PRICING AMERICAN OPTIONS: A VARIANCE REDUCTION TECHNIQUE FOR THE LONGSTAFF-SCHWARTZ ALGORITHM.

NICOLA MORENI

Abstract. In this paper we propose a variance reduction technique based on the Girsanov theorem and importance sampling for the computation of American option prices *via* the Longstaff-Schwartz algorithm. We prove the almost sure convergence of the modified algorithm as well as a central limit theorem. We give details about numerical results and investigate the existence and location of a minimum of Monte Carlo variance.

1. Introduction

It is well known that pricing American options is not an easy task and that difficulties arise as we increase the dimension of the underlying (stock) asset. Performing numerical techniques have been proposed to approximate the unidimensional problem but unfortunately they cannot be extended to high-dimensional cases and up to now the more efficient approach is still a Monte Carlo one. In litterature we find different algorithms involving Monte Carlo techniques[1] and we focus on the one proposed by Longstaff and Schwartz [LS01], which is based on a least squares approach. The aim of this paper is to show that it is possible to develop for this algorithm an efficient importance sampling variance reduction technique by means of theGirsanov theorem.
The paper is organized as follows: in section 2 we summarize basic ideas of the Longstaff-Schwartz algorithm (namely a least squares regression coupled to a Monte Carlo procedure) and introduce notation, while in section 3 we recall Girsanov theorem and explain how to obtain a family of equivalent pricing problems which allows us to evaluate the same mathematical expectation with different mean squares; moreover we prove as in [A03] the existence in the above family of an *optimal* estimator which minimize variance. In section 4, following Clément, Lamberton and Protter [CLP02] we state the convergence of the modified Longstaff-Schwartz algorithm and finally in section 5 we give detailed numerical results and some hints on how to locate and approximate the optimal estimator.

2. The Longstaff-Schwartz algorithm

We refer to [CLP02] for details about the algorithm and proofs about its rate of convergence; here we just recall the general framework and the main results.
First of all we discretise in time the American option problem considering an option which can be exerced only at $M+1$ fixed times $t_0 \equiv 0, t_1, \dots, t_M \equiv T$ between $t=0$ and maturity $t=T$. Then we introduce a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with

Date: 19th November 2003.
Key words and phrases. American Options, Variance Reduction, Importance sampling.
[1]See for instance [BG97],[BPP02],[LS01],[TVR01].

a discrete time filtration $\mathbb{F} = \left(\mathcal{F}_{t_j}\right)_{j=1,\dots,M}$. We choose an $\mathbb{R}^D$ valued and $\mathbb{F}$-adapted Markov chain $\left(X_{t_j}\right)_{j=1,\dots,M}$ as model for the underlying asset, and a payoff function $f : \{t_0, t_1, \dots, t_M\} \text{x} \mathbb{R}^D \to \mathbb{R}^+ \cup \{0\}$ such that the resulting variables $Z_{t_j} \doteq f(t_j, X_{t_j})|_{j=1,\dots,M}$ are square integrable. The initial asset value X_0 is supposed to be deterministic. Given the set $\mathcal{T}_{0,M}$ of all $\mathbb{F}$-stopping times with values in $\{t_0, t_1, \dots, t_M\}$ we define the price U_0 of the American option of payoff f as

$$U_0 = \sup_{\tau \in \mathcal{T}_{0,M}} \mathbb{E}(Z_\tau). \tag{1}$$

We can compute (1) by solving the following dynamic programming problem

$$\begin{cases} U_M = Z_{t_M} \\ U_j = \max\{Z_{t_j}, \mathbb{E}(U_{j+1}|\mathcal{F}_{t_j})\}, \quad 0 \le j \le M-1 \end{cases} \tag{2}$$

and it possible to show that the sequence of stopping times

$$\tau_j = \min\{t_k \ge t_j | U_k = Z_k\}$$

obey the dynamic programming scheme

$$\begin{cases} \tau_M = M \\ \tau_j = j\mathbb{I}_{\{Z_{t_j} \ge \mathbb{E}(U_{j+1}|\mathcal{F}_{t_j})\}} + \tau_{j+1}\mathbb{I}_{\{Z_{t_j} < \mathbb{E}(U_{j+1}|\mathcal{F}_{t_j})\}}, \quad 0 \le j \le M-1. \end{cases} \tag{3}$$

