The Lions and Regnier Monte Carlo algorithm for the pricing of American options

Vlad Bally
Lucia Caramellino
Antonino Zanette

February 5, 2003

Premia 5

Contents

1 Introduction

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

here modeled as a diffusion process:

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

where b and σ denote a vector and a matrix field on $\mathbb{R}^d$ and W is a d-dimensional Brownian motion. Let $\Phi(X_s)$ denote the c ash-flow associated with the option. The

N_{mc}	Price	True Price	Δ	True Δ	CPU
500	4.806		-0.378		4.1
1000	4.795	4.918	-0.386	-0.387	16.2
5000	4.804		-0.384		405.2
10000	4.804		-0.384		405.2

Table 1: Standard American Put, 20 exercise periods

price as seen at time t of such an American option is given by

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

where $\mathcal{T}_{t,T}$ stands for the set of all the stopping times taking values on $[t, T]$ and r denotes the spot rate process, which is here supposed to be deterministic.
provided The solution of this optimal stopping problem has been provided by El Karoui and others [6] by using the theory of the Snell envelopes: the optimal stopping time is given by

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

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

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

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

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

By using the Ito's Lemma, one can prove that $P(t, x)$ solves the following partial differential equation: setting $\mathcal{L}$ as the infinitesimal generator of X (acting on the space variable only), that is

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

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

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

whenever $P(t,x) > \Phi(x)$, with the final condition $P(T,x) = \Phi(x)$. The rigorous formulation of this problem is given by means of variational inequalities (see Bensoussan and Lions, [4]).
Thus, the problem of the pricing of an American option is a strongly nonlinear problem, and there is non hope to find closed formulas. In order to numerically compute the price of an American option, one can use either a deterministic or a stochastic algorithm.
Concerning the deterministic methods, they can be based on finite elements methods for variational inequalities. But here, we are interested in a stochastic approach, that is in a Monte Carlo algorithm for the pricing of American options. It turns out from a dynamic programming principle, but let us firstly start by formalizing better the framework.

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

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

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

Theorem 1.1 *For any $\Delta t = T/n \in (0,1)$, then $\bar{P}_{n\Delta t}(\bar{X}_{k\Delta t}) = \Phi(\bar{X}_{n\Delta t})$ and for any $k = n-1, n-2, \ldots, 1, 0$, one has*

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

As a consequence, one can numerically evaluate the delta $\Delta(t,x)$ of an American option, that is the derivative of the price with respect to the initial value of the underlying asset price: $\Delta(t,x) = \partial_x P(t,x)$. Recall that this is important since it gives the sensibility of the price with respect to the initial underlying asset price and also for the hedging of the option. By considering the case $t = 0$, as in Theorem 1.1, then the following approximation $\bar{\Delta}_0(x)$ of $\Delta(0,x)$ can be stated.

Proposition 1.2 *For any $\Delta t = T/n \in (0,1)$, set*

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

where $\bar{P}_{\Delta t}(\alpha)$ is defined as in Theorem 1.1, that is

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

Then, by setting

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

where ∂_α denotes the gradient, one has

$$\bar{\Delta}_0(x) = \mathbb{E}_x\Big(\bar{\Delta}(\bar{X}_{\Delta t})\Big).$$

Such results state that in order to numerically compute the price $P(0,x)$ and its delta $\Delta(0,x)$, it is sufficient to approximate a family of conditional expectations and of their derivatives, thus allowing one to set up Monte Carlo simulations.

Existing Monte Carlo methods applied to this context, consist in the numerical evaluation of the conditional expectations by means of a stratification of the path space for the approximation of the transition density of the process X_t, as the quantization algorithm by Bally, Pagés, Printemp [10], the algorithm by Broadie and Glassermann [2] or the one by Barraquand and Martineau [3]. Another Monte Carlo approach makes use of regression methods to perform the approximation of the conditional expectation, as made by Longstaff and Schwartz [9] or by Tsitsiklis and VanRoy [11]

Also in order to overcome the problem of the discretization of the path space, another method can be used. It has been introduced by Lions and Regnier [8] and uses formulas allowing to represent conditional expectations like $\mathbb{E}(F(X_t)\,|\,X_s=\alpha)$ and its derivative $\partial_\alpha\mathbb{E}(F(X_t)\,|\,X_s=\alpha)$ written in terms of a suitable ratio of non-conditioned expectations, that is

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

being π_s^α and $\pi_s^{1,\alpha}$ suitable weights, which could also depend on suitable localizing functions. Such representations can be proved by using Malliavin calculus techniques. A review of the main results providing the formulas as in (3) can be found in Section 2, in the framework of the Black and Scholes model; the main reference is the paper by Lions and Regnier [8] but a little different approach (mainly for the multidimensional case) is developed here and can be found in Bally, Caramellino and Zanette [1].

