HAL
open science

Pricing and hedging American options by Monte Carlo methods using a Malliavin calculus approach

Vlad Bally, Lucia Caramellino, Antonino Zanette

To cite this version:

Vlad Bally, Lucia Caramellino, Antonino Zanette. Pricing and hedging American options by Monte Carlo methods using a Malliavin calculus approach. [Research Report] RR-4804, INRIA. 2003. inria-00071782

HAL Id: inria-00071782

https://inria.hal.science/inria-00071782v1

Submitted on 23 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Pricing and hedging American options by Monte Carlo methods using a Malliavin calculus approach

Vlad Bally — Lucia Caramellino — Antonino Zanette

N° 4804

April 2003

THÈME 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4804--FR+ENG

Pricing and hedging American options by Monte Carlo methods using a Malliavin calculus approach

Vlad Bally* , Lucia Caramellino† , Antonino Zanette‡

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Mathfi

Rapport de recherche n° 4804 — April 2003 — 46 pages

Abstract: Following the pioneering papers of Fournié, Lasry, Lebouchoux, Lions and Touzi (see [12] and [13]), an important work concerning the applications of the Malliavin calculus in numerical methods for mathematical finance has come after. One is concerned with two problems: computation of a large number of conditional expectations on one hand and computation of Greeks (sensitivities) on the other hand. A significant test of the power of this approach is given by its application to pricing and hedging American options.
Our paper gives a global and simplified presentation of this topic including the reduction of variance techniques based on localization and control variables. A special interest is given to practical implementation, number of numerical tests are presented and their performances are carefully discussed.

Key-words: American options, Malliavin calculus, COnditional expectations, Sensitivities, Monte Carlo, Reduction of variance, Localization

Acknowledgments. For the kind hospitality at the CERMICS-ENPC and the INRIA, the authors wish to thank Bernard Lapeyre, Agnes Sulem and all the equipe of the PREMIA project (`http://cermics.enpc.fr/~premia`), which partially supported this research.
The authors are also grateful to Bruno Bouchard and Hervé Regnier for useful and interesting discussions.

The work of Antonino Zanette was partially supported by the research project "Progetto di Ricerca Modelli matematici innovativi per lo studio dei rischi finanziari e assicurativi"(Regione Autonoma Friuli-Venezia Giulia, Legge Regionale 3/1998).

* Equipe Statistiques et Processus, Université du Maine, and INRIA Rocquencourt, Project Mathfi, Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex, France
† Dipartimento di Matematica, Università di Roma *Tor Vergata*, via della Ricerca Scientifica, I-00133 Roma, Italy
‡ Dipartimento di Finanza dell'Impresa e dei Mercati Finanziari, Università di Udine, Via Tomadini 30/A, I-33100 Udine, Italy

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 —Télécopie : +33 1 39 63 53 30

Calcul du prix et couverture des options américaines par méthodes de Monte Carlo et calcul de Malliavin

Résumé : A la suite des premiers articles de Fournié, Lasry, Lebouchoux, Lions et Touzi (voir [12] et [13]), un travail considérable sur les applications du calcul de Malliavin pour les méthodes numériques en finance a été développé. Deux problèmes se posent en particulier : le calcul d'espérances conditionnelles d'une part, les sensitivités et le calcul de la couverture de l'autre. Un test significatif de la puissance de cette approche est donné par l'application au calcul de prix d'options américaines.
Notre papier donne une présentation globale et simplifiée, avec l'utilisation de techniques de réduction de la variance, basée sur la localisation et l'usage de variables de contrôle. Un intérêt particulier est donné à l'implémentation concrète ; plusieurs tests numériques ont été réalisés et discutés.

Mots-clés : Options américaines, Calcul de Malliavin, Espérances conditionnelles, Sensitivités, Monte Carlo, Réduction de variance, Localisation

1 Introduction

The theory of Markov Processes and the Stochastic Calculus have provided a probabilistic interpretation for the solutions of linear partial differential equations (shortly, linear PDE's) by means of the Feynman-Kac formula. One of the most striking applications is the emergence of the Monte Carlo method as an alternative to deterministic numerical algorithms for solving PDE's. This method is slower than the analytical ones, but as the dimension increases (more than 3), it is well known that the analytical methods do no more work and so the Monte Carlo method remains the only alternative.
But one has to notice that both the Feynman-Kac formula and the Monte Carlo method are specific to linear problems and collapse when dealing in the non linear case. In the last decay much work has been done in order to extend the probabilistic methods in a non linear frame. On one hand, in a set of papers Pardoux and Peng (see [20], [21]) introduced the Backward Stochastic Differential Equations (BSDE's in short), which generalize the probabilistic representation given by the Feynman-Kac formula to non linear problems. In [14] (see also [1]) this representation was obtained for obstacle problems as well (by means of reflected BSDE's). Applications to mathematical finance have been discussed in [15]. So the theoretical background has been prepared. The second step is to obtain efficient probabilistic algorithms for solving such problems. Recall that one specific point in the Monte Carlo method is that it does not employ grids. More precisely, in order to compute the solution $u(0, x_0)$ of the linear PDE

$$\begin{aligned}&(\partial_t + L)u(t,x) = 0, \quad (t,x) \in [0,T] \times \mathbb{R}^d\\ &u(T,x) = f(x),\end{aligned}$$

one represents $u(0, x_0)$ as an expectation and employes a sample of the underlying diffusion in order to compute this expectation. On the contrary, an analytical method constructs a time-space grid and employs a dynamical programing algorithm in order to compute the solution. This means that in order to compute the solution in one point $(0, x_0)$, the analytical method is obliged to compute it on a whole grid (t_k, x_k^i) while the Monte Carlo method provides directly the solution in $(0, x_0)$. This is a big advantage because in large dimension grids become difficult to handle.

Think now that instead of the above linear problem, one has to solve a PDE of the type

$$(\partial_t + L)u(t,x) + f(t,x,u(t,x)) = 0.$$

Even if a probabilistic representation of u is available (and this is the case), one is obliged to compute the solution on a whole grid for the simple reason that one has to know the value of $u(t,x)$ which comes on in $f(t,x,u(t,x))$. Therefore, the dynamical programing algorithm becomes essential. At this stage, a probabilistic approach is in the same position as an analytical one: one has to construct a grid, then to solve at each time step a linear problem and finally to add the input $f(t,x,u(t,x))$. A terminology difference however: the probabilist would say that he computes a conditional expectation instead of saying that he solves a linear PDE. But the problem remains the same: computing a large number of conditional expectations, and this is not trivial.

The question is now if a probabilistic point of view provides a new approach to this kind of problems. And the answer is affirmative. These last years, essentially motivated by mathematical finance problems - and the most striking one is pricing American options in large dimension - several new ideas appeared in this field. Roughly speaking they may be divided in three families. In the first one, a tree is built up in order to obtain a discretization of the underlying diffusion on a grid. This family includes Broadie and Glasserman algorithm (see [10]), the quantization algorithm (see [2], [3]) as well as Chevance's algorithm (see [11]). The second idea is to use regression on a truncated basis of L^2 in order to compute the conditional expectations. This is done in Longstaff

and Schwartz [18] and in Tsisiklis and Van Roy [22]. Finally in [12], [13], Fournié *et al.* obtain a representation of the conditional expectation using Malliavin's calculus and then employ this representation in order to perform a Monte Carlo method. The specificity of this last approach is that it appears as a pure Monte Carlo method despite the non linearity.
Except for solving PDE's (which amounts to pricing an option in the financial frame), a second problem of interest is to compute the sensitivity of the solution to some parameter (hedging and Greeks, in financial language). It seems that Malliavin calculus is an especially promising tool for solving such a problem. It has been used by Lions and Reigner [17] which follow the third method, as well as in [4] where quantization algorithm is employed.

The present paper deals with the last method based on Malliavin calculus. Although this approach works for a large class of non linear problems, we focus on the American option pricing, which amounts to solve obstacle problems for PDE's. This seems to be a significant test on the efficiency of the method and several papers related to this subject appeared in the last time. Our aim is to give a general view on this topic (with simplified proofs), to present concrete examples and to discuss the numerical results. A special interest is given to the reduction of variance techniques based on localization and control variables and to their impact in concrete simulation.

It worth to mention that as long as one restricts himself to the Black Scholes model (which of course is of interest in mathematical finance) the computations related to Malliavin calculus are rather explicit and elementary. This permits us to give here an approach which is self contained and accessible to readers who are not familiar with the rather heavy machinery of Mallavin calculus. In the case of a general diffusion the same reasonings work as well, but of course one has to employ the Malliavin calculus in all the generality. And consequently, the explicit formulas which are available in the Balck Scholes model become heavier (in particular, they involve the inverse of the Malliavin covariance matrix). The presentation that we give here is based on one more remark: the results obtained in the one dimensional case extend easily (using an elementary geometrical argument) to the multidimensional case. This significantly simplifies the proofs.

The paper is organized as follows. In Section 2 we present the problem and in Section 3 we give the results (representation formulas for conditional expectation and for the strategy) in the one dimensional case. This section also contains the Malliavin calculus approach in the elementary frame which is the ours. In Section 4 we present the multidimensional case and in Section 5 we discuss optimal localization. In Section 6 we give the algorithm based on this approach and we discus the numerical tests.

2 Presentation of the problem

An American option with maturity T, is an option whose holder can exercise his right of option in any time up to T. Let X_s denote the underlying asset price process, here modeled as a diffusion process:

$$\begin{aligned} dX_t &= b(X_t)dt + \sigma(X_t)dW_t \\ X_0 &= x \end{aligned} \tag{1}$$

where b and σ denote a vector and a matrix field on $\mathbb{R}^d$ and W is a d-dimensional Brownian motion. Let $\Phi(X_s)$ denote the cash-flow associated with the option. The price as seen at time t of such an American option is given by

$$P(t, X_t) = \sup_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}\Big(e^{-\int_t^\theta r_s ds} \Phi(X_\theta) \,\Big|\, \mathcal{F}_t \Big)$$

where $\mathcal{T}_{t,T}$ stands for the set of all the stopping times taking values on $[t, T]$ and r denotes the spot rate process, which is here supposed to be deterministic.

The solution of this optimal stopping problem has been provided by using the theory of the Snell envelopes: the first optimal stopping time is given by

$$\theta_t^* = \inf\{s \in [t,T]\,;\, P(s,X_s) = \Phi(X_s)\}.$$

Moreover, the function $P(t,x)$, giving the price of the option, can be characterized as follows. We define the *stopping region*, also called *exercise region*, as

$$\mathcal{E}_s = \{(t,x) \in [0,T] \times \mathbb{R}_+ \,:\, P(t,x) = \Phi(x)\}$$

and the *continuation region* as its complement, that is

$$\mathcal{E}_c = \{(t,x) \in [0,T] \times \mathbb{R}_+ \,:\, P(t,x) > \Phi(x)\}.$$

Roughly speaking, by using the Ito's Lemma, one can see that $P(t,x)$ solves the following partial differential equation: setting $\mathcal{L}$ as the infinitesimal generator of X, that is

$$\mathcal{L}g(x) = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\frac{\partial^2 g}{\partial x_i \partial x_j}(x) + \sum_{i=1}^{d} b_i(x)\frac{\partial g}{\partial x_i}(x)$$

being $a = \sigma\sigma^*$, then

$$\frac{\partial P}{\partial t}(t,x) + \mathcal{L}P(t,x) - r(t)P(t,x) = 0$$

whenever $P(t,x) > \Phi(x)$, with the final condition $P(T,x) = \Phi(x)$. A rigorous reasoning turns out to be much more difficult because generally we have not sufficient regularity for P (that is, C^1 in t and C^2 in x) on one hand and on the other one, the behavior of the solution P in a neighborhood of the free boundary $\{(t,x)\,;\, P(t,x) = \Phi(x)\}$ is rather delicate to describe (it gives a supplementary term in the PDE). This leads to a weak formulation of the problem. The first one was given in terms of variational inequalities by Bensoussan and Lions [6]; in El Karoui *et al.* [14], one can find a formulation in viscosity sense and in [1], Bally *et al.* give a formulation in Sobolev sense.
Thus, the problem of the pricing of an American option is a strongly nonlinear problem, and there is non hope to find closed formulas. In order to numerically compute the price of an American option, one can use either a deterministic or a stochastic algorithm.
Concerning the deterministic methods, they can be based on finite elements methods for variational inequalities. But here, we are interested in a stochastic approach, that is in a Monte Carlo algorithm for the pricing of American options. It turns out from a dynamic programming principle, but let us firstly start by formalizing better the framework.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered probability space where a d-dimensional Browinan motion W is defined and set $\mathcal{F}_t = \sigma(W_s \,:\, s \le t)$. Let X_t denote the underlying asset prices, which evolve according to (1). The price at time t of an associated American option with maturity T and payoff function $\Phi : \mathbb{R}_+^d \to \mathbb{R}$ is then

$$P(t,x) = \sup_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}_{t,x}\Big(e^{-r(\theta-t)}\Phi(X_\theta)\Big) \tag{2}$$

where we have supposed that the spot rate is constant. In order to numerically evaluate $P(0,x)$, that is the price as seen at time 0, it is possible to set up a Bellman dynamic programming principle. Indeed, let $0 = t_0 < t_1 < \ldots < t_n = T$ be a discretization of the time interval $[0,T]$, with step size equal to $\Delta t = T/n$, and let $(\bar{X}_{k\Delta t})_{k=0,1,\ldots,n}$ an approximation of $(X_t)_{t\in[0,T]}$, that is $\bar{X}_{k\Delta t} \simeq X_{k\Delta t}$. The price $P(k\Delta t, \bar{X}_{k\Delta t})$ can be approximated by means of the quantity $\bar{P}_{k\Delta t}(\bar{X}_{k\Delta t})$, given by the following recurrence equality:

Theorem 2.1 *Given $\Delta t = T/n \in (0,1)$, define $\bar{P}_{n\Delta t}(\bar{X}_{k\Delta t}) = \Phi(\bar{X}_{n\Delta t})$ and for any $k = n-1, n-2, \dots, 1, 0$,*

$$\bar{P}_{k\Delta t}(\bar{X}_{k\Delta t}) = \max\Big(\Phi(\bar{X}_{k\Delta t})\,,\, e^{-r\Delta t}\,\mathbb{E}\Big(\bar{P}_{(k+1)\Delta t}(\bar{X}_{(k+1)\Delta t})\,\Big|\,\bar{X}_{k\Delta t}\Big)\Big).$$

Then $\bar{P}_{k\Delta t}(\bar{X}_{k\Delta t}) \simeq P(k\Delta t, X_{k\Delta t})$.

The above statement is heuristic and a rigorous formulation supposes to precise the hypothesis on the diffusion coefficients and on the regularity of the obstacle. This is done by Bally and Pagés [3]. Roughly speaking, the error is of order $\sqrt{\Delta t}$ if one has few regularity and of order Δt if more regularity holds.
As a consequence, one can numerically evaluate the delta $\Delta(t,x)$ of an American option, that is the derivative of the price with respect to the initial value of the underlying asset price: $\Delta(t,x) = \partial_x P(t,x)$. Recall that this is important since it gives the sensibility of the price with respect to the initial underlying asset price and also for the hedging of the option. By considering the case $t = 0$, as in Theorem 2.1, then the following approximation $\bar{\Delta}_0(x)$ of $\Delta(0,x)$ can be stated.

Proposition 2.2 *For any $\Delta t = T/n \in (0,1)$, set*

$$\Gamma_{\Delta t} = \{\alpha \in \mathbb{R}^d\,;\, \bar{P}_{\Delta t}(\alpha) < \Phi(\alpha)\},$$

where $\bar{P}_{\Delta t}(\alpha)$ is defined as in Theorem 2.1, that is

$$\bar{P}_{\Delta t}(\alpha) = \max\Big(\Phi(\alpha)\,,\, e^{-r\Delta t}\,\mathbb{E}\Big(\bar{P}_{2\Delta t}(\bar{X}_{2\Delta t})\,\Big|\,\bar{X}_{\Delta t} = \alpha\Big)\Big).$$

Then, by setting

$$\begin{aligned}\bar{\Delta}(\alpha) &= \partial_\alpha\,\Phi(\alpha)\,\mathbf{1}_{\Gamma_{\Delta t}} + e^{-r\Delta t}\partial_\alpha\,\mathbb{E}\Big(\bar{P}_{(k+1)\Delta t}(\bar{X}_{(k+1)\Delta t})\,\Big|\,\bar{X}_{k\Delta t} = \alpha\Big)\,\mathbf{1}_{\Gamma^c_{\Delta t}} \quad \textit{and}\\ \Delta_0(x) &= \mathbb{E}_x\Big(\bar{\Delta}(\bar{X}_{\Delta t})\Big)\end{aligned}$$

where ∂_α denotes the gradient, one has $\Delta(0,x) \simeq \bar{\Delta}_0(x)$

Such an assertion is heuristic and a rigorous statement, including error bound, turns out to be a more difficult problem. One may find in Bally *et al.* [4] such a bound but it is given in a weaker sense (in $L^2([0,T], dt)$).
Such results state that in order to numerically compute the price $P(0,x)$ and its delta $\Delta(0,x)$, it is sufficient to approximate a family of conditional expectations and of their derivatives, thus allowing one to set up Monte Carlo simulations.

Existing Monte Carlo methods applied to this context, consist in the numerical evaluation of the conditional expectations by means of a stratification of the path space for the approximation of the transition density of the process X_t, as the quantization algorithm by Bally, Pagés and Printems [4], the algorithm by Broadie and Glassermann [10] or the one by Barraquand and Martineau [5]. Another Monte Carlo approach makes use of regression methods to perform the approximation of the conditional expectation, as made by Longstaff and Schwartz [18] or by Tsitsiklis and VanRoy [22]

Also in order to overcome the problem of the discretization of the path space, another method can be used. It has been introduced by Lions and Regnier [17] and uses formulas allowing to represent conditional expectations like $\mathbb{E}(F(X_t)\,|\,X_s = \alpha)$ and its derivative $\partial_\alpha\mathbb{E}(F(X_t)\,|\,X_s = \alpha)$ written in

terms of a suitable ratio of non-conditioned expectations, that is

$$\begin{aligned} \mathbb{E}\Big(F(X_t)\,\Big|\,X_s=\alpha\Big) &= \frac{\mathbb{E}(F(X_t)\,\pi_s^\alpha)}{\mathbb{E}(\pi_s^\alpha)} \\ \partial_\alpha\mathbb{E}\Big(F(X_t)\,\Big|\,X_s=\alpha\Big) &= \frac{\mathbb{E}(F(X_t)\,\pi_s^{1,\alpha})\mathbb{E}(\pi_s^\alpha)-\mathbb{E}(F(X_t)\,\pi_s^\alpha)\mathbb{E}(\pi_s^{1,\alpha})}{\mathbb{E}(\pi_s^\alpha)^2} \end{aligned} \tag{3}$$

being π_s^α and $\pi_s^{1,\alpha}$ suitable weights, which could also depend on suitable localizing functions. Such representations can be proved by using Malliavin calculus techniques. The proofs providing the formulas as in (3) can be found in Section 3 and 4, in the framework of the Black and Scholes model; the original reference is the paper by Lions and Regnier [17] but here we develop a little different approach, mainly for the multidimensional case, allowing to simplify considerably the proofs. Moreover, in Section 5 we discuss an optimality criterium for the choice of the localizing functions, which play an important role for practical purposes, as already studied and observed by Kohatsu-Higa and Petterson [16] and by Bouchard, Ekeland and Touzi [7].
Let us point out that we could employ the "strong" Malliavin calculus for our proofs (that is, Malliavin derivatives, Skorohod integrals etc. - words which do not appear in this paper) but we have opted for a more "friendly" approach, in order to make it readable and comprehensible also to people who are not familiar with the topic.

Representations (3) can be used for the practical purpose of the pricing of American options as follows. In fact, since the weights π_s^α and $\pi_s^{1,\alpha}$ can be written explicitly, expectations like $\mathbb{E}(f(X_t)\,\pi_s^\alpha)$ or $\mathbb{E}(f(X_t)\,\pi_s^{1,\alpha})$ can be approximated through the associated empirical means and used to numerically compute the price $P(0,x)$ and its delta $\Delta(0,x)$ by using Theorem (2.1) and Proposition 2.2, thus avoiding the problem of the approximation of the transition density and of the discretization of the path space. This plan gives also the considerable gain to provide a Monte Carlo algorithm for the evaluation of $P(0,x)$ and $\Delta(0,x)$ which makes use of only one set of simulated trajectories for the computation of any conditional expectation appearing in Theorem 2.1 and in Proposition 2.2. Let us remark that, using this approach, the valuation of the delta is not made through finite difference approximations but it is performed by means of representation formulas written in terms of expectations. We postpone to Section 6 (cfr. §6.1 and §6.2) a comprehensive presentation of such an algorithm. Finally, some numerical results can be found in §6.3.

Finally, here we consider price and delta, then representation formulas for the conditional expectation and its gradient. It is worth to point out that one could also attack the problem of the numerical evaluation of other greeks. Indeed, similar techniques could be applied in order to obtain representation formulas for other derivatives of the conditional expectation: the second derivative (for the gamma), the derivative w.r.t. the spot rate r (for the theta), as well as the derivative w.r.t. the volatility σ (for the vega).