Remark that it can be proved (by induction on j and as $\mathcal{F}_{t_j} \subseteq \mathcal{F}_{t_{j+1}}$) that $U_j = \mathbb{E}(Z_{\tau_j}|\mathcal{F}_{t_j})$ and in particular $U_0 = \mathbb{E}(Z_{\tau_0})$ i.e. the sup of eqn. (1) is indeed a max.
Let us analyse the least squares aproach: Markov property allows us to write

$$\mathbb{E}(U_{j+1}|\mathcal{F}_{t_j}) = \phi_j(X_{t_j}) \tag{4}$$

where $\phi_j(\cdot)$ is formally defined as

$$\phi_j(x) \doteq \mathbb{E}(U_{j+1}|X_{t_j} = x).$$

Hence, by definition of conditional expectation, we have that $\phi_j(X_{t_j})$ is the projection of U_{j+1} on the space $L^2_j \doteq \{\psi : \mathbb{R}^D \to \mathbb{R} \mid \mathbb{E}[\psi^2(X_{t_j})] < +\infty\}$.
The approximation proposed by Longstaff and Schwartz consists in substituting infinite dimensional spaces L^2_j with a finite dimensional one. Thus we consider a set of m linearly independent functions $e_1(\cdot), \dots, e_m(\cdot)$ such that $\mathbb{E}[e_k^2(X_{t_j})] < +\infty$ for all $k = 1$ to m, $j = 1$ to M and we try to approximate ϕ_j's in the space spanned by the $e_k(\cdot)$'s i.e.

$$\phi_j(X_{t_j}) \approx \alpha_j \cdot e(X_{t_j}) \tag{5}$$

with

$$\alpha_j \doteq \arg\min_{a \in \mathbb{R}^m} \mathbb{E}\left[U_{j+1} - a \cdot e(X_{t_j})\right]^2.$$

In terms of stopping time we can now introduce a sequence τ_j^m solution of the problem (3) where we have made approximation (5) i.e.

$$\begin{cases} \tau_M^m = M \\ \tau_j^m = j\mathbb{I}_{\{Z_{t_j} \ge \alpha_j \cdot e(X_{t_j})\}} + \tau_{j+1}^m\mathbb{I}_{\{Z_{t_j} < \alpha_j \cdot e(X_{t_j})\}}, \quad 0 \le j \le M-1. \end{cases} \tag{6}$$

The second step consists in solving least squares regression problem (5) via Monte Carlo sampling paths. If we consider N independent realization of the underlying process, then it is possible to approximate

$$\phi_j(X_{t_j}) \approx \alpha_j^N \cdot e(X_{t_j}) \tag{7}$$

where here

$$\begin{aligned}\alpha_j^N &\doteq \arg\min_{a\in\mathbb{R}^m} \frac{1}{N}\sum_{n=1}^{N}\left(U_{j+1}^{(n)} - a\cdot e(X_{t_j}^{(n)})\right)^2 \\ &= \arg\min_{a\in\mathbb{R}^m} \frac{1}{N}\sum_{n=1}^{N}\left(Z_{\tau_{j+1}^{m,N,n}}^{(n)} - a\cdot e(X_{t_j}^{(n)})\right)^2.\end{aligned} \tag{8}$$

The corresponding optimal stopping time dynamic is

$$\begin{cases}\tau_M^{m,N,n} = M \\ \tau_j^{m,N,n} = j\mathbb{I}_{\{Z_{t_j}^{(n)} \geq \alpha_j^N\cdot e(X_{t_j}^{(n)})\}} + \tau_{j+1}^{m,N,n}\mathbb{I}_{\{Z_{t_j}^{(n)} < \alpha_j^N\cdot e(X_{t_j}^{(n)})\}}, & 0\leq j\leq M-1.\end{cases} \tag{9}$$