Therefore, representation (3) can be used for the practical purpose of the pricing of American options. 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 (1.1) and Proposition 1.2, thus avoiding the problem of the approximation of the transition density and of the discretization of the path space. But this plan gives also another considerable gain, because it provides a Monte Carlo algorithm for the evaluation of $P(0,x)$ and $\Delta(0,x)$ which makes use of only one set of simulated trajectories for the computation of any conditional expectation appearing in Theorem 1.1 and in Proposition 1.2. Let us finally 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 3 a comprehensive presentation of the Lions and Regnier algorithm.

2 Formulas for the conditional expectation and its gradient

In this section we summarize the formulas for the conditional expectation and its first derivatives which will be used in the pricing algorithm. We restrict our attention to the Black and Scholes model.

2.1 The one dimensional case

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

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

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

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

The aim is to study the conditional expectation

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

and its first derivatives, where $0 < s < t$, $\alpha \in \mathbb{R}_+^d$ and F is a function with polynomial growth, that is belonging to

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

where $\mathcal{M}(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} : f \text{ is measurable}\}$. In such a case one can state the following result (Lions and Regnier [8], Lemme 2.1.1 and Lemme 2.1.5):

Theorem 2.1 (Representation formulas I: without localization)
i) For any $0 \le s < t$, $F \in \mathcal{E}_b$ and $\alpha > 0$, one has

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

where

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

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

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

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

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

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

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

Let us point out that, for any fixed function f, the operator $\mathbb{R}_{s,t}[f](\alpha)$ is simply the derivative of $\mathbb{T}_{s,t}[f](\alpha)$: $\mathbb{R}_{s,t}[f](\alpha) = \partial_\alpha \mathbb{T}_{s,t}[f](\alpha)$.
In the proof of Theorem 2.1 it is shown the existence of a process $\pi_{s,t} \in L^2(\Omega)$ such that, roughly speaking,

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

where δ_0 stands for the Dirac measure in 0 (notice that H is actually the distribution function associated with δ_0). Obviously, the Dirac mass is something “very irregular”. In some sense, the Malliavin integration by parts formula allows to regularize it, thus to overcome the problem of handling a Dirac mass, but it is worth to point out that this procedure provides an high variance because of the presence of the Heaviside function H. Thus, it is very useful to give a localization method for the computation of conditional expectations, as made in the following Lemma (similarly to Lions and Regnier [8], Lemme 2.1.1 and Lemme 2.1.6).

Lemma 2.2 *Let $\psi : \mathbb{R} \to [0,+\infty)$ be such that $\int_{\mathbb{R}} \psi(\xi)d\xi = 1$. Then the operators $\mathbb{T}_{s,t}$ and $\mathbb{R}_{s,t}$, defined in (5) and (6) respectively, can be localized as follows:*

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{T}^{\psi}_{s,t}[f](\alpha) \quad \text{and} \quad \mathbb{R}_{s,t}[f](\alpha) = \mathbb{R}^{\psi}_{s,t}[f](\alpha)$$

where

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

and

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

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

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

Theorem 2.3 (Representation formulas II: with localization) *For any $0 \le s < t$, $F \in \mathcal{E}_b$, $\alpha > 0$ and for any $\psi : \mathbb{R} \to [0,+\infty)$, such that $\int_{\mathbb{R}} \psi_i(\xi)d\xi = 1$, one has*

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

and

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

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

We postpone to Section 2.3 a discussion on the choice of the localizing function ψ.

2.2 The multidimensional case

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

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

where:

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

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

$$X_t^i = x_i \exp\Big(h_i t + \sum_{j=1}^{i} \sigma_{ij} W_t^j\Big), \qquad i = 1, \dots, d \tag{9}$$

where from now on we set

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

The aim is to study the conditional expectation

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

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

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

where $\mathcal{M}(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R} : f \text{ is measurable}\}$.

In the case of the geometric Brownian motion, that is whenever X evolves as in (9), it is quite easy to state formulas for the conditional expectation as a direct application of the formulas provided in the one-dimensional case. In few words, to this goal it suffices to consider an auxiliary process $\widetilde{X}$ with independent components for which a formula for the conditional expectation immediately follows as a product. In a second step, such a formula can be adapted to the original process X by means of an (inversible) function giving X from the auxiliary process $\widetilde{X}$. We present here an approach slightly different from the one developed by Lions and Regnier [8], which is detailed in Bally *et al.* [1] (where all the proofs can be found).

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

$$\widetilde{X}_t^i = x_i \exp\Big(h_i\, t + \ell_t^i + \sigma_{ii} W_t^i\Big), \qquad i = 1, \dots, d \tag{10}$$

which obviously satisfies the sde

$$\begin{aligned} d\widetilde{X}_t &= (\hat{r} - \eta + \ell_t')\widetilde{X}_t dt + \bar{\sigma}\widetilde{X}_t dW_t \\ \widetilde{X}_0 &= x \end{aligned} \tag{11}$$

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

$$\widetilde{\sigma}_{ij} = \frac{\sigma_{ij}}{\sigma_{jj}},\ i, j = 1, \dots, d, \quad \text{and} \quad \widehat{\sigma} = \widetilde{\sigma}^{-1} \tag{12}$$

