

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

VLAD BALLY*

LUCIA CAMELLINO†

ANTONINO ZANETTE‡

December 17, 2004

Abstract

Following the pioneering papers of Fournié, Lasry, Lebouchoux, Lions and Touzi, 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. The 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.

Keywords: American options; option pricing and hedging; Malliavin Calculus; Monte Carlo methods

Corresponding author: Lucia Caramellino

*Laboratoire d'Analyse et de Mathématiques Appliquées, Université de Marne-la-Vallée, 77454 Champs-sur-Marne, France; mailto: <bally@univ-mlv.fr>

†Dipartimento di Matematica, Università di Roma *Tor Vergata*, Via della Ricerca Scientifica, I-00133 Roma, Italy; mailto: <caramell@mat.uniroma2.it>.

‡Dipartimento di Finanza dell'Impresa e dei Mercati Finanziari, Università di Udine, Via Tomadini 30/A, I-33100 Udine, Italy; mailto: <antonino.zanette@dfimf.uniud.it>.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 3 |
| 2 | Pricing/hedging American options | 4 |
| 3 | Representation formulas for the conditional expectation and its gradient | 6 |
| 3.1 | Diagonalization procedure and first formulas | 7 |
| 3.2 | Localized formulas | 10 |
| 4 | Optimal localizing functions | 12 |
| 5 | The algorithm for pricing/hedging American options | 18 |
| 5.1 | How to use the formulas in practice | 18 |
| 5.2 | Numerical experiments | 22 |
| 5.3 | Final comments | 23 |
| 6 | Appendix | 25 |
| 6.1 | Proofs of the result in Section 3 | 25 |
| 6.2 | Proof of Proposition 4.5 | 30 |
| | References | 32 |

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).

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 application 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 Monte Carlo methods remain the only alternative.

But one has to notice that both the Feynman-Kac formula and Monte Carlo methods are specific to linear problems and collapse when dealing in the nonlinear case. In the last decay much work has been done in order to extend probabilistic methods in a nonlinear frame. On one hand, in a set of papers Pardoux and Peng (see [22], [23]) introduced the backward stochastic differential equations (BSDE's in short), which generalize the probabilistic representation given by the Feynman-Kac formula to nonlinear problems. In [16] (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 [17]. 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

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

one represents $u(0, x_0)$ as an expectation and employs a sample of the underlying diffusion in order to compute this expectation. On the contrary, an analytical method constructs a time-space grid and uses a dynamical programming algorithm 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 programming 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 problems in mathematical finance - 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 the algorithm by Broadie and Glasserman [11], by Barraquand and Martineau [6], the quantization algorithm in [3] and [4], as well as the Chevance one [13]. The second idea is to use regression on a truncated basis of L^2 in order to compute the conditional

expectations, as done by Longstaff and Schwartz [20] and by Tsisiklis and Van Roy [24]. Finally in [14], [15], [19], [8] and [9], the authors obtain representations for the conditional expectation using Malliavin calculus and then employ them in order to perform a Monte Carlo method. The peculiarity of this last approach is that it appears as a pure Monte Carlo method despite the nonlinearity.

In addition to 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 with respect 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 [19], who follow the third method, as well as in [5] where the 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 nonlinear problems, we focus on the American option pricing/hedging, 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 reduction of variance techniques, based on localization and control variables, and to the 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 allows us to give here an approach which is self contained and accessible to readers who are not familiar with the rather heavy machinery of Malliavin 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 Black 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, in Section 3 we give the results (representation formulas for conditional expectations and for the strategy) and in Section 4 we discuss the optimal localization. Finally, Section 5 is devoted to the description of the algorithm based on this approach and to numerical tests.

2 Pricing/hedging American options

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 denote the process of the underlying asset prices, which is supposed to be a diffusion process on \mathbb{R}^d , and $\Phi(X_s)$ denote the cash-flow associated with the option. The price as seen at time t of such an American option is given by

$$P(t, x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}_{t,x} \left(e^{-\int_t^\tau r_s ds} \Phi(X_\tau) \right) \quad (1)$$

where $\mathcal{T}_{t,T}$ stands for the set of all the stopping times taking values on $[t, T]$ and r is the (deterministic) spot rate.

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

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

and the function $P(t, x)$, giving the price of the option, solves the following (nonlinear) PDE:

$$(\partial_t + L)P(t, x) - rP(t, x) = 0$$

whenever $P(t, x) > \Phi(x)$, with the final condition $P(T, x) = \Phi(x)$, where L is the infinitesimal generator of X .

It is worth to say that a rigorous statement is much more difficult because in general one has 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 [7]; in El Karoui *et al.* [16], one can find a formulation in viscosity sense and in [1], Bally *et al.* give a formulation in Sobolev sense.

In practice, the numerical evaluation of $P(0, x)$, that is the price as seen at time 0, is done by using a Bellman dynamic programming principle. Indeed, let $0 = t_0 < t_1 < \dots < t_n = T$ be a discretization of the time interval $[0, T]$, with step size equal to $\Delta t = T/n$, and let $(\bar{X}_{k\Delta t})_{k=0,1,\dots,n}$ 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 $\bar{P}_{n\Delta t}(\bar{X}_{n\Delta t}) = \Phi(\bar{X}_{n\Delta t})$ and for any $k = n - 1, n - 2, \dots, 1, 0$,*

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

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 [4] (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 $\Delta(t, x) = \partial_x P(t, x)$ of an American option. This is important because $\Delta(t, x)$ gives the sensibility of the price with respect to the initial underlying asset price and also it allows to hedge the option. By considering the case $t = 0$, 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 $\bar{P}_{\Delta t}(\alpha) = \max(\Phi(\alpha), e^{-r\Delta t} \mathbb{E}(\bar{P}_{2\Delta t}(\bar{X}_{2\Delta t}) \mid \bar{X}_{\Delta t} = \alpha))$. Then, by setting

$$\begin{aligned} \bar{\Delta}(\alpha) &= \partial_\alpha \Phi(\alpha) \mathbf{1}_{\Gamma_{\Delta t}} + e^{-r\Delta t} \partial_\alpha \mathbb{E} \left(\bar{P}_{2\Delta t}(\bar{X}_{2\Delta t}) \mid \bar{X}_{\Delta t} = \alpha \right) \mathbf{1}_{\Gamma_{\Delta t}^c} \quad \text{and} \\ \bar{\Delta}_0(x) &= \mathbb{E}_x \left(\bar{\Delta}(\bar{X}_{\Delta t}) \right) \end{aligned}$$

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 bounds, turns out to be a more difficult problem (one may find in Bally *et al.* [5] bounds given in a weak sense, that is 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 their derivatives, thus allowing one to set up Monte Carlo simulations.

In Section 3, we study 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

$$\begin{aligned} \mathbb{E}\left(F(X_t) \middle| X_s = \alpha\right) &= \frac{\mathbb{E}(F(X_t) \pi_s^\alpha)}{\mathbb{E}(\pi_s^\alpha)} \\ \partial_\alpha \mathbb{E}\left(F(X_t) \middle| X_s = \alpha\right) &= \frac{\mathbb{E}(F(X_t) \pi_s^{1,\alpha}) \mathbb{E}(\pi_s^\alpha) - \mathbb{E}(F(X_t) \pi_s^\alpha) \mathbb{E}(\pi_s^{1,\alpha})}{\mathbb{E}(\pi_s^\alpha)^2} \end{aligned} \quad (2)$$

being π_s^α and $\pi_s^{1,\alpha}$ suitable weights, which could also depend on suitable localizing functions. In Section 4 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 [18] and by Bouchard, Ekeland and Touzi [8].

Representations (2) 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. 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 5 a comprehensive description of the algorithm and its numerical behavior.

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 take into account other sensibilities (Greeks). In fact, similar techniques can be applied in order to obtain representation formulas for other derivatives of the conditional expectation: the second derivative (gamma), the derivative w.r.t. the spot rate r (theta), as well as the derivative w.r.t. the volatility σ (vega).

3 Representation formulas for the conditional expectation and its gradient

Let X be the underlying asset price process, driven by the Black and Scholes model, that is

$$dX_t^i = (r - \eta^i) X_t^i dt + \sum_{j=1}^d \sigma_{ij} X_t^i dW_t^j, \quad \text{with } X_0^i = x^i, \quad i = 1, \dots, d$$

where: $x = (x^1, \dots, x^d) \in \mathbb{R}_+^d$ denotes the vector of the initial asset values; r is the (constant) spot rate and $\eta \in \mathbb{R}^d$ being 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

$$X_t^i = x^i \exp \left(h^i t + \sum_{j=1}^i \sigma_{ij} W_t^j \right), \quad i = 1, \dots, d \quad (3)$$

where from now on we set $h^i = r - \eta^i - \frac{1}{2} \sum_{j=1}^i \sigma_{ij}^2$, $i = 1, \dots, d$. The aim is to study the conditional expectation and its gradient, that is

$$\mathbb{E}(\Phi(X_t) | X_s = \alpha) \quad \text{and} \quad \partial_\alpha \mathbb{E}(\Phi(X_t) | X_s = \alpha), \quad \Phi \in \mathcal{E}_b(\mathbb{R}^d),$$

respectively, where $0 < s < t$, $\alpha \in \mathbb{R}_+^d$ and $\mathcal{E}_b(\mathbb{R}^d)$ denotes the class of the measurable functions with polynomial growth, that is $|\Phi(y)| \leq C(1 + |y|^m)$ for some m .

In few words, to this goal it suffices to consider an auxiliary process \tilde{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 (invertible) function giving X from the auxiliary process \tilde{X} . We will study such kind of formulas in Section 3.1 and 3.2, where also a discussion on the connections with the paper by Lions and Regnier [19] is presented. It is worth remarking that the formulas are not new, since they have been studied for example also in Fournié, Lasry, Lebuchoux and Lions [15] and in Bouchard, Ekeland and Touzi [8]. Anyway, we give here a very simple approach and also propose elementary and immediate proofs, which are postponed to the Appendix (see Section 6.1).

3.1 Diagonalization procedure and first formulas

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

$$\tilde{X}_t^i = x^i \exp \left(h^i t + \ell_t^i + \sigma_{ii} W_t^i \right), \quad i = 1, \dots, d. \quad (4)$$

As a first result, we study a transformation allowing to handle the new process \tilde{X} in place of the original process X :

Lemma 3.1 *For any $t \geq 0$ there exists an invertible function $F_t(\cdot) : \mathbb{R}_+^d \rightarrow \mathbb{R}_+^d$ such that $X_t = F_t(\tilde{X}_t)$ and $\tilde{X}_t = F_t^{-1}(X_t)$. In details, F_t and its inverse $G_t = F_t^{-1}$ are given by (set $\prod_{j=1}^0 \stackrel{def}{=} 1$)*

$$F_t^i(y) = e^{-\sum_{j=1}^i \tilde{\sigma}_{ij} \ell_t^j} y^i \prod_{j=1}^{i-1} \left(\frac{y^j}{x^j} e^{-h^j t} \right)^{\tilde{\sigma}_{ij}} \quad \text{and} \quad G_t^i(z) = e^{\ell_t^i} z^i \prod_{j=1}^{i-1} \left(\frac{z^j}{x^j} e^{-h^j t} \right)^{\hat{\sigma}_{ij}}, \quad (5)$$

for $i = 1, \dots, d$ and $y, z \in \mathbb{R}_+^d$, where

$$\tilde{\sigma}_{ij} = \frac{\sigma_{ij}}{\sigma_{jj}}, \quad i, j = 1, \dots, d, \quad \text{and} \quad \hat{\sigma} = \tilde{\sigma}^{-1} \quad (6)$$

Proof. Let t, ℓ, x be fixed. From (4) we have $W_t^i = (\ln(\tilde{X}_t^i/x^i) - h^i t - \ell_t^i)/\sigma_{ii}$, $i = 1, \dots, d$. Inserting this in (3), one obtains

$$X_t^i = x^i \exp \left(h^i t - \sum_{j=1}^i \frac{\sigma_{ij}}{\sigma_{jj}} (h^j t + \ell_t^j) \right) \prod_{j=1}^i \left(\frac{\tilde{X}_t^j}{x^j} \right)^{\sigma_{ij}/\sigma_{jj}}$$

Thus, by setting $\tilde{\sigma}_{ij} = \sigma_{ij}/\sigma_{jj}$, $i, j = 1, \dots, d$ 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 $F_t = (F_t^1, \dots, F_t^d)$ satisfies

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

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

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

Notice that $\hat{\sigma}$ is easy to compute because $\tilde{\sigma}$ is a triangular matrix. Moreover, $\hat{\sigma}$ is itself triangular and $\hat{\sigma}_{ii} = 1$ for any i .