In conclusion, the true price of the option

$$U_0 = \max(Z_{t_0}, \mathbb{E}Z_{\tau_1})$$

is approximated firstly by

$$U_0^m \doteq \max(Z_{t_0}, \mathbb{E}Z_{\tau_1^m})$$

and then by

$$U_0^{m,N} \doteq \max(Z_{t_0}, \frac{1}{N}\sum_{n=1}^{N} Z_{\tau_1^{m,N,n}}^{(n)}).$$

This was the algorithm proposed by Longstaff and Schwartz ([LS01]) in 2001; later Clément, Lamberton and Protter [CLP02] managed to demonstrate, under some quite general hypothesis on f and e_k's, the following convergence results:

$$\begin{aligned} U_0^{m,N} &\xrightarrow[N\to\infty]{a.s.} U_0^m \quad \forall m\in\mathbb{N} \\ U_0^m &\xrightarrow[m\to\infty]{} U_0.\end{aligned} \tag{10}$$

They proved moreover that for every $j = 1$ to M,

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\left(f(\tau_j^{m,N,n}, X_{\tau_j^{m,N,n}}^{(n)}) - \mathbb{E}f(\tau_j^m, X_{\tau_j^m})\right) \xrightarrow[N\to\infty]{w} G \tag{11}$$

where G is a Gaussian vector and by superscript w we mean *weak convergence*.

Remark 1. Let us spend a few words on the very last result: the presence of such a central limit theorem is at the same time a good and a bad news. It is a good one because even if the $\{Z_{\tau_j^{m,N,n}}^{(n)}\}_{n=1,\ldots,N}$ are not independent variables, a central limit theorem holds and ensures us that convergence occurs with the typical MC behaviour $const/\sqrt{N}$. The bad news is that we cannot find any explicit expression for the constant to be put in the formula, i.e. we do not know the variance of G. We could for instance follow the proof in [CLP02] with the *toy exemple* $M = 2$ to see that even in this very simple case it is in practice impossible to write down an expression for *const*. As we will explain in the following sections, this lack of information will require an appropriate way to quantify the rate of convergence of the algorithm.

3. Variance reduction for the unapproximated problem

The task of reducing MC variance of an estimator of U_0 can be achieved by means of the Girsanov theorem and a standard variance reduction technique: importance sampling. In this section we will describe the necessary tools and explain how to develop such a technique.

3.1. **Useful results.**

Let us just recall in this first subsection two results needed in order to develop our variance reduction technique.

3.1.1. *Importance sampling.* It is well known from probability and Monte Carlo estimation theory [PTVF92] that if we are interested in computing mathematical expectations such as $\mathbb{E}Y$, Y square integrable random variable, we can increase precision by means of an importance sampling technique, i.e. we look for a new random variable $\tilde{Y}$ such that

$$\begin{aligned} \mathbb{E}Y = \mathbb{E}\tilde{Y} \quad \text{and} \\ \operatorname{Var} Y > \operatorname{Var} \tilde{Y} \end{aligned} \tag{12}$$

in order to give a standard MC estimate of $\mathbb{E}Y$ as

$$\mathbb{E}Y \approx \frac{1}{N}\sum_{n=1}^{N} \tilde{Y}^{(n)} \pm 1.96\sqrt{\frac{\operatorname{Var}\tilde{Y}}{N}} \tag{13}$$

where $\left(\tilde{Y}^{(n)}\right)_{n=1,\ldots,N}$ are N independent samples drawn along the law of $\tilde{Y}$. This procedure leads to a reduction of a factor $\operatorname{Var}\tilde{Y}/\operatorname{Var} Y$ in the Monte Carlo variance. $\tilde{Y}$ is usually called *estimator for* Y.