It is worth noticing that $\widehat{\sigma}$ is easy to compute because $\widetilde{\sigma}$ is a triangular matrix. Moreover, $\widehat{\sigma}$ is itself triangular and $\widehat{\sigma}_{ii} = 1$ for any i. Thus,

Lemma 2.4 *The function F_t and its inverse $G_t = F_t^{-1}$ such that $X_t = F_t(\widetilde{X}_t)$ and $\widetilde{X}_t = G_t(X_t)$ are given by*

$$\begin{aligned} F_t^i(y) &= e^{-\sum_{j=1}^i \widetilde{\sigma}_{ij}\ell_t^j}\, y_i \prod_{j=1}^{i-1} \Big(\frac{y_j}{x_j}\, e^{-h_j t}\Big)^{\widetilde{\sigma}_{ij}}, \quad i = 1, \dots, d,\ y \in \mathbb{R}_+^d; \\ G_t^i(z) &= e^{\ell_t^i}\, z_i \prod_{j=1}^{i-1} \Big(\frac{z_j}{x_j}\, e^{-h_j t}\Big)^{\widehat{\sigma}_{ij}}, \qquad\qquad i = 1, \dots, d,\ z \in \mathbb{R}_+^d. \end{aligned}$$

Remark 2.5 *In principle one could take ℓ arbitrarily. In Bally et al. [1], a detailed discussion about possible choices for ℓ shows that standard methods (e.g. minimization of the variance) do not allow to optimize with respect to ℓ. Therefore, for practical purposes the simple choice $\ell(t) = \ell\, t$ seems to be good enough. Concerning the (now) constant ℓ, two main possibilities for ℓ can be suggested:*

- *$\ell = 0$: this simplifies the process $\widetilde{X}$;*
- *$\ell = \ell^*$, with ℓ^* chosen such that $\widetilde{\alpha} = G_s(\alpha) = \alpha$ (where α will stand for the value of X at time s, see next Theorem 2.6), that is*

$$\ell_1^* = 0 \quad \textit{and, as } i = 2, \dots, d, \quad \ell_i^* = \sum_{j=1}^{i-1} \widehat{\sigma}_{ij}\Big(h_j - \frac{1}{s}\ln\frac{\alpha_j}{x_j}\Big). \tag{13}$$

  Such a choice gives a formula for the conditional expectation in point of fact identical to that provided by Lions and Reigner [8].

In the following, we refer to the choice $\ell(t) = \ell\, t$. In particular, in such a case the transformation G_t giving $\widetilde{X}_t = G_t(X_t)$ can be rewritten as

$$G_t^i(z) = e^{\ell^i\, t}\, z_i \prod_{j=1}^{i-1} \Big(\frac{z_j}{x_j}\, e^{-h_j t}\Big)^{\widehat{\sigma}_{ij}}, \qquad i = 1, \dots, d,\ z \in \mathbb{R}_+^d. \tag{14}$$

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

Theorem 2.6 (Representation formulas I: without localization)
i) Let $0 \le s < t$ be fixed. For any function $F \in \mathcal{E}_b(\mathbb{R}^d)$ and $\alpha \in \mathbb{R}^d_+$, one has

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

where

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E}\Big(f(X_t)\prod_{i=1}^d \frac{H(\widetilde{X}^i_s-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big) \tag{15}$$

being $\widetilde{X}_s = G_s(X_s)$ and $\widetilde{\alpha} = G_s(\alpha)$ [G_s being defined in (14)], $H(\xi) = 1_{\xi\ge 0}$, $\xi\in\mathbb{R}$, and

$$\Delta W^i_{s,t} = (t-s)(W^i_s + \sigma_{ii}s) - s(W^i_t - W^i_s), \qquad i=1,\dots,d.$$

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

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

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

$$\begin{aligned}\mathbb{R}_{s,t;k}[f](\alpha) = \; & -\mathbb{E}\Big(f(X_t)\frac{H(\widetilde{X}^k_s-\widetilde{\alpha}^k)}{\sigma_{kk}s(t-s)(\widetilde{X}^k_s)^2}\Big[\frac{(\Delta W^k_{s,t})^2}{\sigma_{kk}s(t-s)} + \Delta W^k_{s,t} - \frac{t}{\sigma_{kk}}\Big]\times \\ & \times \prod_{i=1,i\neq k}^d \frac{H(\widetilde{X}^i_s-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big).\end{aligned} \tag{16}$$

Remark 2.7 *If no correlation is assumed among the assets, that is if the volatility matrix σ is diagonal, then $\widehat{\sigma} = Id_{d\times d}$. Thus, the sum appearing for the evaluation of $\partial_{\alpha_j}\mathbb{E}(F(X_t)\,|\,X_s=\alpha)$ reduces to the single term with $k=j$, with coefficient $\widetilde{\alpha}_j/\alpha_j = e^{\ell^j s}$, which in turn is equal to 1 whenever $\ell = 0$.*

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

Lemma 2.8 *Let $\psi(x) = \prod_{i=1}^d \psi_i(x_i)$, $x=(x_1,\dots,x_d)\in\mathbb{R}^d$, with $\psi_i \ge 0$ and $\int_{\mathbb{R}}\psi_i(\xi)d\xi = 1$. Then the operators $\mathbb{T}_{s,t}$ and $\mathbb{R}_{s,t;j}$, defined in (15) and (16) respectively, can be localized as follows:*