3 Representation formulas for the conditional expectation and its gradient: the one dimensional case

Let X be a single underlying asset price process, driven by the Black and Scholes model, i.e. it solves the following stochastic differential equation (sde)

$$\begin{aligned} &dX_t = (r-\eta)X_t dt + \sigma X_t dW_t \\ &X_0 = x \end{aligned}$$

where: $x \in \mathbb{R}_+$; $r, \eta \in \mathbb{R}$, being r the (constant) spot rate and η the dividend of the option; $\sigma > 0$ denotes the volatility; W is a 1-dimensional Brownian motion. Thus,

$$X_t = x \exp\Big(ht + \sigma W_t\Big), \qquad \text{where } h = r - \eta - \frac{1}{2}\sigma^2. \tag{4}$$

The aim is to study the conditional expectation

$$\mathbb{E}(\Phi(X_t) \,|\, X_s = \alpha)$$

where $0 < s < t$, $\alpha \in \mathbb{R}_+$ and Φ is a function with polynomial growth, that is belonging to

$$\mathcal{E}_b(\mathbb{R}) = \{f \in \mathcal{M}(\mathbb{R}) : \text{ there exist } C > 0 \text{ and } m \in \mathbb{N} \text{ such that } f(y) \le C(1 + |y|^m)\}$$

where $\mathcal{M}(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} : f \text{ is measurable}\}$. In such a case one can state the following result (Lions and Regnier [17], Lemme 2.1.1):

Theorem 3.1 (Representation formulas I: without localization)

i) For any $0 \le s < t$, $\Phi \in \mathcal{E}_b$ and $\alpha > 0$, one has

$$\mathbb{E}\Big(\Phi(X_t) \,\Big|\, X_s = \alpha\Big) = \frac{\mathbb{T}_{s,t}[\Phi](\alpha)}{\mathbb{T}_{s,t}[1](\alpha)}$$

where

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E}\Big(f(X_t)\frac{H(X_s - \alpha)}{\sigma s(t-s)X_s}\,\Delta W_{s,t}\Big) \tag{5}$$

being $H(\xi) = \mathbf{1}_{\xi \ge 0}$, $\xi \in \mathbb{R}$, and

$$\Delta W_{s,t} = (t-s)(W_s + \sigma s) - s(W_t - W_s).$$

ii) For any $0 \le s < t$, $\Phi \in \mathcal{E}_b$ and $\alpha > 0$, one has

$$\partial_\alpha \mathbb{E}\Big(\Phi(X_t) \,\Big|\, X_s = \alpha\Big) = \frac{\mathbb{R}_{s,t}[\Phi](\alpha)\mathbb{T}_{s,t}[1](\alpha) - \mathbb{T}_{s,t}[\Phi](\alpha)\mathbb{R}_{s,t}[1](\alpha)}{\mathbb{T}_{s,t}[1](\alpha)^2}$$

where $\mathbb{T}_{s,t}[f]$ is defined above and

$$\mathbb{R}_{s,t}[f](\alpha) = -\mathbb{E}\Big(f(X_t)\frac{H(X_s - \alpha)}{\sigma s(t-s)X_s^2}\Big[\frac{(\Delta W_{s,t})^2}{\sigma s(t-s)} + \Delta W_{s,t} - \frac{t}{\sigma}\Big]\Big). \tag{6}$$

Let us point out that, for any fixed function f, the operator $\mathbb{R}_{s,t}[f](\alpha)$ is simply the derivative of $\mathbb{T}_{s,t}[f](\alpha)$: $\mathbb{R}_{s,t}[f](\alpha) = \partial_\alpha \mathbb{T}_{s,t}[f](\alpha)$.
In the proof of Theorem 3.1 it is shown the existence of a process $\pi_{s,t} \in L^2(\Omega)$ such that, roughly speaking,

$$\text{“ } \mathbb{E}\Big(\Phi(X_t)\delta_0(X_s - \alpha)\Big) \text{ ”} = \mathbb{E}(\Phi(X_t)H(X_s - \alpha)\pi_{s,t})$$

where δ_0 stands for the Dirac measure in 0 (notice that H is actually the distribution function associated with δ_0). Obviously, the Dirac mass is something “very irregular". In some sense, the Malliavin integration by parts formula allows to regularize it, thus to overcome the problem of handling a Dirac mass, but it is worth to point out that this procedure provides an high variance because of the presence of the Heaviside function H. Thus, it is very useful to give a localization method for the computation of conditional expectations, as made in the following Theorem 3.3 (Lions and Regnier [17], Lemme 2.1.1) by introducing the localizing function ψ:

Lemma 3.2 *Let $\psi : \mathbb{R} \to [0, +\infty)$ be such that $\int_{\mathbb{R}} \psi(\xi) d\xi = 1$. Then formulas (5) and (6) hold true with $\mathbb{T}_{s,t}$ and $\mathbb{R}_{s,t}$ replaced by $\mathbb{T}_{s,t} = \mathbb{T}^{\psi}_{s,t}$ and $\mathbb{R}_{s,t} = \mathbb{R}^{\psi}_{s,t}$ respectively, where*

$$\mathbb{T}^{\psi}_{s,t}[f](\alpha) = \mathbb{E}(f(X_t)\psi(X_s - \alpha)) + \mathbb{E}\Big(f(X_t)\frac{H(X_s - \alpha) - \Psi(X_s - \alpha)}{\sigma s(t-s)X_s}\Delta W_{s,t}\Big) \tag{7}$$

and

$$\begin{aligned}\mathbb{R}^{\psi}_{s,t}[f](\alpha) = \quad & -\mathbb{E}\Big(f(X_t)\psi(X_s - \alpha)\frac{\Delta W_{s,t}}{\sigma s(t-s)X_s}\Big) + \\ & -\mathbb{E}\Big(f(X_t)\frac{H(X_s - \alpha) - \Psi(X_s - \alpha)}{\sigma s(t-s)X_s^2}\Big[\frac{(\Delta W_{s,t})^2}{\sigma s(t-s)} + \Delta W_{s,t} - \frac{t}{\sigma}\Big]\Big).\end{aligned} \tag{8}$$

where Ψ denotes the probability distribution function associated with ψ: $\Psi(y) = \int_{-\infty}^{y} \psi(\xi) d\xi$.

By using the localized version for the operators, we can immediately set up localized representation formulas:

Theorem 3.3 (Representation formulas II: with localization) *For any $0 \le s < t$, $F \in \mathcal{E}_b$, $\alpha > 0$ and for any $\psi : \mathbb{R} \to [0, +\infty)$, such that $\int_{\mathbb{R}} \psi_i(\xi) d\xi = 1$, one has*

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

and

$$\partial_\alpha \mathbb{E}\Big(\Phi(X_t)\,\Big|\, X_s = \alpha\Big) = \frac{\mathbb{R}^{\psi}_{s,t}[\Phi](\alpha)\mathbb{T}^{\psi}_{s,t}[1](\alpha) - \mathbb{T}^{\psi}_{s,t}[\Phi](\alpha)\mathbb{R}^{\psi}_{s,t}[1](\alpha)}{\mathbb{T}^{\psi}_{s,t}[1](\alpha)^2},$$

where the operators $\mathbb{T}^{\psi}_{s,t}[f](\alpha)$ and $\mathbb{R}^{\psi}_{s,t}[f](\alpha)$ are defined in (7) and (8) respectively.

We postpone to Section 5 a discussion on the choice of the localizing function ψ.
We give now another localization result due to Kohatsu-Higa and Pettersson [16] in the one-dimensionale case and to Bouchard, Ekeland and Touzi [7] in the multidimensional setting.

Theorem 3.4 (Representation formulas III) *Let $0 \le s \le t$, $\Phi \in \mathcal{E}_b$ and let $\alpha > 0$ be fixed. For any $\tilde{\psi} \in C^1_b(\mathbb{R})$ such that $\tilde{\psi}(\alpha) = 1$ then*

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

where

$$\mathbb{J}^{\tilde{\psi}}_{s,t}[f](\alpha) = \frac{1}{\sigma s(t-s)} \cdot \mathbb{E}\Big(f(X_t)\frac{H(X_s - \alpha)}{X_s}\Big[\tilde{\psi}(X_s)\Delta W_{s,t} + \tilde{\psi}'(X_s)\, X_s\, \sigma\, s(t-s)\Big]\Big)$$

We shall give here an elementary approach to the computation of conditional expectations allowing to prove Theorems 3.1, 3.3 and 3.4. We start next section by proving some results on Malliavin Calculus allowing to prove the above Theorems, whose proofs can be found in the section after the next one.

3.1 Some preliminary results in Malliavin Calculus

The representation formulas stated in the previous section are based on Malliavin Calculus techniques. In the frame of the Black and Scholes model (on which we focus in this paper), these techniques reduce to elementary reasonings based on (standard) integration by parts and so are available to readers who are not familiar with this theory. Therefore, we present in this section the elementary counterpart of Malliavin's integration by parts formula and the way in which one employs this formula for computing conditional expectations. In the case of a general diffusion, the strategy is the same and people familiar with this calculus may easily adapt the arguments.

Let us start by a simple definition.

Definition 3.5 Let F and G denote square integrable random variables. We say that the $IP(F,G)$ (Integration by Parts) property holds if there exists a square integrable weight $\pi_F(G)$ such that

$$\mathbb{E}(\phi'(F)\,G) = \mathbb{E}(\phi(F)\,\pi_F(G)), \qquad \text{for any } \phi \in C_b^\infty(\mathbb{R}) \tag{9}$$

(here, as usual, $C_b^\infty(\mathbb{R})$ stands for the set of bounded and infinitely differentiable paths).

One has

Lemma 3.6 *(i) Suppose that $IP(F,1)$ holds. Then the law of the random variable F has a probability density p given by*

$$p(\alpha) = \mathbb{E}(H(F-\alpha)\,\pi_F(1)).$$

(ii) Suppose that both $IP(F,1)$ and $IP(F,G)$ hold. Then

$$\mathbb{E}(G\,|\,F=\alpha) = \frac{\mathbb{E}(H(F-\alpha)\,\pi_F(G))}{\mathbb{E}(H(F-\alpha)\,\pi_F(1))},$$

with the convention that the above quantity is equal to 0 whenever $\mathbb{E}(H(F-\alpha)\,\pi_F(1)) = 0$.

Proof.
(i) We employ regularization functions constructed as follows. Take a smooth, symmetric, non-negative function ϕ with support contained in $[-1,1]$ and such that $\int_{\mathbb{R}} \phi(t)\,dt = 1$. Then consider $\phi_\delta(x) = \phi(x/\delta)/\delta$ and $\Phi_\delta(x) = \int_{-\infty}^x \phi_\delta(t)\,dt$. Note that $\Phi'_\delta = \phi_\delta$ and also that $\lim_{\delta\to 0}\Phi_\delta(x) = \widetilde{\mathbf{1}}_{[0,+\infty)}(x) = \mathbf{1}_{(0,+\infty)}(x) + \delta_0(x)/2$. Take now G_δ as a random variable, independent of F, whose law is $\phi_\delta(x)\,dx$. Clearly $\lim_{\delta\to 0}\mathbb{E}(f(F-G_\delta)) = \mathbb{E}(f(F))$, for any continuous function f. Then we can write

$$\mathbb{E}(f(F-G_\delta)) = \int_{\mathbb{R}}\int_{\mathbb{R}} f(u-v)\,\phi_\delta(v)\,dv\,d\mathbb{P}\circ F^{-1}(u)$$
$$= \int_{\mathbb{R}} f(z)\int_{\mathbb{R}} \phi_\delta(u-z)\,d\mathbb{P}\circ F^{-1}(u)dz = \int_{\mathbb{R}} f(z)\mathbb{E}(\phi_\delta(F-z))\,dz.$$

Since the $IP(F,1)$ property holds, we have

$$\mathbb{E}(\phi_\delta(F-z)) = \mathbb{E}(\Phi'_\delta(F-z)) = \mathbb{E}(\Phi_\delta(F-z)\,\pi_F(1)),$$

so that by using Lebesgue's Theorem we obtain

$$\lim_{\delta\to 0}\mathbb{E}(f(F-G_\delta)) = \int_{\mathbb{R}} f(z)\mathbb{E}(\widetilde{\mathbf{1}}_{[0,+\infty)}(F-z)\,\pi_F(1))\,dz.$$

Thus, for any continuous function f we have

$$\mathbb{E}(f(F)) = \int_{\mathbb{R}} f(z)\mathbb{E}(\widetilde{\mathbf{1}}_{[0,+\infty)}(F-z)\,\pi_F(1))\,dz,$$

which allows to state that the law of F has a probability density given by $p(z) = \mathbb{E}(\widetilde{\mathbf{1}}_{[0,+\infty)}(F-z)\,\pi_F(1)) = \mathbb{E}(\mathbf{1}_{[z,+\infty)}(F)\,\pi_F(1)) = \mathbb{E}(H(F-\alpha)\,\pi_F(1))$ (we can replace $\widetilde{\mathbf{1}}_{[0,+\infty)}(F-z)$ with $\mathbf{1}_{[0,+\infty)}(F-z)$ because we already know that the law of F is absolutely continuous).

(ii) Let us denote

$$\theta(\alpha) = \frac{\mathbb{E}(H(F-\alpha)\,\pi_F(G))}{\mathbb{E}(H(F-\alpha)\,\pi_F(1))}.$$

We only have to check that for any bounded and continuous function f it holds that $\mathbb{E}(f(F)\,G) = \mathbb{E}(f(F)\,\theta(F))$. By using again the regularization function introduced above, the validity of the $IP(F,G)$ property and what proved in part (i) allow to state that

$$\begin{aligned}
\mathbb{E}(f(F)\,G) &= \mathbb{E}\Big(G\,\lim_{\delta\to 0}\int_{\mathbb{R}} f(z)\,\phi_\delta(F-z)\,dz\Big) = \lim_{\delta\to 0}\int_{\mathbb{R}} f(z)\,\mathbb{E}(G\,\phi_\delta(F-z))\,dz \\
&= \lim_{\delta\to 0}\int_{\mathbb{R}} f(z)\,\mathbb{E}(G\,\Phi'_\delta(F-z))\,dz = \lim_{\delta\to 0}\int_{\mathbb{R}} f(z)\,\mathbb{E}(\Phi_\delta(F-z)\,\pi_F(G))\,dz \\
&= \int_{\mathbb{R}} f(z)\,\mathbb{E}(\mathbf{1}_{[0,+\infty)}(F-z)\,\pi_F(G))\,dz = \int_{\mathbb{R}} f(z)\,\theta(z)\,p(z)\,dz = \mathbb{E}(f(F)\,\theta(F)),
\end{aligned}$$

that is the statement holds.

□

We give now a localized version of the result presented above, which will be used in order to prove Theorem 3.4.

Definition 3.7 Let F and G denote square integrable random variables and let $D \subset \mathbb{R}$ be an open set. We say that the $IP_D(F,G)$ (Integration by Parts localized on D) property holds if there exists a square integrable weight $\pi_F^D(G)$ such that

$$\mathbb{E}(\phi'(F)\,G) = \mathbb{E}(\phi(F)\,\pi_F^D(G)), \qquad \text{for any } \phi \in C_b^\infty(\mathbb{R}) \text{ such that } \operatorname{supp}(\phi') \subset D. \tag{10}$$

Moreover, we say that F has a local density p on D if

$$\mathbb{E}(f(F)) = \int f(x)\,p(x)\,dx, \qquad \text{for any } f \in C_b(\mathbb{R}) \text{ such that } \operatorname{supp}(f) \subset D.$$

Finally, we say that $\mathbb{E}(G\,|\,F=\alpha) = \theta(\alpha)$ for $\alpha \in D$ if

$$\mathbb{E}(f(F)\,G) = \mathbb{E}(\theta(F)\,G) \qquad \text{for any } f \in C_b(\mathbb{R}) \text{ such that } \operatorname{supp}(f) \subset D.$$

With these "localized" definitions, we obtain the exact analogous of Lemma 3.6, whose proof is dropped since it is actually identical to the proof of Lemma 3.6:

Lemma 3.8 *(i) Suppose that $IP_D(F,1)$ holds. Then the law of the random variable F has a local probability density p on D given by*

$$p(\alpha) = \mathbb{E}(H(F-\alpha)\,\pi_F^D(1)).$$

(ii) Suppose that both $IP_D(F,1)$ and $IP_D(F,G)$ hold. Then

$$\mathbb{E}(G\,|\,F=\alpha)=\frac{\mathbb{E}(H(F-\alpha)\,\pi_F^D(G))}{\mathbb{E}(H(F-\alpha)\,\pi_F^D(1))},$$

with the convention that the above quantity is equal to 0 whenever $\mathbb{E}(H(F-\alpha)\,\pi_F^D(1))=0$.

Let us prove now that the following lemma, allowing to state the IP and IP_D properties in a suitable gaussian case:

Lemma 3.9 *Let $\Delta\sim \mathrm{N}(0,\delta)$.*

(a) For every $f,g\in C^1$, one has

$$\mathbb{E}(f'(\Delta)\,g(\Delta))=\mathbb{E}\Big(f(\Delta)\,\Big(g(\Delta)\,\frac{\Delta}{\delta}-g'(\Delta)\Big)\Big).$$

(b) Setting $X=xe^{\mu+\sigma\,\Delta}$, one has

$$\mathbb{E}(f'(X)\,g(X))=\mathbb{E}\Big(f(X)\,\Big(\frac{g(X)}{\sigma\,X}\,\Big(\frac{\Delta}{\delta}+\sigma\Big)-g'(X)\Big)\Big),$$

with $f,g\in C^1$.

Proof. (a) By using the standard integration by parts formula, one has

$$\begin{aligned}\mathbb{E}(f'(\Delta)\,g(\Delta))&=\int f'(y)\,g(y)\,\frac{1}{\sqrt{2\pi\delta}}\,e^{-y^2/(2\delta)}\,dy=\int f(y)\Big(g(y)\,\frac{y}{\delta}-g'(y)\Big)\,\frac{1}{\sqrt{2\pi\delta}}\,e^{-y^2/(2\delta)}\,dy\\&=\mathbb{E}\Big(f(\Delta)\,\Big(g(\Delta)\,\frac{\Delta}{\delta}-g'(\Delta)\Big)\Big).\end{aligned}$$

Concerning (b), notice that X is a function of Δ. Thus, we can use (a): setting $f(X)=f_1(\Delta)$ and $g(X)=g_1(\Delta)$, one has $f_1'(\Delta)=f'(X)\,\partial_\Delta X=f'(X)\,\sigma X$ and $g_1'(\Delta)=g'(X)\,\partial_\Delta X=g'(X)\,\sigma X$. Therefore, by (a),

$$\begin{aligned}\mathbb{E}(f'(X)\,g(X))&=\mathbb{E}\Big(f_1'(\Delta)\,\frac{g_1(\Delta)}{\sigma X}\Big)=\mathbb{E}\Big(f_1(\Delta)\,\Big(\frac{g_1(\Delta)}{\sigma X}\,\frac{\Delta}{\delta}-\partial_\Delta\frac{g_1(\Delta)}{\sigma X}\Big)\Big)\\&=\mathbb{E}\Big(f(X)\,\Big(\frac{g(X)}{\sigma\,X}\,\Big(\frac{\Delta}{\delta}+\sigma\Big)-g'(X)\Big)\Big).\end{aligned}$$

□

We can then state the following

Proposition 3.10 *Let X evolve as in (4), let $s<t$ and $g:\mathbb{R}\to\mathbb{R}$ with polynomial growth.*

(i) The $IP(X_s,g(X_t))$ property holds, that is

$$\mathbb{E}(\phi'(X_s)\,g(X_t))=\mathbb{E}(\phi(X_s)\,\pi_s[g](X_t))$$

for any $\phi\in C_b^\infty(\mathbb{R})$, with

$$\pi_s[g](X_t)=g(X_t)\,\frac{1}{\sigma s(t-s)\,X_s}\,\Delta W_{s,t},$$

being $\Delta W_{s,t}=tW_s-sW_t+\sigma s(t-s)$.

(ii) *For a fixed* $\alpha > 0$*, let* $\phi \in C_b^1(\mathbb{R})$ *such that* $\mathrm{supp}(\phi') \subset B_\varepsilon(\alpha) = (\alpha - \varepsilon, \alpha + \varepsilon)$*, with* $\varepsilon > 0$*. Let* $\tilde{\psi} \in C_b^1(\mathbb{R})$ *such that* $\tilde{\psi}\Big|_{B_\varepsilon(\alpha)} = 1$*. Then,*

$$\mathbb{E}(\phi'(X_s)\, g(X_t)) = \mathbb{E}(\phi(X_s)\, \pi_s^{\tilde{\psi}}[g](X_t)),$$

with

$$\pi_s^{\tilde{\psi}}[g](X_t) = \frac{g(X_t)}{\sigma s(t-s)\, X_s}\, \Delta_{\tilde{\psi}} W_{s,t},$$

being

$$\begin{aligned}\Delta_{\tilde{\psi}} W_{s,t} &= \tilde{\psi}(X_s)\Big(tW_s - sW_t + \sigma s(t-s)\Big) + \tilde{\psi}'(X_s)\, X_s\, \sigma\, s(t-s)\\ &= \tilde{\psi}(X_s)\Delta W_{s,t} + \tilde{\psi}'(X_s)\, X_s\, \sigma\, s(t-s).\end{aligned}$$

In particular, the $IP_D(X_s, g(X_t))$ *property holds, with* $D = B_\varepsilon(\alpha)$*, for any fixed* α*.*

(iii) *The* $IP(X_s, g(X_t)\frac{\Delta W_{s,t}}{\sigma s(t-s)X_s})$ *property holds, that is*

$$\mathbb{E}\Big(\phi'(X_s)\, g(X_t)\frac{\Delta W_{s,t}}{\sigma s(t-s)X_s}\Big) = \mathbb{E}(\phi(X_s)\, \bar{\pi}_s[g](X_t))$$

for any $\phi \in C_b^\infty(\mathbb{R})$*, with*

$$\bar{\pi}_s[g](X_t) = g(X_t)\, \frac{1}{\sigma s(t-s)\, X_s^2}\, \Big(\frac{(\Delta W_{s,t})^2}{\sigma s(t-s)} + \Delta W_{s,t} - \frac{t}{\sigma}\Big).$$

Proof. We shall prove parts (ii) and (iii): the statement in (i) follows by using arguments similar (almost identical) to the ones used for proving (ii).

Proof of (ii). The key point of the localization formula, that is the formula in (ii), is that $\phi'\tilde{\psi} = \phi'$, so that

$$\mathbb{E}\Big(\phi'(X_s)\, g(X_t)\Big) = \mathbb{E}\Big(\phi'(X_s)\, \tilde{\psi}(X_s)\, g(X_t)\Big).$$

Since $X_t = X_s\, Y$, with $Y = e^{h(t-s)+\sigma(W_t - W_s)}$ independent of X_s, we can write

$$\mathbb{E}(\phi'(X_s)\, \tilde{\psi}(X_s)\, g(X_t)) = \mathbb{E}\Big(\mathbb{E}(\phi'(X_s)\, \tilde{\psi}(X_s)\, g(y\, X_s))\Big|_{y=Y}\Big).$$

Now, $X_s = x\, e^{hs+\sigma W_s}$, so that by using (b) of Lemma 3.9, one has

$$\begin{aligned}&\mathbb{E}(\phi'(X_s)\, \tilde{\psi}(X_s)\, g(y\, X_s))\\ &= \mathbb{E}\Big(\phi(X_s)\Big(\frac{\tilde{\psi}(X_s)\, g(y\, X_s)}{\sigma\, X_s}\, \Big(\frac{W_s}{s} + \sigma\Big) - \tilde{\psi}'(X_s)\, g(y\, X_s) - \tilde{\psi}(X_s)\, y\, g'(y\, X_s)\Big)\Big)\\ &= \mathbb{E}\Big(\phi(X_s)\Big(\frac{\tilde{\psi}(X_s)\, g(y\, X_s)}{\sigma\, X_s}\, \Big(\frac{W_s}{s} + \sigma\Big) - \tilde{\psi}'(X_s)\, g(y\, X_s)\Big)\Big) - \mathbb{E}\Big(y\, \phi(X_s)\, \tilde{\psi}(X_s)\, g'(y\, X_s)\Big).\end{aligned}$$

By using again the independence between X_s and Y, one has

$$\begin{aligned}&\mathbb{E}(\phi'(X_s)\, \tilde{\psi}(X_s)\, g(X_t))\\ &= \mathbb{E}\Big(\phi(X_s)\Big(\frac{\tilde{\psi}(X_s)\, g(X_t)}{\sigma\, X_s}\, \Big(\frac{W_s}{s} + \sigma\Big) - \tilde{\psi}'(X_s)\, g(X_t)\Big)\Big) - \mathbb{E}\Big(Y\, \phi(X_s)\, \tilde{\psi}(X_s)\, g'(X_t)\Big).\end{aligned}$$

Let us consider the second addendum of the r.h.s.: we will use the integration by parts formula based on Y in order to remove the derivative of g. By conditioning with respect to X_s, one can write

$$\begin{aligned}\mathbb{E}\Big(Y\,\phi(X_s)\,\tilde\psi(X_s)\,g'(X_t)\Big) &= \mathbb{E}\Big(\mathbb{E}\Big(\phi(z)\,\tilde\psi(z)\,Y\,g'(z\,Y)\Big)\Big|_{z=X_s}\Big)\\ &= \mathbb{E}\Big(\phi(z)\,\tilde\psi(z)\mathbb{E}\Big(Y\,g'(z\,Y)\Big)\Big|_{z=X_s}\Big).\end{aligned}$$

By using (b) of Lemma 3.9, it holds

$$\mathbb{E}\Big(g'(z\,Y)\,Y\Big) = \mathbb{E}\Big(\frac{g(z\,Y)}{z}\Big(\frac{Y}{\sigma\,Y}\Big(\frac{W_t-W_s}{t-s}+\sigma\Big)-1\Big)\Big) = \mathbb{E}\Big(\frac{g(z\,Y)}{z}\,\frac{W_t-W_s}{t-s}\Big)$$

and by inserting this in the above equality one obtains

$$\mathbb{E}\Big(Y\,\phi(X_s)\,\tilde\psi(X_s)\,g'(X_t)\Big) = \mathbb{E}\Big(\phi(X_s)\,\tilde\psi(X_s)\,\frac{g(X_t)}{X_s}\,\frac{W_t-W_s}{t-s}\Big).$$

We can finally write

$$\begin{aligned}&\mathbb{E}(\phi'(X_s)\,\tilde\psi(X_s)\,g(X_t))\\ &= \mathbb{E}\Big(\phi(X_s)\Big(\frac{\tilde\psi(X_s)\,g(X_t)}{\sigma\,X_s}\Big(\frac{W_s}{s}+\sigma\Big)-\tilde\psi'(X_s)\,g(X_t)\Big)\Big)-\mathbb{E}\Big(\phi(X_s)\,\tilde\psi(X_s)\,\frac{g(X_t)}{X_s}\,\frac{W_t-W_s}{t-s}\Big)\end{aligned}$$

and straightforward computations allow to conclude. Let us finally observe that to achieve this representation one has implicitly assumed that g is regular (C^1), which is not true in general. But this is not really a problem: one can regularize g with some suitable mollifier and by using density arguments, the statement follows.