□

Remark 3.2 *It is worth to stress that the introduction of \tilde{X} is done in order to be able to handle a process with independent components. Therefore, \tilde{X} has no interesting practical meaning in the one-dimensional case, where one simply has $\tilde{X}_t = e^{\ell_t} X_t$ and $X_t = e^{-\ell_t} \tilde{X}_t$.*

Theorem 3.3 (Representation formulas I: without localization) *Let $0 < s < t$, $\Phi \in \mathcal{E}_b(\mathbb{R}^d)$ and $\alpha \in \mathbb{R}_+^d$ be fixed. Set: $\tilde{X}_s = G_s(X_s)$ and $\tilde{\alpha}_s = G_s(\alpha)$, G_s being defined in (5), $H(\xi) = \mathbf{1}_{\xi \geq 0}$, $\xi \in \mathbb{R}$, $\hat{\sigma}$ as in (6) and*

$$\Delta W_{s,t}^i = (t-s)(W_s^i + \sigma_{ii}s) - s(W_t^i - W_s^i), \quad i = 1, \dots, d. \quad (7)$$

i) *The following representation formula for the conditional expectation holds:*

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

where

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E}\left(f(X_t) \prod_{i=1}^d \frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right). \quad (8)$$

ii) *The following representation formula for the gradient of the conditional expectation holds: for $j = 1, \dots, d$,*

$$\partial_{\alpha_j} \mathbb{E}\left(\Phi(X_t) \mid X_s = \alpha\right) = \sum_{k=1}^j \hat{\sigma}_{kj} \frac{\tilde{\alpha}_s^k}{\alpha^j} \times \frac{\mathbb{G}_{s,t;k}[\Phi](\alpha) \mathbb{T}_{s,t}[1](\alpha) - \mathbb{T}_{s,t}[\Phi](\alpha) \mathbb{G}_{s,t;k}[1](\alpha)}{\mathbb{T}_{s,t}[1](\alpha)^2},$$

where, as $k = 1, \dots, d$,

$$\begin{aligned} \mathbb{G}_{s,t;k}[f](\alpha) = & -\mathbb{E}\left(f(X_t) \frac{H(\tilde{X}_s^k - \tilde{\alpha}_s^k)}{\sigma_{kk}s(t-s)(\tilde{X}_s^k)^2} \left[\frac{(\Delta W_{s,t}^k)^2}{\sigma_{kk}s(t-s)} + \Delta W_{s,t}^k - \frac{t}{\sigma_{kk}} \right] \times \right. \\ & \left. \times \prod_{i=1, i \neq k}^d \frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right). \end{aligned} \quad (9)$$

Let us now give some observations:

Remark 3.4 • When $d = 1$, the choice $\ell = 0$, which means simply to take $\tilde{X} = X$, simplifies the above formulas for the operator $\mathbb{T}_{s,t}[f](\alpha)$ and $\mathbb{G}_{s,t}[f](\alpha)$ as follows:

$$\begin{aligned}\mathbb{T}_{s,t}[f](\alpha) &= \mathbb{E}\left(f(X_t) \frac{H(X_s - \alpha)}{\sigma s(t-s)X_s} \Delta W_{s,t}\right) \\ \mathbb{G}_{s,t}[f](\alpha) &= -\mathbb{E}\left(f(X_t) \frac{H(X_s - \alpha)}{\sigma s(t-s)X_s^2} \left[\frac{(\Delta W_{s,t})^2}{\sigma s(t-s)} + \Delta W_{s,t} - \frac{t}{\sigma}\right]\right).\end{aligned}$$

- If no correlation is assumed among the assets, that is if the volatility matrix σ is diagonal, then $\hat{\sigma} = Id_{d \times d}$. 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 $\tilde{\alpha}_s^j / \alpha^j = e^{\ell_s^j}$, which in turn is equal to 1 whenever $\ell = 0$.

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

A way to choose the drift ℓ is in order to have $G_s(\alpha) = \alpha$, which is the implicit choice made by Lions and Regnier in [19]. The corresponding ℓ is computed in the following

Proposition 3.5 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 (set $\sum_{j=1}^0 \stackrel{\text{def}}{=} 0$)

$$\ell_s^{*i} = \sum_{j=1}^{i-1} \hat{\sigma}_{ij} \left(h^j s - \ln \frac{\alpha^j}{x^j} \right), \quad i = 1, \dots, d,$$

$\hat{\sigma}$ being defined in (6). In particular, if $\ell_t = \rho t$ then $G_s(\alpha) = \alpha$ if and only if $\rho = \rho_s^*$, where

$$\rho_s^{*i} = \sum_{j=1}^{i-1} \hat{\sigma}_{ij} \left(h^j - \frac{1}{s} \ln \frac{\alpha^j}{x^j} \right), \quad i = 1, \dots, d.$$

Proof. The proof is immediate: by (5) one can shortly write $\ln G_s(z) = \ell_s + \hat{\sigma} \ln z + (I - \hat{\sigma})(\ln x + hs)$, thus $\alpha = G_s(\alpha)$ if and only if (set $\ln(\alpha/x)$ as the vector whose i^{th} entry is $\ln(\alpha^i/x^i)$)

$$\ell_s = \ell_s^* = (\hat{\sigma} - I) \left(h s - \ln \frac{\alpha}{x} \right).$$

□

Remark 3.6 It is worth noticing that if originally σ is diagonal, then $\tilde{\sigma} = I$, so that $\alpha = G_s(\alpha)$ if and only if $\ell_s^* = 0$, as it must obviously follow. In any case, $\alpha = G_s(\alpha)$ gives always $\ell_s^{*1} = 0$, because of the fact that $\tilde{\sigma}_{11} = 1$. Moreover, the choice $\ell = \ell^*$ gives a considerable simplification of the function F_t and its inverse G_t , which become

$$F_t^i(y) = y_i \prod_{j=1}^{i-1} \left(\frac{y^j}{x^j} \left(\frac{x^j}{\alpha^j} \right)^{t/s} \right)^{\tilde{\sigma}_{ij}} \quad \text{and} \quad G_t^i(z) = z^i \prod_{j=1}^{i-1} \left(\frac{z^j}{x^j} \left(\frac{x^j}{\alpha^j} \right)^{t/s} \right)^{\hat{\sigma}_{ij}}, \quad i = 1, \dots, d.$$

Now, the formulas given by Lions and Regnier [19] take into account the process \hat{X} whose components are defined as

$$\hat{X}_t^i = x^i \exp \left(h^i t + \sum_{j=1}^{i-1} \sigma_{ij} w_t^j + \sigma_{ii} W_t^i \right), \quad i = 1, \dots, d$$

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

$$\sum_{k=1}^i \sigma_{ik} w_t^k = \ln \alpha^i - \ln x^i - h^i t, \quad i = 1, \dots, d. \quad (10)$$

It is worth remarking that \widehat{X} has independent components, as well as \widetilde{X} . The formula given in [19] for the conditional expectation states the following:

$$\mathbb{E}\left(\Phi(X_t) \mid X_s = \alpha\right) = \frac{\mathbb{L}_{s,t}[\Phi](\alpha)}{\mathbb{L}_{s,t}[1](\alpha)}, \quad \text{where} \quad \mathbb{L}_{s,t}[f](\alpha) = \mathbb{E}\left(f(X_t) \prod_{i=1}^d \frac{H(\widehat{X}_s^i - \alpha^i)}{\sigma_{ii} s(t-s) \widehat{X}_s^i} \Delta W_{s,t}^i\right). \quad (11)$$

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 (10)). Furthermore, the formulas giving the operators $\mathbb{T}_{s,t}$ and $\mathbb{L}_{s,t}$, allowing to write down the conditional expectation, can be summarized as follows: setting

$$\mathbb{A}_{s,t}[f](\alpha) = \mathbb{E}\left(f(X_t) \prod_{i=1}^d \frac{H(Y^i - g^i(\alpha))}{\sigma_{ii} s(t-s) Y^i} \Delta W_{s,t}^i\right),$$

then

- 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, the connection between the two approaches is simple: if one sets $\ell_t^1 = 0$ and for $i = 2, \dots, d$, $\ell_t^i = \sum_{j=1}^{i-1} \sigma_{ij} w_t^j$, then it is straightforward to see that condition (10) equals to the condition studied in Proposition 3.5, that is $\widetilde{\alpha}_s = G_s(\alpha) = \alpha$. This means that if ℓ_s is chosen as in Proposition 3.5, the two formulas are actually identical. Therefore, the approach studied here is some more general than the one developed by Lions and Regnier.

Another way to choose ℓ might be done in order to minimize the variance (or the integrated one) of the random variable whose expectation gives the operator $\mathbb{T}_{s,t}[f](\alpha)$ or $\mathbb{G}_{s,t;k}[f]$. Unfortunately, this procedure gives a dependence on ℓ in such a way that the percentage ratio between the mean and the square root of the variance is independent of ℓ . In terms of the Central Limit Theorem, this means that this kind of optimization gives a nonsense. For details, see [2].

Let us resume the above observations in the following

Remark 3.7 *In principle one could take ℓ arbitrarily, so that for practical purposes the simple choice $\ell(t) = \rho t$ seems to be good enough. Concerning the (now) constant ρ , up to now two main choices for ℓ can be suggested:*

- $\rho = 0$: this simplifies the process \widetilde{X} ;
- $\rho = \rho^*$, with ρ^* as in Proposition 3.5: as observed, this gives a formula for the conditional expectation in point of fact identical to the one provided by Lions and Reigner in [19].

3.2 Localized formulas

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 $\mathbb{G}_{s,t;j}[f](\alpha)$ and then for the conditional expectation and its gradient. In fact, one first has

Lemma 3.8 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 $\mathbb{G}_{s,t;j}$, defined in (8) and (9) respectively, can be localized as follows:

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

where

$$\mathbb{T}_{s,t}^\psi[f](\alpha) = \mathbb{E}\left(f(X_t) \prod_{i=1}^d \left[\psi_i(X_s - \alpha) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i \right]\right) \quad (12)$$

and

$$\begin{aligned} \mathbb{G}_{s,t;k}^\psi[f](\alpha) = & -\mathbb{E}\left(f(X_t) \left[\psi_k(\tilde{X}_s^k - \tilde{\alpha}_s^k) \frac{\Delta W_{s,t}^k}{\sigma_{kk}s(t-s)\tilde{X}_s^k} + \right. \right. \\ & \left. \left. + \frac{(H - \Psi_k)(\tilde{X}_s^k - \tilde{\alpha}_s^k)}{\sigma_{kk}s(t-s)(\tilde{X}_s^k)^2} \left(\frac{(\Delta W_{s,t}^k)^2}{\sigma_{kk}s(t-s)} + \Delta W_{s,t}^k - \frac{t}{\sigma_{kk}} \right) \right] \right) \times \\ & \times \prod_{i=1, i \neq k}^d \left[\psi_i(\tilde{X}_s^i - \tilde{\alpha}_s^i) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i \right]. \end{aligned} \quad (13)$$

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

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

Theorem 3.9 (Representation formulas II: with localization) For any $0 \leq s < t$, $\Phi \in \mathcal{E}_b$, $\alpha \in \mathbb{R}_+^d$ and for any $\psi \in \mathcal{L}_d$, one has

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

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

$$\partial_{\alpha^j} \mathbb{E}\left(\Phi(X_t) \mid X_s = \alpha\right) = \sum_{k=1}^j \hat{\sigma}_{kj} \frac{\tilde{\alpha}_s^k}{\alpha^j} \times \frac{\mathbb{G}_{s,t;k}^\psi[\Phi](\alpha) \mathbb{T}_{s,t}^\psi[1](\alpha) - \mathbb{T}_{s,t}^\psi[\Phi](\alpha) \mathbb{G}_{s,t;k}^\psi[1](\alpha)}{\mathbb{T}_{s,t}^\psi[1](\alpha)^2},$$

where the operators $\mathbb{T}_{s,t}^\psi[f](\alpha)$ and $\mathbb{G}_{s,t;k}^\psi[f](\alpha)$ are defined in (12) and (13) respectively.

Remark 3.10 In principle, one could take different localizing functions for each operator, that is:

$$\begin{aligned} \mathbb{E}\left(\Phi(X_t) \mid X_s = \alpha\right) &= \frac{\mathbb{T}_{s,t}^{\psi_1}[\Phi](\alpha)}{\mathbb{T}_{s,t}^{\psi_2}[1](\alpha)} \\ \partial_{\alpha^j} \mathbb{E}\left(\Phi(X_t) \mid X_s = \alpha\right) &= \sum_{k=1}^j \hat{\sigma}_{kj} \frac{\tilde{\alpha}_s^k}{\alpha^j} \times \frac{\mathbb{G}_{s,t;k}^{\psi_3}[\Phi](\alpha) \mathbb{T}_{s,t}^{\psi_4}[1](\alpha) - \mathbb{T}_{s,t}^{\psi_5}[\Phi](\alpha) \mathbb{G}_{s,t;k}^{\psi_6}[1](\alpha)}{\mathbb{T}_{s,t}^{\psi_7}[1](\alpha)^2}. \end{aligned}$$

See next section for a discussion on localizing functions. Furthermore, what observed in Remark 3.4 holds here as well: if σ is diagonal, the sum giving $\partial_{\alpha^j} \mathbb{E}(\Phi(X_t) \mid X_s = \alpha)$ reduces to the single term with $k = j$, with coefficient $\tilde{\alpha}_s^j / \alpha^j = e^{\ell^j s}$, which is equal to 1 if $\ell = 0$.

4 Optimal localizing functions

Let us discuss here on the choice of the localizing functions.

By referring to Theorem 3.9, in order to compute $\mathbb{E}(\Phi(X_t) | X_s = \alpha)$ one has to evaluate

$$\mathbb{T}_{s,t}^\psi[f](\alpha) = \mathbb{E}\left(f(X_t) \prod_{i=1}^d \left[\psi_i(\tilde{X}_s^i - \tilde{\alpha}_s^i) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i \right] \right),$$

with $f = \Phi$ and $f = 1$. Such an expectation is practically evaluated by means of the empirical mean obtained through many independent replications. The aim is now to choose the localizing function ψ allowing to reduce the variance. To this purpose, we follow the optimization criterium introduced in the one-dimensional case by Kohatsu-Higa and Petterson [18], which has been used also by Bouchard, Ekeland and Touzi [8]. It consists in looking for the localizing function ψ which minimizes the integrated variance, given by

$$I_d^f(\psi) = \int_{\mathbb{R}^d} \mathbb{E}\left(f^2(X_t) \prod_{i=1}^d \left[\psi_i(\tilde{X}_s^i - \tilde{\alpha}^i) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i \right]^2 \right) d\tilde{\alpha}, \quad (14)$$

up to the constant (with respect to the localizing function ψ) term coming out from $\mathbb{T}_{s,t}^\psi[f](\alpha) = \mathbb{T}_{s,t}[f](\alpha)$. Then the following result holds:

Proposition 4.1 *Set $\mathcal{L}_1 = \{\psi : \mathbb{R} \rightarrow [0, +\infty); \psi \in C^1(\mathbb{R}), \psi(+\infty) = 0 \text{ and } \int_{\mathbb{R}} \psi(t) dt = 1\}$, and $\mathcal{L}_d = \{\psi : \mathbb{R}^d \rightarrow [0, +\infty); \psi(x) = \prod_{i=1}^d \psi_i(x^i), \text{ where } \psi_i \in \mathcal{L}_1, \text{ for any } i\}$. Then*

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

where $\psi^*(x) = \prod_{j=1}^d \psi_j^*(x^j)$, with $\psi_j^*(\xi) = \lambda_j^* e^{-\lambda_j^* |\xi|/2}$ is a Laplace probability density function on \mathbb{R} and $\lambda_j^* = \lambda_j^*[f]$ enjoys the following system of nonlinear equations:

$$\lambda_j^{*2} = \frac{\mathbb{E}\left(f^2(X_t) \Theta_{s,t;j}^2 \prod_{i:i \neq j} [\lambda_i^{*2} + \Theta_{s,t;i}^2]\right)}{\mathbb{E}\left(f^2(X_t) \prod_{i:i \neq j} [\lambda_i^{*2} + \Theta_{s,t;i}^2]\right)}, \quad j = 1, \dots, d, \quad (15)$$

where $\Theta_{s,t;i} = \Delta W_{s,t}^i / (\sigma_{ii}s(t-s)\tilde{X}_s^i)$, $i = 1, \dots, d$.