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

where

$$\mathbb{T}^{\psi}_{s,t}[f](\alpha) = \mathbb{E}(f(X_t)\prod_{i=1}^d\Big[\psi_i(X_s-\alpha)) + \frac{H(\widetilde{X}^i_s-\widetilde{\alpha}_i) - \Psi_i(\widetilde{X}^i_s-\widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big) \tag{17}$$

and

$$\mathbb{R}^{\psi}_{s,t;k}[f](\alpha) = -\mathbb{E}\Big(f(X_t)\Big[\,\psi_k(\widetilde{X}^k_s - \widetilde{\alpha}^k)\frac{\Delta W^k_{s,t}}{\sigma_{kk}s(t-s)\widetilde{X}^k_s} + \\ + \frac{H(\widetilde{X}^k_s - \widetilde{\alpha}^k) - \Psi_k(\widetilde{X}^k_s - \widetilde{\alpha}^k)}{\sigma_{kk}s(t-s)(\widetilde{X}^k_s)^2}\Big(\frac{(\Delta W^k_{s,t})^2}{\sigma_{kk}s(t-s)} + \Delta W^k_{s,t} - \frac{t}{\sigma_{kk}}\Big)\Big]\times \\ \times \prod_{i=1,i\neq k}^{d} \frac{H(\widetilde{X}^i_s - \widetilde{\alpha}_i) - \Psi_i(\widetilde{X}^i_s - \widetilde{\alpha}_i)}{\sigma_{ii}s(t-s)\widetilde{X}^i_s}\,\Delta W^i_{s,t}\Big). \tag{18}$$

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

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

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

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

where the operators $\mathbb{T}^{\psi}_{s,t}[f](\alpha)$ and $\mathbb{R}^{\psi}_{s,t;k}[f](\alpha)$ are defined in (17) and (18) respectively.

Remark 2.10 *In principle, one could take different localizing functions for each operator, that is:*

$$\mathbb{E}\Big(F(X_t)\,\Big|\,X_s = \alpha\Big) = \frac{\mathbb{T}^{\psi_1}_{s,t}[F](\alpha)}{\mathbb{T}^{\psi_2}_{s,t}[1](\alpha)}$$

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

Furthermore, what observed in Remark 2.7 holds here as well: when σ is diagonal, the sum giving $\partial_{\alpha_j}\mathbb{E}(F(X_t)\,|\,X_s = \alpha)$ reduces to the single term with $k = j$, with coefficient $\widetilde{\alpha}_j/\alpha_j = e^{\ell^j s}$, which in turn is equal to 1 if $\ell = 0$.

2.3 A short discussion on the localizing functions

Let us conclude this "theoretical" section with a brief analysis on the choice of the localizing functions.

Let us first discuss the one dimensional case. By referring to Theorem 2.3, in order to compute $\mathbb{E}(F(X_t)\,|\,X_s = \alpha)$ one has to evaluate

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

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

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

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

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

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

Proposition 2.11 *Setting* $\mathcal{L}_1 = \{\psi : \mathbb{R} \to [0,+\infty)\,;\, \psi \in C^1(\mathbb{R}),\ \psi(+\infty) = 0$ *and* $\int_{\mathbb{R}} \psi(t)\,dt = 1\}$, *then*

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

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

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

Remark 2.12 *The optimal value of the parameter* λ *corresponding to* $f = 1$ *can be explicitly written (see Bally et al. [1] - recall that* x *denotes the starting underlying asset price):*

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

The above optimization criterium has been introduced by Kohatsu-Higa and Petterson [7]. In principle, one could consider a measure more general than the Lebesgue one, namely to replace $d\alpha$ with $\rho(d\alpha)$ in the expression for $I_1^f(\psi)$, but in such a case it is not possible to write down explicitly the optimal localizing function.
Such an approach can be generalized to the multidimensional case [5], where the functional I_1^f has to be obviously replaced by

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

Then the following result holds:

Proposition 2.13 *Setting* $\mathcal{L}_d = \{\psi : \mathbb{R}^d \to [0,+\infty)\,;\, \psi(x) = \prod_{i=1}^d \psi_i(x_i)$, *where* $\psi_i \in \mathcal{L}_1$, *for any* $i\}$, *then*

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

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

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

Remark 2.14 *If* $\ell = 0$ *(see Remark 2.5 for details), for* $f = 1$ *the corresponding optimal values of the parameters* λ_j *are given by (recall that* $x_1, \dots, x_d$ *are the starting underlying asset prices)*

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

It is worth to point out that similar arguments could be used in order to handle the problem of minimizing the variance coming out from the expectation giving the operator $\mathbb{R}_{s,t}^{\psi}[f](\alpha)$ (and $\mathbb{R}_{s,t;k}^{\psi}[f](\alpha)$, see (7) and (17)). Anyway, for practical purposes, numerical evidences show that the choice $\lambda^* = 1/\sqrt{s}$ works good enough (thus avoiding to weight the algorithm with the computation of further expectations).