Proof of (iii). We write once again $X_t = y\,X_s$, $g(X_t) = g(y\,X_s)$ and $\Delta W_{s,t} = v(X_s, X_t)$, where

$$v(X_s,X_t) = \frac{t-s}{\sigma}\Big(\ln X_s - \ln x - hs + \sigma^2 s\Big) - \frac{s}{\sigma}\Big(\ln X_t - \ln X_s - h(t-s)\Big).$$

Then by using (b) of Lemma 3.9, we obtain

$$\begin{aligned}\mathbb{E}\Big(\phi'(X_s-\alpha)g(X_t)\,\tfrac{\Delta W_{s,t}}{\sigma s(t-s)X_s}\Big) = \mathbb{E}\Big(\phi(X_s-\alpha)\Big[&\frac{g(X_t)\,v(X_s,X_t)}{\sigma^2 s^2(t-s)X_s^2}(W_s+\sigma s)\\ &-Y g'(X_t)\,\tfrac{v(X_s,X_t)}{\sigma s(t-s)X_s} - g(X_t)\,\frac{v_x(X_s,X_t)+Y v_y(X_s,X_t)}{\sigma s(t-s)X_s} + \frac{g(X_t)\,v(X_s,X_t)}{\sigma s(t-s)X_s^2}\Big]\Big)\end{aligned}$$

Consider now the term containing g', that is $\mathbb{E}\Big(\phi(X_s-\alpha)Y g'(X_t)\,\frac{v(X_s,X_t)}{\sigma s(t-s)X_s}\Big)$: by conditioning first on X_s, one has to evaluate:

$$\mathbb{E}\Big(\phi(x-\alpha)Y g'(xY)\,\frac{v(x,xY)}{\sigma s(t-s)x}\Big) = \frac{\phi(x-\alpha)}{\sigma s(t-s)x}\,\mathbb{E}(Y g'(xY)\,v(x,xY)).$$

By recalling that $Y = \exp(h(t-s)+\sigma(W_t-W_s))$ and by using again part (b) of Lemma 3.9, one has

$$\begin{aligned}\mathbb{E}(Y g'(xY)\,v(x,xY)) =\ & \mathbb{E}\Big(\frac{g(xY)}{x}\Big[\frac{Y v(x,xY)}{\sigma Y}\Big(\frac{W_t-W_s}{t-s}+\sigma\Big) - v(x,xY) - Y\,v_y(x,xY)\,x\Big]\Big)\\ =\ & \mathbb{E}\Big(\frac{g(xY)}{\sigma(t-s)x}\,v(x,xY)\,(W_t-W_s) - Y\,g(xY)\,v_y(x,xY)\Big)\end{aligned}$$

so that

$$\mathbb{E}\Big(\phi(X_s-\alpha)Yg'(X_t)\,\frac{v(X_s,X_t)}{\sigma s(t-s)X_s}\Big)$$
$$=\mathbb{E}\Big(\phi(X_s-\alpha)\frac{g(X_t)}{\sigma s(t-s)X_s}\Big[\frac{v(X_s,X_t)}{\sigma(t-s)X_s}(W_t-W_s)-Y\,v_y(X_s,X_t)\Big]\Big).$$

By inserting such a quantity in the expression above, straightforward computations allow to obtain

$$\mathbb{E}\Big(\phi'(X_s-\alpha)g(X_t)\,\tfrac{\Delta W_{s,t}}{\sigma s(t-s)X_s}\Big)=-\mathbb{E}\Big(\phi(X_s-\alpha)\frac{g(X_t)}{\sigma^2 s^2(t-s)X_s^2}[v(X_s,X_t)(t-s)(W_s+\sigma s)$$
$$-v(X_s,X_t)s(W_t-W_s)+v(X_s,X_t)\sigma s(t-s)-v_x(X_s,X_t)\sigma s(t-s)X_s]\Big).$$

Now, since $v_x(X_s,X_t)=\frac{t}{\sigma X_s}$, one can rewrite the above expression as

$$\mathbb{E}\Big(\phi'(X_s-\alpha)g(X_t)\,\frac{\Delta W_{s,t}}{\sigma s(t-s)X_s}\Big)=\mathbb{E}\Big(g(X_t)\frac{\phi(X_s-\alpha)}{\sigma s(t-s)X_s^2}\Big[\frac{(\Delta W_{s,t})^2}{\sigma s(t-s)}+\Delta W_{s,t}-\frac{t}{\sigma}\Big]\Big).$$

□

Remark 3.11 *The key point in the above lemma was that we used* two *integration by parts formulas: the first with respect to X_s and the second with respect to $Y=X_t/X_s=e^{h(t-s)+\sigma(W_t-W_s)}$. From a numerical point of view, it is the integration by parts with respect to Y which is "dangerous". In fact, the noise contained in Y is W_t-W_s, so, if $t-s$ is small, we have "few noise". As a consequence, the weight which appears in the integration by parts contains $1/\sqrt{t-s}$, which blows up if $t-s$ is small. Think now to our dynamical programming principle, where $s=t_{k-1}$ and $t=t_k$, so that $t-s=t_k-t_{k-1}$. If one wants to take a very rich time-grid (which is necessary in order to get a good approximation of an American option with a Bermudean one), then one has to face the problem of $1/\sqrt{t_k-t_{k-1}}$. This represents the specific difficulty of this type of algorithm: one has to handle the balance between the analytical error, which is of order t_k-t_{k-1}, and the variance of the weights which appear in the integration by parts. In dimension $d=1$, this is of order $(t_k-t_{k-1})^{-1/2}$ but in dimension d, we will have $(t_k-t_{k-1})^{-d/2}$, and this is the way in which the dimension comes on in this algorithm. The localization technique allows to reduce this variance, and this is crucial.*

3.2 Proofs

Proof of Theorem 3.1. Part $i)$ is is an immediate consequence of Lemma 3.6 and Proposition 3.10. Concerning part $ii)$, it suffices to prove that

$$\partial_\alpha\mathbb{T}_{s,t}[f](\alpha)=\mathbb{R}_{s,t}[f](\alpha)=-\mathbb{E}\Big(f(X_t)\frac{H(X_s-\alpha)}{\sigma s(t-s)X_s^2}\,\Big[\frac{(\Delta W_{s,t})^2}{\sigma s(t-s)}+\Delta W_{s,t}-\frac{t}{\sigma}\Big]\Big),$$

for any $f\in\mathcal{E}_b(\mathbb{R})$. First, let us take h_δ as a C^∞ probability density function weakly convergent, as $\delta\to 0$, to the Dirac mass in 0. This means that the associated distribution function H_δ converges, as $\delta\to 0$, to H. Let us set

$$\mathbb{T}^\delta_{s,t}[f](\alpha)=-\mathbb{E}\Big(f(X_t)\frac{H_\delta(X_s-\alpha)}{\sigma s(t-s)X_s}\,\Delta W_{s,t}\Big).$$

Obviously, $\mathbb{T}^\delta_{s,t}[f](\alpha)\to\mathbb{T}_{s,t}[f](\alpha)$ as $\delta\to 0$, but also, by using (almost) standard density arguments, one can easily show that

$$\lim_{\delta\to 0}\partial_\alpha\mathbb{T}^\delta_{s,t}[f](\alpha)=\partial_\alpha\mathbb{T}_{s,t}[f](\alpha).$$

So, let us work with $\partial_\alpha \mathbb{T}^\delta_{s,t}[f](\alpha)$. One has

$$\partial_\alpha \mathbb{T}^\delta_{s,t}[f](\alpha) = -\mathbb{E}\Big(f(X_t)\, h_\delta(X_s-\alpha)\, \frac{\Delta W_{s,t}}{\sigma s(t-s) X_s}\Big).$$

By using part (iii) of Proposition 3.10, one has

$$\partial_\alpha \mathbb{T}^\delta_{s,t}[f](\alpha) = -\mathbb{E}\Big(f(X_t) \frac{H_\delta(X_s-\alpha)}{\sigma s(t-s) X_s^2} \Big[\frac{(\Delta W_{s,t})^2}{\sigma s(t-s)} + \Delta W_{s,t} - \frac{t}{\sigma}\Big]$$

(recall that $h_\delta = H'_\delta$) and by taking the limit as $\delta \to 0$, the statement follows.

□

Proof of Lemma 3.2. Let us denote for simplicity

$$F = X_s, \quad G = f(X_t), \quad \pi_F(G) = f(X_t)\, \frac{1}{\sigma s(t-s)\, X_s}\, \Delta W_{s,t}, \tag{11}$$

so that we can write shortly $\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E}(H(F-\alpha)\, \pi_F(G))$, where $H(x) = \mathbf{1}_{x\geq 0}$. Take $\psi \geq 0$ such that $\int_{\mathbb{R}} \psi(t)\, dt = 1$ and set $\Psi(x) = \int_{-\infty}^x \psi(t)\, dt$. Since the $IP(F,G)$ holds (recall (11) and (i) of Proposition 3.10), one has

$$\begin{aligned}
\mathbb{E}(H(F-\alpha)\, \pi_F(G)) &= \mathbb{E}(\psi(F-\alpha)\, G) - \mathbb{E}(\psi(F-\alpha)\, G) + \mathbb{E}(H(F-\alpha)\, \pi_F(G)) \\
&= \mathbb{E}(\psi(F-\alpha)\, G) + \mathbb{E}(H(F-\alpha)\, \pi_F(G)) - \mathbb{E}(\Psi'(F-\alpha)\, G) \\
&= \mathbb{E}(\psi(F-\alpha)\, G) + \mathbb{E}(H(F-\alpha)\, \pi_F(G)) - \mathbb{E}(\Psi(F-\alpha)\, \pi_F(G)) \\
&= \mathbb{E}(\psi(F-\alpha)\, G) + \mathbb{E}((H(F-\alpha) - \Psi(F-\alpha))\, \pi_F(G)).
\end{aligned}$$

Inserting now the exact values for F, G and $\pi_F(G)$ as in (11), one obtains

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E}\Big(f(X_t)\, \psi(X_s-\alpha)\Big) + \mathbb{E}\Big(f(X_t)\, \frac{(H-\Psi)(X_s-\alpha)}{\sigma s(t-s)\, X_s}\, \Delta W_{s,t}\Big) = \mathbb{T}^\psi_{s,t}[f](\alpha).$$

Concerning $\mathbb{R}_{s,t}$, we set

$$F = X_s,\ \bar G = f(X_t)\, \frac{1}{\sigma s(t-s)\, X_s}\, \Delta W_{s,t},\ \pi_F(\bar G) = f(X_t)\, \frac{1}{\sigma s(t-s)\, X_s^2} \Big(\frac{(\Delta W_{s,t})^2}{\sigma s(t-s)} + \Delta W_{s,t} - \frac{t}{\sigma}\Big), \tag{12}$$

so that$\mathbb{R}_{s,t}[f](\alpha) = \mathbb{E}(H(F-\alpha)\, \pi_F(\bar G))$. Take $\psi \geq 0$ such that $\int_{\mathbb{R}} \psi(t)\, dt = 1$ and set $\Psi(x) = \int_{-\infty}^x \psi(t)\, dt$. By using the $IP(F,\bar G)$ property proved in (iii) of Proposition 3.10, one has

$$\begin{aligned}
\mathbb{E}(H(F-\alpha)\, \pi_F(\bar G)) &= \mathbb{E}(\psi(F-\alpha)\, \bar G) + \mathbb{E}(H(F-\alpha)\, \pi_F(\bar G)) - \mathbb{E}(\Psi'(F-\alpha)\, \bar G) \\
&= \mathbb{E}(\psi(F-\alpha)\, \bar G) + \mathbb{E}(H(F-\alpha)\, \pi_F(\bar G)) - \mathbb{E}(\Psi(F-\alpha)\, \pi_F(\bar G)) \\
&= \mathbb{E}(\psi(F-\alpha)\, \bar G) + \mathbb{E}((H-\Psi)(F-\alpha)\, \pi_F(\bar G)).
\end{aligned}$$

The conclusion follows by inserting the exact values for F, $\bar G$ and $\pi_F(\bar G)$ as in (12).

□

Proof of Theorem 3.4. We recall that α is fixed and $\tilde\psi(\alpha) = 1$. In a first step, let us assume the stronger condition $\tilde\psi\Big|_{B_\varepsilon(\alpha)} = 1$, for some $\varepsilon > 0$. Then, by using (ii) of Proposition 3.10, we have

that the $IP_{B_\varepsilon(\alpha)}(F,G)$ property holds, with $F = X_s$ and $G = \Phi(X_t)$, and by using Lemma 3.8 we obtain

$$\mathbb{E}(\Phi(X_t)\,|\,X_s = \beta) = \frac{\mathbb{E}\Big(H(X_s-\beta)\,\pi^{\tilde\psi}_{X_s}(\Phi(X_s))\Big)}{\mathbb{E}\Big(H(X_s-\beta)\,\pi^{\tilde\psi}_{X_s}(1)\Big)}, \qquad \text{for any } \beta \in B_\varepsilon(\alpha), \tag{13}$$

and in particular for $\beta = \alpha$.

Assume now that we just have $\tilde\psi(\alpha) = 1$. Then we may construct a sequence $\{\tilde\psi_n\}_n$ such that $\tilde\psi_n = 1$ on $B_{1/n}(\alpha)$, $\tilde\psi_n = \tilde\psi$ on $B^c_{2/n}(\alpha)$ and $|\tilde\psi_n| \le 1 \wedge \tilde\psi$ for any n. We have $\tilde\psi_n(\beta) \to \tilde\psi(\beta)$ for any β and $\tilde\psi_n'(\beta) \to \tilde\psi'(\beta)$ for any $\beta \neq \alpha$. It then follows that $\pi^{\tilde\psi_n}_{X_s}(\Phi(X_s)) \to \pi^{\tilde\psi}_{X_s}(\Phi(X_s))$ and $\pi^{\tilde\psi_n}_{X_s}(1) \to \pi^{\tilde\psi}_{X_s}(1)$ as $n \to \infty$ a.s. By using Lebesgue's dominated convergence theorem, one can pass to the limit in the expectations in (13), written with $\tilde\psi_n$, and to obtain the validity of (13) for such a $\tilde\psi$.

□

4 Representation formulas for the conditional expectation and its gradient: the multidimensional case

Let X be the underlying asset price process, driven by the Black and Scholes model, i.e. it solves the following stochastic differential equation (sde)

$$\begin{aligned} &dX_t = (\hat r - \eta)X_t dt + \sigma X_t dW_t \\ &X_0 = x \end{aligned}$$

where:

- $x \in \mathbb{R}^d_+$;
- $\hat r, \eta \in \mathbb{R}^d$, with $\hat r_i = r$ for any $i = 1,\dots,d$, being r the (constant) spot rate, and with η the vector of the dividends of the option;
- σ denotes the $d \times d$ volatility matrix which we suppose to be non-degenerate;
- W is a d-dimensional correlated Brownian motion.

Without loss of generality, one can suppose that σ is a sub-triangular matrix, that is $\sigma_{ij} = 0$ whenever $i < j$, and that W is a standard d-dimensional Brownian motion. Thus, any component of X_t can be written as

$$X^i_t = x_i \exp\Big(h_i t + \sum_{j=1}^{i} \sigma_{ij} W^j_t\Big), \qquad i = 1,\dots,d \tag{14}$$

where from now on we set

$$h_i = r_i - \eta_i - \frac{1}{2}\sum_{j=1}^{i} \sigma^2_{ij}, \qquad i = 1,\dots,d.$$

The aim is to study the conditional expectation

$$\mathbb{E}(\Phi(X_t)\,|\,X_s = \alpha)$$

where $0 < s < t$, $\alpha \in \mathbb{R}^d_+$ and $\Phi \in \mathcal{E}_b(\mathbb{R}^d)$ the class of the measurable functions with polynomial growth, i.e $|\Phi(y)| \le C(1+|y|^m)$, in order to derive representation formulas in dimension d using the ones in dimension $d=1$.
In few words, to this goal it suffices to consider an auxiliary process $\widetilde{X}$ with independent components for which a formula for the conditional expectation immediately follows as a product. In a second step, such a formula can be adapted to the original process X by means of an (inversible) function giving X from the auxiliary process $\widetilde{X}$. We will discuss later the connections with what has been developed in the paper of Lions and Regnier [17].

To our purposes, let $\ell_t = (\ell^1_t, \dots, \ell^d_t)$ be a fixed C^1 function and let us set

$$\widetilde{X}^i_t = x_i \exp\Big(h_i\, t + \ell^i_t + \sigma_{ii} W^i_t\Big), \qquad i = 1, \dots, d \tag{15}$$

which obviously satisfies the sde

$$\begin{aligned} &d\widetilde{X}_t = (\hat{r} - \eta + \ell'_t)\widetilde{X}_t dt + \bar{\sigma}\widetilde{X}_t dW_t \\ &\widetilde{X}_0 = x \end{aligned} \tag{16}$$

being $\bar{\sigma}$ the diagonal matrix whose entries are given by σ_{ii}, $i = 1, \dots, d$. As a first result, we study a transformation allowing to handle the new process $\widetilde{X}$ in place of the original process X:

Lemma 4.1 *For any $t \ge 0$ there exists a function $F_t(\cdot) : \mathbb{R}^d_+ \to \mathbb{R}^d_+$ such that F_t is inversible and*

$$X_t = F_t(\widetilde{X}_t) \qquad\qquad \widetilde{X}_t = F_t^{-1}(X_t).$$

Proof. Let t, ℓ, x be fixed. From (15) we have

$$W^i_t = \frac{1}{\sigma_{ii}}\Big(\ln \frac{\widetilde{X}^i_t}{x^i} - h_i\, t - \ell^i_t\Big).$$

Inserting this in (14), one obtains

$$X^i_t = x_i \exp\Big(h_i t - \sum_{j=1}^{i} \frac{\sigma_{ij}}{\sigma_{jj}}(h_j\, t + \ell^j_t)\Big) \prod_{j=1}^{i} \Big(\frac{\widetilde{X}^j_t}{x_j}\Big)^{\sigma_{ij}/\sigma_{jj}}$$

Thus, by setting

$$\widetilde{\sigma}_{ij} = \frac{\sigma_{ij}}{\sigma_{jj}}, \qquad i, j = 1, \dots, d.$$

and by using the notation $\ln \xi = (\ln \xi_1, \dots, \ln \xi_d)$, for $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$, then $F_t = (F^1_t, \dots, F^d_t)$ satisfies

$$\ln F_t(y) = -\widetilde{\sigma}\, \ell_t + \widetilde{\sigma}\, \ln y + (I - \widetilde{\sigma})(\ln x + ht)$$

and, by setting $\widehat{\sigma} = \widetilde{\sigma}^{-1}$, its inverse function is given by

$$\ln F_t^{-1}(z) = \ell_t + \widehat{\sigma}\, \ln z + (I - \widehat{\sigma})(\ln x + ht).$$

□

Let us summarize for later use the previous result. We set

$$\widetilde{\sigma}_{ij} = \frac{\sigma_{ij}}{\sigma_{jj}},\ i, j = 1, \dots, d, \quad \text{and} \quad \widehat{\sigma} = \widetilde{\sigma}^{-1} \tag{17}$$

It is worth noticing that $\widehat{\sigma}$ is easy to compute because $\widetilde{\sigma}$ is a triangular matrix. Moreover, $\widehat{\sigma}$ is itself triangular and $\widehat{\sigma}_{ii} = 1$ for any i. Thus, the function F_t and its inverse $G_t = F_t^{-1}$ such that

$$X_t = F_t(\widetilde{X}_t) \quad \text{and} \quad \widetilde{X}_t = G_t(X_t)$$

are given by

$$F_t^i(y) = e^{-\sum_{j=1}^i \widetilde{\sigma}_{ij}\ell_t^j}\, y_i \prod_{j=1}^{i-1} \Big(\frac{y_j}{x_j}\, e^{-h_j t}\Big)^{\widetilde{\sigma}_{ij}}, \qquad i = 1, \dots, d,\ y \in \mathbb{R}_+^d \tag{18}$$

and

$$G_t^i(z) = e^{\ell_t^i}\, z_i \prod_{j=1}^{i-1} \Big(\frac{z_j}{x_j}\, e^{-h_j t}\Big)^{\widehat{\sigma}_{ij}}, \qquad i = 1, \dots, d,\ z \in \mathbb{R}_+^d. \tag{19}$$

By using the process $\widetilde{X}$, mainly the fact that its components are independent, we can easily obtain a first formula for the conditional expectation starting from the one-dimensional one.

Theorem 4.2 (Representation formulas I: without localization)

i) Let $0 \le s < t$ be fixed. For any function $\Phi \in \mathcal{E}_b(\mathbb{R}^d)$ and $\alpha \in \mathbb{R}_+^d$, one has

$$\mathbb{E}\Big(\Phi(X_t) \,|\, X_s = \alpha\Big) = \frac{\mathbb{T}_{s,t}[\Phi](\alpha)}{\mathbb{T}_{s,t}[1](\alpha)}$$

where

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E}\Big(f(X_t) \prod_{i=1}^d \frac{H(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii} s(t-s)\widetilde{X}_s^i}\, \Delta W_{s,t}^i\Big) \tag{20}$$

being $\widetilde{X}_s = G_s(X_s)$ and $\widetilde{\alpha} = G_s(\alpha)$ [G_s being defined in (19)], $H(\xi) = \mathbf{1}_{\xi \ge 0}$, $\xi \in \mathbb{R}$, and

$$\Delta W_{s,t}^i = (t-s)(W_s^i + \sigma_{ii} s) - s(W_t^i - W_s^i), \qquad i = 1, \dots, d.$$

ii) Let $0 \le s < t$ be fixed. For any function $\Phi \in \mathcal{E}_b(\mathbb{R}^d)$, one has, for $j = 1, \dots, d$,

$$\partial_{\alpha_j} \mathbb{E}\Big(\Phi(X_t) \,|\, X_s = \alpha\Big) = \sum_{k=1}^j \widehat{\sigma}_{kj} \frac{\widetilde{\alpha}_k}{\alpha_j} \times \frac{\mathbb{R}_{s,t;k}[\Phi](\alpha)\mathbb{T}_{s,t}[1](\alpha) - \mathbb{T}_{s,t}[\Phi](\alpha)\mathbb{R}_{s,t;k}[1](\alpha)}{\mathbb{T}_{s,t}[1](\alpha)^2},$$

where $\mathbb{T}_{s,t}[f](\alpha)$ is defined above and, as $k = 1, \dots, d$,

$$\begin{aligned} \mathbb{R}_{s,t;k}[f](\alpha) = \quad & -\mathbb{E}\Big(f(X_t) \frac{H(\widetilde{X}_s^k - \widetilde{\alpha}^k)}{\sigma_{kk} s(t-s)(\widetilde{X}_s^k)^2} \Big[\frac{(\Delta W_{s,t}^k)^2}{\sigma_{kk} s(t-s)} + \Delta W_{s,t}^k - \frac{t}{\sigma_{kk}}\Big] \times \\ & \times \prod_{i=1, i\ne k}^d \frac{H(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii} s(t-s)\widetilde{X}_s^i}\, \Delta W_{s,t}^i\Big). \end{aligned} \tag{21}$$

Remark 4.3 *If no correlation is assumed among the assets, that is if the volatility matrix σ is diagonal, then $\widehat{\sigma} = Id_{d\times d}$. Thus, the sum appearing for the evaluation of $\partial_{\alpha_j}\mathbb{E}(F(X_t) \,|\, X_s = \alpha)$ reduces to the single term with $k = j$, with coefficient $\widetilde{\alpha}_j/\alpha_j = e^{\ell_s^j}$, which in turn is equal to 1 whenever $\ell = 0$.*

Proof of Theorem 4.2. *i)* Let us set $\widetilde{\Phi}_t(y) \equiv \widetilde{\Phi}(y) = \Phi \circ F_t(y)$, $y \in \mathbb{R}^d_+$, being F_t defined in (18). Since $X_t = F_t(\widetilde{X}_t)$ for any t, one obviously has

$$\mathbb{E}\Big(\Phi(X_t)\,\Big|\,X_s = \alpha\Big) = \mathbb{E}\Big(\widetilde{\Phi}(\widetilde{X}_t)\,\Big|\,\widetilde{X}_s = G_s(\alpha)\Big),$$