Proof. Notice first that we can write $I_d^f(\psi) = I_d^f(\psi_1, \dots, \psi_d)$, with $\psi_1, \dots, \psi_d \in \mathcal{L}_1$. In order to compute the derivative in the direction $(\hat{\psi}_1, 0, \dots, 0)$, where $\hat{\psi}_1 = \hat{\psi}_1(x_1)$ is some arbitrary function, we reduce the computation to the one-dimensional case. In fact, for a fixed f , let us put $\tilde{f}_t(y) \equiv \tilde{f}(y) = f \circ F_t(y)$, $y \in \mathbb{R}_+^d$ and F_t being defined in (5). Setting

$$\tilde{f}_1^2(x^1) = \int_{\mathbb{R}^{d-1}} d\tilde{\alpha}^2 \dots d\tilde{\alpha}^d \mathbb{E}\left(\tilde{f}^2(x^1, \tilde{X}_t^2, \dots, \tilde{X}_t^d) \prod_{i=2}^d \left[\psi_i(\tilde{X}_s^i - \tilde{\alpha}^i) + (H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}^i) \Theta_{s,t;i} \right]^2 \right)$$

then

$$I_d^f(\psi) = \int_{\mathbb{R}} d\tilde{\alpha}^1 \mathbb{E}\left(\tilde{f}_1^2(\tilde{X}_t^1) \left[\psi_1(\tilde{X}_s^1 - \tilde{\alpha}^1) + (H - \Psi_1)(\tilde{X}_s^1 - \tilde{\alpha}^1) \Theta_{s,t;1} \right]^2 \right) = I_1^{f_1}(\psi_1).$$

By interchanging the order of integration and by considering the change of variable $\beta = \tilde{X}_s^1 - \tilde{\alpha}^1$, one has

$$I_1^{f_1}(\psi_1) = \mathbb{E} \int_{\mathbb{R}} f_1^2(\tilde{X}_t^1) \left(\psi_1(\beta) + (H - \Psi_1)(\beta) \Theta_{s,t;1} \right)^2 d\beta.$$

Now, set $\varepsilon \in \mathbb{R}$ and let $\hat{\psi}_1 \in L^1(\mathbb{R})$ be such that for any small ε then $\psi_1 + \varepsilon \hat{\psi}_1 \in \mathcal{L}_1$. Setting $\hat{\Psi}_1(x) = \int_{-\infty}^x \hat{\psi}_1(\xi) d\xi$, one has

$$\begin{aligned} \frac{\partial I_d^f(\psi)}{\partial \psi_1}(\hat{\psi}_1) &= (I_1^{f_1})'(\psi_1)(\hat{\psi}_1) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left(I_1^{f_1}(\psi_1 + \varepsilon \hat{\psi}_1) - I_1^{f_1}(\psi_1) \right) \\ &= 2\mathbb{E} \int_{\mathbb{R}} f_1^2(\tilde{X}_t^1) \left(\psi_1(\beta) - \hat{\Psi}_1(\beta) \Theta_{s,t;1} \right) \left(\psi_1(\beta) + (H - \Psi_1)(\beta) \Theta_{s,t;1} \right) d\beta. \end{aligned}$$

We look for a function ψ_1^* in \mathcal{L}_1 such that $(I_1^{f_1})'(\psi_1^*)(\hat{\psi}_1) = 0$ for any $\hat{\psi}_1$ satisfying the conditions above. Consider the first term of the (last) r.h.s.: by using the standard integration by parts formula (recall that $\hat{\Psi}_1' = \hat{\psi}_1$), we can write

$$\begin{aligned} \int_{\mathbb{R}} \hat{\psi}_1(\beta) f_1^2(\tilde{X}_t^1) \left(\psi_1(\beta) + (H - \Psi_1)(\beta) \Theta_{s,t;1} \right) d\beta &= \hat{\Psi}_1(\beta) \tilde{f}_1^2(\tilde{X}_t^1) \left(\psi_1(\beta) + (H - \Psi_1)(\beta) \Theta_{s,t;1} \right) \Big|_{-\infty}^{+\infty} \\ &\quad - \int_{\mathbb{R}} \hat{\Psi}_1(\beta) \tilde{f}_1^2(\tilde{X}_t^1) \partial_\beta \left(\psi_1(\beta) + (H - \Psi_1)(\beta) \Theta_{s,t;1} \right) d\beta \end{aligned}$$

Now, since $\hat{\Psi}_1(\beta) \rightarrow 0$ as $\beta \rightarrow -\infty$ and ψ_1 is a (quite smooth) probability density function, then $\psi_1(\beta) + (H - \Psi_1)(\beta) \rightarrow 0$ as $\beta \rightarrow +\infty$ and the first term nullifies. Moreover, $\partial_\beta(\psi_1(\beta) + (H - \Psi_1)(\beta) \Theta_{s,t;1}) = \psi_1'(\beta) + (\delta_0 - \psi_1)(\beta) \Theta_{s,t;1}$, being δ_0 the Dirac mass in 0. Thus, we obtain

$$\begin{aligned} (I_1^{f_1})'(\psi_1)(\hat{\psi}_1) &= -2\mathbb{E} \int_{\mathbb{R}} \hat{\Psi}_1(\beta) \tilde{f}_1^2(\tilde{X}_t^1) \left(\psi_1'(\beta) + (H - \Psi_1)(\beta) \Theta_{s,t;1}^2 \right) d\beta - 2\hat{\Psi}_1(0) \mathbb{E}(\tilde{f}_1^2(\tilde{X}_t^1) \Theta_{s,t;1}) \\ &= -2 \int_{\mathbb{R}} \hat{\Psi}_1(\beta) \mathbb{E} \left(\tilde{f}_1^2(\tilde{X}_t^1) \left(\psi_1'(\beta) + (H - \Psi_1)(\beta) \Theta_{s,t;1}^2 \right) \right) d\beta, \end{aligned}$$

where we have used $\mathbb{E}(\tilde{f}_1^2(\tilde{X}_t^1) \Theta_{s,t;1}) = 0$, proved in *a*) of Corollary 6.2. Since $\hat{\Psi}_1$ is the primitive function of an almost arbitrarily integrable function, we can say that $(I_1^{f_1})'(\psi_1)(\hat{\psi}_1) = 0$ for any $\hat{\psi}_1$ if and only if

$$\mathbb{E} \left(\tilde{f}_1^2(\tilde{X}_t^1) \left((H - \Psi_1)(\beta) \Theta_{s,t;1}^2 + \psi_1'(\beta) \right) \right) = 0$$

for (almost) any β . This gives rise to an ordinary differential equation: setting

$$\lambda_1^{*2} = \frac{\mathbb{E}(\tilde{f}_1^2(\tilde{X}_t^1) \Theta_{s,t;1}^2)}{\mathbb{E}(\tilde{f}_1^2(X_t))}$$

and $v(\beta) = \Psi_1(\beta)$, then it must be: $v''(\beta) - \lambda_1^{*2} v(\beta) + \lambda_1^{*2} H(\beta) = 0$. The solution is simply $v(\beta) = c_1 e^{\lambda_1^* \beta} + c_2 e^{-\lambda_1^* \beta} + 1$ if $\beta > 0$ and $v(\beta) = c_3 e^{\lambda_1^* \beta} + c_4 e^{-\lambda_1^* \beta}$ if $\beta < 0$, with $c_1, \dots, c_4 \in \mathbb{R}$ and $\lambda_1^* = \sqrt{\lambda_1^{*2}}$. Now, by imposing that $v(+\infty) = 1$, $v(-\infty) = 0$ and $v'(\beta)$ has to be a continuous probability density function, we finally get the solution v^* : $v^*(\beta) = 1 - \frac{1}{2} e^{-\lambda_1^* \beta}$ if $\beta > 0$ and $v^*(\beta) = \frac{1}{2} e^{\lambda_1^* \beta}$ if $\beta < 0$.

Since everything is symmetric, we obtain similar equations and solutions for the remaining coordinates $2, \dots, d$, so that the following system hold: for $j = 1, \dots, d$,

$$\psi_j^*(\beta) = \frac{\lambda_j^*}{2} e^{-\lambda_j^* |\beta|}, \quad \beta \in \mathbb{R},$$

$$\lambda_j^{*2} = \frac{\int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-j} \mathbb{E} \left(f^2(X_t) \Theta_{s,t;j}^2 \prod_{i:i \neq j} \left[\psi_i^*(\tilde{X}_s^i - \tilde{\alpha}^i) + (H - \Psi_i^*)(\tilde{X}_s^i - \tilde{\alpha}^i) \Theta_{s,t;i} \right]^2 \right)}{\int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-j} \mathbb{E} \left(f^2(X_t) \prod_{i:i \neq j} \left[\psi_i^*(\tilde{X}_s^i - \tilde{\alpha}^i) + (H - \Psi_i^*)(\tilde{X}_s^i - \tilde{\alpha}^i) \Theta_{s,t;i} \right]^2 \right)}$$

where “ $d\tilde{\alpha}_{-j}$ ” means that one has to integrate with respect to all the variables except for $\tilde{\alpha}^j$. Let us now show that ψ^* is actually a minimum. Straightforward computations allow to write

$$\sum_{k,j=1}^d \frac{\partial^2 I_d^f}{\partial \hat{\psi}_k \partial \hat{\psi}_j}(\psi^*)(\hat{\psi}_k, \hat{\psi}_j) = 4\mathbb{E} \int_{\mathbb{R}^d} \tilde{f}(\tilde{X}_t)^2 \left(\sum_{m=1}^d (\hat{\psi}_m(\beta_m) - \hat{\Psi}_m(\beta_m) \Theta_{s,t;m}) \right)^2 \times$$

$$\times \prod_{i=1}^d \left(\psi_i^*(\beta_i) + (H - \Psi_i^*)(\beta_i) \Theta_{s,t;i} \right)^2 d\beta$$

which is positive, so that ψ^* is a minimum.

It remains now to give a better representation of the integrals giving the optimal values λ^* . Notice that one has to handle something like

$$\int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-j} \mathbb{E} \left(f^2(X_t) \pi_{s,t;j}^2 \prod_{i:i \neq j} \left[\psi_i^*(\tilde{X}_s^i - \tilde{\alpha}^i) + (H - \Psi_i^*)(\tilde{X}_s^i - \tilde{\alpha}^i) \Theta_{s,t;i} \right]^2 \right)$$

$$= \mathbb{E} \left(f^2(X_t) \pi_{s,t;j}^2 \int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-j} \prod_{i:i \neq j} \left[\psi_i^*(\tilde{X}_s^i - \tilde{\alpha}^i) + (H - \Psi_i^*)(\tilde{X}_s^i - \tilde{\alpha}^i) \Theta_{s,t;i} \right]^2 \right),$$

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 $(H - \Psi_i^*)(\xi) = \text{sign}(\xi) e^{-\lambda_i^* |\xi|/2}$, so that

$$\int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-j} \left[\psi_i^*(\tilde{X}_s^i - \tilde{\alpha}^i) + (H - \Psi_i^*)(\tilde{X}_s^i - \tilde{\alpha}^i) \Theta_{s,t;i} \right]^2$$

$$= \prod_{i:i \neq j} \int_{\mathbb{R}} d\tilde{\alpha}^i \frac{1}{4} e^{-2\lambda_i^* |\tilde{X}_s^i - \tilde{\alpha}^i|} \left[\lambda_i^* + \text{sign}(\tilde{X}_s^i - \tilde{\alpha}^i) \Theta_{s,t;i} \right]^2 = \prod_{i:i \neq j} \frac{1}{4\lambda_i^*} [\lambda_i^{*2} + \Theta_{s,t;i}^2]$$

By inserting everything, one obtains

$$\int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-j} \mathbb{E} \left(f^2(X_t) \pi_{s,t;j}^2 \prod_{i:i \neq j} \left[\psi_i^*(\tilde{X}_s^i - \tilde{\alpha}^i) + (H - \Psi_i^*)(\tilde{X}_s^i - \tilde{\alpha}^i) \Theta_{s,t;i} \right]^2 \right)$$

$$= \frac{1}{4^d \lambda_1^* \dots \lambda_d^*} \mathbb{E} \left(f^2(X_t) \pi_{s,t;j}^2 \prod_{i:i \neq j} [\lambda_i^{*2} + \Theta_{s,t;i}^2] \right),$$

and the statement follows straightforwardly. \square

In the above optimization criterium, in principle one could consider a measure more general than the Lebesgue one, namely to replace $d\tilde{\alpha}$ with $\mu(d\tilde{\alpha})$ in (14). For example, since $\tilde{\alpha}$ represents a possible value for \tilde{X}_s , then it would be more significant to consider $\mu(d\tilde{\alpha})$ as a lognormal measure, or at least a normal one (with suitable mean and covariance matrix). Then, it is worth to say that everything runs in the same way. But unfortunately, one arrives to an ordinary differential equation (as in the proof of Proposition 4.1) which has not an explicit solution, that is for a general μ it is not possible to write down explicitly the optimal localizing function.