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

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

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

3 The algorithm for the pricing of American options

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

3.1 How to use the formulas in practice

The algorithm is devoted to the numerical evaluation of the price $P(0,x)$ and the delta $\Delta(0,x)$ of an American option with payoff function Φ and maturity T, on underlying assets whose price X evolves following the Black-Scholes model, that is as in (9). It has been briefly described in the Introduction, let us now go into the details.

Let $0 = t_0 < t_1 < \ldots < t_n = T$ be a discretization of the time interval $[0, T]$, with step size equal to $\varepsilon = T/n$. By using Theorem 1.1 and Proposition 1.2, the price $P(0, x)$ is approximated by means of $\bar{P}_0(x)$, where $\bar{P}_{k\varepsilon}(X_{k\varepsilon})$, as $k = 0, 1, \ldots, n$, is iteratively defined as:

$$\begin{aligned} & \bar{P}_{n\varepsilon}(X_{n\varepsilon}) = \Phi(X_{n\varepsilon}) \equiv \Phi(X_T) \\ k = n-1, \ldots, 1, 0: \quad & \bar{P}_{k\varepsilon}(X_{k\varepsilon}) = \max\Big\{\Phi(X_{k\varepsilon})\,,\, e^{-r\varepsilon}\,\mathbb{E}\Big(\bar{P}_{(k+1)\varepsilon}(X_{(k+1)\varepsilon})\,\Big|\,X_{k\varepsilon}\Big)\Big\} \end{aligned} \tag{19}$$

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

$$\begin{aligned} \text{setting} \quad \bar{\Delta}(X_\varepsilon) &= \begin{cases} \partial_\alpha\,\Phi(\alpha)\Big|_{\alpha=X_\varepsilon} & \text{if} \quad \bar{P}_\varepsilon(X_\varepsilon) < \Phi(X_\varepsilon) \\ e^{-r\varepsilon}\,\partial_\alpha\,\mathbb{E}\Big(\bar{P}_{2\varepsilon}(X_{2\varepsilon})\,\Big|\,X_\varepsilon = \alpha\Big)\Big|_{\alpha=X_\varepsilon} & \text{if} \quad \bar{P}_\varepsilon(X_\varepsilon) > \Phi(X_\varepsilon) \end{cases} \\ \text{then} \quad \bar{\Delta}_0(x) &= \mathbb{E}_x\Big(\bar{\Delta}(X_\varepsilon)\Big). \end{aligned} \tag{20}$$

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

Remark 3.1 *In the context of the geometric Brownian motion, the process X can be exactly simulated at each instant $t_k = k\varepsilon$. So, in this particular case we do not need an approximation $\bar{X}_{k\varepsilon}$ of $X_{k\varepsilon}$, and thus we write directly $X_{k\varepsilon}$. Furthermore, it is worth remarking that the algorithm allows to use the same sample in order to simulate all the involved conditional expectations, as it will follows from the next description.*

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

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

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

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

which gives $X_{n\varepsilon}$. Now, in order to simulate $X_{(n-1)\varepsilon}$, we need $W_{(n-1)\varepsilon}$, which in turn can be simulated by using the Brownian bridge: since $W_{n\varepsilon}$ is known, $W_{(n-1)\varepsilon}$ can be simulated by using the conditional law of $W_{(n-1)\varepsilon}$ given the observed value for $W_{n\varepsilon}$. It is well known that the law of W_s given that $W_t = y$ for $0 < s < t$ is given by a gaussian law with mean $s/t\,y$ and variance $s(t-s)/t\,I$. Thus,

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

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

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

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

$$\mathcal{U} = \{U^{i,q}_k\,;\, k = 1, \ldots, n \text{ (time)},\, i = 1, \ldots, d \text{ (dimension)},\, q = 1, \ldots, N \text{ (sample)}\} \tag{21}$$

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

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

$$\begin{aligned} &\text{for } k=n: \quad W^{i,q}_{n\varepsilon}=\sqrt{n\varepsilon}\,U^{i,q}_n, \quad i=1,\dots,d, \text{ and}\\ &\text{for } k=n-1,\dots,1: \quad W^{i,q}_{k\varepsilon}=\frac{k}{k+1}\,W^{i,q}_{(k+1)\varepsilon}+\sqrt{\frac{k}{k+1}\varepsilon}\,U^{i,q}_k, \quad i=1,\dots,d. \end{aligned} \tag{22}$$

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

$$X^{i,q}_{k\varepsilon}=x_i\,e^{(r-\eta_i-\frac{1}{2}\sum_{j=1}^i\sigma_{ij}^2)k\varepsilon+\sum_{j=1}^i\sigma_{ij}W^{i,q}_{k\varepsilon}}, \quad i=1,\dots,d \tag{23}$$

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

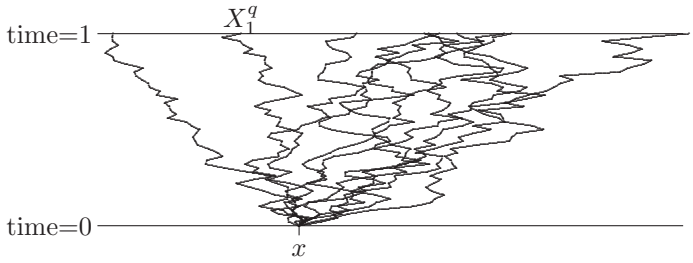

Figure 1 An example of the tree turning out by simulating 10 paths of the process X on $[0,1]$.