(recall that $G_s = F_s^{-1}$). Thus, setting $\widetilde{\alpha}_s \equiv \widetilde{\alpha} = G_s(\alpha)$, it is sufficient to prove that

$$\mathbb{E}\Big(\widetilde{\Phi}(\widetilde{X}_t)\,\Big|\,\widetilde{X}_s = \widetilde{\alpha}\Big) = \frac{\widetilde{\mathbb{T}}_{s,t}[\widetilde{\Phi}](\widetilde{\alpha})}{\widetilde{\mathbb{T}}_{s,t}[1](\widetilde{\alpha})} \tag{22}$$

where

$$\widetilde{\mathbb{T}}_{s,t}[f](\widetilde{\alpha}) = \prod_{i=1}^d \frac{1}{\sigma_{ii}s(t-s)} \cdot \mathbb{E}\Big(f(\widetilde{X}_t)\prod_{i=1}^d \frac{H(\widetilde{X}^i_s - \widetilde{\alpha}_i)}{\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big)$$

Now, let us firstly suppose that $\widetilde{\Phi}(y) = \widetilde{\Phi}_1(y_1)\cdots\widetilde{\Phi}_d(y_d)$, that is $\widetilde{\Phi}$ can be separated in the product of d functions each one depending only on a single variable and belonging to $\mathcal{E}_b(\mathbb{R})$. In such a case, one obviously has

$$\mathbb{E}\Big(\widetilde{\Phi}(\widetilde{X}_t)\,\Big|\,\widetilde{X}_s = \widetilde{\alpha}\Big) = \prod_{i=1}^d \mathbb{E}\Big(\widetilde{\Phi}_i(\widetilde{X}^i_t)\,\Big|\,\widetilde{X}^i_s = \widetilde{\alpha}_i\Big).$$

Now, it is easy to see that for each $\widetilde{X}^i_t$ Theorem 3.1 can be applied (indeed, here the drift of $\widetilde{X}^i_t$ is of the form $\rho^i_t\,\widetilde{X}^i_t$ while the process handled in the one-dimensional case has $\rho^i_t = const.$, but this is not a difficulty). Thus, recalling that $\widetilde{X}$ has independent components, by Theorem 3.1 one has

$$\mathbb{E}\Big(\widetilde{\Phi}(\widetilde{X}_t)\,\Big|\,\widetilde{X}_s = \widetilde{\alpha}\Big) = \prod_{i=1}^d \mathbb{E}\Big(\widetilde{\Phi}_i(\widetilde{X}^i_t)\,\Big|\,\widetilde{X}^i_s = \widetilde{\alpha}_i\Big) = \prod_{i=1}^d \frac{\mathbb{T}^i_{s,t}[\widetilde{\Phi}_i](\widetilde{\alpha}_i)}{\mathbb{T}^i_{s,t}[1](\widetilde{\alpha}_i)}$$

being

$$\mathbb{T}^i_{s,t}[g](\widetilde{\alpha}_i) = \frac{1}{\sigma_{ii}s(t-s)} \cdot \mathbb{E}\Big(g(\widetilde{X}^i_t)\frac{H(\widetilde{X}^i_s - \widetilde{\alpha}_i)}{\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big).$$

By using again the independence for the components of $\widetilde{X}$, it easily follows that

$$\widetilde{\mathbb{T}}_{s,t}[\widetilde{\Phi}](\widetilde{\alpha}) = \prod_{i=1}^d \mathbb{T}^i_{s,t}[\widetilde{\Phi}_i](\widetilde{\alpha}_i) \qquad \text{and} \qquad \widetilde{\mathbb{T}}_{s,t}[1](\widetilde{\alpha}) = \prod_{i=1}^d \mathbb{T}^i_{s,t}[1](\widetilde{\alpha}_i),$$

so that (22) holds when $\widetilde{\Phi}(y) = \widetilde{\Phi}_1(y_1)\cdots\widetilde{\Phi}_d(y_d)$. In the general case, the statement holds by using a density argument: for any $\widetilde{\Phi} \in \mathcal{E}_b(\mathbb{R}^d)$ there exists a sequence of functions $\{\widetilde{\Phi}^n\}_n \subset \mathcal{E}_b(\mathbb{R}^d)$ such that $\widetilde{\Phi}^n(\widetilde{X}_t) \to \widetilde{\Phi}(\widetilde{X}_t)$ in L^2 and such that each $\widetilde{\Phi}^n$ is a linear combination of functions which separate the variables as above. Since representation (22) holds for any $\widetilde{\Phi}^n$, it finally holds for $\widetilde{\Phi}$ as well, as it immediately follows by passing to the limit.

ii) First, notice that, by (19), $\widehat{\sigma}_{kj}\widetilde{\alpha}_k/\alpha_j = \partial_{\alpha_j}G^k_s(\alpha) = \partial_{\alpha_j}\widetilde{\alpha}_k$. Thus, by considering $\mathbb{T}_{s,t}[f](\alpha)$ as a function (as it is!) of $\widetilde{\alpha}$, that is $\mathbb{T}_{s,t}[f](\alpha) = \widetilde{\mathbb{T}}_{s,t}[f](\widetilde{\alpha})$, then we only have to show that

$$\mathbb{R}_{s,t;k}[f](\alpha) = \partial_{\widetilde{\alpha}_k}\widetilde{\mathbb{T}}_{s,t}[f](\widetilde{\alpha}) = \partial_{\widetilde{\alpha}_k}\mathbb{E}\Big(\widetilde{f}(\widetilde{X}_t)\prod_{i=1}^d \frac{H(\widetilde{X}^i_s - \widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big),$$

where we have set $\widetilde{f}(\widetilde{X}_t) = f \circ F_t(\widetilde{X}_t)$. By conditioning w.r.t. all the coordinates of $\widetilde{X}_t$ except for the kth one, and by recalling that $\widetilde{X}$ has independent components, one has

$$\widetilde{\mathbb{T}}_{s,t}[f](\widetilde{\alpha}) = \mathbb{E}\Big(\mathbb{E}\Big(\widetilde{f}_{-k}(\widetilde{X}_t^k)\frac{H(\widetilde{X}_s^k - \widetilde{\alpha}_k)}{\sigma_{kk}s(t-s)\widetilde{X}_s^k}\,\Delta W_{s,t}^k\Big)\Big|_{\widetilde{x}_j=\widetilde{X}_t^j, j\neq k} \times \prod_{i=1, i\neq k}^{d} \frac{H(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big),$$

being $\widetilde{f}_{-k}(\widetilde{X}_t^k) = \widetilde{f}(\widetilde{x}_1, \dots, \widetilde{x}_{k-1}, \widetilde{X}_t^k, \widetilde{x}_{k+1}, \dots \widetilde{x}_d)$. Thus,

$$\partial_{\widetilde{\alpha}_k}\widetilde{\mathbb{T}}_{s,t}[f](\widetilde{\alpha}) = \mathbb{E}\Big(\partial_{\widetilde{\alpha}_k}\mathbb{E}\Big(\widetilde{f}_{-k}(\widetilde{X}_t^k)\tfrac{H(\widetilde{X}_s^k-\widetilde{\alpha}_k)}{\sigma_{kk}s(t-s)\widetilde{X}_s^k}\,\Delta W_{s,t}^k\Big)\Big|_{\widetilde{x}_j=\widetilde{X}_t^j, j\neq k}\times$$
$$\times \prod_{i=1, i\neq k}^{d} \frac{H(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big).$$

Now, by using the one dimensional results (see Theorem 3.1 and (6) in particular), one has

$$\partial_{\widetilde{\alpha}_k}\mathbb{E}\Big(\widetilde{f}_{-k}(\widetilde{X}_t^k)\frac{H(\widetilde{X}_s^k - \widetilde{\alpha}_k)}{\sigma_{kk}s(t-s)\widetilde{X}_s^k}\,\Delta W_{s,t}^k\Big) = \mathbb{E}\Big(\widetilde{f}_{-k}(\widetilde{X}_t^k)\frac{H(\widetilde{X}_s^k - \widetilde{\alpha}_k)}{\sigma_{kk}s(t-s)(\widetilde{X}_s^k)^2}\Big[\frac{\Delta W_{s,t}^k}{\sigma_{kk}s(t-s)} + \Delta W_{s,t}^k - \frac{t}{\sigma}\Big]\Big)$$

By rearranging everything, the formula holds.

□

Let us give some commentary remarks about the significance of the drift-function ℓ_t. Setting

$$Q_t^i(z) = e^{-\ell_t^i}\,G_t^i(z) = z_i \prod_{j=1}^{i-1}\Big(\frac{z_j}{x_j}\,e^{-h_j t}\Big)^{\widehat{\sigma}_{ij}},$$

then

$$H(G_s^i(X_s) - G_s^i(\alpha)) = \mathbf{1}_{G_s^i(X_s)\geq G_s^i(\alpha)} = \mathbf{1}_{Q_s^i(X_s)\geq Q_s^i(\alpha)} = H(Q_s^i(X_s) - Q_s^i(\alpha)),$$

so that we obtain

$$\mathbb{T}_{s,t}[f](\alpha) = e^{-\sum_{i=1}^d \ell_s^i}\prod_{i=1}^{d}\frac{1}{\sigma_{ii}s(t-s)}\cdot\mathbb{E}\Big(f(X_t)\prod_{i=1}^{d}\frac{H(Q_s^i(X_s) - Q_s^i(\alpha))}{Q_s^i(X_s)}\,\Delta W_{s,t}^i\Big). \tag{23}$$

Now, by recalling that $Q_t^i(z)$ does no more depend on ℓ, the only way in which ℓ_t comes out in our formula is $\sum_{i=1}^d \ell_t^i$, so that we have just one degree of freedom. The problem is how to choose ℓ, let us now discuss about this.

The main application of the formula is addressed to Monte Carlo simulations, therefore, as usual, one way is to optimize (here, minimize) the variance or the integrated variance of the Monte Carlo estimator with respect to ℓ. More precisely, by using Theorem 4.2, the conditional expectation $m(\alpha) = \mathbb{E}(\Phi(X_t)\,|\,X_s = \alpha)$ is numerically evaluated in a Monte Carlo setting by means of

$$m_N(\alpha) = \frac{\mathbb{T}_{s,t}^N[\Phi](\alpha)}{\mathbb{T}_{s,t}^N[1](\alpha)}$$

where

$$\mathbb{T}_{s,t}^N[f](\alpha) = \prod_{i=1}^{d}\frac{1}{\sigma_{ii}s(t-s)}\cdot\frac{1}{N}\sum_{k=1}^{N} f(X_t^{(k)})\prod_{i=1}^{d}\frac{H(\widetilde{X}_s^{i,(k)} - \widetilde{\alpha}_i)}{\widetilde{X}_s^{i,(k)}}\,\Delta W_{s,t}^{i,(k)}$$

or equivalently

$$\mathbb{T}^N_{s,t}[f](\alpha) = \prod_{i=1}^d \frac{1}{\sigma_{ii} s(t-s)} \cdot \frac{1}{N} \sum_{k=1}^N f(X_t^{(k)}) \prod_{i=1}^d \frac{H(G_s^i(X_s^{(k)}) - G_s^i(\alpha))}{G_s^i(X_s^{(k)})} \Delta W_{s,t}^{i,(k)},$$

where $(W^{(k)}, X^{(k)})$, as $k = 1, \dots, N$ stand for independent copies of (W, X). By using (23), we can also write

$$\mathbb{T}^N_{s,t}[f](\alpha) = e^{-\sum_{i=1}^d \ell_s^i} \prod_{i=1}^d \frac{1}{\sigma_{ii} s(t-s)} \cdot \frac{1}{N} \sum_{k=1}^N f(X_t^{(k)}) \prod_{i=1}^d \frac{H(Q_s^i(X_s^{(k)}) - Q_s^i(\alpha))}{Q_s^i(X_s^{(k)})} \Delta W_{s,t}^{i,(k)},$$

that is, in our context the Monte Carlo estimator $\mathbb{T}^N_{s,t}[f](\alpha)$ is the empirical mean, up to a constant term, of independent copies of the random variable

$$\mathcal{H}^{\ell_s}_{s,t}[f](\alpha) = e^{-\sum_{i=1}^d \ell_s^i} \prod_{i=1}^d \frac{1}{\sigma_{ii} s(t-s)} \cdot f(X_t) \prod_{i=1}^d \frac{H(Q_s^i(X_s) - Q_s^i(\alpha))}{Q_s^i(X_s)} \Delta W_{s,t}^i.$$

Thus, one could ask for ℓ_s to be set in order to minimize the variance of $\mathcal{H}^{\ell_s}_{s,t}[f](\alpha)$, or its integrated variance, that is

$$v[f](\ell_s) = \mathrm{Var}\Big(\mathcal{H}^{\ell_s}_{s,t}[f](\alpha)\Big) \quad \text{or} \quad v[f](\ell_s) = \mathrm{Var}\Big(\int_{\mathbb{R}^d} \mathcal{H}^{\ell_s}_{s,t}[f]\mu(d\alpha)\Big).$$

By recalling that the Q_s^i's do not depend on ℓ, we can write

$$v[f](\ell_s) = e^{-2\sum_{i=1}^d \ell_i(s)}\, v[f]$$

where $v[f]$ is independent of ℓ_s. Thus, the percentage ratios

$$\frac{\mathbb{E}\Big(\mathcal{H}^{\ell_s}_{s,t}[f](\alpha)\Big)}{\mathrm{Var}\Big(\mathcal{H}^{\ell_s}_{s,t}[f](\alpha)\Big)^{1/2}} \quad \text{and} \quad \frac{\mathbb{E}\Big(\int_{\mathbb{R}^d} \mathcal{H}^{\ell_s}_{s,t}[f](\alpha)\mu(d\alpha)\Big)}{\mathrm{Var}\Big(\int_{\mathbb{R}^d} \mathcal{H}^{\ell_s}_{s,t}[f](\alpha)\mu(d\alpha)\Big)^{1/2}}$$

do not depend on ℓ, that is the minimization of the variance with respect to ℓ is a nonsense in this context.

Another way to choose the drift ℓ is in order in order to have $G_s(\alpha) = \alpha$, which is the choice made by Lions and Regnier in [17]. The corresponding ℓ is computed in the following

Proposition 4.4 *α is a fixed point for the transformation G_s if and only if $\ell = \ell^*$, where ℓ^* is any path having value at time s given by*

$$\ell_i^*(s) = \sum_{j=1}^{i-1} \widehat{\sigma}_{ij} \Big(h_j\, s - \ln \frac{\alpha_j}{x_j}\Big), \quad i = 1, \dots, d,$$

$\widehat{\sigma}$ being defined in (17). In particular, if $\ell_t = \ell\, t$ then $G_s(\alpha) = \alpha$ if and only if $\ell = \ell^*$, where

$$\ell_i^* = \sum_{j=1}^{i-1} \widehat{\sigma}_{ij} \Big(h_j - \frac{1}{s} \ln \frac{\alpha_j}{x_j}\Big), \quad i = 1, \dots, d.$$

Proof. The proof is immediate. By (19) one can shortly write $\ln G_s(z) = \ell_s + \widehat{\sigma}\,\ln z + (I-\widehat{\sigma})(\ln x + hs)$, thus $\alpha = G_s(\alpha)$ if and only if

$$\ell_s = \ell_s^* = \Big(\widehat{\sigma} - I\Big)\Big(h\,s - \ln\frac{\alpha}{x}\Big)$$

where the symbol $\ln\frac{\alpha}{x}$ stands for a vector whose i^{th} entry is given by $\ln\frac{\alpha_i}{x_i}$.

□

Remark 4.5 *It is worth noticing that if originally σ is diagonal, then $\alpha = G_s(\alpha)$ if and only if $\ell^* = 0$. In fact, in such a case $\widetilde{\sigma} = I$ so that $\widehat{\sigma} - I = 0$, which ensures that $\ell^* = 0$, as it must obviously follow. In any case, one always has $\ell_1^* = 0$, because of the fact that $\widetilde{\sigma}_{11} = 1$. Moreover, the choice $\ell = \ell^*$ gives a considerable simplification of the function F_t and its inverse G_t. Indeed, since $\ln F_t(y) = -\widetilde{\sigma}\ell t + \widetilde{\sigma}\ln y + (I-\widetilde{\sigma})(\ln x + ht)$, one easily obtains*

$$F_t^i(y) = y_i \prod_{j=1}^{i-1}\left(\frac{y_j}{x_j}\Big(\frac{x_j}{\alpha_j}\Big)^{t/s}\right)^{\widetilde{\sigma}_{ij}}, \quad i = 1,\dots,d; \tag{24}$$

moreover, since $\ln G_s(z) = \ell s + \widehat{\sigma}\,\ln z + (I-\widehat{\sigma})(\ln x + hs)$, straightforward computations allow to conclude that

$$G_t^i(z) = z_i \prod_{j=1}^{i-1}\left(\frac{z_j}{x_j}\Big(\frac{x_j}{\alpha_j}\Big)^{t/s}\right)^{\widehat{\sigma}_{ij}}, \quad i = 1,\dots,d. \tag{25}$$

Let us now discuss about the connection with the formula given by Lions and Regnier [17], which takes into account the process $\widehat{X}$ defined as

$$\widehat{X}_t = x_i \exp\Big(h_i\,t + \sum_{j=1}^{i-1}\sigma_{ij}w_t^j + \sigma_{ii}W_t^i\Big)$$

where (as usual, $\sum_{j=1}^{0}(\cdot) := 0$ and) w_t solves the system

$$\sum_{k=1}^{i}\sigma_{ik}w_t^k = \ln\alpha_i - \ln x_i - h_i\,t, \quad i = 1,\dots,d. \tag{26}$$

It is worth remarking that $\widehat{X}$ has independent components, as well as $\widetilde{X}$. The formula given in [17] states the following: setting

$$\mathbb{L}_{s,t}[f](\alpha) = \prod_{i=1}^{d}\frac{1}{\sigma_{ii}s(t-s)}\cdot\mathbb{E}\Big(f(X_t)\prod_{i=1}^{d}\frac{H(\widehat{X}_s^i - \alpha_i)}{\widehat{X}_s^i}\,\Delta W_{s,t}^i\Big)$$

then

$$\mathbb{E}\Big(\Phi(X_t)\,|\,X_s = \alpha\Big) = \frac{\mathbb{L}_{s,t}[\Phi](\alpha)}{\mathbb{L}_{s,t}[1](\alpha)} \tag{27}$$

Thus, both formulas depend on an auxiliary process with independent components ($\widetilde{X}$ and $\widehat{X}$ respectively), which in turn is determined by the introduction of a new drift (ℓ and w respectively, the latter being defined through (26)). Furthermore, the formulas giving the operators $\mathbb{T}_{s,t}$ and

$\mathbb{L}_{s,t}$, allowing to write down the conditional expectation, are really very similar except for the point α. More precisely, both operators can be rewritten as

$$\mathbb{A}_{s,t}[f](\alpha) = \prod_{i=1}^{d} \frac{1}{\sigma_{ii} s(t-s)} \cdot \mathbb{E}\Big(f(X_t) \prod_{i=1}^{d} \frac{H(Y^i - g_i(\alpha))}{Y^i} \, \Delta W^i_{s,t} \Big),$$

in which

- choose $Y = \widetilde{X}_s$ and $g(\alpha) = G_s(\alpha)$ in order to obtain $\mathbb{A}_{s,t}[f](\alpha) = \mathbb{T}_{s,t}[f](\alpha)$;
- choose $Y = \widehat{X}_s$ and $g(\alpha) = \alpha$ in order to obtain $\mathbb{A}_{s,t}[f](\alpha) = \mathbb{L}_{s,t}[f](\alpha)$.

Thus: what is the connection between the two approaches? It is simple: if one sets $\ell^1_t = 0$ and for $i = 2, \dots, d$, $\ell^i_t = \sum_{j=1}^{i-1} \sigma_{ij} w^j_t$, then it is straightforward to see that condition (26) equals to the condition studied in Proposition 4.4, that is $\widetilde{\alpha} = G_t(\alpha) = \alpha$. This means that if ℓ_s is fixed in order to have $G_s(\alpha) = \alpha$, as it has been made in Proposition 4.4, the two formulas are actually identical. Therefore, the approach studied here is some more general than the one developed by Lions and Regnier.

Let us resume the above observations in the following

Remark 4.6 *In principle one could take ℓ arbitrarily, so that for practical purposes the simple choice $\ell(t) = \ell\, t$ seems to be good enough. Concerning the (now) constant ℓ, up to now two main choices for ℓ can be suggested:*

- *$\ell = 0$: this simplifies the process $\widetilde{X}$; the operator allowing to write down the expectation becomes in this case*

$$\mathbb{T}_{s,t}[f](\alpha) = \prod_{i=1}^{d} \frac{1}{\sigma_{ii} s(t-s)} \cdot \mathbb{E}\Big(f(X_t) \prod_{i=1}^{d} \frac{H_i(X_s, \alpha)}{X^i_s \prod_{j=1}^{i-1} \left(\frac{X^j_s}{x_j} e^{-h_j s} \right)^{\widehat{\sigma}_{ij}}} \, \Delta W^i_{s,t} \Big),$$

  where $H_i(y, \alpha) = \mathbf{1}_{\prod_{j=1}^{i} (X^j_s/\alpha_j)^{\widehat{\sigma}_{ij}} \geq 1}$.

- *$\ell = \ell^*$, with ℓ^* as in Proposition 4.4: as observed, this gives a formula for the conditional expectation in point of fact identical to the one provided by Lions and Reigner and the operator $\mathbb{T}_{s,t}[f](\alpha)$ is here*

$$\mathbb{T}_{s,t}[f](\alpha) = \prod_{i=1}^{d} \frac{e^{-\ell^*_i s}}{\sigma_{ii} s(t-s)} \cdot \mathbb{E}\Big(f(X_t) \prod_{i=1}^{d} \frac{H_i(X_s - \alpha)}{X^i_s \prod_{j=1}^{i-1} \left(\frac{X^j_s}{x_j} e^{-h_j s} \right)^{\widehat{\sigma}_{ij}}} \, \Delta W^i_{s,t} \Big),$$

  which can be rewritten as

$$\mathbb{T}_{s,t}[f](\alpha) = \prod_{i=1}^{d} \frac{\prod_{j=1}^{i-1} \alpha_j^{\widehat{\sigma}_{ij}}}{\sigma_{ii} s(t-s)} \cdot \mathbb{E}\Big(f(X_t) \prod_{i=1}^{d} \frac{H_i(X_s - \alpha)}{\prod_{j=1}^{i} (X^j_s)^{\widehat{\sigma}_{ij}}} \, \Delta W^i_{s,t} \Big).$$

Let us now discuss formulas involving localization functions. If we restrict our attention to product-type localizing function, then we can first state a localized formula for the operators $\mathbb{T}_{s,t}[f](\alpha)$ and $\mathbb{R}_{s,t;j}[f](\alpha)$ and then for the conditional expectation and its gradient. In fact, one first has

Lemma 4.7 *Let $\psi(x) = \prod_{i=1}^{d} \psi_i(x_i)$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, with $\psi_i \geq 0$ and $\int_{\mathbb{R}} \psi_i(\xi) d\xi = 1$. Then the operators $\mathbb{T}_{s,t}$ and $\mathbb{R}_{s,t;j}$, defined in (20) and (21) respectively, can be localized as follows:*

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{T}^{\psi}_{s,t}[f](\alpha) \quad \textit{and} \quad \mathbb{R}_{s,t;k}[f](\alpha) = \mathbb{R}^{\psi}_{s,t;j}[f](\alpha), \quad k = 1, \dots, d,$$

where

$$\mathbb{T}^{\psi}_{s,t}[f](\alpha) = \mathbb{E}\Big(f(X_t)\prod_{i=1}^{d}\Big[\psi_i(X_s-\alpha)) + \frac{H(\widetilde{X}^i_s-\widetilde{\alpha}_i)-\Psi_i(\widetilde{X}^i_s-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big]\Big) \tag{28}$$

and

$$\begin{aligned}\mathbb{R}^{\psi}_{s,t;k}[f](\alpha) = -\mathbb{E}\Big(f(X_t)\Big[\ \psi_k(\widetilde{X}^k_s-\widetilde{\alpha}^k)\frac{\Delta W^k_{s,t}}{\sigma_{kk}s(t-s)\widetilde{X}^k_s}+\\ +\frac{H(\widetilde{X}^k_s-\widetilde{\alpha}^k)-\Psi_k(\widetilde{X}^k_s-\widetilde{\alpha}^k)}{\sigma_{kk}s(t-s)(\widetilde{X}^k_s)^2}\Big(\frac{(\Delta W^k_{s,t})^2}{\sigma_{kk}s(t-s)}+\Delta W^k_{s,t}-\frac{t}{\sigma_{kk}}\Big)\Big]\times\\ \times\prod_{i=1,i\neq k}^{d}\Big[\psi_i(\widetilde{X}^i_s-\widetilde{\alpha}_i)+\frac{H(\widetilde{X}^i_s-\widetilde{\alpha}_i)-\Psi_i(\widetilde{X}^i_s-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big]\Big).\end{aligned} \tag{29}$$

where Ψ_i denotes the probability distribution function associated with ψ_i: $\Psi_i(y)=\int_{-\infty}^{y}\psi_i(\xi)d\xi$.