Remark 4.2 *The nonlinear system giving the optimal parameters $\lambda_j^* = \lambda_j^*[f]$ can be rewritten as*

$$\lambda_j^*[f]^2 = \frac{\mathbb{E}\left(f^2(X_t) \frac{t + \sigma_{jj}^2 s(t-s)}{\sigma_{jj}^2 s(t-s)(\tilde{X}_s^j)^2} \prod_{i:i \neq j} \left[\lambda_i^*[f]^2 + \frac{t + \sigma_{ii}^2 s(t-s)}{\sigma_{ii}^2 s(t-s)(\tilde{X}_s^i)^2}\right]\right)}{\mathbb{E}\left(f^2(X_t) \prod_{i:i \neq j} \left[\lambda_i^*[f]^2 + \frac{t + \sigma_{ii}^2 s(t-s)}{\sigma_{ii}^2 s(t-s)(\tilde{X}_s^i)^2}\right]\right)}, \quad j = 1, \dots, d. \quad (16)$$

Indeed, consider first the expectation in the numerator of (15). By conditioning with respect to W^j , one has to evaluate $\mathbb{E}(\bar{f}_j^2(\tilde{X}_t^j) \Theta_{s,t;j}^2)$, being

$$\bar{f}_j^2(x^j) = \mathbb{E}\left(\bar{f}^2(\tilde{X}_t^1, \dots, \tilde{X}_t^{j-1}, x^j, \tilde{X}_t^{j+1}, \dots, \tilde{X}_t^d) \prod_{i:i \neq j} [\lambda_i^{*2} + \Theta_{s,t;i}^2]\right).$$

By using b) of Corollary 6.2, one has

$$\mathbb{E}(\bar{f}_j^2(\tilde{X}_t^j) \Theta_{s,t;j}^2) = \mathbb{E}\left(\bar{f}_j^2(\tilde{X}_t^j) \frac{t + \sigma_{jj}^2 s(t-s)}{\sigma_{jj}^2 s(t-s)(\tilde{X}_s^j)^2}\right),$$

so that

$$\mathbb{E}\left(f^2(X_t) \Theta_{s,t;j}^2 \prod_{i:i \neq j} [\lambda_i^{*2} + \Theta_{s,t;i}^2]\right) = \mathbb{E}\left(f^2(X_t) \frac{t + \sigma_{jj}^2 s(t-s)}{\sigma_{jj}^2 s(t-s)(\tilde{X}_s^j)^2} \prod_{i:i \neq j} [\lambda_i^{*2} + \Theta_{s,t;i}^2]\right).$$

Now, by conditioning with respect to W^k as $k \neq j$ and by iterating this procedure, one obtains

$$\mathbb{E}\left(f^2(X_t) \Theta_{s,t;j}^2 \prod_{i:i \neq j} [\lambda_i^{*2} + \Theta_{s,t;i}^2]\right) = \mathbb{E}\left(f^2(X_t) \frac{t + \sigma_{jj}^2 s(t-s)}{\sigma_{jj}^2 s(t-s)(\tilde{X}_s^j)^2} \prod_{i:i \neq j} \left[\lambda_i^{*2} + \frac{t + \sigma_{ii}^2 s(t-s)}{\sigma_{ii}^2 s(t-s)(\tilde{X}_s^i)^2}\right]\right).$$

Similarly,

$$\mathbb{E}\left(f^2(X_t) \prod_{i:i \neq j} [\lambda_i^{*2} + \Theta_{s,t;i}^2]\right) = \mathbb{E}\left(f^2(X_t) \prod_{i:i \neq j} \left[\lambda_i^{*2} + \frac{t + \sigma_{ii}^2 s(t-s)}{\sigma_{ii}^2 s(t-s)(\tilde{X}_s^i)^2}\right]\right),$$

so that formula (16) holds.

Remark 4.3 *For $f = 1$ the corresponding optimal values of the parameters λ_j can be exactly computed and are given by (recall that x^1, \dots, x^d are the initial underlying asset prices)*

$$\lambda_j^*[1] = \frac{e^{-(h^j s + l_s^j) + \sigma_{jj}^2 s}}{x^j} \sqrt{\frac{t + \sigma_{jj}^2 s(t-s)}{\sigma_{jj}^2 s(t-s)}}, \quad j = 1, \dots, d.$$

Indeed, whenever $f = 1$, by Remark 4.2, one has (recall that the \tilde{X}^i 's are independent)

$$\lambda_j^*[1]^2 = \mathbb{E}\left(\frac{t + \sigma_{jj}^2 s(t-s)}{\sigma_{jj}^2 s(t-s)(\tilde{X}_s^j)^2}\right), \quad j = 1, \dots, d.$$

Now, since $\tilde{X}_s^j = x^j \exp(h^j s + \ell_s^j + \sigma_{jj} W_s^j)$, one has $\mathbb{E}((\tilde{X}_s^j)^{-2}) = (x^j)^{-2} e^{-2(h^j s + \ell_s^j)} \mathbb{E}(e^{-2\sigma_{jj} W_s^j}) = (x^j)^{-2} e^{-2(h^j s + \ell_s^j)} e^{2\sigma_{jj}^2 s}$, and the above formula immediately follows.

For practical purposes, numerical evidence shows that the choice $\lambda^* = 1/\sqrt{t-s}$ works good enough, thus avoiding to weight the algorithm with the computation of further expectations. When $f = 1$, this kind of behaviour is clear from Remark (4.3). In the general case, we can state the following

Proposition 4.4 *For any $j = 1, \dots, d$, one has $\lambda_j^*[f] = O(1/\sqrt{t-s})$ as $t \rightarrow s$. Moreover, if f is continuous, then*

$$\lim_{\sigma \rightarrow 0} \lim_{t \rightarrow s} \frac{\lambda_j^*[f]}{\lambda_j^*[1]} = 1.$$

Proof. Let us set:

$$\nu_j^* = \lambda_j^*[f]^2 \sigma_{jj}^2 s(t-s) \quad \text{and} \quad \bar{\nu}_j = \lambda_j^*[1]^2 \sigma_{jj}^2 s(t-s),$$

as $j = 1, \dots, d$. By (16), $\nu^* = (\nu_1^*, \dots, \nu_d^*)$ is such that for any $j = 1, \dots, d$,

$$\nu_j^* = \frac{\mathbb{E}\left(f^2(X_t) \frac{t + \sigma_{jj}^2 s(t-s)}{(\tilde{X}_s^j)^2} \prod_{i:i \neq j} \left[\nu_i^* + \frac{t + \sigma_{ii}^2 s(t-s)}{(\tilde{X}_s^i)^2}\right]\right)}{\mathbb{E}\left(f^2(X_t) \prod_{i:i \neq j} \left[\nu_i^* + \frac{t + \sigma_{ii}^2 s(t-s)}{(\tilde{X}_s^i)^2}\right]\right)}$$

and since the above quantity is bounded as $(t-s) \rightarrow 0$, it then follows that $\nu_j^* = O(1/(t-s))$, or equivalently $\lambda_j^*[f] = O(1/\sqrt{t-s})$.

Let us now discuss what happens whenever also $\sigma \rightarrow 0$. One has $(\tilde{X}_s^j)^{-2} = (x^j)^{-2} \exp(-2h^j s - 2\ell_s^j - 2\sigma_{jj} W_s^j) = (x^j)^{-2} \exp(-2h^j s - 2\ell_s^j + 2\sigma_{jj}^2 s) \exp(-2\sigma_{jj} W_s^j - 2\sigma_{jj} W_s^j)$. Setting $\exp(-2\sigma_{jj} W_s^j - 2\sigma_{jj} W_s^j) = 1 + Y_s^j$, we can write, as $j = 1, \dots, d$, (see Remark 4.3)

$$\frac{t + \sigma_{jj}^2 s(t-s)}{(\tilde{X}_s^j)^2} = \bar{\nu}_j (1 + Y_s^j),$$

so that

$$\nu_j^* = \bar{\nu}_j \frac{\mathbb{E}\left(f^2(X_t) (1 + Y_s^j) \prod_{i:i \neq j} \left[\nu_i^* + \frac{t + \sigma_{ii}^2 s(t-s)}{(\tilde{X}_s^i)^2}\right]\right)}{\mathbb{E}\left(f^2(X_t) \prod_{i:i \neq j} \left[\nu_i^* + \frac{t + \sigma_{ii}^2 s(t-s)}{(\tilde{X}_s^i)^2}\right]\right)}$$

and

$$\frac{\nu_j^*}{\bar{\nu}_j} = 1 + \frac{\mathbb{E}\left(f^2(X_t) Y_s^j \prod_{i:i \neq j} \left[\nu_i^* + \frac{t + \sigma_{ii}^2 s(t-s)}{(\tilde{X}_s^i)^2}\right]\right)}{\mathbb{E}\left(f^2(X_t) \prod_{i:i \neq j} \left[\nu_i^* + \frac{t + \sigma_{ii}^2 s(t-s)}{(\tilde{X}_s^i)^2}\right]\right)}.$$

Now, by applying twice the Lebesgue dominated convergence theorem, one has (recall that f is continuous and $Y_s^j \rightarrow 0$ a.s. as $\sigma \rightarrow 0$)

$$\lim_{\sigma \rightarrow 0} \lim_{t \downarrow s} \frac{\mathbb{E}\left(f^2(X_t) Y_s^j \prod_{i: i \neq j} \left[\nu_i^* + \frac{t + \sigma_{ii}^2 s(t-s)}{(\tilde{X}_s^i)^2}\right]\right)}{\mathbb{E}\left(f^2(X_t) \prod_{i: i \neq j} \left[\nu_i^* + \frac{t + \sigma_{ii}^2 s(t-s)}{(\tilde{X}_s^i)^2}\right]\right)} = 0,$$

and the statement holds. \square

Similar arguments can be used in order to handle the problem of minimizing the variance coming out from the expectation giving the operators $\mathbb{G}_{s,t;k}^\psi[f](\alpha)$ (see (12)). To this purpose, let us recall that

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

where

$$\Theta_{s,t;k} = \frac{\Delta W_{s,t}^k}{\sigma_{kk} s(t-s) \tilde{X}_s^k}, \quad \Upsilon_{s,t;k} = \frac{1}{\sigma_{kk} s(t-s) (\tilde{X}_s^k)^2} \left(\frac{(\Delta W_{s,t}^k)^2}{\sigma_{kk} s(t-s)} + \Delta W_{s,t}^k - \frac{t}{\sigma_{kk}} \right) \quad (17)$$

and

$$\pi_{s,t;k}((\tilde{X}_s - \tilde{\alpha})_{-k}) = \prod_{i=1, i \neq k}^d \left(\psi_i(\tilde{X}_s^i - \tilde{\alpha}^i) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}^i)}{\sigma_{ii} s(t-s) \tilde{X}_s^i} \Delta W_{s,t}^i \right)$$

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 k^{th} one. In the one dimensional case, one simply has to drop the subscript k , take $\tilde{\alpha} = \alpha$ and set $\pi_{s,t;k}(\tilde{\alpha}_{-k}) \equiv 1$. Thus, if we set

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

then the following result holds:

Proposition 4.5 *Let $\Upsilon_{s,t;k}$ and $\Theta_{s,t;k}$ be as in (17). Setting $\mathcal{L}_d = \{\psi : \mathbb{R}^d \rightarrow [0, +\infty); \psi(x) = \prod_{i=1}^d \psi_i(x^i), \text{ where } \psi_i \in \mathcal{L}_1, \text{ for any } i\}$, then*

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

where $\psi^{*;k}(x) = \prod_{j=1}^d \psi_{j;k}^*(x^j)$, with $\psi_{j;k}^*(\xi) = \frac{\mu_{j;k}^*}{2} e^{-\mu_{j;k}^* |\xi|}$ is a Laplace probability density function and $\mu_{j;k}^* = \mu_{j;k}^*[f]$ are given by:

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

$$\mu^* = \mu^*[f] = \left(\frac{\mathbb{E}\left(f^2(X_t) \Upsilon_{s,t}^2\right)}{\mathbb{E}\left(f^2(X_t) \Theta_{s,t}^2\right)} \right)^{1/2};$$

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

$$\begin{aligned} (\mu_{k;k}^*)^2 &= \frac{\mathbb{E}\left(f^2(X_t) \Upsilon_{s,t;k}^2 \prod_{i=1, i \neq k}^d \left[(\mu_{i;k}^*)^2 + \Theta_{s,t;j}^2\right]\right)}{\mathbb{E}\left(f^2(X_t) \Theta_{s,t;i}^2 \prod_{i=1, i \neq k}^d \left[(\mu_{i;k}^*)^2 + \Theta_{s,t;i}^2\right]\right)} \quad \text{and for } j \neq k : \\ (\mu_{j;k}^*)^2 &= \frac{\mathbb{E}\left(f^2(X_t) \left((\mu_{k;k}^*)^2 \Theta_{s,t;k}^2 + \Upsilon_{s,t;k}^2\right) \Theta_{s,t;j}^2 \prod_{i=1, i \neq j, k}^d \left[(\mu_{i;k}^*)^2 + \Theta_{s,t;i}^2\right]\right)}{\mathbb{E}\left(f^2(X_t) \left((\mu_{k;k}^*)^2 \Theta_{s,t;k}^2 + \Upsilon_{s,t;k}^2\right) \prod_{i=1, i \neq j, k}^d \left[(\mu_{i;k}^*)^2 + \Theta_{s,t;i}^2\right]\right)}. \end{aligned}$$

The proof of Proposition 4.5 uses the same technique as in the proof Proposition 4.1 and is postponed to Section 6.2 of the Appendix. Again, numerical evidences show that the choice $\mu_{j;k}^* = 1/\sqrt{t-s}$ works good enough.