3. Computation of $(\widetilde{X}^{i,q}_{k\varepsilon})_{i=1,\dots,d;\,k=0,\dots,n}$, $q=1,\dots,N$, allowing to numerically evaluate the conditional expectations involved in (19). In order to do this, if $d>1$ one needs to introduce the drift ℓ, which could vary according to the time interval of interest $[k\varepsilon,(k+1)\varepsilon]$, that is one has something like ℓ^i_k, for $k=1,\dots,n-1$ (time) and $i=1,\dots,d$ (dimension). As a first stage, the ℓ^i_k's can be chosen arbitrarily. But since one could also choose $\ell_k=(\ell^1_k,\dots,\ell^d_k)$ depending on the position of X at time $k\varepsilon$, the weight ℓ_k could also depend on the q^{th} sample, as suggested in Remark 2.5. So let us consider the set

$$\mathcal{L}=\{\ell^{i,q}_k;\ k=1,\dots,n-1 \text{ (time)}, i=1,\dots,d \text{ (dimension)}, q=1,\dots,N \text{ (sample)}\}. \tag{24}$$

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

$$\widetilde{X}^{i,q}_{k\varepsilon}=x_i\,e^{(r-\eta_i-\frac{1}{2}\sum_{j=1}^i\sigma_{ij}^2)k\varepsilon+\ell^{i,q}_k\,k\varepsilon+\sigma_{ii}\,W^{i,q}_{k\varepsilon}}, \quad i=1,\dots,d, \tag{25}$$

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

Let us point out some remarks:

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

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

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

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

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

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

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

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

$$\Delta W^{i,q}_{k}=\varepsilon W^{i,q}_{k\varepsilon}-k\varepsilon(W^{i,q}_{(k+1)\varepsilon}-W^{i,q}_{k\varepsilon})+k\varepsilon^2\sigma_{ii},\quad 1,\dots,d. \tag{26}$$

Once we have all the previous ingredients, we can proceed to the computation of $\bar{P}_{k\varepsilon}(X_{k\varepsilon})$ given by (19). To this purpose, let us set

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

where

$$X^{q}_{k\varepsilon}=(X^{1,q}_{k\varepsilon},\dots,X^{d,q}_{k\varepsilon})=\text{the } q^{\text{th}}\text{ sample, given by (23).}$$

which has to be considered here as a *datum*. Notice that for each fixed (time) k, we have a random space grid $X^{q}_{k\varepsilon}=(X^{1,q}_{k\varepsilon},\dots,X^{d,q}_{k\varepsilon})\in\mathbb{R}^d$ and we compute $\mathcal{P}_k[F](X^{q}_{k\varepsilon})$ for each point of the grid. By Theorem 2.6, $\mathcal{P}_k[F](X^{q}_{k\varepsilon})$ is given by

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

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

$$\mathcal{P}_k[F](X^{q}_{k\varepsilon})=\frac{\displaystyle\sum_{q'=1}^{N}F(X^{q'}_{(k+1)\varepsilon})\prod_{i=1}^{d}\frac{H(\widetilde{X}^{i,q'}_{k\varepsilon}-\widetilde{X}^{i,q}_{k\varepsilon})}{\widetilde{X}^{i,q'}_{k\varepsilon}}\,\Delta W^{i,q'}_{k}}{\displaystyle\sum_{q'=1}^{N}\prod_{i=1}^{d}\frac{H(\widetilde{X}^{i,q'}_{k\varepsilon}-\widetilde{X}^{i,q}_{k\varepsilon})}{\widetilde{X}^{i,q'}_{k\varepsilon}}\,\Delta W^{i,q'}_{k}} \tag{27}$$

Here, $\widetilde{X}^{i,q'}_{k\varepsilon}$ and $\Delta W^{i,q'}_{k}$ are computed through (25) and (26) respectively. Finally, we can set up the dynamic programming principle:

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

Obviously, u_k gives the Monte Carlo estimate for $\bar{P}_{k\varepsilon}$ in (19) and finally the price $\bar{P}_0$ is approximated by $u_0(X_0^q) \equiv u_0(x) = \max(\Phi(x), \mathcal{P}_0[u_1](X_\varepsilon^q))$, where in practice we set

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

This procedure gives the price. Concerning the delta, everything starts at the final steps, that is when time ε is considered. Indeed, by (20) we can approximate $\bar{\Delta}_0(x) = (\bar{\Delta}_{0;1}(x), \dots, \bar{\Delta}_{0;d}(x))$ through its Monte Carlo estimate $v_0(x) = (v_{0;1}(x), \dots, v_{0;d}(x))$ given by, for $j = 1, \dots, d$,