Proof. Set $\widetilde{f}(\widetilde{X}_t)=f\circ F_t(\widetilde{X}_t)$. By conditioning w.r.t. all the coordinates of $\widetilde{X}_t$ except for the first one, and by recalling that $\widetilde{X}$ has independent components, one has

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E}\Big(\mathbb{E}\Big(\widetilde{f}_{-1}(\widetilde{X}^1_t)\frac{H(\widetilde{X}^1_s-\widetilde{\alpha}_1)}{\sigma_{11}s(t-s)\widetilde{X}^1_s}\,\Delta W^1_{s,t}\Big)\Big|_{\widetilde{x}_j=\widetilde{X}^j_t,j\neq 1}\times\prod_{i=1,i\neq 1}^{d}\frac{H(\widetilde{X}^i_s-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big),$$

where we have set $\widetilde{f}_{-1}(\widetilde{X}^k_t)=\widetilde{f}(\widetilde{X}^1_t,\widetilde{x}_2,\dots\widetilde{x}_d)$. By using the one dimensional results (see Lemma 3.2), one has

$$\mathbb{E}\Big(\widetilde{f}_{-1}(\widetilde{X}^1_t)\frac{H(\widetilde{X}^1_s-\widetilde{\alpha}_1)}{\sigma_{11}s(t-s)\widetilde{X}^1_s}\,\Delta W^1_{s,t}\Big) = \mathbb{E}\Big(\widetilde{f}_{-1}(\widetilde{X}^1_t)\Big[\psi_1(X^1_s-\widetilde{\alpha}_1)+\frac{(H-\Psi)(\widetilde{X}^1_s-\widetilde{\alpha}_1)}{\sigma_{11}s(t-s)\widetilde{X}^1_s}\,\Delta W^1_{s,t}\Big]\Big),$$

so that

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E}\Big(f(X_t)\Big[\psi_1(X^1_s-\widetilde{\alpha}_1)+\frac{(H-\Psi)(\widetilde{X}^1_s-\widetilde{\alpha}_1)}{\sigma_{11}s(t-s)\widetilde{X}^1_s}\,\Delta W^1_{s,t}\Big]\prod_{i=1,i\neq 1}^{d}\frac{H(\widetilde{X}^i_s-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big).$$

Now, by considering the conditioning w.r.t. $\widetilde{X}^1_t,\widetilde{X}^3_t,\dots,\widetilde{X}^d_t$, one has

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E}\Big(f(X_t)\prod_{j=1}^{2}\Big[\psi_j(X^j_s-\widetilde{\alpha}_j)+\frac{(H-\Psi)(\widetilde{X}^j_s-\widetilde{\alpha}_j)}{\sigma_{jj}s(t-s)\widetilde{X}^j_s}\,\Delta W^j_{s,t}\Big]\prod_{i=1,i\neq 1,2}^{d}\frac{H(\widetilde{X}^i_s-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big).$$

By iterating a similar procedure on the remaining components, one arrives to the end.
Concerning $\mathbb{R}_{s,t;k}$, the statement can be proved in the same way, by using the one dimensional localization formula for $\mathbb{R}_{s,t}$ in Lemma 3.2. Indeed, by conditioning w.r.t. all the coordinates except for the kth one, one first arrives to

$$\begin{aligned}\mathbb{R}_{s,t;k}[f](\alpha) = -\mathbb{E}\Big(f(X_t)\Big[\ \psi_k(\widetilde{X}^k_s-\widetilde{\alpha}^k)\frac{\Delta W^k_{s,t}}{\sigma_{kk}s(t-s)\widetilde{X}^k_s}+\\ +\frac{H(\widetilde{X}^k_s-\widetilde{\alpha}^k)-\Psi_k(\widetilde{X}^k_s-\widetilde{\alpha}^k)}{\sigma_{kk}s(t-s)(\widetilde{X}^k_s)^2}\Big(\frac{(\Delta W^k_{s,t})^2}{\sigma_{kk}s(t-s)}+\Delta W^k_{s,t}-\frac{t}{\sigma_{kk}}\Big)\Big]\times\\ \times\prod_{i=1,i\neq k}^{d}\frac{H(\widetilde{X}^i_s-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big]\Big).\end{aligned}$$

Now, by conditioning w.r.t. all the coordinates except for the jth one, with $j \neq k$, and by using the localization properties for $\mathbb{T}_{s,t}$ as in Lemma 3.2, the final formula can be achieved.

□

By using the localized version for the operators, the localized representation formulas for the conditional expectation and its gradient immediately follows:

Theorem 4.8 (Representation formulas II: with localization) *For any $0 \leq s < t$, $F \in \mathcal{E}_b$, $\alpha \in \mathbb{R}^d_+$ and for any $\psi \in \mathcal{L}_d$, one has*

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

and, as $j = 1, \ldots, d$,

$$\partial_{\alpha_j}\mathbb{E}\Big(F(X_t)\,\Big|\,X_s = \alpha\Big) = \sum_{k=1}^{j} \widehat{\sigma}_{kj}\frac{\widetilde{\alpha}_k}{\alpha_j} \times \frac{\mathbb{R}^{\psi}_{s,t;k}[F](\alpha)\mathbb{T}^{\psi}_{s,t}[1](\alpha) - \mathbb{T}^{\psi}_{s,t}[F](\alpha)\mathbb{R}^{\psi}_{s,t;k}[1](\alpha)}{\mathbb{T}^{\psi}_{s,t}[1](\alpha)^2},$$

where the operators $\mathbb{T}^{\psi}_{s,t}[f](\alpha)$ and $\mathbb{R}^{\psi}_{s,t;k}[f](\alpha)$ are defined in (28) and (29) respectively.

Remark 4.9 *In principle, one could take different localizing functions for each operator, that is:*

$$\begin{aligned}
\mathbb{E}\Big(F(X_t)\,\Big|\,X_s = \alpha\Big) &= \frac{\mathbb{T}^{\psi_1}_{s,t}[F](\alpha)}{\mathbb{T}^{\psi_2}_{s,t}[1](\alpha)} \\
\partial_{\alpha_j}\mathbb{E}\Big(F(X_t)\,\Big|\,X_s = \alpha\Big) &= \sum_{k=1}^{j} \widehat{\sigma}_{kj}\frac{\widetilde{\alpha}_k}{\alpha_j} \times \frac{\mathbb{R}^{\psi_3}_{s,t;k}[F](\alpha)\mathbb{T}^{\psi_4}_{s,t}[1](\alpha) - \mathbb{T}^{\psi_5}_{s,t}[F](\alpha)\mathbb{R}^{\psi_6}_{s,t;k}[1](\alpha)}{\mathbb{T}^{\psi_7}_{s,t}[1](\alpha)^2}.
\end{aligned}$$

See next section for a discussion on localizing functions. Furthermore, what observed in Remark 4.3 holds here as well: when σ is diagonal, the sum giving $\partial_{\alpha_j}\mathbb{E}(F(X_t)\,|\,X_s = \alpha)$ reduces to the single term with $k = j$, with coefficient $\widetilde{\alpha}_j/\alpha_j = e^{\ell^j s}$, which in turn is equal to 1 if $\ell = 0$.

5 Optimal localizing functions

Let us conclude this "theoretical" part with an analysis on the choice of the localizing functions. Let us first discuss the one dimensional case. By referring to Theorem 3.3, in order to compute $\mathbb{E}(\Phi(X_t)\,|\,X_s = \alpha)$ one has to evaluate

$$\mathbb{T}^{\psi}_{s,t}[f](\alpha) = \mathbb{E}\Big(f(X_t)\Big[\psi(X_s - \alpha) + \frac{H(X_s - \alpha) - \Psi(X_s - \alpha)}{\sigma s(t-s)X_s}\,\Delta W_{s,t}\Big]\Big),$$

with $f = \Phi$ and $f = 1$. Such an expectation is practically evaluated by means of the empirical mean obtained through many independent replications:

$$\mathbb{T}^{\psi}_{s,t}[f](\alpha) \simeq \frac{1}{N}\sum_{q=1}^{N} f(X_t^{(q)})\Big[\psi(X_s^{(q)} - \alpha) + \frac{H(X_s^{(q)} - \alpha) - \Psi(X_s^{(q)} - \alpha)}{\sigma s(t-s)X_s^{(q)}}\,\Delta W_{s,t}^{(q)}\Big].$$

The aim is now to choose the localizing function ψ in order to reduce the variance as well as possible. To this purpose, let us introduce the quantity

$$I_1^f(\psi) = \int_{\mathbb{R}} \mathbb{E}\Big(f^2(X_t)\Big[\psi(X_s - \alpha) + \frac{H(X_s - \alpha) - \Psi(X_s - \alpha)}{\sigma s(t-s)X_s}\,\Delta W_{s,t}\Big]^2\Big)d\alpha,$$

which gives the integrated variance up to the constant (with respect to ψ) term $\mathbb{T}^{\psi}_{s,t}[f](\alpha) = \mathbb{T}_{s,t}[f](\alpha)$. Then one has

Proposition 5.1 *Setting* $\mathcal{L}_1 = \{\psi : \mathbb{R} \to [0,+\infty)\,;\, \psi \in C^1(\mathbb{R}),\ \psi(+\infty) = 0$ *and* $\int_{\mathbb{R}} \psi(t)\,dt = 1\}$, *then*

$$\inf_{\psi \in \mathcal{L}_1} I_1^f(\psi) = I_1^f(\psi^*),$$

where $\psi^* = \psi^*(\xi)$, $\xi \in \mathbb{R}$, *is a Laplace-type probability density function:*

$$\psi^*(\xi) = \frac{\lambda^*}{2}\, e^{-\lambda^*|\xi|}, \qquad \textit{with}\ \ \lambda^* = \lambda^*[f] = \left(\frac{\mathbb{E}\Big(f^2(X_t)\left(\frac{1}{\sigma s(t-s)\,X_s}\,\Delta W_{s,t}\right)^2\Big)}{\mathbb{E}\Big(f^2(X_t)\Big)} \right)^{1/2}.$$

Proof. Let us set

$$\pi_{s,t} = \frac{\Delta W_{s,t}}{\sigma s(t-s) X_s}.$$

Fix a localizing function $\psi \in \mathcal{L}_1$. First, by interchanging the order of integration and by considering the change of variable $\beta = X_s - \alpha$, one has

$$I_1^f(\psi) = \mathbb{E}\int_{\mathbb{R}} f^2(X_t)\Big(\psi(\beta) + (H-\Psi)(\beta)\,\pi_{s,t}\Big)^2\, d\beta.$$

Now, set $\varepsilon \in \mathbb{R}$ and let $\hat{\psi} \in L^1(\mathbb{R})$ such that for any small ε then $\psi + \varepsilon\hat{\psi} \in \mathcal{L}_1$. Setting $\hat{\Psi}(x) = \int_{-\infty}^{x} \hat{\psi}(t)\,dt$, one has

$$\begin{aligned}(I_1^f)'(\psi)(\hat{\psi}) &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\Big(I_1^f(\psi+\varepsilon\hat{\psi}) - I_1^f(\psi)\Big)\\ &= 2\mathbb{E}\int_{\mathbb{R}} f^2(X_t)\Big(\hat{\psi}(\beta) - \hat{\Psi}(\beta)\pi_{s,t}\Big)\Big(\psi(\beta) + (H-\Psi)(\beta)\pi_{s,t}\Big)\, d\beta.\end{aligned}$$

We show now that ψ^* is the only function in $\mathcal{L}_1$ such that $(I_1^f)'(\psi^*)(\hat{\psi}) = 0$ for any $\hat{\psi}$ satisfying the conditions above. Consider the first term of the (last) r.h.s.: by using the standard integration by parts formula (recall that $\hat{\Psi}' = \hat{\psi}$), we can write

$$\begin{aligned}&\int_{\mathbb{R}} \hat{\psi}(\beta)\, f^2(X_t)\Big(\psi(\beta) + (H-\Psi)(\beta)\pi_{s,t}(1)\Big)\Big)\, d\beta\\ &= \hat{\Psi}(\beta)\, f^2(X_t)\Big(\psi(\beta) + (H-\Psi)(\beta)\pi_{s,t}\Big)\Big|_{-\infty}^{+\infty} - \int_{\mathbb{R}} \hat{\Psi}(\beta)\, f^2(X_t)\,\partial_\beta\Big(\psi(\beta) + (H-\Psi)(\beta)\pi_{s,t}\Big)\, d\beta\end{aligned}$$

Now, since $\hat{\Psi}(\beta) \to 0$ as $\beta \to -\infty$ and ψ is a (quite smooth) probability density function, so that in particular $\psi(\beta) + (H-\Psi)(\beta) \to 0$ as $\beta \to +\infty$, the first term nullifies. Moreover, $\partial_\beta\Big(\psi(\beta) + (H-\Psi)(\beta)\pi_{s,t}\Big) = \psi'(\beta) - \psi(\beta)\pi_{s,t}$ for almost any β. Thus, we obtain

$$\begin{aligned}(I_1^f)'(\psi)(\hat{\psi}) &= -2\mathbb{E}\int_{\mathbb{R}} \hat{\Psi}(\beta)\, f^2(X_t)\left(\psi'(\beta) - \psi(\beta)\pi_{s,t} + \psi(\beta)\,\pi_{s,t} + (H-\Psi)(\beta)\pi_{s,t}^2\right) d\beta\\ &= -2\int_{\mathbb{R}} \hat{\Psi}(\beta)\,\mathbb{E}\Big(f^2(X_t)\left(\psi'(\beta) + (H-\Psi)(\beta)\pi_{s,t}^2\right)\Big)\, d\beta.\end{aligned}$$

Since $\hat{\Psi}$ is the primitive function of an almost arbitrarily integrable function, we can say that $(I_1^f)'(\psi)(\hat{\psi}) = 0$ for any $\hat{\psi}$ if and only if

$$\mathbb{E}\Big(f^2(X_t)\left((H-\Psi)(\beta)\pi_{s,t}^2 + \psi'(\beta)\right)\Big) = 0$$

for (almost) any β. Let us denote

$$a^2 = \frac{\mathbb{E}(f^2(X_t)\,\pi_{s,t}^2)}{\mathbb{E}(f^2(X_t))}$$

and $v(\beta) = \Psi(\beta)$, so that $v'(\beta) = \psi(\beta)$ and $v''(\beta) = \psi'(\beta)$. We therefore achieve the following ordinary differential equation:

$$v''(\beta) - a^2\, v(\beta) + a^2\, H(\beta) = 0,$$

whose solution is simply

$$v(\beta) = \begin{cases} c_1\, e^{a\beta} + c_2\, e^{-a\beta} + 1 & \text{if } \beta > 0 \\ c_3\, e^{a\beta} + c_4\, e^{-a\beta} & \text{if } \beta < 0, \end{cases}$$

with $c_1, \ldots, c_4 \in \mathbb{R}$ and $a = \sqrt{a^2}$. Now, by imposing that $v(+\infty) = 1$, $v(-\infty) = 0$ and $v'(\beta)$ has to be a continuous probability density function, we finally get

$$v^*(\beta) = \begin{cases} 1 - \frac{1}{2}\, e^{-a\beta} & \text{if } \beta > 0 \\ \frac{1}{2}\, e^{a\beta} & \text{if } \beta < 0, \end{cases}$$

or equivalently

$$\psi^*(\beta) = \partial_\beta v^*(\beta) = \frac{a}{2}\, e^{-a|\beta|}.$$

Thus, $(I_1^f)'(\psi^*)(\hat{\psi}) = 0$ for any $\hat{\psi} \in L^1(\mathbb{R})$. Now, in order to show that ψ^* gives the minimum, notice that for any φ such that $\psi^* + \varphi \in \mathcal{L}_1$ one has

$$I(\psi^* + \varphi) - I(\psi^*) = \mathbb{E} \int_{\mathbb{R}} f^2(X_t)(\varphi(\beta) - \Phi(\beta)\pi_{s,t})^2\, d\beta + (I_1^f)'(\psi^*)(\varphi),$$

where $\Phi(x) = \int_{-\infty}^x \varphi(z)\, dz$. Since $(I_1^f)'(\psi^*)(\varphi) = 0$ for any $\varphi \in L^1(\mathbb{R})$, then

$$I(\psi^* + \varphi) - I(\psi^*) = \mathbb{E} \int_{\mathbb{R}} f^2(X_t)(\varphi(\beta) - \Phi(\beta)\pi_{s,t})^2\, d\beta \geq 0,$$

which proves that ψ^* is actually the minimizing function.

□

Remark 5.2 *The optimal value of the parameter λ corresponding to $f = 1$ can be explicitly written (recall that x denotes the starting underlying asset price):*

$$\lambda^*[1] = x^{-1}\, e^{-hs+\sigma^2 s} \sqrt{\frac{t + \sigma^2 s(t-s)}{\sigma^2 s(t-s)}}.$$

In fact, by Proposition 5.1 one has

$$\lambda^*[1] = \frac{1}{\sigma s(t-s)} \mathbb{E}\Big(X_s^{-2}\, (\Delta W_{s,t})^2 \Big)^{1/2},$$

which can be computed. Indeed, by using the representation of X_s, it holds

$$\mathbb{E}\Big(X_s^{-2}\, (\Delta W_{s,t})^2 \Big) = x^{-2} e^{-2hs} \mathbb{E}\Big(e^{-2\sigma W_s} \Big((t-s)W_s - s(W_t - W_s) + \sigma s(t-s) \Big)^2 \Big)$$

$$= x^{-2} e^{-2hs} \mathbb{E}\Big(\mathbb{E}\Big(e^{-2\sigma W_s} \Big((t-s)W_s - s\,y + \sigma s(t-s) \Big)^2 \Big)\Big|_{y=W_t-W_s} \Big).$$

Now, since $\mathbb{E}(e^{-2\sigma W_s}) = e^{2\sigma^2 s}$, $\mathbb{E}(W_s\, e^{-2\sigma W_s}) = -2\sigma s e^{2\sigma^2 s}$ *and* $\mathbb{E}(W_s^2\, e^{-2\sigma W_s}) = s(1+4\sigma^2 s)e^{2\sigma^2 s}$, *one has*

$$\mathbb{E}\Big(e^{-2\sigma W_s} \Big((t-s)W_s - s\,y + \sigma s(t-s) \Big)^2 \Big)$$
$$= e^{2\sigma^2 s} \Big(s(t-s)^2(1+4\sigma^2 s) + s^2(\sigma(t-s)-y)^2 - 4\sigma s(t-s)(\sigma(t-s)-y) \Big).$$

By inserting $y = W_t - W_s$ *and by computing the turning out expectation, it follows that*

$$\mathbb{E}\Big(X_s^{-2}\, (\Delta W_{s,t})^2 \Big) = x^{-2} e^{-2hs+2\sigma^2 s} s(t-s)(t+\sigma^2 s(t-s)),$$

so that

$$\lambda^*[1] = x^{-1}\, e^{-hs+\sigma^2 s} \sqrt{\frac{t+\sigma^2 s(t-s)}{\sigma^2 s(t-s)}}.$$

The above optimization criterium has been introduced by Kohatsu-Higa and Petterson [16]. In principle, one could consider a measure more general than the Lebesgue one, namely to replace $d\alpha$ with $\rho(d\alpha)$ in the expression for $I_1^f(\psi)$, but in such a case it is not possible to write down explicitly the optimal localizing function.
Such an approach can be generalized to the multidimensional case, where the functional I_1^f has to be obviously replaced by

$$I_d^f(\psi) = \int_{\mathbb{R}^d} \mathbb{E}\Big(f^2(X_t) \prod_{i=1}^d \Big[\psi_i(\widetilde{X}_s^i - \widetilde{\alpha}_i) + \frac{(H-\Psi_i)(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii} s(t-s)\widetilde{X}_s^i}\, \Delta W_{s,t}^i \Big]^2 \Big)\, d\widetilde{\alpha}.$$

Then the following result holds:

Proposition 5.3 *Setting* $\mathcal{L}_d = \{\psi : \mathbb{R}^d \to [0,+\infty)\,;\, \psi(x) = \prod_{i=1}^d \psi_i(x_i),$ *where* $\psi_i \in \mathcal{L}_1$, *for any* $i\}$, *then*

$$\inf_{\psi \in \mathcal{L}_d} I_d^f(\psi) = I_d^f(\psi^*)$$

where $\psi^*(\xi) = \prod_{j=1}^d \psi_j^*(\xi_j)$, $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$, *with* $\psi_j^*(\xi_j) = \frac{\lambda_j^*}{2}\, e^{-\lambda_j^* |\xi_j|}$, $\xi_j \in \mathbb{R}$ *and* $\lambda_j^* = \lambda_j^*[f]$, *enjoys the following system of nonlinear equations:*

$$\lambda_j^{*2} = \frac{\mathbb{E}\Big(f^2(X_t) \Big[\frac{\Delta W_{s,t}^j}{\sigma_{jj} s(t-s)\widetilde{X}_s^j} \Big]^2 \prod_{i\,:\,i\neq j} \Big[\lambda_i^{*2} + \Big(\frac{\Delta W_{s,t}^i}{\sigma_{ii} s(t-s)\widetilde{X}_s^i} \Big)^2 \Big] \Big)}{\mathbb{E}\Big(f^2(X_t) \prod_{i\,:\,i\neq j} \Big[\lambda_i^{*2} + \Big(\frac{\Delta W_{s,t}^i}{\sigma_{ii} s(t-s)\widetilde{X}_s^i} \Big)^2 \Big] \Big)}, \quad j = 1, \dots, d.$$

Proof. First, we can say that $I_d^f(\psi) = I_d^f(\psi_1, \dots, \psi_d)$, with $\psi_1, \dots, \psi_d \in \mathcal{L}_1$. We will then compute the derivative in the direction $(\widehat{\psi}_1, 0, \dots, 0)$, where $\widehat{\psi}_1 = \widehat{\psi}_1(x_1)$ is some arbitrary function, by reducing the computation to the one-dimensional case, as in Proposition 5.1. For a fixed f, let us put $\widetilde{f}_t(y) \equiv \widetilde{f}(y) = f \circ F_t(y)$, $y \in \mathbb{R}_+^d$ and F_t being defined in (18). Setting

$$\widetilde{f}_1^2(x_1) = \int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_2 \dots d\widetilde{\alpha}_d \mathbb{E}\Big(\widetilde{f}^2(x_1, \widetilde{X}_t^2, \dots, \widetilde{X}_t^d) \prod_{i=2}^d \Big[\psi_i(\widetilde{X}_s^i - \widetilde{\alpha}_i) + \frac{(H-\Psi_i)(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii} s(t-s)\widetilde{X}_s^i}\, \Delta W_{s,t}^i \Big]^2 \Big)$$

then

$$I_d^f(\psi) = \int_{\mathbb{R}} d\widetilde{\alpha}_1\, \mathbb{E}\Big(\widetilde{f}_1^2(\widetilde{X}_t^1)\, \Big[\psi_1(\widetilde{X}_s^1 - \widetilde{\alpha}_1) + \frac{(H-\Psi_1)(\widetilde{X}_s^1 - \widetilde{\alpha}_1)}{\sigma_{11} s(t-s)\widetilde{X}_s^1}\, \Delta W_{s,t}^1\Big]^2\Big) = I_1^{f_1}(\psi_1).$$