Finally, let us conclude the discussion on the optimal localizing function 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 rewritten as

$$\mathbb{E}(\Phi(X_t) | X_s = \alpha) = \mathbb{E}\left(\Phi(X_t) \frac{\pi_{s,t}^\psi}{\mathbb{E}(\pi_{s,t}^\psi)}\right), \quad \pi_{s,t}^\psi = \psi(X_s - \alpha) + \frac{(H - \Psi)(X_s - \alpha)}{\sigma s(t-s)X_s} \Delta W_{s,t}$$

(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 $\pi_{s,t}^\psi / \mathbb{E}(\pi_{s,t}^\psi)$. But since it is written in terms of a ratio, at this stage it does not seem feasible to obtain results giving the associated optimal localizing function ψ .

5 The algorithm for pricing/hedging American options

We give here a detailed presentation of the use of the representation formulas in the applied context of the pricing and hedging of American options. Afterwards, we show some numerical results coming out by applying the algorithm in practice.

5.1 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 (3).

Let $0 = t_0 < t_1 < \dots < t_n = T$ 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 $k = 0, 1, \dots, n$, is iteratively defined as:

$$\begin{aligned} \bar{P}_{n\varepsilon}(X_{n\varepsilon}) &= \Phi(X_{n\varepsilon}) \equiv \Phi(X_T) \quad \text{and as } k = n-1, \dots, 1, 0 \\ \bar{P}_{k\varepsilon}(X_{k\varepsilon}) &= \max \left\{ \Phi(X_{k\varepsilon}), \underbrace{e^{-r\varepsilon} \mathbb{E}\left(\bar{P}_{(k+1)\varepsilon}(X_{(k+1)\varepsilon}) \mid X_{k\varepsilon}\right)}_{\clubsuit} \right\} \end{aligned} \quad (18)$$

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

$$\bar{\Delta}(X_\varepsilon) = \begin{cases} \partial_\alpha \Phi(\alpha) \Big|_{\alpha=X_\varepsilon} & \text{if } \bar{P}_\varepsilon(X_\varepsilon) < \Phi(X_\varepsilon) \\ e^{-r\varepsilon} \underbrace{\partial_\alpha \mathbb{E} \left(\bar{P}_{2\varepsilon}(X_{2\varepsilon}) \Big| X_\varepsilon = \alpha \right)}_{\clubsuit} \Big|_{\alpha=X_\varepsilon} & \text{if } \bar{P}_\varepsilon(X_\varepsilon) > \Phi(X_\varepsilon) \end{cases} \quad (19)$$

then $\bar{\Delta}_0(x) = \mathbb{E}_x \left(\bar{\Delta}(X_\varepsilon) \right)$.

The conditional expectations (terms \clubsuit) and their derivatives (terms \spadesuit) will be computed through the formulas previously given, by means of unconditioned expectations which in turn are numerically evaluated by averaging on N simulated paths.

Remark 5.1 *In the context of the geometric Brownian motion, the process X can be exactly simulated. So, we do not need any 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 compute all the conditional expectations, as it will follows.*

Since the algorithm is backward, for the simulation we consider the Brownian bridge law (recall that for $0 < s < t$, the law of W_s given $W_t = y$ is a gaussian law with mean $s/t y$ and variance $s(t-s)/t$). Therefore,

- at time $T = n\varepsilon$, we simulate $W_{n\varepsilon}$ as usual: $W_{n\varepsilon} = \sqrt{n\varepsilon} U_n$, with $U_n = (U_n^1, \dots, U_n^d)$, $U_n^i \sim \mathcal{N}(0, 1)$, $i = 1, \dots, d$, all independent;
- as $k = n-2, \dots, 1$, we simulate by the Brownian bridge property: $W_{k\varepsilon} = k/(k+1) W_{(k+1)\varepsilon} + \sqrt{k\varepsilon/(k+1)} U_k$, with $U_k = (U_k^1, \dots, U_k^d)$, $U_k^i \sim \mathcal{N}(0, 1)$, $i = 1, \dots, d$, independent.

Thus, the basic data in the algorithm are given by

$$\mathcal{U} = \{ \underbrace{U_k^{i,q}}_{\text{time}} ; \underbrace{k=1, \dots, n}_{\text{dimension}}, \underbrace{i=1, \dots, d}_{\text{sample}}, \underbrace{q=1, \dots, N} \} \quad (20)$$

and \mathcal{U} gives all the samples we are interested in, that is:

- $(W_{k\varepsilon}^{i,q})_{i=1, \dots, d; k=1, \dots, n, q=1, \dots, N}$: as $i = 1, \dots, d$,

$$\begin{aligned} \text{for } k = n : \quad & W_{n\varepsilon}^{i,q} = \sqrt{n\varepsilon} U_n^{i,q}, \text{ and} \\ \text{for } k = n-1, \dots, 1 : \quad & W_{k\varepsilon}^{i,q} = \frac{k}{k+1} W_{(k+1)\varepsilon}^{i,q} + \sqrt{\frac{k}{k+1}} \varepsilon U_k^{i,q}, \end{aligned} \quad (21)$$

- $(X_{k\varepsilon}^{i,q})_{i=1, \dots, d; k=0, \dots, n, q=1, \dots, N}$: for $k = n, n-1, \dots, 1$,

$$X_{k\varepsilon}^{i,q} = x_i e^{(r-\eta_i - \frac{1}{2} \sum_{j=1}^i \sigma_{ij}^2)k\varepsilon + \sum_{j=1}^i \sigma_{ij} W_{k\varepsilon}^{j,q}}, \quad i = 1, \dots, d \quad (22)$$

and $X_0^{i,q} = x_i$, $i = 1, \dots, d$ (as an example, Figure 1 shows a set of simulated paths of X);

- $\{\Delta W_k^{i,q}\}_{i=1, \dots, d; k=1, \dots, n-1, q=1, \dots, N}$ (giving a sample for $\Delta W_{s,t}$): for $k = n, n-1, \dots, 1$,

$$\Delta W_k^{i,q} = \varepsilon (W_{k\varepsilon}^{i,q} + \sigma_{ii} k\varepsilon) - k\varepsilon (W_{(k+1)\varepsilon}^{i,q} - W_{k\varepsilon}^{i,q}), \quad 1, \dots, d; \quad (23)$$

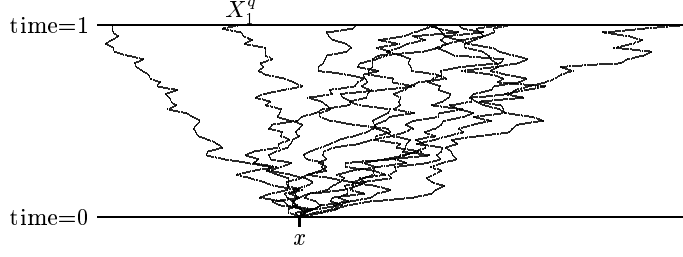


Figure 1 An example of the tree turning out by simulating 10 paths of the process X on $[0, 1]$.

- $(\tilde{X}_{k\varepsilon}^{i,q})_{i=1,\dots,d; k=0,\dots,n, q=1,\dots,N}$: for $k = n, \dots, 1$,

$$\tilde{X}_{k\varepsilon}^{i,q} = x_i e^{(r-\eta_i - \frac{1}{2} \sum_{j=1}^i \sigma_{ij}^2)k\varepsilon + \ell_k^{i,q} k\varepsilon + \sigma_{ii} W_{k\varepsilon}^{i,q}}, \quad i = 1, \dots, d, \quad (24)$$

and $\tilde{X}_0^{i,q} = x_i$, $i = 1, \dots, d$. Obviously, if $d > 1$ one needs to introduce the drift ℓ , which could vary according to the time interval of interest $[k\varepsilon, (k+1)\varepsilon]$ and also on the position of X at time $k\varepsilon$ (see Remark 3.7). So, $\tilde{X}_{k\varepsilon}^{i,q}$ can be simulated once the following set \mathcal{L} is chosen:

$$\mathcal{L} = \{\ell_k^{i,q}; \underbrace{k = 1, \dots, n-1}_{\text{time}}, \underbrace{i = 1, \dots, d}_{\text{dimension}}, \underbrace{q = 1, \dots, N}_{\text{sample}}\}.$$

The pricing algorithm can be summarized step by step as follows.

- **Step n** : simulation of $W_{n\varepsilon}^q$ and $X_{n\varepsilon}^q$ as in (21) and (22); then, computation of $\bar{P}_{t_n}(X_{n\varepsilon}^q)$:

$$\bar{P}_{t_n}(X_{n\varepsilon}^q) = \Phi(X_{n\varepsilon}^q) \quad q = 1, \dots, N.$$

- **Step $n-1$** : simulation of $W_{(n-1)\varepsilon}^q$, $X_{(n-1)\varepsilon}^q$, ΔW_{n-1}^q and $\tilde{X}_{(n-1)\varepsilon}^q$ as in (21), (22), (23) and (24); then, computation of $\bar{P}_{(n-1)\varepsilon}(X_{(n-1)\varepsilon}^q)$:

$$\bar{P}_{(n-1)\varepsilon}(X_{(n-1)\varepsilon}^q) = \max \left(\Phi(X_{(n-1)\varepsilon}^q), e^{-r\varepsilon} \mathbb{E}(\bar{P}_{n\varepsilon}(X_{n\varepsilon}) | X_{(n-1)\varepsilon} = X_{(n-1)\varepsilon}^q) \right)$$

for $q = 1, \dots, N$, where the expectation is evaluated by the formulas in the previous sections, for example

$$\mathbb{E}(\bar{P}_{n\varepsilon}(X_{n\varepsilon}) | X_{(n-1)\varepsilon} = X_{(n-1)\varepsilon}^q) = \frac{\mathbb{T}^{n-1}[\bar{P}_{n\varepsilon}](X_{(n-1)\varepsilon}^q)}{\mathbb{T}^{n-1}[1](X_{(n-1)\varepsilon}^q)}$$

where, for $f = \bar{P}_{n\varepsilon}, 1$,

$$\mathbb{T}^{n-1}[f](X_{(n-1)\varepsilon}^q) = \mathbb{E} \left(f(X_{n\varepsilon}) \prod_{i=1}^d \frac{H(\tilde{X}_{(n-1)\varepsilon}^i - \tilde{\alpha}^i)}{\tilde{X}_{(n-1)\varepsilon}^i} \Delta W_{(n-1)\varepsilon, n\varepsilon}^i \right) \Big|_{\tilde{\alpha} = X_{(n-1)\varepsilon}^q}$$

Notice that $\mathbb{T}^{n-1}[f](X_{(n-1)\varepsilon}^q)$ is the mean of a random variable for which one has N trials, so that it is numerically computed by the law of large numbers:

$$\mathbb{T}^{n-1}[f](X_{(n-1)\varepsilon}^q) = \frac{1}{N} \sum_{q'=1}^N f(X_{n\varepsilon}^{q'}) \prod_{i=1}^d \frac{H(\tilde{X}_{(n-1)\varepsilon}^{i,q'} - \tilde{X}_{(n-1)\varepsilon}^{i,q})}{\tilde{X}_{(n-1)\varepsilon}^{i,q'}} \Delta W_{n-1}^{i,q'}.$$

- **Step** k , $k = n - 2, \dots, 1$: as in step $n - 1$, with k in place of $n - 1$.

At step 1, from (19) one has to add the following computation for the delta:

$$\bar{\Delta}(X_\varepsilon^q) = \begin{cases} \partial_\alpha \Phi(\alpha) \Big|_{\alpha=X_\varepsilon^q} & \text{if } \bar{P}_\varepsilon(X_\varepsilon^q) < \Phi(X_\varepsilon^q) \\ e^{-r\varepsilon} \partial_\alpha \mathbb{E} \left(\bar{P}_{2\varepsilon}(X_{2\varepsilon}) \Big| X_\varepsilon = \alpha \right) \Big|_{\alpha=X_\varepsilon^q} & \text{if } \bar{P}_\varepsilon(X_\varepsilon^q) > \Phi(X_\varepsilon^q) \end{cases}$$

where $\partial_\alpha \mathbb{E}(\bar{P}_{2\varepsilon}(X_{2\varepsilon}) | X_\varepsilon = \alpha) \Big|_{\alpha=X_\varepsilon^q}$ is computed by using the formulas in the previous sections, for example:

$$\begin{aligned} & \partial_{\alpha^j} \mathbb{E} \left(F(X_{2\varepsilon}) = \Big| X_\varepsilon = \alpha \right) \Big|_{\alpha=X_\varepsilon^q} \\ &= \sum_{m=1}^j \hat{\sigma}_{mj} \frac{\tilde{X}_\varepsilon^{m,q}}{X_\varepsilon^{j,q}} \times \frac{\mathbb{G}_{s,t;m}[F](X_\varepsilon^q) \mathbb{T}_{s,t}[1](X_\varepsilon^q) - \mathbb{T}_{s,t}[F](X_\varepsilon^q) \mathbb{G}_{s,t;m}[1](X_\varepsilon^q)}{\mathbb{T}_{s,t}[1](X_\varepsilon^q)^2}, \end{aligned} \quad (25)$$

with $s = \varepsilon$ and $t = 2\varepsilon$. $\mathbb{T}_{s,t}$ and $\mathbb{G}_{s,t;m}$, given by (8) and (9) respectively, are weighted expectations of random variables for which we have N samples, so they are evaluated in practice by means of the associated empirical mean.

At the end, the samples $(\bar{P}_\varepsilon(X_\varepsilon^q), \bar{\Delta}(X_\varepsilon^q))_{q=1, \dots, N}$ are available.

- **Step** 0: computation of the price and the delta:

$$\bar{P}_0(x) = \max \left(\Phi(x), e^{-r \Delta t} \frac{1}{N} \sum_{q=1}^N \bar{P}_\varepsilon(X_\varepsilon^q) \right) \quad \text{and} \quad \bar{\Delta}_0(x) = \frac{1}{N} \sum_{q=1}^N \bar{\Delta}(X_\varepsilon^q).$$

Let us point out that, for the sake of simplicity, the description takes into account the non localized formulas. In practice, it is much better to use localizing functions, so one should use the formulas coming from Theorem 3.9. Obviously, nothing changes and for the choice of the localizing functions, we refer to the discussion in Section 4.