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

where $v_{1;j}(X_\varepsilon^q) = \partial_{\alpha_j}\Phi(\alpha)|_{\alpha=X_\varepsilon^q}$ if $u_\varepsilon(X_\varepsilon^q) < \Phi(X_\varepsilon^q)$ and $v_{1;j}(X_\varepsilon^q) = e^{-r\varepsilon}\partial_{\alpha_j}\mathbb{E}(u_2(X_{2\varepsilon})\,|\,X_\varepsilon = \alpha)|_{\alpha=X_\varepsilon^q}$ if $u_\varepsilon(X_\varepsilon^q) > \Phi(X_\varepsilon^q)$ (recall that u_2 is the estimate for $\bar{P}_{2\varepsilon}$). The gradient $\partial_\alpha\Phi(\alpha)$ should obviously be considered as given in *input*. Concerning the gradient of the conditional expectation, by using Theorem 2.6 one has to evaluate

$$\begin{aligned}
&\mathcal{H}_j[u_2](X_\varepsilon^q) \equiv \partial_{\alpha_j}\mathbb{E}\Big(u_2(X_{2\varepsilon})\,\Big|\,X_\varepsilon = \alpha\Big)\Big|_{\alpha=X_\varepsilon^q} \\
&= \sum_{m=1}^{j} \widehat{\sigma}_{mj}\frac{\widetilde{X}_\varepsilon^{m,q}}{X_\varepsilon^{j,q}} \times \frac{\mathbb{R}_{s,t;m}[u_2](X_\varepsilon^q)\mathbb{T}_{s,t}[1](X_\varepsilon^q) - \mathbb{T}_{s,t}[u_2](X_\varepsilon^q)\mathbb{R}_{s,t;m}[1](X_\varepsilon^q)}{\mathbb{T}_{s,t}[1](X_\varepsilon^q)^2}
\end{aligned} \tag{28}$$

where $\mathbb{T}_{s,t}$ and $\mathbb{R}_{s,t;m}$ are given by (15) and (16) respectively. Now, since they are weighted expectations of random variables for which we have N samples, they can be practically evaluated by means of the associated empirical mean: by taking into account (15) and (16), we write, for $f = u_2$ or $f = 1$ (more precisely, by replacing $f(X_{2\varepsilon}^{q'}) = u_2(X_{2\varepsilon}^{q'})$ or $f(X_{2\varepsilon}^{q'}) = 1$ in the formulas below),

$$\begin{aligned}
\mathbb{T}_{s,t}[f](X_\varepsilon^q) = \;& \frac{1}{N}\sum_{q'=1}^{N} f(X_{2\varepsilon}^{q'})\prod_{i=1}^{d}\frac{H(\widetilde{X}_\varepsilon^{i,q'} - \widetilde{X}_\varepsilon^{i,q})}{\sigma_{ii}\varepsilon^2\,\widetilde{X}_\varepsilon^{i,q'}}\,\Delta W_1^{i,q'} \\
\mathbb{R}_{s,t;m}[f](X_\varepsilon^q) = \;& -\frac{1}{N}\sum_{q'=1}^{N} f(X_{2\varepsilon}^{q'})\frac{H(\widetilde{X}_\varepsilon^{m,q'} - \widetilde{X}_\varepsilon^{m,q})}{\sigma_{mm}s(t-s)(\widetilde{X}_\varepsilon^{m,q'})^2}\Big[\frac{(\Delta W_1^{m,q'})^2}{\sigma_{mm}s(t-s)} + \Delta W_1^{m,q'} - \frac{t}{\sigma_{mm}}\Big]\times \\
& \times \prod_{i=1,i\neq m}^{d}\frac{H(\widetilde{X}_\varepsilon^{i,q'} - \widetilde{X}_\varepsilon^{i,q})}{\sigma_{ii}s(t-s)\widetilde{X}_\varepsilon^{i,q'}}\,\Delta W_1^{i,q'}
\end{aligned} \tag{29}$$

This concludes the analysis of the pricing/hedging algorithm.

Let us point out that, for the sake of simplicity, in the above description we have taken into account the non localized formulas. In practice, it is much better to use localizing functions in order to reduce the variance, so one should use the formulas coming from Theorem 2.9. Obviously, nothing changes except for the choice of the localizing functions, for which we refer to the discussion in Section 2.3.

Finally, let us observe that one could use a further technique allowing to reduce the variance: the introduction of a control variable. Unfortunately, there is not a standard way to proceed in this direction. For example, one could use as a control variable the price of the associated European

option. The idea is the following. For a fixed initial time t and underlying asset price x, let us set $P^{\text{am}}(t,x)$ 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

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

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

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

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

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

respectively. Notice that the new obstacle has to be evaluated at each time step: in order to set up this program, it should be possible to compute the price/delta of an European option on Φ. This happens for some call or put options, for which prices and deltas are known in closed form. But one could think also to proceed by simulation for their computation, by using the formulas given in Theorem 2.1 and Theorem 2.6.