By employing the one dimensional results, we get

$$\frac{\partial I_d^f(\psi)}{\partial \widehat{\psi}_1}(\widehat{\psi}_1, 0, \dots, 0) = \frac{\partial I_1^{f_1}(\psi_1)}{\partial \widehat{\psi}_1}(\widehat{\psi}_1)$$
$$= -2 \int_{\mathbb{R}} d\beta\, \widehat{\Psi}_1(\beta)\, \mathbb{E}\Big(\widetilde{f}_1^2(\widetilde{X}_t^1)\, \Big(\psi_1'(\beta) + (H-\Psi_1)(\beta)\, \Big[\frac{\Delta W_{s,t}^1}{\sigma_{11} s(t-s)\widetilde{X}_s^1}\Big]^2\Big)\Big),$$

where $\widehat{\Psi}_1(y) = \int_{-\infty}^y \widehat{\psi}_1(\xi)\, d\xi$. We solve the equation

$$\frac{\partial I_d^f(\psi)}{\partial \widehat{\psi}_1}(\widehat{\psi}_1, 0, \dots, 0) = 0, \qquad \text{for any } \widehat{\psi}_1 :$$

since $\widehat{\psi}_1$ is arbitrary, $\widehat{\Psi}_1$ is also arbitrary, so that the above equation gives

$$\mathbb{E}\Big(\widetilde{f}_1^2(\widetilde{X}_t^1)\, \Big(\psi_1'(\beta) + (H-\Psi_1)(\beta)\, \Big[\frac{\Delta W_{s,t}^1}{\sigma_{11} s(t-s)\widetilde{X}_s^1}\Big]^2\Big)\Big) = 0, \qquad \text{for any } \beta.$$

This gives rise to an ordinary differential equation which has been studied in the proof of Proposition 5.1: the solution is then

$$\psi_1^*(\beta) = \frac{\lambda_1^*}{2}\, e^{-\lambda_1^*\, |\beta|}, \qquad \beta \in \mathbb{R},$$

with

$$\lambda_1^{*2} = \lambda_1^*(f_1)^2 = \frac{\mathbb{E}\Big(\widetilde{f}_1^2(\widetilde{X}_t^1)\, \Big[\frac{\Delta W_{s,t}^1}{\sigma_{11} s(t-s)\widetilde{X}_s^1}\Big]^2\Big)}{\mathbb{E}\Big(\widetilde{f}_1^2(\widetilde{X}_t^1)\Big)}$$
$$= \frac{\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_2 \dots d\widetilde{\alpha}_d \mathbb{E}\Big(f^2(X_t)\, \Big[\frac{\Delta W_{s,t}^1}{\sigma_{11} s(t-s)\widetilde{X}_s^1}\Big]^2 \prod_{i=2}^d \Big[\psi_i(\widetilde{X}_s^i - \widetilde{\alpha}_i) + \frac{(H-\Psi_i)(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii} s(t-s)\widetilde{X}_s^i}\, \Delta W_{s,t}^i\Big]^2\Big)}{\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_2 \dots d\widetilde{\alpha}_d \mathbb{E}\Big(f^2(X_t) \prod_{i=2}^d \Big[\psi_i(\widetilde{X}_s^i - \widetilde{\alpha}_i) + \frac{(H-\Psi_i)(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii} s(t-s)\widetilde{X}_s^i}\, \Delta W_{s,t}^i\Big]^2\Big)}$$

Now, since everything is symmetric, we obtain similar equations for the remaining coordinates $2, \dots, d$. Now, if we want all these equations to hold simultaneously, we obtain the following system:

$$\psi_j^*(\beta) = \tfrac{\lambda_j^*}{2}\, e^{-\lambda_j^*\, |\beta|}, \qquad \beta \in \mathbb{R},$$
$$\lambda_j^{*2} = \frac{\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-j}\, \mathbb{E}\Big(f^2(X_t)\, \Big[\frac{\Delta W_{s,t}^j}{\sigma_{jj} s(t-s)\widetilde{X}_s^j}\Big]^2 \prod_{i\,:\,i\neq j} \Big[\psi_i^*(\widetilde{X}_s^i - \widetilde{\alpha}_i) + \frac{(H-\Psi_i^*)(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii} s(t-s)\widetilde{X}_s^i}\, \Delta W_{s,t}^i\Big]^2\Big)}{\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-j}\, \mathbb{E}\Big(f^2(X_t) \prod_{i\,:\,i\neq j} \Big[\psi_i^*(\widetilde{X}_s^i - \widetilde{\alpha}_i) + \frac{(H-\Psi_i^*)(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii} s(t-s)\widetilde{X}_s^i}\, \Delta W_{s,t}^i\Big]^2\Big)}$$

where the notation $d\widetilde{\alpha}_{-j}$ means that one has to integrate with respect to all the variables except for $\widetilde{\alpha}_j$. It remains now to give a better representation of the above integrals. They can be rewritten in the following simpler form: setting

$$\Theta_{s,t;i} = \frac{\Delta W_{s,t}^i}{\sigma_{ii} s(t-s)\widetilde{X}_s^i},$$

then one has to handle something like

$$\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-j}\, \mathbb{E}\Big(f^2(X_t)\,\pi^2_{s,t;j} \prod_{i\,:\,i\neq j} \Big[\psi^*_i(\widetilde{X}^i_s-\widetilde{\alpha}_i)+(H-\Psi^*_i)(\widetilde{X}^i_s-\widetilde{\alpha}_i)\Theta_{s,t;i}\Big]^2\Big)$$
$$= \mathbb{E}\Big(f^2(X_t)\,\pi^2_{s,t;j} \int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-j} \prod_{i\,:\,i\neq j} \Big[\psi^*_i(\widetilde{X}^i_s-\widetilde{\alpha}_i)+(H-\Psi^*_i)(\widetilde{X}^i_s-\widetilde{\alpha}_i)\Theta_{s,t;i}\Big]^2\Big),$$

with $\pi_{s,t;j}=\Theta_{s,t;j}$ or $\pi_{s,t;j}=1$. Let us consider the integral inside the expectation: by recalling that $\psi^*_i(\xi)=\lambda^*_i\, e^{-\lambda^*_i\,|\xi|}/2$, then $(H-\Psi^*_i)(\xi)=\mathrm{sign}(\xi)\, e^{-\lambda^*_i\,|\xi|}/2$, so that

$$\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-j} \Big[\psi^*_i(\widetilde{X}^i_s-\widetilde{\alpha}_i)+(H-\Psi^*_i)(\widetilde{X}^i_s-\widetilde{\alpha}_i)\Theta_{s,t;i}\Big]^2$$
$$= \prod_{i\,:\,i\neq j} \int_{\mathbb{R}} d\widetilde{\alpha}_i\, \frac{1}{4}\, e^{-2\lambda^*_i\,|\widetilde{X}^i_s-\widetilde{\alpha}_i|} \Big[\lambda^*_i+\mathrm{sign}(\widetilde{X}^i_s-\widetilde{\alpha}_i)\,\Theta_{s,t;i}\Big]^2 = \prod_{i\,:\,i\neq j} \frac{1}{4\lambda^*_i}[{\lambda^*_i}^2+\Theta^2_{s,t;i}]$$

By inserting everything, one obtains

$$\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-j}\, \mathbb{E}\Big(f^2(X_t)\,\pi^2_{s,t;j} \prod_{i\,:\,i\neq j} \Big[\psi^*_i(\widetilde{X}^i_s-\widetilde{\alpha}_i)+(H-\Psi^*_i)(\widetilde{X}^i_s-\widetilde{\alpha}_i)\Theta_{s,t;i}\Big]^2\Big)$$
$$= \mathbb{E}\Big(f^2(X_t)\,\pi^2_{s,t;j} \prod_{i\,:\,i\neq j} \frac{1}{4\lambda^*_i}[{\lambda^*_i}^2+\Theta^2_{s,t;i}]\Big)$$

and the statement immediately follows.

□

Remark 5.4 *If $\ell=0$ (see Remark 4.6 for details), for $f=1$ the corresponding optimal values of the parameters λ_j are given by (recall that $x_1,\dots,x_d$ are the starting underlying asset prices)*

$$\lambda^*_j[1]=x_j^{-1}\, e^{-h_j s+\sigma^2_{jj}s}\sqrt{\frac{t+\sigma^2_{jj}s(t-s)}{\sigma^2_{jj}s(t-s)}}, \quad j=1,\dots,d.$$

This is an immediate consequence of the computations made in Remark 5.2.

It is worth to point out that similar arguments could be used in order to handle the problem of minimizing the variance coming out from the expectation giving the operator $\mathbb{R}^{\psi}_{s,t}[f](\alpha)$ and $\mathbb{R}^{\psi}_{s,t;k}[f](\alpha)$ (see (7) and (28)). To this purpose, let us recall that

$$\mathbb{R}^{\psi}_{s,t;k} = -\mathbb{E}\Big(f(X_t)\Big[\psi_k(\widetilde{X}^k_s-\widetilde{\alpha}^k)\Theta_{s,t;k}+(H-\Psi_k)(\widetilde{X}^k_s-\widetilde{\alpha}^k)\Upsilon_{s,t;k}\Big]\pi_{s,t;k}((\widetilde{X}_s-\widetilde{\alpha})_{-k})\Big),$$

where

$$\Theta_{s,t;k}=\frac{\Delta W^k_{s,t}}{\sigma_{kk}s(t-s)\widetilde{X}^k_s}, \quad \Upsilon_{s,t;k}=\frac{1}{\sigma_{kk}s(t-s)(\widetilde{X}^k_s)^2}\Big(\frac{(\Delta W^k_{s,t})^2}{\sigma_{kk}s(t-s)}+\Delta W^k_{s,t}-\frac{t}{\sigma_{kk}}\Big) \tag{30}$$

and

$$\pi_{s,t;k}((\widetilde{X}_s-\widetilde{\alpha})_{-k}) = \prod_{i=1,i\neq k}^{d} \Big(\psi_i(\widetilde{X}^i_s-\widetilde{\alpha}_i)+\frac{H(\widetilde{X}^i_s-\widetilde{\alpha}_i)-\Psi_i(\widetilde{X}^i_s-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big)$$

where, for $\beta \in \mathbb{R}^d$, the notation β_{-k} means the vector on $\mathbb{R}^{d-1}$ having the same coordinates of β except for the kth one. In the one dimensional case, one simply has to drop the subscript k, take $\widetilde{\alpha} = \alpha$ and set $\pi_{s,t;k}(\widetilde{\alpha}_{-k}) \equiv 1$. Thus, if we set

$$J_d^{f;k}(\psi) = \int_{\mathbb{R}^d} \mathbb{E}\Big(f^2(X_t)\Big[\psi_k(\widetilde{X}_s^k - \widetilde{\alpha}^k)\Theta_{s,t;k} + (H - \Psi_k)(\widetilde{X}_s^k - \widetilde{\alpha}^k)\Upsilon_{s,t;k}\Big]^2 \pi^2_{s,t;k}((\widetilde{X}_s - \widetilde{\alpha})_{-k})\Big)\, d\widetilde{\alpha}$$

then the following result holds:

Proposition 5.5 *Setting $\mathcal{L}_d = \{\psi : \mathbb{R}^d \to [0, +\infty)\,;\, \psi(x) = \prod_{i=1}^d \psi_i(x_i),$ where $\psi_i \in \mathcal{L}_1$, for any $i\}$, then*

$$\inf_{\psi \in \mathcal{L}_d} J_d^{f;k}(\psi) = J_d^{f;k}(\psi^{*;k})$$

where $\psi^{*;k}(\xi) = \prod_{j=1}^d \psi^*_{j;k}(\xi_j)$, $\xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d$, with $\psi^*_{j;k}(\xi_j) = \frac{\mu^*_{j;k}}{2}\, e^{-\mu^*_{j;k}\,|\xi_j|}$, $\xi_j \in \mathbb{R}$ and $\mu^*_{j;k} = \mu^*_{j;k}[f]$, are given by:

i) in the one dimensional case ($k = j = 1$), one has:

$$\mu^* = \mu^*[f] = \left(\frac{\mathbb{E}\Big(f^2(X_t)\,\Upsilon^2_{s,t}\Big)}{\mathbb{E}\Big(f^2(X_t)\,\Theta^2_{s,t}\Big)}\right)^{1/2}$$

where $\Upsilon_{s,t}$ and $\Theta_{s,t}$ are as in (30);

ii) in the multidimensional case, for any fixed $k = 1, \ldots, d$, the $\mu^*_{j;k}$'s solve the nonlinear system, as $j = 1, \ldots, d$:

$$(\mu^*_{k;k})^2 = \frac{\mathbb{E}\Big(f^2(X_t)\,\Upsilon^2_{s,t;k}\,\prod_{i=1, i\neq k}^d \Big[(\mu^*_{i;k})^2 + \Theta^2_{s,t;j}\Big]\Big)}{\mathbb{E}\Big(f^2(X_t)\,\Theta^2_{s,t;i}\,\prod_{i=1, i\neq k}^d \Big[(\mu^*_{i;k})^2 + \Theta^2_{s,t;i}\Big]\Big)} \quad \textit{and for } j \neq k:$$

$$(\mu^*_{j;k})^2 = \frac{\mathbb{E}\Big(f^2(X_t)\,\Big((\mu^*_{k;k})^2\Theta^2_{s,t;k} + \Upsilon^2_{s,t;k}\Big)\Theta^2_{s,t;j}\,\prod_{i=1, i\neq j,k}^d \Big[(\mu^*_{i;k})^2 + \Theta^2_{s,t;i}\Big]\Big)}{\mathbb{E}\Big(f^2(X_t)\Big((\mu^*_{k;k})^2\Theta^2_{s,t;k} + \Upsilon^2_{s,t;k}\Big)\prod_{i=1, i\neq j,k}^d \Big[(\mu^*_{i;k})^2 + \Theta^2_{s,t;i}\Big]\Big)}$$

where $\Upsilon_{s,t;k}$ and $\Theta_{s,t;k}$ are defined in (30).

In the proof, the same technique as in the proof Proposition 5.1 and 5.3 is used. Since it is also quite tedious, it is postponed to the Appendix.
Anyway, for practical purposes, numerical evidences show that the choice $\lambda^* = 1/\sqrt{t-s}$ works good enough (thus avoiding to weight the algorithm with the computation of further expectations).

Finally, let us conclude with a short consideration. For simplicity, let us consider the one dimensional case. The main problem is a good estimate of $\mathbb{E}(\Phi(X_t)\,|\,X_s = \alpha)$, which can be written as the ratio between $\mathbb{T}^{\psi}_{s,t}[\Phi](\alpha)$ and $\mathbb{T}^{\psi}_{s,t}[1](\alpha)$ but also in the following way:

$$\mathbb{E}(\Phi(X_t)\,|\,X_s = \alpha) = \mathbb{E}\Big(\Phi(X_t)\frac{\pi^{\psi}_{s,t}}{\mathbb{E}(\pi^{\psi}_{s,t})}\Big), \qquad \pi^{\psi}_{s,t} = \psi(X_s - \alpha) + \frac{H(X_s - \alpha) - \Psi(X_s - \alpha)}{\sigma s(t-s)X_s}\,\Delta W_{s,t}$$

(one could also complicate things by considering two different localizing functions in the above ratio...). So, another reasonable way to proceed might take into account the variance coming out from the weight $\pi^{\psi}_{s,t}$. But since it is written in terms of a ratio, at this stage it does not seem reasonably feasible to obtain results giving the associated optimal localizing function ψ.

6 The algorithm for the pricing of American options

We give here first a detailed presentation of the use of the representation formulas in the applied context of the pricing and hedging of American options. Secondly, we summarize the pricing/hedging algorithm.

6.1 How to use the formulas in practice

The algorithm is devoted to the numerical evaluation of the price $P(0,x)$ and the delta $\Delta(0,x)$ of an American option with payoff function Φ and maturity T, on underlying assets whose price X evolves following the Black-Scholes model, that is as in (14). It has been briefly described in the Introduction, let us now go into the details.

Let $0 = t_0 < t_1 < \ldots < t_n = T$ be a discretization of the time interval $[0,T]$, with step size equal to $\varepsilon = T/n$. By using Theorem 2.1 and Proposition 2.2, the price $P(0,x)$ is approximated by means of $\bar{P}_0(x)$, where $\bar{P}_{k\varepsilon}(X_{k\varepsilon})$, as $k = 0,1,\ldots,n$, is iteratively defined as:

$$\begin{aligned} & \bar{P}_{n\varepsilon}(X_{n\varepsilon}) = \Phi(X_{n\varepsilon}) \equiv \Phi(X_T) \\ k = n-1,\ldots,1,0: \qquad & \bar{P}_{k\varepsilon}(X_{k\varepsilon}) = \max\Big\{\Phi(X_{k\varepsilon})\,,\, e^{-r\varepsilon}\,\mathbb{E}\Big(\bar{P}_{(k+1)\varepsilon}(X_{(k+1)\varepsilon})\,\Big|\,X_{k\varepsilon}\Big)\Big\} \end{aligned} \tag{31}$$

and the delta $\Delta(0,x)$ is approximated by using the following plan:

$$\begin{aligned} \text{setting} \quad \bar{\Delta}(X_\varepsilon) &= \begin{cases} \partial_\alpha\,\Phi(\alpha)\Big|_{\alpha=X_\varepsilon} & \text{if} \quad \bar{P}_\varepsilon(X_\varepsilon) < \Phi(X_\varepsilon) \\ e^{-r\varepsilon}\,\partial_\alpha\,\mathbb{E}\Big(\bar{P}_{2\varepsilon}(X_{2\varepsilon})\,\Big|\,X_\varepsilon = \alpha\Big)\Big|_{\alpha=X_\varepsilon} & \text{if} \quad \bar{P}_\varepsilon(X_\varepsilon) > \Phi(X_\varepsilon) \end{cases} \\ \text{then} \quad \bar{\Delta}_0(x) &= \mathbb{E}_x\Big(\bar{\Delta}(X_\varepsilon)\Big). \end{aligned} \tag{32}$$

The conditional expectation $\mathbb{E}(\bar{P}_{(k+1)\varepsilon}(X_{(k+1)\varepsilon})\,|\,X_{k\varepsilon})$ and the derivative $\partial_\alpha\,\mathbb{E}(\bar{P}_{2\varepsilon}(X_{2\varepsilon})\,|\,X_\varepsilon = \alpha)|_{\alpha=X_\varepsilon}$ will be computed through the formulas given in the previous section, by means of suitable empirical means evaluated over N simulated paths.

Remark 6.1 *In the context of the geometric Brownian motion, the process X can be exactly simulated at each instant $t_k = k\varepsilon$. So, in this particular case we do not need an approximation $\bar{X}_{k\varepsilon}$ of $X_{k\varepsilon}$, and thus we write directly $X_{k\varepsilon}$. Furthermore, it is worth remarking that the algorithm allows to use the same sample in order to simulate all the involved conditional expectations, as it will follows from the next description.*

As $k = n, n-1, \ldots, 0$ we need

$$X^i_{k\varepsilon} = x_i\, e^{(r-\eta_i-\frac{1}{2}\sum_{j=1}^i \sigma_{ij}^2)k\varepsilon + \sum_{j=1}^i \sigma_{ij} W^j_{k\varepsilon}}, \quad i = 1,\ldots,d.$$

In order to have $X_{k\varepsilon}$, we simply need $W_{k\varepsilon}$. Since the algorithm is of backward-type, for the above simulation we consider the following backward approach, which uses the Brownian bridge law. Indeed, at time $T = n\varepsilon$, we can simulate $W_{n\varepsilon}$ in the classical way:

$$W_{n\varepsilon} = \sqrt{n\varepsilon}\,U_n, \quad \text{with } U_n = (U_n^1,\ldots,U_n^d), \quad U_n^i \sim \mathrm{N}(0,1), i = 1,\ldots,d, \text{ independent},$$

which gives $X_{n\varepsilon}$. Now, in order to simulate $X_{(n-1)\varepsilon}$, we need $W_{(n-1)\varepsilon}$, which in turn can be simulated by using the Brownian bridge: since $W_{n\varepsilon}$ is known, $W_{(n-1)\varepsilon}$ can be simulated by using the conditional law of $W_{(n-1)\varepsilon}$ given the observed value for $W_{n\varepsilon}$. It is well known that the law

of W_s given that $W_t = y$ for $0 < s < t$ is given by a gaussian law with mean $s/t\, y$ and variance $s(t-s)/t\, I$. Thus,

$$W_{(n-1)\varepsilon} = \frac{n-1}{n} W_{n\varepsilon} + \sqrt{\frac{n-1}{n}\varepsilon}\, U_{n-1}$$

with $U_{n-1} = (U^1_{n-1}, \dots, U^d_{n-1})$, $U^i_{n-1} \sim \mathrm{N}(0,1)$, $i = 1, \dots, d$, are all independent. Obviously, we can proceed similarly for the simulation of $W_{k\varepsilon}$, as $k = n-2, \dots, 1$:

$$W_{k\varepsilon} = \frac{k}{k+1} W_{(k+1)\varepsilon} + \sqrt{\frac{k}{k+1}\varepsilon}\, U_k,$$

with $U_k = (U^1_k, \dots, U^d_k)$, $U^i_k \sim \mathrm{N}(0,1)$, $i = 1, \dots, d$, independent. Thus, the basic data in the algorithm are given by

$$\mathcal{U} = \{U^{i,q}_k\,;\, k = 1, \dots, n \text{ (time)},\, i = 1, \dots, d \text{ (dimension)},\, q = 1, \dots, N \text{ (sample)}\} \tag{33}$$

and the simulation algorithm can be summarized step by step as follows.

1. Computation of the samples $(W^{i,q}_{k\varepsilon})_{i=1,\dots,d;\, k=1,\dots,n}$, $q = 1, \dots, N$: for any fixed sample $q = 1, \dots, N$, set

$$\begin{aligned} &\text{for } k = n: \quad W^{i,q}_{n\varepsilon} = \sqrt{n\varepsilon}\, U^{i,q}_n, \quad i = 1, \dots, d, \text{ and} \\ &\text{for } k = n-1, \dots, 1: \quad W^{i,q}_{k\varepsilon} = \frac{k}{k+1} W^{i,q}_{(k+1)\varepsilon} + \sqrt{\frac{k}{k+1}\varepsilon}\, U^{i,q}_k, \quad i = 1, \dots, d. \end{aligned} \tag{34}$$

2. Computation of the samples $(X^{i,q}_{k\varepsilon})_{i=1,\dots,d;\, k=0,\dots,n}$, $q = 1, \dots, N$: for any fixed sample $q = 1, \dots, N$, set for $k = n, n-1, \dots, 1$

$$X^{i,q}_{k\varepsilon} = x_i\, e^{(r-\eta_i-\frac{1}{2}\sum_{j=1}^i \sigma^2_{ij})k\varepsilon + \sum_{j=1}^i \sigma_{ij} W^{i,q}_{k\varepsilon}}, \quad i = 1, \dots, d \tag{35}$$

and $X^{i,q}_0 = x_i$, $i = 1, \dots, d$. As an example, Figure 1 shows a set of simulated paths of X.

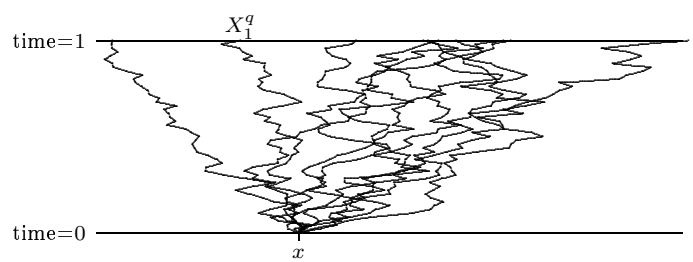

Figure 1 An example of the tree turning out by simulating 10 paths of the process X on $[0,1]$.