Furthermore, in order to reduce the variance, one could use a control variable. Unfortunately, there is not a standard way to proceed in this direction but in the case of American option, it is quite natural to think to the associated European one. The idea is the following.

For a fixed initial time t and underlying asset price x , let us set $P^{\text{am}}(t, x)$ and $P^{\text{eu}}(t, x)$ as the price of an American and European option respectively, with the same payoff Φ and maturity T . We define

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

Then it is easy to see that

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

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

$$\hat{\Phi}(t, x) = \Phi(x) - P^{\text{eu}}(t, x)$$

(notice the obstacle $\hat{\Phi}(t, x)$ is now dependent on the time variable also, and is such that $\hat{\Phi}(T, x) = 0$). 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 Φ is replaced by $\hat{\Phi}(t, x)$.

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

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

respectively. Notice that the new obstacle has to be evaluated at each time step: in order to set up this program, one has to compute the price/delta of an European option on Φ . This can be done exactly for some call or put options, for which prices and deltas are known in closed form, otherwise one can proceed by simulation for their computation.

5.2 Numerical experiments

Here, we numerically study the behavior of the pricing/hedging algorithm described in Section 5.1. In order to have comparable results, a symmetric case is considered, where the deltas are all equal. Thus, the initial values are assumed to be all equal, and the same for the volatilities; we also set equal to zero both the dividend rates and the correlations among the assets. The parameters are: initial values $x_1 = \dots = x_d = 100$; volatilities $\sigma_{11} = \dots = \sigma_{dd} = 0.2$; exercise price $K = 100$; maturity $T = 1$ year; risk-free interest rate $r = \ln(1.1)$.

We consider the following examples.

- **Standard one dimensional American put**, payoff $\Phi(x) = (K - x)_+$, see Table 1.

The “true” reference price and delta are issued by the binomial Black-Scholes Richardson extrapolation tree-method [10] (BBSR- P and BBSR- Δ in the table), with 1000 steps.

- **Put on the minimum of 2 assets**, payoff $\Phi(x) = (K - \min(x_1, x_2))_+$, see Table 2.

The “true” reference price and deltas are obtained by the Villeneuve-Zanette finite difference algorithm [25] (VZ- P and VZ- Δ in the table), with 500 time-space steps.

- **Put on the geometric mean of 5 assets**: payoff $\Phi(x) = (K - (\prod_{i=1}^5 x_i)^{1/5})_+$, see Table 3.

The peculiarity of the payoff allows one to benchmark the results with the one-dimensional American BBSR tree-method [10] (BBSR- P and BBSR- Δ in the table), with 1000 steps.

We have considered $N_{\text{mc}} = 500, 1000, 5000, 10000, 20000$ number of Monte Carlo iterations and varying time periods n ($n = 10, 20, 50$ in dimension $d = 1, 2$ and $n = 5, 10, 20$ for $d = 5$). In the tables, \bar{P} and $\bar{\Delta}$ denote the approximating price and delta respectively, and the “true” reference price and delta are reported. 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-Martineau [6] and the Longstaff-Schwartz [20] algorithm¹ are reported (prices and deltas are denoted as BM- P , BM- Δ and LS- P , LS- Δ , respectively).

In any experiment, the price of the associated European 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 4). In order to relax the computational executing time, we have always used the parameter of the localizing function equal to $1/\sqrt{\varepsilon}$, 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 as in Proposition 3.5 (see Remark 3.7 for details). Actually, the results are not influenced by the two possibilities: the outcomes are quite identical.

¹The 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 |
|-----------------|----------|-----------|-----------|---------|---------|----------------|----------------|--------------|--------------|
| 10 time periods | 500 | 4.807 | | 5.611 | 4.914 | -0.378 | | -0.316 | -0.284 |
| | 1000 | 4.795 | | 5.147 | 4.714 | -0.387 | | -0.325 | -0.232 |
| | 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 |
| 20 time periods | 500 | 4.896 | | 5.469 | 5.025 | -0.398 | | -0.284 | -0.237 |
| | 1000 | 4.896 | | 5.198 | 4.903 | -0.385 | | -0.261 | -0.251 |
| | 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 |
| 50 time periods | 500 | 4.936 | | 5.347 | 5.165 | -0.384 | | -0.261 | -0.258 |
| | 1000 | 4.951 | | 5.904 | 5.118 | -0.384 | | -0.270 | -0.3018 |
| | 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 |
|-----------------|----------|-----------|---------|------------------|------------------|--------------|
| 10 time periods | 500 | 10.298 | | -0.288 | -0.296 | |
| | 1000 | 10.283 | | -0.295 | -0.294 | |
| | 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 | |
| 20 time periods | 500 | 10.388 | | -0.302 | -0.294 | |
| | 1000 | 10.371 | | -0.293 | -0.292 | |
| | 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 | |
| 50 time periods | 500 | 10.511 | | -0.306 | -0.287 | |
| | 1000 | 10.443 | | -0.288 | -0.289 | |
| | 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

5.3 Final comments

The numerical values for the price and delta in Table 1, 2 and 3, show good accordance with the “true” values. Notice that this holds for a not high number of both time periods ($n = 10$) and Monte Carlo iterations ($N_{mc} = 500, 1000$). Moreover, the tables suggest that, as the dimension increases, the convergence in terms of N_{mc} 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 [9] (Theorem 6.3) have proved that, in a very similar simulation scheme (in particular, using localizing functions giving formulas rather different from the ones as in Lemma 3.8), 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 $\varepsilon \rightarrow 0$ (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 $\varepsilon \rightarrow 0$. This would suggest that as the dimension increases, one should increase also the number of Monte Carlo iterations in order to achieve good results.

| | N_{mc} | \bar{P} | BBSR- P | $\bar{\Delta}_1$ | $\bar{\Delta}_2$ | $\bar{\Delta}_3$ | $\bar{\Delta}_4$ | $\bar{\Delta}_5$ | BBSR- Δ |
|-----------------|----------|-----------|-----------|------------------|------------------|------------------|------------------|------------------|----------------|
| 5 time periods | 500 | 1.525 | | -0.0752 | -0.0741 | -0.0824 | -0.0767 | -0.0708 | |
| | 1000 | 1.535 | | -0.0743 | -0.0739 | -0.0744 | -0.0687 | -0.0784 | |
| | 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 | |
| 10 time periods | 500 | 1.716 | | -0.0781 | -0.0766 | -0.0868 | -0.0934 | -0.0708 | |
| | 1000 | 1.684 | | -0.0763 | -0.0756 | -0.0839 | -0.0786 | -0.0823 | |
| | 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 | |
| 20 time periods | 500 | 1.800 | | -0.0777 | -0.0754 | -0.0681 | -0.0827 | -0.0643 | |
| | 1000 | 1.846 | | -0.0851 | -0.0872 | -0.0753 | -0.0720 | -0.0768 | |
| | 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

Table 4, referring to the simplest case of the one dimensional American put, concerns what happens if one does not consider the localization function and/or the control variable.

| N_{mc} | \bar{P} : | yes localization no control variable | no localization yes control variable | no localization no control variable |
|----------|-------------|---|---|--|
| 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, Table 4 points out that the algorithm numerically does not work if the localization is not taken into account. On the contrary, the use of the localization, even without any control variables, gives unstable but rather reasonable prices. Thus, Table 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 ($N_{mc} = 500, 1000$).

Let us discuss about the CPU time arising from the running of the algorithm. The 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 one dimensional American put option. 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.

We can conclude that this method is interesting but uncompetitive with respect to other Monte Carlo methods, such as the Barraquand-Martineau [6] and the Longstaff-Schwartz [20] ones, in terms of computing time. Nevertheless, differently from these procedures (which do not provide an *ad hoc* way for the Greeks: they are computed simply by finite differences), this method allows to efficiently obtain the delta. In [12], a suitable combination of this procedure with other Monte Carlo ones is developed, giving interesting numerical results also in terms of computing time costs.

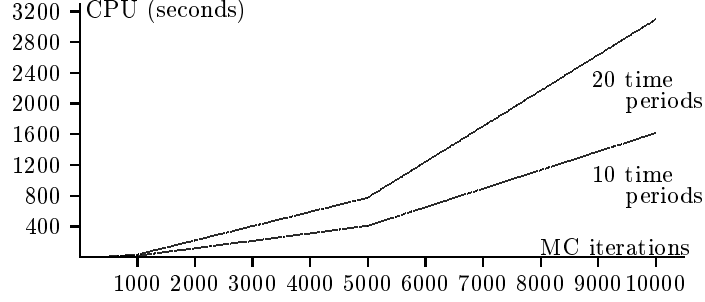


Figure 2 Required CPU time (in seconds) in terms of Monte Carlo iterations.

6 Appendix

6.1 Proofs of the result in Section 3

Let us first consider the following result, concerning the one-dimensional case:

Lemma 6.1 *Let $X_t = x e^{\mu t + \sigma W_t}$, being μ a deterministic drift. Suppose $f, g : \mathbb{R} \rightarrow \mathbb{R}$, where f has a polynomial growth and g has a continuous derivative, and $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}$ with continuous first derivatives. Then for any $0 < s < t$ one has:*

$$\begin{aligned} & \mathbb{E}(f(X_t) g'(X_s) \varphi(X_s, W_s, W_t)) \\ &= \mathbb{E}\left(f(X_t) g(X_s) \left[\varphi(X_s, W_s, W_t) \frac{\Delta W_{s,t}}{\sigma s(t-s) X_s} - \varphi_x(X_s, W_s, W_t) - \frac{\varphi_y(X_s, W_s, W_t)}{\sigma X_s} \right]\right), \end{aligned}$$

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

As a consequence, for any fixed $\alpha \in \mathbb{R}$, the following formulas hold:

$$\begin{aligned} i) \quad & \mathbb{E}(f(X_t) g'(X_s - \alpha)) = \mathbb{E}\left(f(X_t) \frac{g(X_s - \alpha)}{\sigma s(t-s) X_s} \Delta W_{s,t}\right); \\ ii) \quad & \mathbb{E}\left(f(X_t) \frac{g'(X_s - \alpha)}{\sigma s(t-s) X_s} \Delta W_{s,t}\right) = \mathbb{E}\left(f(X_t) \frac{g(X_s - \alpha)}{\sigma s(t-s) X_s^2} \left[\frac{(\Delta W_{s,t})^2}{\sigma s(t-s)} + \Delta W_{s,t} - \frac{t}{\sigma} \right]\right). \end{aligned}$$

Proof.

i) The proof consists in applying twice the Malliavin Integration by Parts (MIbP) formula, first on the time interval $[0, s]$ and secondly over $[s, t]$.

1) Use of the MIbP formula over $[0, s]$.

If D_r denotes the Malliavin derivative (recall that $D_r g(X_s) = \partial_{\Delta W_r} g(X_s) = g'(X_s) D_r X_s$), one has $D_r g(X_s) = g'(X_s) \sigma X_s$ for any $r < s$. Therefore, $g'(X_s) = \int_0^s D_r g(X_s) / (\sigma s X_s) dr$ and

$$\begin{aligned} \mathbb{E}(f(X_t) g'(X_s) \varphi(X_s, W_s, W_t)) &= \mathbb{E}\left(\int_0^s D_r g(X_s) \cdot \frac{f(X_t)}{\sigma s X_s} \varphi(X_s, W_s, W_t) dr\right) \\ &= \mathbb{E}\left(g(X_s) \int_0^s \frac{f(X_t)}{\sigma s X_s} \varphi(X_s, W_s, W_t) dW_r\right), \end{aligned}$$

where the latter equality comes from the application of the MIbP formula over $[0, s]$ (notice that the stochastic integral above is a Skorohod one). Now, recalling the property $\int_0^s G dW_r = G W_s - \int_0^s D_r G dr$, one obtains

$$\int_0^s \frac{f(X_t)}{\sigma s X_s} \varphi(X_s, W_s, W_t) dW_r = \frac{f(X_t)}{\sigma s X_s} \varphi(X_s, W_s, W_t) W_s - \int_0^s D_r \left(\frac{f(X_t)}{\sigma s X_s} \varphi(X_s, W_s, W_t) \right) dr$$

$$= f(X_t) \left[\varphi(X_s, W_s, W_t) \frac{W_s + \sigma s}{\sigma s X_s} - \frac{1}{\sigma X_s} \left(\varphi_x(X_s, W_s, W_t) \sigma X_s + \varphi_y(X_s, W_s, W_t) + \varphi_z(X_s, W_s, W_t) \right) \right] \\ - f'(X_t) \varphi(X_s, W_s, W_t) \frac{X_t}{X_s},$$

so that

$$\mathbb{E}(f(X_t)g'(X_s)) = \mathbb{E} \left(f(X_t)g(X_s) \left[\varphi(X_s, W_s, W_t) \frac{W_s + \sigma s}{\sigma s X_s} - \frac{1}{\sigma X_s} (\varphi_x(X_s, W_s, W_t) \sigma X_s + \varphi_y(X_s, W_s, W_t) + \varphi_z(X_s, W_s, W_t)) \right] \right) \\ - \mathbb{E} \left(f'(X_t)g(X_s) \varphi(X_s, W_s, W_t) \frac{X_t}{X_s} \right).$$

We have obtained a term which is “bad” because of the presence of the derivative of f : we are now going to drop it.

