n\eta^i} \mathbb{E} \int_{B_n^{2d}} \{ \bar{U}_i(x, y) + \eta_i(\xi) \bar{V}_i(x, y) \}^2 dx dy , \end{aligned}$$

where

$$\begin{aligned} \bar{U}_i(x, y) &= H_x(X_\theta) H_y(X_{\theta+\delta}) S^g \circ S^h (\varphi_{(-i)}(X_\theta - x) \psi(X_{\theta+\delta} - y)) , \\ \bar{V}_i(x, y) &:= H_x(X_\theta) H_y(X_{\theta+\delta}) S^g \circ S_{(-i)}^h (\varphi_{(-i)}(X_\theta - x) \psi(X_{\theta+\delta} - y)) . \end{aligned}$$

By Jensen's inequality, this provides :

$$J^{g,h}(\eta, \rho) \geq \frac{1}{2\eta^i (2n)^{-2d+1}} \left\{ \int_{B_n^{2d}} \mathbb{E} [\bar{U}_i(x, y) + \eta^i \bar{V}_i(x, y)] dx dy \right\}^2 . \quad (7.8)$$

Now observe that

$$\mathbb{E} [\bar{U}(x, y)] = \mathbb{E} [\varepsilon_x(X_\theta) \varepsilon_y(X_{\theta+\delta})] = f(x, y)$$

by Theorem 3.1, and

$$\mathbb{E} [\bar{V}(x, y)] = \mathbb{E} [H_{x^i}(X_\theta^i) \varepsilon_{x^{(-i)}}(X_\theta^{(-i)}) \varepsilon_y(X_{\theta+\delta})]$$

by Remark 3.2. In particular, $\mathbb{E} [\bar{U}(x, y)]$ and $\mathbb{E} [\bar{V}(x, y)]$ are positive and independent of the pair (η, ρ) . We then deduce from (7.8) that for all constant $C > 0$, there is a constant $c > 0$ such that :

$$J^{g,h}(\eta, \rho) \leq C \implies c^{-1} \leq \eta^i \leq c .$$

By following the same line of argument, we obtain the same result for the parameter ρ , completing the proof of (7.7). \square

We are now ready for the proof of Theorem 4.1.

Proof of Theorem 4.1 We first show that $v^{g,h} = w^{g,h}$ implying that the solution of the problem $w^{g,h}$ solves $v^{g,h}$. Then, we will show that this solution is unique and derive its characterization as in the statement of the theorem.

1. Let $(\varphi_n, \psi_n)_n \subset \mathcal{L}_+^s \times \mathcal{L}_+^s$ be a minimizing sequence of the problem $v^{g,h}$. Using repeatedly Lemma 7.2, we pass to an exponential minimizing sequence $(\varphi_{\eta_n}, \varphi_{\rho_n})$. This shows that $v^{g,h} \leq w^{g,h}$. The reverse inequality is trivial by definition of $w^{g,h}$.

2. The partial minimization problems (7.1)-(7.3) consist in minimizing a strictly convex functional on a convex domain. This proves that the only possible solution of the first order conditions for the integrated mse problem is of the exponential type, as in the definition of the problem $w^{g,h}$. We then deduce that the integrated mse problem has a unique solution characterized by the first order condition (7.2)-(7.4). The characterization reported in the statement of the theorem is easily obtained from Corollary 7.1, by observing that $\hat{\varphi}'(x^i) = -\hat{\eta}^i \hat{\varphi}(x^i)$ and $\hat{\psi}'(x^i) = -\hat{\rho}^i \hat{\psi}(x^i)$, and performing explicitly a partial integration with respect to each variable x^i and y^i . \square

References

- [1] V. Bally and G. Pagès (2002). A quantization algorithm for solving discrete time multi-dimensional optimal stopping problems, *Bernoulli* 9(6), 1003-1049, 2003.
- [2] V. Bally, G. Pagès and J. Printems (2002). A quantization method for pricing and hedging multi-dimensional American style options, *Mathematical Finance*, 15(1), 119-168, 2005.
- [3] D. Bosq (1998). *Non-parametric Statistics for Stochastic Processes*, Springer Verlag, New York.
- [4] B. Bouchard (2000). *Contrôle Stochastique Appliqué à la Finance*, PhD thesis, Université Paris Dauphine.
- [5] B. Bouchard, I. Ekeland and N. Touzi (2004). On the Malliavin approach to Monte Carlo approximation of conditional expectations, *Finance and Stochastics*, 8, 45-71.
- [6] B. Bouchard and N. Touzi (2004). Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations, *Stochastic Processes and their Applications*, 111, 175-206.

- [7] M. Broadie and P. Glasserman (1997). Pricing American-Style Securities Using Simulation, *Journal of Economic Dynamics and Control*, **21**, n8-9, 1323-1352.
- [8] J. F. Carrière (1996). Valuation of the Early-Exercise Price for Options using Simulations and Nonparametric Regression, *Insurance : mathematics and Economics*, **19**, 19-30.
- [9] P. Cohort (2000). Sur quelques problèmes de quantification, PhD thesis, Université Paris VI.
- [10] E. Fournié, J.-M. Lasry, J. Lebuchoux and P.-L. Lions (2001). Applications of Malliavin calculus to Monte Carlo methods in finance II, *Finance and Stochastics* **5**, 201-236.
- [11] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions and N. Touzi (1999). Applications of Malliavin calculus to Monte Carlo methods in finance, *Finance and Stochastics* **3**, 391-412.
- [12] A. Kohatsu-Higa and R. Pettersson (2002). Variance reduction methods for simulation of densities on Wiener space, *SIAM Journal on Numerical Analysis* **40**, 431-450.
- [13] P.-L. Lions and H. Regnier (2001). Calcul du prix et des sensibilités d'une option américaine par une méthode de Monte Carlo, preprint.
- [14] F. A. Longstaff and R. S. Schwartz (2001). Valuing American Options By Simulation : A simple Least-Square Approach, *Review of Financial Studies* **14**, 113-147.
- [15] D. Nualart (1995). *The Malliavin Calculus and Related Topics* Springer Verlag, Berlin.
- [16] J.N. Tsitsiklis and B. Van Roy (1999). Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives, *IEEE Trans. Automat. Control*, **44**, n10, 1840-1851.