3. Computation of $(\widetilde{X}^{i,q}_{k\varepsilon})_{i=1,\dots,d;\, k=0,\dots,n}$, $q = 1, \dots, N$, allowing to numerically evaluate the conditional expectations involved in (31). In order to do this, if $d > 1$ one needs to introduce the drift ℓ, which could vary according to the time interval of interest $[k\varepsilon, (k+1)\varepsilon]$, that is one has something like ℓ^i_k, for $k = 1, \dots, n-1$ (time) and $i = 1, \dots, d$ (dimension). As a first stage, the ℓ^i_k's can be chosen arbitrarily. But since one could also choose $\ell_k = (\ell^1_k, \dots, \ell^d_k)$ depending on the position of X at time $k\varepsilon$, the weight ℓ_k could also depend on the qth sample, as suggested in Remark 4.6. So let us consider the set

$$\mathcal{L} = \{\ell^{i,q}_k\,;\; k = 1, \dots, n-1 \text{ (time)},\, i = 1, \dots, d \text{ (dimension)},\, q = 1, \dots, N \text{ (sample)}\}. \tag{36}$$

Once $\mathcal{L}$ is given or computed, one can compute the sample $(\widetilde{X}^{i,q}_{k\varepsilon})_{i=1,\dots,d;\,k=0,\dots,n}$: for any fixed sample $q=1,\dots,N$, set for $k=n,n-1,\dots,1$

$$\widetilde{X}^{i,q}_{k\varepsilon} = x_i\, e^{(r-\eta_i-\frac{1}{2}\sum_{j=1}^{i}\sigma_{ij}^2)k\varepsilon+\ell_k^{i,q}\,k\varepsilon+\sigma_{ii}\,W^{i,q}_{k\varepsilon}}, \quad i=1,\dots,d, \tag{37}$$

and $\widetilde{X}^{i,q}_0 = x_i$, $i=1,\dots,d$.

Let us point out some remarks:

(a) in the one dimensional case, the auxiliary process $\widetilde{X}$ does not need, so one can drop this computation or else set $\widetilde{X}^q_{k\varepsilon} = X^q_{k\varepsilon}$;

(b) each $\widetilde{X}^{i,q}_k$ has to be evaluated in terms of its corresponding $\ell^{i,q}_k$, $q=1,\dots,N$;

(c) there are two main differences between $X^{i,q}_{k\varepsilon}$ and $\widetilde{X}^{i,q}_{k\varepsilon}$:

  i. $\widetilde{X}^{i,q}_{k\varepsilon}$ does not contain $\sum_{j=1}^{i}\sigma_{ij}W^j_{k\varepsilon}$ but only $\sigma_{ii}W^i_{k\varepsilon}$. As a consequence, $\widetilde{X}^{1,q}_{k\varepsilon},\dots,\widetilde{X}^{d,q}_{k\varepsilon}$ are independent, for any fixed q and k;

  ii. $\widetilde{X}^{i,q}_{k\varepsilon}$ contains the term ℓ^i_k which does not appear in $X^{i,q}_{k\varepsilon}$.

4. Computation of the weights $\{\Delta W^{i,q}_k\}_{i=1,\dots,d;\,k=1,\dots,n-1}$, $q=1,\dots,N$, defined as

$$\Delta W^{i,q}_k := \Delta W^{i,q}_{t_k t_{k+1}} = (t_{k+1}-t_k)W^{i,q}_{t_k} - t_k(W^{i,q}_{t_{k+1}} - W^{i,q}_{t_k}) + t_k(t_{k+1}-t_k)\sigma_{ii}.$$

So, for any fixed sample $q=1,\dots,N$, set for $k=n-1,\dots,1$

$$\Delta W^{i,q}_k = \varepsilon W^{i,q}_{k\varepsilon} - k\varepsilon(W^{i,q}_{(k+1)\varepsilon} - W^{i,q}_{k\varepsilon}) + k\varepsilon^2\sigma_{ii}, \quad 1,\dots,d. \tag{38}$$

Once we have all the previous ingredients, we can proceed to the computation of $\bar{P}_{k\varepsilon}(X_{k\varepsilon})$ given by (31). To this purpose, let us set

$$\mathcal{P}_k[F](X^q_{k\varepsilon}) = \mathbb{E}\Big(F(X_{(k+1)\varepsilon})\,\Big|\,X_{k\varepsilon}=\alpha\Big)\Big|_{\alpha=X^q_{k\varepsilon}}$$

where

$$X^q_{k\varepsilon} = (X^{1,q}_{k\varepsilon},\dots,X^{d,q}_{k\varepsilon}) = \text{the } q^{\text{th}} \text{ sample, given by (35).}$$

which has to be considered here as a *datum*. Notice that for each fixed (time) k, we have a random space grid $X^q_{k\varepsilon} = (X^{1,q}_{k\varepsilon},\dots,X^{d,q}_{k\varepsilon}) \in \mathbb{R}^d$ and we compute $\mathcal{P}_k[F](X^q_{k\varepsilon})$ for each point of the grid. By Theorem 4.2, $\mathcal{P}_k[F](X^q_{k\varepsilon})$ is given by

$$\mathcal{P}_k[F](X^q_{k\varepsilon}) = \left.\frac{\mathbb{E}\Big(F(X_{(k+1)\varepsilon})\displaystyle\prod_{i=1}^{d}\frac{H(\widetilde{X}^i_{k\varepsilon}-\widetilde{\alpha}_i)}{\widetilde{X}^i_{k\varepsilon}}\,\Delta W^i_{k\varepsilon\,(k+1)\varepsilon}\Big)}{\mathbb{E}\Big(\displaystyle\prod_{i=1}^{d}\frac{H(\widetilde{X}^i_{k\varepsilon}-\widetilde{\alpha}_i)}{\widetilde{X}^i_{k\varepsilon}}\,\Delta W^i_{k\varepsilon\,(k+1)\varepsilon}\Big)}\right|_{\widetilde{\alpha}=\widetilde{X}^q_{k\varepsilon}}$$

where, for $\xi\in\mathbb{R}$, $H(\xi)=1$ if $\xi\geq 0$ and $h(\xi)=0$ otherwise. In this practical context, such expectations are computed and thus replaced by the associated empirical means, that is we set in practice

$$\mathcal{P}_k[F](X^q_{k\varepsilon}) = \frac{\displaystyle\sum_{q'=1}^{N}F(X^{q'}_{(k+1)\varepsilon})\prod_{i=1}^{d}\frac{H(\widetilde{X}^{i,q'}_{k\varepsilon}-\widetilde{X}^{i,q}_{k\varepsilon})}{\widetilde{X}^{i,q'}_{k\varepsilon}}\,\Delta W^{i,q'}_k}{\displaystyle\sum_{q'=1}^{N}\prod_{i=1}^{d}\frac{H(\widetilde{X}^{i,q'}_{k\varepsilon}-\widetilde{X}^{i,q}_{k\varepsilon})}{\widetilde{X}^{i,q'}_{k\varepsilon}}\,\Delta W^{i,q'}_k} \tag{39}$$

Here, $\widetilde{X}^{i,q'}_{k\varepsilon}$ and $\Delta W^{i,q'}_k$ are computed through (37) and (38) respectively. Finally, we can set up the dynamic programming principle:

$$u_n(X^q_{n\varepsilon}) = \Phi(X^q_{n\varepsilon}),\ q = 1,\dots,N$$
$$\text{for } k = n-1,\dots,1,0 \text{ then} \quad u_k(X^q_{k\varepsilon}) = \max\Big(\Phi(X^q_{k\varepsilon}), \mathcal{P}_k[u_{k+1}](X^q_{k\varepsilon})\Big),\ q = 1,\dots,N.$$

Obviously, u_k gives the Monte Carlo estimate for $\bar{P}_{k\varepsilon}$ in (31) and finally the price $\bar{P}_0$ is approximated by $u_0(X^q_0) \equiv u_0(x) = \max(\Phi(x), \mathcal{P}_0[u_1](X^q_\varepsilon))$, where in practice we set

$$\mathcal{P}_0[u_1](x) = \frac{1}{N}\sum_{q=1}^N u_1(X^q_\varepsilon).$$

This procedure gives the price. Concerning the delta, everything starts at the final steps, that is when time ε is considered. Indeed, by (32) we can approximate $\bar{\Delta}_0(x) = (\bar{\Delta}_{0;1}(x),\dots,\bar{\Delta}_{0;d}(x))$ through its Monte Carlo estimate $v_0(x) = (v_{0;1}(x),\dots,v_{0;d}(x))$ given by, for $j = 1,\dots,d$,

$$v_{0;j}(x) = \frac{1}{N}\sum_{q=1}^N v_{1;j}(X^q_\varepsilon)$$

where $v_{1;j}(X^q_\varepsilon) = \partial_{\alpha_j}\Phi(\alpha)|_{\alpha=X^q_\varepsilon}$ if $u_\varepsilon(X^q_\varepsilon) < \Phi(X^q_\varepsilon)$ and $v_{1;j}(X^q_\varepsilon) = e^{-r\varepsilon}\partial_{\alpha_j}\mathbb{E}(u_2(X_{2\varepsilon})\,|\,X_\varepsilon = \alpha)|_{\alpha=X^q_\varepsilon}$ if $u_\varepsilon(X^q_\varepsilon) > \Phi(X^q_\varepsilon)$ (recall that u_2 is the estimate for $\bar{P}_{2\varepsilon}$). The gradient $\partial_\alpha\Phi(\alpha)$ should obviously be considered as given in *input*. Concerning the gradient of the conditional expectation, by using Theorem 4.2 one has to evaluate

$$\begin{aligned}\mathcal{H}_j[u_2](X^q_\varepsilon) &\equiv \partial_{\alpha_j}\mathbb{E}\Big(u_2(X_{2\varepsilon})\,\Big|\,X_\varepsilon = \alpha\Big)\Big|_{\alpha=X^q_\varepsilon}\\ &= \sum_{m=1}^j \widehat{\sigma}_{mj}\frac{\widetilde{X}^{m,q}_\varepsilon}{X^{j,q}_\varepsilon} \times \frac{\mathbb{R}_{s,t;m}[u_2](X^q_\varepsilon)\mathbb{T}_{s,t}[1](X^q_\varepsilon) - \mathbb{T}_{s,t}[u_2](X^q_\varepsilon)\mathbb{R}_{s,t;m}[1](X^q_\varepsilon)}{\mathbb{T}_{s,t}[1](X^q_\varepsilon)^2}\end{aligned} \tag{40}$$

where $\mathbb{T}_{s,t}$ and $\mathbb{R}_{s,t;m}$ are given by (20) and (21) respectively. Now, since they are weighted expectations of random variables for which we have N samples, they can be practically evaluated by means of the associated empirical mean: by taking into account (20) and (21), we write, for $f = u_2$ or $f = 1$ (more precisely, by replacing $f(X^{q'}_{2\varepsilon}) = u_2(X^{q'}_{2\varepsilon})$ or $f(X^{q'}_{2\varepsilon}) = 1$ in the formulas below),

$$\begin{aligned}\mathbb{T}_{s,t}[f](X^q_\varepsilon) &= \frac{1}{N}\sum_{q'=1}^N f(X^{q'}_{2\varepsilon})\prod_{i=1}^d \frac{H(\widetilde{X}^{i,q'}_\varepsilon - \widetilde{X}^{i,q}_\varepsilon)}{\sigma_{ii}\varepsilon^2\,\widetilde{X}^{i,q'}_\varepsilon}\,\Delta W^{i,q'}_1\\ \mathbb{R}_{s,t;m}[f](X^q_\varepsilon) &= -\frac{1}{N}\sum_{q'=1}^N f(X^{q'}_{2\varepsilon})\frac{H(\widetilde{X}^{m,q'}_\varepsilon - \widetilde{X}^{m,q}_\varepsilon)}{\sigma_{mm}s(t-s)(\widetilde{X}^{m,q'}_\varepsilon)^2}\Big[\frac{(\Delta W^{m,q'}_1)^2}{\sigma_{mm}s(t-s)} + \Delta W^{m,q'}_1 - \frac{t}{\sigma_{mm}}\Big]\times\\ &\qquad\times\prod_{i=1,i\neq m}^d \frac{H(\widetilde{X}^{i,q'}_\varepsilon - \widetilde{X}^{i,q}_\varepsilon)}{\sigma_{ii}s(t-s)\widetilde{X}^{i,q'}_\varepsilon}\,\Delta W^{i,q'}_1\end{aligned} \tag{41}$$

This concludes the analysis of the pricing/hedging algorithm.

Let us point out that, for the sake of simplicity, in the above description we have taken into account the non localized formulas. In practice, it is much better to use localizing functions in order to reduce the variance, so one should use the formulas coming from Theorem 4.8. Obviously, nothing

changes except for the choice of the localizing functions, for which we refer to the discussion in Section 5.

Finally, let us observe that one could use a further technique allowing to reduce the variance: the introduction of a control variable. Unfortunately, there is not a standard way to proceed in this direction. For example, one could use as a control variable the price of the associated European option. The idea is the following. For a fixed initial time t and underlying asset price x, let us set $P^{\mathrm{am}}(t,x)$ and $P^{\mathrm{eu}}(t,x)$ as the price of an American and European option respectively, with the same payoff Φ and maturity T. We define

$$P(t,x) = P^{\mathrm{am}}(t,x) - P^{\mathrm{eu}}(t,x).$$

Then it is easy to see that

$$P(t,X_t) = \sup_{\theta\in\mathcal{T}_{t,T}} \mathbb{E}\Big(e^{-r(\theta-t)}\widehat{\Phi}(\theta,X_\theta)\,\Big|\,\mathcal{F}_t\Big)$$

where $\mathcal{T}_{t,T}$ stands for the set of all the stopping times taking values on $[t,T]$ and $\widehat{\Phi}$ is defined by

$$\widehat{\Phi}(t,x) = \Phi(x) - P^{\mathrm{eu}}(t,x)$$

(notice the obstacle $\widehat{\Phi}(t,x)$ is now dependent on the time variable also, and is such that $\widehat{\Phi}(T,x) = 0$). Thus, for the numerical valuation of $P(0,x)$, one can set up a dynamic programming principle in point of fact identical to the one previously described, provided that the obstacle Φ is replaced by the new obstacle $\widehat{\Phi}(t,x)$. Once the estimated "price" $\bar{P}_0(x)$ and "delta" $\bar{\Delta}_0(x)$ are computed, the approximation of the price and delta of the American option is then given by

$$\bar{P}_0^{\mathrm{am}}(x) = \bar{P}_0(x) + P^{\mathrm{eu}}(0,x) \quad \text{and} \quad \bar{\Delta}_0^{\mathrm{am}}(x) = \bar{\Delta}_0(x) + \Delta^{\mathrm{eu}}(0,x)$$

respectively. Notice that the new obstacle has to be evaluated at each time step: in order to set up this program, it should be possible to compute the price/delta of an European option on Φ. This happens for some call or put options, for which prices and deltas are known in closed form. But one could think also to proceed by simulation for their computation, by using the formulas given in Theorem 3.1 and Theorem 4.2.

6.2 Sketch of the pricing/hedging algorithm

The algorithm itself can be stated as follows. We refer here to the simplest case: we do not consider localizing functions (but we underline in footnote where they should be) and control variables.

- Set the diffusion and option parameters: the dimension d, the starting point x, the interest rate r and the dividends η_i, the volatility matrix σ and the auxiliary matrices $\widetilde{\sigma}$ and $\widehat{\sigma}$, the maturity T, the payoff Φ and its first derivatives $\partial_j\Phi$, $j = 1,\dots d$.
- Choose n and set $\varepsilon = T/n$.
- **STEP n**
  - Produce $U_n^q \sim \mathrm{N}(0,I)$, $q = 1,\dots,N$.
  - Using (34), compute $W_{n\varepsilon}^q$, $q = 1,\dots,N$.
  - Using (35), compute $X_{n\varepsilon}^q$, $q = 1,\dots,N$.
  - Initialization: set $u_n(X_{n\varepsilon}^q) = \Phi(X_{n\varepsilon}^q)$.

- **STEP k, for k=n-1,...,1**
  - Produce $U_k^q \sim \mathrm{N}(0, I)$, $q = 1, \dots, N$.
  - Using (34), compute $W_{k\varepsilon}^q$, $q = 1, \dots, N$.
  - Using (35), compute $X_{k\varepsilon}^q$, $q = 1, \dots, N$.
  - Do the following[1].
    - Choose ℓ_k^q, $q = 1, \dots, N$. For example:
      choice 1: $\ell_k^q = 0$;
      choice 2: $\ell_k^q = \ell^*(X_{k\varepsilon}^q)$ as in (4.4) (with $\alpha = X_{k\varepsilon}^q$ and $s = k\varepsilon$):

$$\ell_k^{1,q} = 0 \quad \text{and for } i = 2, \dots, d: \qquad \ell_k^{i,q} = \sum_{j=1}^{i-1} \widehat{\sigma}_{ij} \Big(h_j - \frac{1}{k\varepsilon} \ln \frac{X_{k\varepsilon}^{j,q}}{x_j} \Big).$$

    - Using (37), compute $\widetilde{X}_{k\varepsilon}^q$, $q = 1, \dots, N$.
  - Using (38), compute ΔW_k^q, $q = 1, \dots, N$.
  - Using (39)[2], compute $\mathcal{P}_k[u_{k+1}](X_{k\varepsilon}^q)$, $q = 1, \dots, N$.
  - Compute for any $q = 1, \dots, N$,

$$u_k(X_{k\varepsilon}^q) = \max \Big(\Phi(X_{k\varepsilon}^q), e^{-r\varepsilon} \mathcal{P}_k[u_{k+1}](X_{k\varepsilon}^q) \Big)$$

  - If **k=1** then add the following[3].
    - Using (41)[4] and (40), compute $\mathcal{H}_j[u_2](X_\varepsilon^q)$, $j = 1, \dots, d$, $q = 1, \dots, N$.
    - Compute for any $j = 1, \dots, d$ and $q = 1, \dots, N$,

$$v_{1;j}(X_\varepsilon^q) = \partial_j \Phi(X_\varepsilon^q) \, \mathbf{1}_{\{u_1(X_\varepsilon^q) < \Phi(X_\varepsilon^q)\}} + e^{-r\varepsilon} \mathcal{H}_j[u_2](X_\varepsilon^q) \, \mathbf{1}_{\{u_1(X_\varepsilon^q) > \Phi(X_\varepsilon^q)\}}.$$

Remark. For $q = 1, \dots, N$, U_{k+1}^q, $W_{(k+1)\varepsilon}^q$, $X_{(k+1)\varepsilon}^q$, ℓ_k^q, $\widetilde{X}_{k\varepsilon}^q$, ΔW_k^q and $u_{k+1}(X_{k\varepsilon}^q)$, will not be employed anymore.

- **STEP 0**:
  - Compute

$$\mathcal{P}_0[u_1](x) = \frac{1}{N} \sum_{q=1}^{N} u_1(X_\varepsilon^q).$$

    and set

$$u_0(x) = \max \Big(\Phi(x), e^{-r\varepsilon} \mathcal{P}_0[u_1](x) \Big),$$

    which finally approximates the price $P(0, x)$.
  - Compute for any $j = 1, \dots, d$

$$v_{0;j}(x) = \frac{1}{N} \sum_{q=1}^{N} v_{1;j}(X_\varepsilon^q),$$

    which finally approximates the jth component of the delta vector $\Delta(0, x)$.

[1] This step is devoted to the multidimensional case: recall that $\widetilde{X} \equiv X$ in the one dimensional case.

[2] Or: first choose the localizing (=probability density) functions ψ_i, $i = 1, \dots, d$, compute the Ψ_i's as the associated probability distribution functions and after use the localized version of (39), coming from part *i*) of Theorem 4.8 (or Theorem 3.3 if $d = 1$).

[3] This step is devoted to the computation of the delta.

[4] Or: first choose the localizing (=probability density) functions ψ_i, $i = 1, \dots, d$, compute the Ψ_i's as the associated probability distribution functions and after use the localized version of (41), coming from part *ii*) of Theorem 4.8 (or Theorem 3.3 if $d = 1$).

6.3 Numerical results

In this section, the Malliavin-Monte Carlo method is numerically illustrated on three test problems. We first study a one-dimensional and two dimensional American put problem in order to assess the numerical behavior of the algorithm. We then investigate the methodology by pricing and hedging American put options on the geometric mean of five stocks.

6.3.1 Numerical experiments

Standard American put options. We consider a one-dimensional American put option with payoff $\Phi(x) = (K - x)_+$, exercise price $K = 100$, volatility $\sigma = 0.2$, maturity $T = 1$ year, instantaneous interest rate $r = \ln(1.1)$ and dividend yield rate $\eta = 0$. We take as the "true" reference price and delta, the one issued of the Binomal Black Scholes Richardson extrapolation tree-method [9] (BBSR-P and BBSR-Δ in the next tables) with 1000 step. Results are reported for problems with $n = 10, 20, 50$ time periods in Table 1, where N_{mc} is the number of Monte Carlo iterations.

Put on the minimum of 2 assets. We evaluate an American put option on the minimum of two underlying assets with payoff $\Phi(x) = (K - \min(x_1, x_2))_+$. We assume that the initial values of the stock prices are $x_1 = x_2 = 100$, the exercise price is $K = 100$, the volatility is $\sigma_1 = \sigma_2 = 0.2$ and the value of the correlation is $\rho = 0$ (this means that the volatility matrix σ is diagonal: $\sigma = \mathrm{diag}[\sigma_1, \sigma_2]$), the interest rate is $r = \ln(1.05)$, while the continuous dividend rates are $\eta_1 = \eta_2 = 0$. We take as the "true" reference price and deltas, the one issued of the Villeneuve-Zanette finite difference algorithm [23] (VZ-P and VZ-Δ in the next tables) with 500 time-space steps. Results are reported for problems with $n = 10, 20, 50$ time periods in Table 2.

Put on the geometric mean of 5 assets. We compute the American put option on the the geometric mean of 5 assets with payoff $\Phi(x) = (K - (\prod_{i=1}^{5} x_i)^{1/5})_+$. We assume that the initial values of the stock prices are $x_1 = \ldots = x_5 = 100$, the exercise price is $K = 100$, the volatility $\sigma_1 = \ldots = \sigma_5 = 0.2$ and the correlations are null (again, this means that $\sigma = \mathrm{diag}[\sigma_1, \ldots, \sigma_5]$), the interest rate is $r = \ln(1.1)$, while the continuous dividend rates are $\eta_1 = \ldots = \eta_5 = 0$. With this payoff, the simulation results are benchmarked with the one-dimensional American BBSR tree-method [9] with 1000 step. Results are reported for problems with $n = 10, 20, 50$ time periods in Table 3.

Tables of results. The tables refer to the examples previously introduced. Here, $\bar{P}$ and $\bar{\Delta}$ denotes the approximating price and delta respectively. In any experiment, the price of the associated Europen option has been used as a control variable. Moreover, also the localization has been taken into account, as a Laplace-type probability density function (see Section 5). In order to relax the computational executing time, we have always used the parameter of the localizing function equal to $\sqrt{\varepsilon}$, being $\varepsilon = T/n$, where T stands for the maturity and n is the number of time periods. Concerning the auxiliary drift ℓ, we have chosen both $\ell = 0$ and ℓ giving α as a fixed point for G (see Remark 4.6 for details). Actually, the results are not influenced by the two possibilities: we obtain really similar outcomes.

For the sake of comparison with other existing Monte Carlo methods, in the case of the standard American put (Table 1) also the price and delta provided by the Barraquand and Martineau [5] and Longstaff-Schwartz [18] algorithms[5] are reported (prices and deltas are denoted as BM-P, BM-Δ and LS-P, LS-Δ, respectively).

[5]It is worth noticing that the Barraquand and Martineau algorithm takes into account a space-grid, while the Longstaff and Schwartz algorithm uses least-squares approximation techniques: we have considered the ones implemented in the software PREMIA, see `http://cermics.enpc.fr/~premia`

	N_{mc}	$\bar{P}$	BBSR-P	BM-P	LS-P	$\bar{\Delta}$	BBSR-Δ	BM-Δ	LS-Δ
	500	4.807		5.611	4.914	-0.378		-0.316	-0.284
10 time	1000	4.795		5.147	4.714	-0.387		-0.325	-0.232
periods	5000	4.804	4.918	5.038	4.710	-0.384	-0.387	-0.287	-0.258
	10000	4.818		4.973	4.747	-0.387		-0.279	-0.254
	20000	4.823		4.928	4.816	-0.388		-0.270	0.265
	500	4.896		5.469	5.025	-0.398		-0.284	-0.237
20 time	1000	4.896		5.198	4.903	-0.385		-0.261	-0.251
periods	5000	4.864	4.918	5.086	4.765	-0.385	-0.387	-0.281	-0.256
	10000	4.873		5.076	4.843	-0.392		-0.272	-0.262
	20000	4.875		5.057	4.868	-0.387		-0.278	-0.269
	500	4.936		5.347	5.165	-0.384		-0.261	-0.258
50 time	1000	4.951		5.904	5.118	-0.384		-0.270	-0.3018
periods	5000	4.897	4.918	5.184	4.880	-0.386	-0.387	-0.283	-0.263
	10000	4.904		5.181	4.859	-0.392		-0.273	-0.263
	20000	4.910		5.149	4.907	-0.389		-0.281	-0.278

Table 1: Standard American put.