2) Use of MIBP formula over $[s, t]$

By using arguments similar to the ones developed above, we can write

$$\mathbb{E} \left(f'(X_t)g(X_s) \varphi(X_s, W_s, W_t) \frac{X_t}{X_s} \right) = \mathbb{E} \left(\int_s^t \frac{g(X_s) \varphi(X_s, W_s, W_t)}{\sigma(t-s)X_s} D_r f(X_t) dr \right) \\ = \mathbb{E} \left(f(X_t) \int_s^t \frac{g(X_s) \varphi(X_s, W_s, W_t)}{\sigma(t-s)X_s} dW_r \right) = \mathbb{E} \left(f(X_t) \frac{g(X_s)}{\sigma(t-s)X_s} \int_s^t \varphi(X_s, W_s, W_t) dW_r \right)$$

(the term $g(X_s)/(\sigma(t-s)X_s)$ has been put outside the integral because it is \mathcal{F}_s -measurable). Again using the property $\int_0^s G dW_r = GW_s - \int_0^s D_r G dr$, one obtains

$$\mathbb{E} \left(f'(X_t)g(X_s) \varphi(X_s, W_s, W_t) \frac{X_t}{X_s} \right) \\ = \mathbb{E} \left(f(X_t)g(X_s) \left[\varphi(X_s, W_s, W_t) \frac{W_t - W_s}{\sigma(t-s)X_s} - \frac{\varphi_z(X_s, W_s, W_t)}{\sigma X_s} \right] \right).$$

By inserting this term, in conclusion we obtain

$$\mathbb{E}(g'(X_s) f(X_t) \varphi(X_s, W_s, W_t)) \\ = \mathbb{E} \left(f(X_t)g(X_s) \left[\varphi(X_s, W_s, W_t) \frac{\Delta W_{s,t}}{\sigma s(t-s)X_s} - \varphi_x(X_s, W_s, W_t) - \frac{\varphi_y(X_s, W_s, W_t)}{\sigma X_s} \right] \right),$$

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

Finally, equality in *i*) holds by taking $\varphi \equiv 1$ and *ii*) follows by considering $\varphi(X_s, W_s, W_t) = \Delta W_{s,t}$.

□

As a consequence, we can state the following result, which comes out to be useful in the proof of the results in Section 4:

Corollary 6.2 *Let $X_t = x e^{\mu t + \sigma W_t}$ be as in Lemma 6.1. Then for any $f : \mathbb{R} \rightarrow \mathbb{R}$ with polynomial growth and $0 < s < t$, one has*

$$a) \mathbb{E} \left(f(X_t) \frac{\Delta W_{s,t}}{\sigma s(t-s)X_s} \right) = 0;$$

$$b) \mathbb{E}\left(f(X_t) \left[\frac{\Delta W_{s,t}}{\sigma s(t-s)X_s}\right]^2\right) = \mathbb{E}\left(f(X_t) \frac{t + \sigma^2 s(t-s)}{\sigma^2 s(t-s)(X_s)^2}\right).$$

Proof. Statement *a*) immediately follows by taking $g = 1$ and $\varphi = 1$ in Lemma 6.1. Concerning *b*), first taking $g = 1$ and $\varphi = 1$ in Lemma 6.1, one has

$$\mathbb{E}\left(f(X_t) \left[\frac{\Delta W_{s,t}}{\sigma s(t-s)X_s}\right]^2\right) = -\mathbb{E}\left(f(X_t) \left[\frac{\Delta W_{s,t}}{\sigma s(t-s)(X_s)^2} - \frac{t}{\sigma^2 s(t-s)(X_s)^2}\right]\right).$$

Now, by taking $\varphi = 1$ and $g(x) = 1/x$ in Lemma 6.1, one has

$$\mathbb{E}\left(f(X_t) \frac{1}{X_s^2}\right) = -\mathbb{E}\left(f(X_t) \frac{\Delta W_{s,t}}{\sigma s(t-s)(X_s)^2}\right).$$

By inserting this term in the equality above, statement *b*) immediately follows. \square

We are now ready to prove the results in Section 3.

Proof of Theorem 3.3. *i*) Let us set $\tilde{\Phi}_t(y) \equiv \tilde{\Phi}(y) = \Phi \circ F_t(y)$, $y \in \mathbb{R}_+^d$, being F_t defined in (5). Since $X_t = F_t(\tilde{X}_t)$ for any t , one obviously has

$$\mathbb{E}\left(\Phi(X_t) \mid X_s = \alpha\right) = \mathbb{E}\left(\tilde{\Phi}(\tilde{X}_t) \mid \tilde{X}_s = G_s(\alpha)\right),$$

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

$$\mathbb{E}\left(\tilde{\Phi}(\tilde{X}_t) \mid \tilde{X}_s = \tilde{\alpha}_s\right) = \frac{\tilde{\mathbb{T}}_{s,t}[\tilde{\Phi}](\tilde{\alpha})}{\tilde{\mathbb{T}}_{s,t}[1](\tilde{\alpha})} \quad (26)$$

where

$$\tilde{\mathbb{T}}_{s,t}[f](\tilde{\alpha}) = \mathbb{E}\left(f(\tilde{X}_t) \prod_{i=1}^d \frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right)$$

Now, let us firstly suppose that $\tilde{\Phi}(y) = \tilde{\Phi}_1(y_1) \cdots \tilde{\Phi}_d(y_d)$, that is $\tilde{\Phi}$ 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

$$\mathbb{E}\left(\tilde{\Phi}(\tilde{X}_t) \mid \tilde{X}_s = \tilde{\alpha}_s\right) = \prod_{i=1}^d \mathbb{E}\left(\tilde{\Phi}_i(\tilde{X}_t^i) \mid \tilde{X}_s^i = \tilde{\alpha}_s^i\right).$$

Now, let us consider $\mathbb{E}\left(\tilde{\Phi}_i(\tilde{X}_t^i) \mid \tilde{X}_s^i = \tilde{\alpha}_s^i\right)$, for any fixed $i = 1, \dots, d$. Let h_δ be a C^∞ density function on \mathbb{R} which weakly converges to the Dirac mass in 0 as $\delta \rightarrow 0$. Then one can write

$$\mathbb{E}\left(\tilde{\Phi}_i(\tilde{X}_t^i) \mid \tilde{X}_s^i = \tilde{\alpha}_s^i\right) = \lim_{\delta \rightarrow 0} \frac{\mathbb{E}(\tilde{\Phi}_i(\tilde{X}_t^i) h_\delta(\tilde{X}_s^i - \tilde{\alpha}_s^i))}{\mathbb{E}(h_\delta(\tilde{X}_s^i - \tilde{\alpha}_s^i))}.$$

Setting H_δ the probability distribution function associated with h_δ , we have to handle something like $\mathbb{E}(f(\tilde{X}_t^i) H_\delta'(\tilde{X}_s^i - \tilde{\alpha}_s^i))$. Since the process \tilde{X}^i is of the same type studied in Lemma 6.1, we can apply its part *i*):

$$\mathbb{E}(f(\tilde{X}_t^i) H_\delta'(\tilde{X}_s^i - \tilde{\alpha}_s^i)) = \mathbb{E}\left(f(\tilde{X}_t^i) \frac{H_\delta(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right),$$

where $\Delta W_{s,t}^i = (t-s)(W_s^i + \sigma_{ii}s) - (t-s)(W_t^i - W_s^i)$. By using the Lebesgue dominated convergence theorem, one has

$$\mathbb{E}\left(\tilde{\Phi}_i(\tilde{X}_t^i) \mid \tilde{X}_s^i = \tilde{\alpha}_s^i\right) = \lim_{\delta \rightarrow 0} \frac{\mathbb{E}\left(\tilde{\Phi}_i(\tilde{X}_t^i) \frac{H_\delta(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right)}{\mathbb{E}\left(\frac{H_\delta(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right)} = \frac{\mathbb{E}\left(\tilde{\Phi}_i(\tilde{X}_t^i) \frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right)}{\mathbb{E}\left(\frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right)},$$

where $H(\xi) = \lim_{\delta \rightarrow 0} H_\delta(\xi) = \mathbf{1}_{\xi \geq 0}$. Therefore,

$$\begin{aligned} \mathbb{E}\left(\tilde{\Phi}(\tilde{X}_t) \mid \tilde{X}_s = \tilde{\alpha}_s\right) &= \prod_{i=1}^d \mathbb{E}\left(\tilde{\Phi}_i(\tilde{X}_t^i) \mid \tilde{X}_s^i = \tilde{\alpha}_s^i\right) \\ &= \prod_{i=1}^d \frac{\mathbb{E}\left(\tilde{\Phi}_i(\tilde{X}_t^i) \frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right)}{\mathbb{E}\left(\frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right)} = \frac{\tilde{\mathbb{T}}_{s,t}[\tilde{\Phi}](\tilde{\alpha})}{\tilde{\mathbb{T}}_{s,t}[1](\tilde{\alpha})}, \end{aligned}$$

so that (26) holds when $\tilde{\Phi}(y) = \tilde{\Phi}_1(y_1) \cdots \tilde{\Phi}_d(y_d)$. In the general case, the statement holds by using a density argument: for any $\tilde{\Phi} \in \mathcal{E}_b(\mathbb{R}^d)$ there exists a sequence of functions $\{\tilde{\Phi}^n\}_n \subset \mathcal{E}_b(\mathbb{R}^d)$ such that $\tilde{\Phi}^n(\tilde{X}_t) \rightarrow \tilde{\Phi}(\tilde{X}_t)$ in L^2 and such that each $\tilde{\Phi}^n$ is a linear combination of functions which separate the variables as above. Since representation (26) holds for any $\tilde{\Phi}^n$, it finally holds for $\tilde{\Phi}$ as well, as it immediately follows by passing to the limit.

ii) First, notice that, by (5), $\hat{\sigma}_{kj}\tilde{\alpha}_s^k/\alpha^j = \partial_{\alpha^j} G_s^k(\alpha) = \partial_{\alpha^j} \tilde{\alpha}^k$. Thus, by considering $\mathbb{T}_{s,t}[f](\alpha)$ as a function of $\tilde{\alpha}_s$ (as it is!), that is $\mathbb{T}_{s,t}[f](\alpha) = \tilde{\mathbb{T}}_{s,t}[f](\tilde{\alpha})|_{\tilde{\alpha}=\tilde{\alpha}_s}$, then we only have to show that

$$\mathbb{G}_{s,t;k}[f](\alpha) = \partial_{\tilde{\alpha}^k} \tilde{\mathbb{T}}_{s,t}[f](\tilde{\alpha}) = \partial_{\tilde{\alpha}^k} \mathbb{E}\left(\tilde{f}(\tilde{X}_t) \prod_{i=1}^d \frac{H(\tilde{X}_s^i - \tilde{\alpha}^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right),$$

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

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

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

$$\begin{aligned} \partial_{\tilde{\alpha}^k} \tilde{\mathbb{T}}_{s,t}[f](\tilde{\alpha}) &= \mathbb{E}\left(\partial_{\tilde{\alpha}^k} \mathbb{E}\left(\tilde{f}_{-k}(\tilde{X}_t^k) \frac{H(\tilde{X}_s^k - \tilde{\alpha}^k)}{\sigma_{kk}s(t-s)\tilde{X}_s^k} \Delta W_{s,t}^k\right) \Big|_{\tilde{x}^j = \tilde{X}_t^j, j \neq k}\right) \times \\ &\quad \times \prod_{i=1, i \neq k}^d \frac{H(\tilde{X}_s^i - \tilde{\alpha}^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i. \end{aligned}$$

It remains to give a representation for the derivative inside the expectation.

Let us take h_δ as a C^∞ probability density function weakly convergent, as $\delta \rightarrow 0$, to the Dirac mass in 0. This means that the associated distribution function H_δ weakly converges, as $\delta \rightarrow 0$, to H . Let us set

$$\tilde{\mathbb{T}}_{s,t}^{k,\delta}[f](\tilde{\alpha}) = -\mathbb{E}\left(\tilde{f}_{-k}(\tilde{X}_t^k) \frac{H(\tilde{X}_s^k - \tilde{\alpha}^k)}{\sigma_{kk}s(t-s)\tilde{X}_s^k} \Delta W_{s,t}^k\right).$$

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

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

So, let us work with $\partial_{\tilde{\alpha}^k} \tilde{\mathbb{T}}_{s,t}^{k,\delta}[f](\tilde{\alpha})$: by using the one dimensional results given by *ii*) of Lemma 6.1, one has

$$\partial_{\tilde{\alpha}^k} \tilde{\mathbb{T}}_{s,t}^{k,\delta}[f](\tilde{\alpha}) = \mathbb{E} \left(\tilde{f}_{-k}(\tilde{X}_t^k) \frac{H_\delta(\tilde{X}_s^k - \tilde{\alpha}^k)}{\sigma_{kk}s(t-s)(\tilde{X}_s^k)^2} \left[\frac{(\Delta W_{s,t}^k)^2}{\sigma_{kk}s(t-s)} + \Delta W_{s,t}^k - \frac{t}{\sigma} \right] \right)$$

By passing to the limit and rearranging everything, the formula holds. \square

We can now prove the localized version of the operators giving the conditional expectation and its gradient.