3.2 Sketch of the pricing/hedging algorithm

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

- Set the diffusion and option parameters: the dimension d, the starting point x, the interest rate r and the dividends η_i, the volatility matrix σ and the auxiliary matrices $\widetilde{\sigma}$ and $\widehat{\sigma}$, the maturity T, the payoff Φ and its first derivatives $\partial_j\Phi$, $j = 1,\dots d$.
- Choose n and set $\varepsilon = T/n$.
- **STEP n**
  - Produce $U_n^q \sim \mathrm{N}(0,I)$, $q = 1,\dots,N$.
  - Using (22), compute $W_{n\varepsilon}^q$, $q = 1,\dots,N$.
  - Using (23), compute $X_{n\varepsilon}^q$, $q = 1,\dots,N$.
  - Initialization: set $u_n(X_{n\varepsilon}^q) = \Phi(X_{n\varepsilon}^q)$.
- **STEP k, for k=n-1,...,1**

– Produce $U_k^q \sim \mathrm{N}(0, I)$, $q = 1, \dots, N$.
– Using (22), compute $W_{k\varepsilon}^q$, $q = 1, \dots, N$.
– Using (23), compute $X_{k\varepsilon}^q$, $q = 1, \dots, N$.
– Do the following[1].
  * Choose ℓ_k^q, $q = 1, \dots, N$. For example:
    choice 1: $\ell_k^q = 0$;
    choice 2: $\ell_k^q = \ell^*(X_{k\varepsilon}^q)$ as in (13) (with $\alpha = X_{k\varepsilon}^q$ and $s = k\varepsilon$):

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

  * Using (25), compute $\widetilde{X}_{k\varepsilon}^q$, $q = 1, \dots, N$.
– Using (26), compute ΔW_k^q, $q = 1, \dots, N$.
– Using (27)[2], compute $\mathcal{P}_k[u_{k+1}](X_{k\varepsilon}^q)$, $q = 1, \dots, N$.
– Compute for any $q = 1, \dots, N$,

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

– If **k=1** then add the following[3].
  * Using (29)[4] and (28), compute $\mathcal{H}_j[u_2](X_\varepsilon^q)$, $j = 1, \dots, d$, $q = 1, \dots, N$.
  * Compute for any $j = 1, \dots, d$ and $q = 1, \dots, N$,

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

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

- **STEP 0**:
  – Compute

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

  and set

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

  which finally approximates the price $P(0, x)$.

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

[2] Or: first choose the localizing (=probability density) functions ψ_i,, $i = 1, \dots, d$, compute the Ψ_i's as the associated probability distribution functions and after use the localized version of (27), coming from part i) of Theorem 2.9 (or Theorem 2.3 if $d = 1$).

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

[4] Or: first choose the localizing (=probability density) functions ψ_i, $i = 1, \dots, d$, compute the Ψ_i's as the associated probability distribution functions and after use the localized version of (29), coming from part ii) of Theorem 2.9 (or Theorem 2.3 if $d = 1$).

– Compute for any $j = 1, \ldots, d$

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

which finally approximates the j^{th} component of the delta vector $\Delta(0, x)$.

References

[1] V. Bally, L. Caramellino and A. Zanette (2003). Pricing American options by Monte Carlo methods using a Malliavin Calculus approach. Technical report, INRIA. 4, 7, 8, 11

[2] M. Broadie and P. Glassermann (1997). Pricing American-style securities using simulation. *Jourbal of Economic Dynamics and Control,* **21**, 1323-1352. 4

[3] J. Barraquand and D. Martineau (1995). Numerical valuation of high dimensiona multivariate American securities. *Journal of Finance and Quantitative Analysis*, **30**, 383-405. 4

[4] A. Bensoussan and J.L. Lions (1984). *Impulse Control and Quasivariational Inequalities.* Gauthiers-Villars. 3

[5] B. Bouchard 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 `www.proba.jussieu.fr/mathdoc/preprints/index.html` 11

[6] N. El Karoui C.Kapoudjan E.Pardoux S.Peng M.C.Quenez reflected solutions of Backward SDE's and related obstacle problems for PDE's. *The Annals of Probability 1997,* **Vol. 25, No 2** , 702-73. 2

[7] A. Kohatsu-Higa and R. Pettersson (2001). Variance reduction methods for simulation of densities on Wiener space. Preprint. 11

[8] P.L. Lions and H. Regnier (2001). Calcul du prix et des sensibilités d'une option américaine par une méthode de Monte Carlo. Preprint. 4, 5, 6, 7, 8

[9] F.A. Longstaff and E.S. Schartz (1998). Valuing American options by simulations: a simple least squares approach. *Working paper, Anderson Graduate School of Management, University of California,* **25**. 4

[10] V. Bally,G. Pags,J. Printems(2003). First-orders Schemes in the Numerical Quantization Method *Math Finance,***Vol 13, 1 , Janvier 2003, pp 1-16** 4

[11] **J.N. Tsitsiklis and B. VanRoy (2000). Regression methods for pricing complex American-style options.** *Working paper, MIT,*1-22. 4