	N_{mc}	$\bar{P}$	VZ-P	$\bar{\Delta}_1$	$\bar{\Delta}_2$	VZ-Δ
	500	10.298		-0.288	-0.296	
10 time	1000	10.283		-0.295	-0.294	
periods	5000	10.271	10.306	-0.295	-0.297	-0.295
	10000	10.277		-0.297	-0.297	
	20000	10.263		-0.296	-0.295	
	500	10.388		-0.302	-0.294	
20 time	1000	10.371		-0.293	-0.292	
periods	5000	10.350	10.306	-0.295	-0.296	-0.295
	10000	10.349		-0.296	-0.297	
	20000	10.327		-0.296	-0.297	
	500	10.511		-0.306	-0.287	
50 time	1000	10.443		-0.288	-0.289	
periods	5000	10.416	10.306	-0.300	-0.299	-0.295
	10000	10.424		-0.297	-0.299	
	20000	10.403		-0.298	-0.297	

Table 2: American put on the minimum of 2 assets (here, VZ-Δ denotes the common value of VZ-Δ_1 and VZ-Δ_2).

	N_{mc}	$\bar{P}$	BBSR-P	$\bar{\Delta}_1$	$\bar{\Delta}_2$	$\bar{\Delta}_3$	$\bar{\Delta}_4$	$\bar{\Delta}_5$	BBSR-Δ
	500	1.525		-0.0752	-0.0741	-0.0824	-0.0767	-0.0708	
5 time	1000	1.535		-0.0743	-0.0739	-0.0744	-0.0687	-0.0784	
periods	5000	1.528	1.583	-0.0738	-0.0794	-0.0750	-0.0727	0.0761	-0.0755
	10000	1.537		-0.0775	-0.0765	-0.0786	-0.0759	-0.0780	
	20000	1.541		-0.0777	-0.0777	-0.0797	0.0779	-0.0783	
	500	1.716		-0.0781	-0.0766	-0.0868	-0.0934	-0.0708	
10 time	1000	1.684		-0.0763	-0.0756	-0.0839	-0.0786	-0.0823	
periods	5000	1.740	1.583	-0.0878	-0.0864	-0.0870	-0.0825	-0.0863	-0.0755
	10000	1.737		-0.0864	-0.0842	-0.0872	-0.0865	-0.0878	
	20000	1.727		-0.0838	0.0871	-0.0867	-0.084	-0.0842	
	500	1.800		-0.0777	-0.0754	-0.0681	-0.0827	-0.0643	
20 time	1000	1.846		-0.0851	-0.0872	-0.0753	-0.0720	-0.0768	
periods	5000	1.907	1.583	-0.0864	-0.0833	0.0919	-0.0933	-0.0983	-0.0755
	10000	1.880		0.0918	-0.0882	-0.0890	-0.0829	0.0849	
	20000	1.879		-0.0847	-0.0913	-0.0911	-0.0907	-0.0872	

Table 3: American put on the geometric mean of 5 assets (here, BBSR-Δ denotes the common value of BBSR-$\Delta_1, \ldots,$BBSR-Δ_5).

6.3.2 Final comments

The numerical values for the price and delta in tables 1, 2 and 3, show good accordance with the "true" values. It is worth noticing that this holds for a not high number of both time periods (10) and Monte Carlo iterations (500, 1000). Moreover, from Table 3 one may deduce that, as the dimension increases, the convergence (in terms of the number of iterations) is slower when the number of time periods is high. This could be explained if one knew the theoretical error, for which at the moment there are no results. However, Bouchard and Touzi in [8] (Theorem 6.3) have proved that, in a simulation scheme very similar to the one developed here (in particular, using localizing functions which generalize the one given in Theorem 3.4 in the one dimensional case), the maximum of all the L^p distances between the true conditional expectations and the associated regression estimators is of order $\varepsilon^{-d/(4p)} N^{-1/(2p)}$ as $\varepsilon \to 0$ (recall that ε is the discretization time-step, N is the number of simulations and d stands for the dimension - notice that there is a dependence on the dimension as well). In particular, the maximum of the variances is of order $\varepsilon^{-d/4} N^{-1/2}$ as $\varepsilon \to 0$. This would suggest that as the dimension increases, one should increase also the number of Monte Carlo iterations in order to achieve good results.

We also tested what happens if one does not consider the localization function and/or the control variable. The results are shown in Table 4, which refers to the simplest case of the one dimensional American put previously introduced.

N_{mc}	$\bar{P}$: with loc. without contr. var.	without loc. with contr. var.	without loc. without contr. var.
500	5.056	4.615	4.024
1000	4.873	4.725	4.272
5000	4.717	10.417	124.148
10000	4.773	14.330	10437.562
20000	4.829	10.224	219.259

Table 4: Standard American Put, 10 time periods ("true" price: 4.918)

First of all, it is worth to point out that the algorithm numerically does not work if the localization is not taken into account. On the contrary, as Table 4 shows, the use of the localization, even without any control variables, gives unstable but rather reasonable prices. Thus, the results in tables 1, 2, 3 and 4 allow to deduce that the introduction of the control variable together with the localization then brings to more stable results, with quite satisfactory precision both for prices and deltas also with a small number of Monte Carlo iterations (500, 1000).

Let us analyze the CPU time arising from the running of the algorithm. Our computations have been performed in double precision on a PC Pentium IV 1.8 GHz with 256 Mb of RAM. Figure 2 shows the computational time spent for the pricing of the one dimensional American put option as in the above example. As expected, this Monte Carlo procedure has a rather high CPU time cost, which empirically comes out to be quadratic with respect to the number of simulations.
Figure 2 allows to conclude that this method is interesting for further developments but is uncompetitive with respect to other Monte Carlo methods, such as Barraquand-Martineau [5] and Longstaff-Schwartz [18], in terms of computing time. Nevertheless, differently from these other procedures (which do not provide an *ad hoc* way for the Greeks: they are computed simply and poorly by finite differences), this method allows to efficiently obtain the delta. Therefore, as a further research development, one could think to combine this procedure with some other one in order to achieve better results, mainly cheaper in terms of CPU time costs, as it has been suggested by H. Regnier in private communications.

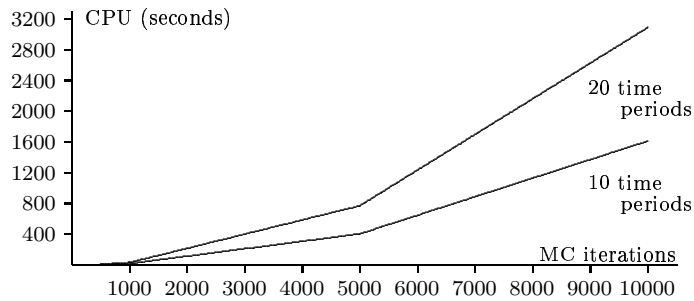

Figure 2 Required CPU time (in seconds) in terms of Monte Carlo iterations.

Appendix

Proof of Proposition 5.5. We give here only a sketch of the proof, since it is quite similar to the proof of Propositions 5.1 and 5.3.
First, suppose $d = 1$. Take $\hat{\psi} \in L^1(\mathbb{R})$ such that for any small ε then $\psi + \varepsilon\hat{\psi} \in \mathcal{L}_1$. Setting $\hat{\Psi}(x) = \int_{-\infty}^{x} \hat{\psi}(t)\, dt$, one has (as in the proof of Proposition 5.1)

$$\begin{aligned}(J_1^f)'(\psi)(\hat{\psi}) &= 2\mathbb{E}\int_{\mathbb{R}} f^2(X_t)\Big(\hat{\psi}(\beta)\Theta_{s,t} - \hat{\Psi}(\beta)\Upsilon_{s,t}\Big)\Big(\psi(\beta)\Theta_{s,t} + (H-\Psi)(\beta)\Upsilon_{s,t}\Big)\, d\beta.\\ &= -2\int_{\mathbb{R}} \hat{\Psi}(\beta)\,\mathbb{E}\Big(f^2(X_t)\,\Big(\psi'(\beta)\Theta_{s,t}^2 + (H-\Psi)(\beta)\Upsilon_{s,t}^2\Big)\Big)\, d\beta.\end{aligned}$$

By setting

$$\mu^{*2} = \frac{\mathbb{E}(f^2(X_t)\,\Upsilon_{s,t}^2)}{\mathbb{E}(f^2(X_t)\,\Theta_{s,t}^2)}$$

and $v(\beta) = \Psi(\beta)$, one has to solve the ordinary differential equation $v''(\beta) - a^2\, v(\beta) + a^2\, H(\beta) = 0$. This gives $\psi^*(\beta) = \partial_\beta v(\beta) = \mu^*\, e^{-\mu^*|\beta|}/2$. Now, it is immediate to see that ψ^* gives the minimum, so the statement follows in dimension 1.

Consider now the case $d > 1$. For simplicity, let us assume $k = 1$: by symmetry arguments, the general case will be clear. Also, let us set $f_t(y) \equiv \widetilde{f}(y) = f \circ F_t(y)$, $y \in \mathbb{R}_+^d$ and F_t being defined in (18).
First, we compute the derivative of $J_d^{f;1}(\psi)$ in the direction $(\hat{\psi}_1, 0, \ldots, 0)$. Setting

$$\widetilde{f}_1^2(x_1) = \int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-1}\mathbb{E}\Big(\widetilde{f}^2(x_1, \widetilde{X}_t^2, \ldots, \widetilde{X}_t^d)\prod_{i=2}^{d}\Big[\psi_i(\widetilde{X}_s^i - \widetilde{\alpha}_i) + \frac{(H-\Psi_i)(\widetilde{X}_s^i - \widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big]^2\Big)$$

then

$$J_d^{f;1}(\psi) = \int_{\mathbb{R}} d\widetilde{\alpha}_1\,\mathbb{E}\Big(\widetilde{f}_1^2(\widetilde{X}_t^1)\,\Big[\psi_1(\widetilde{X}_s^1 - \widetilde{\alpha}^1)\Theta_{s,t;1} + (H-\Psi_1)(\widetilde{X}_s^1 - \widetilde{\alpha}^1)\Upsilon_{s,t;1}\Big]^2\Big) = J_1^{\widetilde{f}_1}(\psi_1).$$

Now, since we are here interested in the behavior in the direction $(\hat{\psi}_1, 0, \ldots, 0)$, we can employ the one dimensional result, getting immediately the solution:

$$\psi_1^*(\beta) = \frac{\mu_1^*}{2}\, e^{-\mu_1^*|\beta|}, \qquad \beta \in \mathbb{R},$$

with

$$\mu_1^{*2} = \frac{\mathbb{E}\Big(\widetilde{f}_1^2(\widetilde{X}_t^1)\,\Upsilon_{s,t;1}^2\Big)}{\mathbb{E}\Big(\widetilde{f}_1^2(\widetilde{X}_t^1)\,\Theta_{s,t;1}^2\Big)}$$

$$= \frac{\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-1}\mathbb{E}\Big(f^2(X_t)\,\Upsilon_{s,t;1}^2\,\prod_{i=2}^d\Big[\psi_i(\widetilde{X}_s^i-\widetilde{\alpha}_i)+\frac{(H-\Psi_i)(\widetilde{X}_s^i-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big]^2\Big)}{\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-1}\mathbb{E}\Big(f^2(X_t)\,\Theta_{s,t;1}^2\,\prod_{i=2}^d\Big[\psi_i(\widetilde{X}_s^i-\widetilde{\alpha}_i)+\frac{(H-\Psi_i)(\widetilde{X}_s^i-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big]^2\Big)}$$

We consider now the behavior in the direction $(0,\hat{\psi}_2,0,\dots,0)$. Setting

$$\widetilde{f}_2^2(x_2) = \int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-2}\mathbb{E}\Big(\widetilde{f}^2(\widetilde{X}_t^1,x_2,\dots,\widetilde{X}_t^d)\Big[\psi_1(\widetilde{X}_s^1-\widetilde{\alpha}^1)\Theta_{s,t;1}+(H-\Psi_1)(\widetilde{X}_s^1-\widetilde{\alpha}^1)\Upsilon_{s,t;1}\Big]^2\times$$
$$\times\prod_{i=3}^d\Big[\psi_i(\widetilde{X}_s^i-\widetilde{\alpha}_i)+\frac{(H-\Psi_i)(\widetilde{X}_s^i-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big]^2\Big)$$

then

$$J_d^{f;1}(\psi) = \int_{\mathbb{R}} d\widetilde{\alpha}_2\,\mathbb{E}\Big(\widetilde{f}_2^2(\widetilde{X}_t^2)\Big[\psi_2(\widetilde{X}_s^2-\widetilde{\alpha}^2)+(H-\Psi_2)(\widetilde{X}_s^2-\widetilde{\alpha}^2)\frac{\Delta W_{s,t}^2}{\sigma_{22}s(t-s)\widetilde{X}_s^2}\Big]^2\Big) = I_1^{\widetilde{f}_2}(\psi_2),$$

where $I_1^{\widetilde{f}}(\cdot)$ is the one handled in Proposition 5.1. By using similar arguments, one obtains

$$\psi_2^*(\beta) = \frac{\mu_2^*}{2}\,e^{-\mu_2^*\,|\beta|}, \qquad \beta\in\mathbb{R},$$

with

$$\mu_2^{*2} = \frac{\mathbb{E}\Big(\widetilde{f}_2^2(\widetilde{X}_t^2)\Big[\frac{\Delta W_{s,t}^2}{\sigma_{22}s(t-s)\widetilde{X}_s^2}\Big]^2\Big)}{\mathbb{E}\Big(\widetilde{f}_2^2(\widetilde{X}_t^2)\Big)}$$

$$= \frac{\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-2}\mathbb{E}\Big(f^2(X_t)\,\Gamma(\widetilde{X}_s^1-\widetilde{\alpha}_1)^2\Big[\frac{\Delta W_{s,t}^2}{\sigma_{22}s(t-s)\widetilde{X}_s^2}\Big]^2\prod_{i=3}^d\Big[\psi_i(\widetilde{X}_s^i-\widetilde{\alpha}_i)+\frac{(H-\Psi_i)(\widetilde{X}_s^i-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big]^2\Big)}{\int_{\mathbb{R}^{d-1}} d\widetilde{\alpha}_{-2}\mathbb{E}\Big(f^2(X_t)\Gamma(\widetilde{X}_s^1-\widetilde{\alpha}_1)^2\prod_{i=3}^d\Big[\psi_i(\widetilde{X}_s^i-\widetilde{\alpha}_i)+\frac{(H-\Psi_i)(\widetilde{X}_s^i-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big]^2\Big)}$$

where

$$\Gamma(\widetilde{X}_s^1-\widetilde{\alpha}_1)\equiv\Gamma^{\psi_1}(\widetilde{X}_s^1-\widetilde{\alpha}_1) = \psi_1(\widetilde{X}_s^1-\widetilde{\alpha}^1)\Theta_{s,t;1}+(H-\Psi_1)(\widetilde{X}_s^1-\widetilde{\alpha}^1)\Upsilon_{s,t;1}$$

Now, for the remaining coordinates one can apply similar arguments. Thus, by summarizing, one has that $\psi^*(\xi)=\prod_{j=1}^d\psi_j^*(\xi_j)$, $\xi=(\xi_1,\dots,\xi_d)\in\mathbb{R}^d$, with $\psi_j^*(\xi_j)=\frac{\mu_j^*}{2}\,e^{-\mu_j^*\,|\xi_j|}$, $\xi_j\in\mathbb{R}$ and $\mu_j^*=\mu_j^*[f]$, are given by:

$$\mu_1^{*2} = \frac{\int_{\mathbb{R}^{d-1}} d\beta_{-1}\mathbb{E}\Big(f^2(X_t)\,\Upsilon_{s,t;1}^2\,\prod_{i=2}^d\Big[\psi_i^*(\beta_i)+\frac{(H-\Psi_i^*)(\beta_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big]^2\Big)}{\int_{\mathbb{R}^{d-1}} d\beta_{-1}\mathbb{E}\Big(f^2(X_t)\,\Theta_{s,t;1}^2\,\prod_{i=2}^d\Big[\psi_i^*(\beta_i)+\frac{(H-\Psi_i^*)(\beta_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big]^2\Big)} \quad \text{and for } j=2,\dots,d:$$

$$\mu_j^{*2} = \frac{\int_{\mathbb{R}^{d-1}} d\beta_{-j}\mathbb{E}\Big(f^2(X_t)\,\Gamma^*(\beta_1)^2\Big[\frac{\Delta W_{s,t}^j}{\sigma_{jj}s(t-s)\widetilde{X}_s^j}\Big]^2\prod_{i=2,i\neq j}^d\Big[\psi_i^*(\beta_i)+\frac{(H-\Psi_i^*)(\beta_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big]^2\Big)}{\int_{\mathbb{R}^{d-1}} d\beta_{-j}\mathbb{E}\Big(f^2(X_t)\Gamma^*(\beta_1)^2\prod_{i=2,i\neq j}^d\Big[\psi_i^*(\beta_i)+\frac{(H-\Psi_i^*)(\beta_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big]^2\Big)}$$

where $\Gamma^*(\beta_1) \equiv \Gamma^{\psi_1^*}(\beta_1) = \psi_1^*(\beta_1)\Theta_{s,t;1} + (H - \Psi_1^*)(\beta_1)\Upsilon_{s,t;1}$. Now, in order to give a more interesting representation for the μ^*'s, it is easy to show that (see also the proof of Proposition 5.3)

$$\int_{\mathbb{R}} \Big(\psi_i^*(\beta_i) + \frac{(H-\Psi_i^*)(\beta_i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\,\Delta W_{s,t}^i\Big)^2 d\beta_i = \frac{1}{4\mu_i^*}\Big(\mu_i^{*2} + \Big[\frac{\Delta W_{s,t}^i}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\Big]^2\Big),$$
$$\int_{\mathbb{R}} \Gamma^*(\beta_1)^2\, d\beta_1 = \frac{1}{4\mu_1^*}\Big(\mu_1^{*2}\Theta_{s,t;1}^2 + \Upsilon_{s,t;1}^2\Big),$$

so that the μ^*'s have to solve the nonlinear system

$$\mu_1^{*2} = \frac{\mathbb{E}\Big(f^2(X_t)\,\Upsilon_{s,t;1}^2 \prod_{i=2}^d \Big[\mu_i^{*2} + \Big[\frac{\Delta W_{s,t}^i}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\Big]^2\Big]\Big)}{\mathbb{E}\Big(f^2(X_t)\,\Theta_{s,t;1}^2 \prod_{i=2}^d \Big[\mu_i^{*2} + \Big[\frac{\Delta W_{s,t}^i}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\Big]^2\Big]\Big)} \quad \text{and for } j = 2,\dots,d:$$
$$\mu_j^{*2} = \frac{\mathbb{E}\Big(f^2(X_t)\,\Big(\mu_1^{*2}\Theta_{s,t;1}^2 + \Upsilon_{s,t;1}^2\Big)\Big[\frac{\Delta W_{s,t}^j}{\sigma_{jj}s(t-s)\widetilde{X}_s^j}\Big]^2 \prod_{i=2,i\neq j}^d \Big[\mu_i^{*2} + \Big[\frac{\Delta W_{s,t}^i}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\Big]^2\Big]\Big)}{\mathbb{E}\Big(f^2(X_t)\Big(\mu_1^{*2}\Theta_{s,t;1}^2 + \Upsilon_{s,t;1}^2\Big) \prod_{i=2,i\neq j}^d \Big[\mu_i^{*2} + \Big[\frac{\Delta W_{s,t}^i}{\sigma_{ii}s(t-s)\widetilde{X}_s^i}\Big]^2\Big]\Big)}.$$

Finally, it is straightforward to see that ψ^* actually gives the minimum.

□

References

[1] V. Bally, M.E. Caballero, B. Fernandez, N. El Karoui (2002) Reflected BSDE's, PDE's and variational inequalities. *Preprint INRIA*.

[2] V. Bally and G. Pagés (2000). A quantization algorithm for solving multi-dimensional optimal stopping problems. Prépublication du Laboratoire de Probabilités & Modèles Aléatoires, n. 628, Université Paris VI.

[3] V. Bally and G. Pagés (2002). Error analysis of a quantization algorithm for obstacle problems. To appear on *Stoch. Proc. and their Appl.*.

[4] V. Bally, G. Pagés and J. Printems (2003). First-orders Schemes in the Numerical Quantization Method. *Mathematical Finance,* **13**,1-16.

[5] J. Barraquand and D. Martineau (1995). Numerical valuation of high dimensiona multivariate American securities. *Journal of Finance and Quantitative Analysis*, **30**, 383-405.

[6] A. Bensoussan and J.L. Lions (1984). *Impulse Control and Quasivariational Inequalities.* Gauthiers-Villars.

[7] B. Bouchard, I. Ekeland and N. Touzi (2002). On the Malliavin approach to Monte Carlo approximation of conditional expectations. Prépublication du Laboratoire de Probabilités & Modèles Aléatoires, n. 709; available at `http://www.proba.jussieu.fr/mathdoc/preprints/index.html`

[8] B. Bouchard and N. Touzi (2002). Discrete time approximation and Monte-Carlo simulation of Backward Stochastic Differential Equations. Preprint.

[9] M. Broadie and J. Detemple (1996). American option valuation: new bounds, approximations, and a comparison of existing methods securities using simulation. *The Review of Financial Studies,* **9**, 1221-1250.

[10] M. Broadie and P. Glassermann (1997). Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control,* **21**, 1323-1352.

[11] D. Chevance (1997). Thèse de l'Université de Provence & INRIA.

[12] É. Fournié, J.M. Lasry, J. Lebouchoux, P.L. Lions, N. Touzi (1999). Applications of Malliavin calculus to Monte carlo methods in Finance. *Finance & Stochastics,***3**, 391-412.

[13] É. Fournié, J.M. Lasry, J. Lebouchoux, P.L. Lions (2001). Applications of Malliavin calculus to Monte carlo methods in Finance II. *Finance & Stochastics,***5**, 201-236.

[14] N. El Karoui, C. Kapoudjan, E. Pardoux, S. Peng and M.C. Quenez (1997). Reflected solutions of Backward SDE's and related obstacle problems for PDE's. *The Annals of Probability*, **25**, 702-73.

[15] N. El Karoui, S. Peng and M.C. Quenez (1994). Backward stochastic differential equations in finance. *Mathematical Finance*, **25**, 702-73.

[16] A. Kohatsu-Higa and R. Pettersson (2001). Variance reduction methods for simulation of densities on Wiener space. *SIAM Journal of Numerical Analysis*, **4**, 431-450.

[17] P.L. Lions and H. Regnier (2001). Calcul du prix et des sensibilités d'une option américaine par une méthode de Monte Carlo. Preprint.

[18] F.A. Longstaff and E.S. Schwartz (2001). Valuing American options by simulations: a simple least squares approach. *The Review of Financial Studies,* **14**, 113-148.

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

[20] E. Pardoux and S. Peng (1990). Adapted solution of a backward stochastic differential equation. *Systems and Control Letters,* **14**, 55-61.

[21] E. Pardoux and S. Peng (1990). Backward stochastic differential equations and quasilinear parabolic partial differential equations and their applications. In *Lect. Notes Control Inf. Sci.*, Vol. 176, 200-217.

[22] J.N. Tsitsiklis and B. VanRoy (1999). Optimal stopping for Markov processes: Hilbert space theory, approximation algorithm and an application to pricing high-dimensional financial derivatives. *IEEE Trans. Autom. Controll,* **44**, 1840-1851.

[23] S. Villeneuve and A. Zanette (2002). Parabolic ADI methods for pricing American options on two stocks. *Mathematics of Operations Research*, **27**, 121-149.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399