Proof of Lemma 3.8. Set $\tilde{f}(\tilde{X}_t) = f \circ F_t(\tilde{X}_t)$. By conditioning w.r.t. all the coordinates of \tilde{X}_t except for the first one, and by recalling that \tilde{X} has independent components, one has

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E} \left(\mathbb{E} \left(\tilde{f}_{-1}(\tilde{X}_t^1) \frac{H(\tilde{X}_s^1 - \tilde{\alpha}_s^1)}{\sigma_{11}s(t-s)\tilde{X}_s^1} \Delta W_{s,t}^1 \right) \Big|_{\tilde{x}^j = \tilde{X}_t^j, j \neq 1} \times \prod_{i=1, i \neq 1}^d \frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i \right),$$

where we have set $\tilde{f}_{-1}(\tilde{X}_t^k) = \tilde{f}(\tilde{X}_t^1, \tilde{x}^2, \dots, \tilde{x}^d)$. Now, we can write

$$\begin{aligned} \mathbb{E} \left(\tilde{f}_{-1}(\tilde{X}_t^1) \frac{H(\tilde{X}_s^1 - \tilde{\alpha}_s^1)}{\sigma_{11}s(t-s)\tilde{X}_s^1} \Delta W_{s,t}^1 \right) &= \mathbb{E} \left(\tilde{f}_{-1}(\tilde{X}_t^1) \frac{\Psi_1(\tilde{X}_s^1 - \tilde{\alpha}_s^1)}{\sigma_{11}s(t-s)\tilde{X}_s^1} \Delta W_{s,t}^1 \right) \\ &\quad + \mathbb{E} \left(\tilde{f}_{-1}(\tilde{X}_t^1) \frac{(H - \Psi_1)(\tilde{X}_s^1 - \tilde{\alpha}_s^1)}{\sigma_{11}s(t-s)\tilde{X}_s^1} \Delta W_{s,t}^1 \right). \end{aligned}$$

By *i*) of Lemma 6.1,

$$\mathbb{E} \left(\tilde{f}_{-1}(\tilde{X}_t^1) \frac{\Psi_1(\tilde{X}_s^1 - \tilde{\alpha}_s^1)}{\sigma_{11}s(t-s)\tilde{X}_s^1} \Delta W_{s,t}^1 \right) = \mathbb{E}(\tilde{f}_{-1}(\tilde{X}_t^1) \psi_1(\tilde{X}_s^1 - \tilde{\alpha}_s^1)),$$

so that

$$\mathbb{E} \left(\tilde{f}_{-1}(\tilde{X}_t^1) \frac{H(\tilde{X}_s^1 - \tilde{\alpha}_s^1)}{\sigma_{11}s(t-s)\tilde{X}_s^1} \Delta W_{s,t}^1 \right) = \mathbb{E} \left(\tilde{f}_{-1}(\tilde{X}_t^1) \left[\psi_1(\tilde{X}_s^1 - \tilde{\alpha}_s^1) + \frac{(H - \Psi)(\tilde{X}_s^1 - \tilde{\alpha}_s^1)}{\sigma_{11}s(t-s)\tilde{X}_s^1} \Delta W_{s,t}^1 \right] \right)$$

and thus

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E} \left(f(X_t) \left[\psi_1(X_s^1 - \tilde{\alpha}_s^1) + \frac{(H - \Psi_1)(\tilde{X}_s^1 - \tilde{\alpha}_s^1)}{\sigma_{11}s(t-s)\tilde{X}_s^1} \Delta W_{s,t}^1 \right] \prod_{i=1, i \neq 1}^d \frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i \right).$$

Now, by considering the conditioning w.r.t. $\tilde{X}_t^1, \tilde{X}_t^3, \dots, \tilde{X}_t^d$ and by using similar arguments, one obtains

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E} \left(f(X_t) \prod_{j=1}^2 \left[\psi_j(X_s^j - \tilde{\alpha}_s^j) + \frac{(H - \Psi_j)(\tilde{X}_s^j - \tilde{\alpha}_s^j)}{\sigma_{jj}s(t-s)\tilde{X}_s^j} \Delta W_{s,t}^j \right] \prod_{i=1, i \neq 1, 2}^d \frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i \right).$$

By iterating the procedure on the remaining components, one arrives to the end.

Concerning $\mathbb{G}_{s,t;k}$, the statement can be proved in the same way, by using the statement *ii*) of Lemma 6.1. Indeed, by conditioning w.r.t. all the coordinates except for the k^{th} one, one first arrives to

$$\begin{aligned} \mathbb{G}_{s,t;k}[f](\alpha) = & -\mathbb{E}\left(f(X_t) \left[\psi_k(\tilde{X}_s^k - \tilde{\alpha}_s^k) \frac{\Delta W_{s,t}^k}{\sigma_{kk}s(t-s)\tilde{X}_s^k} + \right. \right. \\ & \left. \left. + \frac{H(\tilde{X}_s^k - \tilde{\alpha}_s^k) - \Psi_k(\tilde{X}_s^k - \tilde{\alpha}_s^k)}{\sigma_{kk}s(t-s)(\tilde{X}_s^k)^2} \left(\frac{(\Delta W_{s,t}^k)^2}{\sigma_{kk}s(t-s)} + \Delta W_{s,t}^k - \frac{t}{\sigma_{kk}} \right) \right] \times \right. \\ & \left. \times \prod_{i=1, i \neq k}^d \frac{H(\tilde{X}_s^i - \tilde{\alpha}_s^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i \right]. \end{aligned}$$

Now, by conditioning w.r.t. all the coordinates except for the j^{th} one, with $j \neq k$, and by using now part *i*) of Lemma 6.1, the final formula can be achieved. \square

6.2 Proof of Proposition 4.5

Proof of Proposition 4.5. We give here only a sketch of the proof, since it is quite similar to the proof of Proposition 4.1.

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 4.1)

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

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 $v''(\beta) - \mu^{*2}v(\beta) + \mu^{*2}H(\beta) = 0$. 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 $\tilde{f}_t(y) \equiv \tilde{f}(y) = f \circ F_t(y)$, $y \in \mathbb{R}_+^d$ and F_t as in (5).

First, we compute the derivative of $J_d^{f;1}(\psi)$ in the direction $(\hat{\psi}_1, 0, \dots, 0)$. Setting

$$\tilde{f}_1^2(x^1) = \int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-1} \mathbb{E} \left(\tilde{f}^2(x^1, \tilde{X}_t^2, \dots, \tilde{X}_t^d) \prod_{i=2}^d \left[\psi_i(\tilde{X}_s^i - \tilde{\alpha}^i) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i \right]^2 \right)$$

then

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

Now, since we are here interested in the behavior in the direction $(\hat{\psi}_1, 0, \dots, 0)$, we can employ the one dimensional result, getting immediately the solution: $\psi_1^*(\beta) = \mu_1^* e^{-\mu_1^* |\beta|/2}$, $\beta \in \mathbb{R}$, with

$$\begin{aligned} \mu_1^{*2} &= \frac{\mathbb{E}\left(\tilde{f}_1^2(\tilde{X}_t^1) \Upsilon_{s,t;1}^2\right)}{\mathbb{E}\left(\tilde{f}_1^2(\tilde{X}_t^1) \Theta_{s,t;1}^2\right)} \\ &= \frac{\int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-1} \mathbb{E}\left(f^2(X_t) \Upsilon_{s,t;1}^2 \prod_{i=2}^d \left[\psi_i(\tilde{X}_s^i - \tilde{\alpha}^i) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right]^2\right)}{\int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-1} \mathbb{E}\left(f^2(X_t) \Theta_{s,t;1}^2 \prod_{i=2}^d \left[\psi_i(\tilde{X}_s^i - \tilde{\alpha}^i) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right]^2\right)} \end{aligned}$$

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

$$\begin{aligned} \tilde{f}_2^2(x^2) &= \int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-2} \mathbb{E}\left(\tilde{f}^2(\tilde{X}_t^1, x^2, \dots, \tilde{X}_t^d) \left[\psi_1(\tilde{X}_s^1 - \tilde{\alpha}^1) \Theta_{s,t;1} + (H - \Psi_1)(\tilde{X}_s^1 - \tilde{\alpha}^1) \Upsilon_{s,t;1}\right]^2 \times \right. \\ &\quad \left. \times \prod_{i=3}^d \left[\psi_i(\tilde{X}_s^i - \tilde{\alpha}^i) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right]^2\right) \end{aligned}$$

then

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

where $I_1^{\tilde{f}}(\cdot)$ is the one handled in Proposition 4.1. By using similar arguments, one obtains $\psi_2^*(\beta) = \mu_2^* e^{-\mu_2^* |\beta|/2}$, $\beta \in \mathbb{R}$, with

$$\begin{aligned} \mu_2^{*2} &= \frac{\mathbb{E}\left(\tilde{f}_2^2(\tilde{X}_t^2) \left[\frac{\Delta W_{s,t}^2}{\sigma_{22}s(t-s)\tilde{X}_s^2}\right]^2\right)}{\mathbb{E}\left(\tilde{f}_2^2(\tilde{X}_t^2)\right)} \\ &= \frac{\int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-2} \mathbb{E}\left(f^2(X_t) \Gamma(\tilde{X}_s^1 - \tilde{\alpha}^1)^2 \left[\frac{\Delta W_{s,t}^2}{\sigma_{22}s(t-s)\tilde{X}_s^2}\right]^2 \prod_{i=3}^d \left[\psi_i(\tilde{X}_s^i - \tilde{\alpha}^i) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right]^2\right)}{\int_{\mathbb{R}^{d-1}} d\tilde{\alpha}_{-2} \mathbb{E}\left(f^2(X_t) \Gamma(\tilde{X}_s^1 - \tilde{\alpha}^1)^2 \prod_{i=3}^d \left[\psi_i(\tilde{X}_s^i - \tilde{\alpha}^i) + \frac{(H - \Psi_i)(\tilde{X}_s^i - \tilde{\alpha}^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right]^2\right)} \end{aligned}$$

where

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

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 $\psi_j^*(\xi_j) = \mu_j^* e^{-\mu_j^* |\xi_j|/2}$, $\xi_j \in \mathbb{R}$ and $\mu_j^* = \mu_j^*[f]$, are given by:

$$\begin{aligned} \mu_1^{*2} &= \frac{\int_{\mathbb{R}^{d-1}} d\beta_{-1} \mathbb{E}\left(f^2(X_t) \Upsilon_{s,t;1}^2 \prod_{i=2}^d \left[\psi_i^*(\beta^i) + \frac{(H - \Psi_i^*)(\beta^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right]^2\right)}{\int_{\mathbb{R}^{d-1}} d\beta_{-1} \mathbb{E}\left(f^2(X_t) \Theta_{s,t;1}^2 \prod_{i=2}^d \left[\psi_i^*(\beta^i) + \frac{(H - \Psi_i^*)(\beta^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right]^2\right)} \quad \text{and for } j = 2, \dots, d: \\ \mu_j^{*2} &= \frac{\int_{\mathbb{R}^{d-1}} d\beta_{-j} \mathbb{E}\left(f^2(X_t) \Gamma^*(\beta^1)^2 \left[\frac{\Delta W_{s,t}^j}{\sigma_{jj}s(t-s)\tilde{X}_s^j}\right]^2 \prod_{i=2, i \neq j}^d \left[\psi_i^*(\beta^i) + \frac{(H - \Psi_i^*)(\beta^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right]^2\right)}{\int_{\mathbb{R}^{d-1}} d\beta_{-j} \mathbb{E}\left(f^2(X_t) \Gamma^*(\beta^1)^2 \prod_{i=2, i \neq j}^d \left[\psi_i^*(\beta^i) + \frac{(H - \Psi_i^*)(\beta^i)}{\sigma_{ii}s(t-s)\tilde{X}_s^i} \Delta W_{s,t}^i\right]^2\right)} \end{aligned}$$

where $\Gamma^*(\beta^1) \equiv \Gamma^{\psi^*}(\beta^1) = \psi_1^*(\beta^1)\Theta_{s,t;1} + (H - \Psi_1^*)(\beta^1)\Upsilon_{s,t;1}$. 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 4.1)

$$\int_{\mathbb{R}} \left(\psi_i^*(\beta^i) + \frac{(H - \Psi_i^*)(\beta^i)}{\sigma_{ii}s(t-s)\widetilde{X}_s^i} \Delta W_{s,t}^i \right)^2 d\beta^i = \frac{1}{4\mu_i^*} \left(\mu_i^{*2} + \left[\frac{\Delta W_{s,t}^i}{\sigma_{ii}s(t-s)\widetilde{X}_s^i} \right]^2 \right),$$

$$\int_{\mathbb{R}} \Gamma^*(\beta^1)^2 d\beta^1 = \frac{1}{4\mu_1^*} \left(\mu_1^{*2}\Theta_{s,t;1}^2 + \Upsilon_{s,t;1}^2 \right),$$

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

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

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

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, L. Caramellino, A. Zanette (2003). Pricing and hedging American options y Monte Carlo methods by using a Malliavin calculus approach. *Preprint INRIA, n. 4804, April 2003*.
- [3] 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.
- [4] V. Bally and G. Pagés (2002). Error analysis of a quantization algorithm for obstacle problems. To appear on *Stoch. Proc. and their Appl.*
- [5] V. Bally, G. Pagés and J. Printems (2003). First-orders Schemes in the Numerical Quantization Method. *Mathematical Finance*, **13**,1-16.
- [6] J. Barraquand and D. Martineau (1995). Numerical valuation of high dimensiona multivariate American securities. *Journal of Finance and Quantitative Analysis*, **30**, 383-405.
- [7] A. Bensoussan and J.L. Lions (1984). *Impulse Control and Quasivariational Inequalities*. Gauthiers-Villars.
- [8] 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>

- [9] B. Bouchard and N. Touzi (2002). Discrete time approximation and Monte-Carlo simulation of Backward Stochastic Differential Equations. Preprint.
- [10] 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.
- [11] M. Broadie and P. Glassermann (1997). Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control*, **21**, 1323-1352.
- [12] L. Caramellino and A. Zanette (2004). Monte Carlo methods for pricing and hedging American options in high dimension. Preprint.
- [13] D. Chevance (1997). Thèse de l'Université de Provence & INRIA.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] N. El Karoui, S. Peng and M.C. Quenez (1997). Backward stochastic differential equations in finance. *Mathematical Finance*, **7**, 1-71.
- [18] 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.
- [19] P.L. Lions and H. Regnier (2001). Calcul du prix et des sensibilités d'une option américaine par une méthode de Monte Carlo. Preprint.
- [20] 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.
- [21] D. Nualart (1995). *The Malliavin Calculus and Related Topics*. Springer Verlag, Berlin.
- [22] E. Pardoux and S. Peng (1990). Adapted solution of a backward stochastic differential equation. *Systems and Control Letters*, **14**, 55-61.
- [23] 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.
- [24] 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. Control*, **44**, 1840-1851.
- [25] S. Villeneuve and A. Zanette (2002). Parabolic ADI methods for pricing American options on two stocks. *Mathematics of Operations Research*, **27**, 121-149.