3.1.2. *The Girsanov theorem for Brownian motion.*

Theorem (Girsanov). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathbb{F} = \left(\mathcal{F}_t\right)_{0\leq t\leq T}$ and let W_t be a d-dimensional $\mathbb{F}$-Brownian motion. Consider an adapted d-dimensional process $\left(\theta_t\right)_{0\leq t\leq T}$ such that*

$$\int_0^t ds||\theta_s||^2 < +\infty \text{ a.s.} \,\forall\, 0 \leq t \leq T$$

and that

$$L_t \doteq \exp\{-\int_0^t \theta_s \cdot dW_s - \frac{1}{2}\int_0^t ||\theta_s||^2 ds\}$$

is a martingale $\forall\, 0 \leq t \leq T$ ($||\cdot||$ is the $\mathbb{R}^d$ norm). Then under the probability measure $\mathbb{P}^\theta$ defined by $L_T = d\mathbb{P}^\theta/d\mathbb{P}$, the stochastic process

$$W_t^\theta \doteq W_t + \int_0^t \theta_s ds \tag{14}$$

is a $\mathbb{F}$-brownian motion.

Remark 2. W_t under $\mathbb{P}$ and W_t^θ under $\mathbb{P}^\theta$ are both brownian motion with respect to the **same** filtration $\mathbb{F}$.

3.2. **Estimators for the optimal stopping time problem.** Now let us go back to the initial problem (1) and restrict to the case of a D-dimensional Black and Scholes model for the asset evolution:

$$\begin{cases} \dfrac{dX_t^i}{X_t^i} = \mu^i dt + \sigma dW_t \quad i = 1, \dots, D \\ X_0 = x \in \mathbb{R}^D \end{cases} \tag{15}$$

with σ a Dxd definite positive matrix and W_t a d-dimensional standard brownian motion. This diffusion equation has an explicit solution and one can write X_{t_j} as a function of the Brownian motion W_{t_j} and pose $f(t_j, X_{t_j}) \doteq \tilde{f}(t_j, W_{tj})$.
Now as in [GHS99] let us choose a constant vector $\theta \in \mathbb{R}^d$ and build a *drifted* Brownian motion

$$W_t^\theta \doteq \theta t + W_t. \tag{16}$$

the Girsanov theorem ensures us that there exists a probability measure $\mathbb{P}^\theta$ defined by $d\mathbb{P}^\theta/d\mathbb{P} = L_T^\theta$ where

$$L_t^\theta \doteq \exp\{-\theta \cdot W_t - \frac{1}{2}||\theta||^2 t\} = \exp\{-\theta \cdot W_t^\theta + \frac{1}{2}||\theta||^2 t\} \tag{17}$$

such that W_t^θ is a Brownian motion under $\mathbb{P}^\theta$ and with respect to the same filtration $\mathbb{F}$. Moreover it is easy to see that L_t^θ is a $\mathbb{P}$-martingale.
Once our framework has been set we can state the following

Proposition 3.1. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space enriched with a filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \le t \le T}$ and let W_t be a d-dimensional $\mathbb{F}$-Brownian motion. Then $\forall\, \theta \in \mathbb{R}^d$, $\forall$ $\mathbb{F}$-stopping time $\tau \in \mathcal{T}_{0,M}$ and for all function $\tilde{f} : \{t_0, t_1, \dots, t_M\}x\mathbb{R}^D \to \mathbb{R}^+ \cup \{0\}$ such that $\tilde{f}(t_i, W_{t_i})\big|_{i=1,\dots,M}$ are integrable variables, we have*

$$\mathbb{E}\tilde{f}(\tau, W_\tau) = \mathbb{E}L_\tau^\theta \tilde{f}(\tau, W_\tau + \theta\tau) \tag{18}$$

and

$$\sup_{\tau \in \mathcal{T}_{0,M}} \mathbb{E}\tilde{f}(\tau, W_\tau) = \sup_{\tau \in \mathcal{T}_{0,M}} \mathbb{E}L_\tau^\theta \tilde{f}(\tau, W_\tau + \theta\tau) \tag{19}$$

where L_τ^θ is defined according to (17).

Remark 3. It is worth to stress that as the definition of stopping time [2] involves only the notion of filtration, all change in probability measure leading from $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\Omega, \mathcal{A}, \mathbb{P}^\theta)$ leaves $\mathbb{F}$ unchanged and thus does not affect the set of $\mathbb{F}$-stopping times and in particular $\mathcal{T}_{0,M}$.

Proof. Proof is a direct application of the Girsanov theorem once we introduce the probability measure $\mathbb{P}^\theta$ under which the process W_t^θ defined in (16) is a Brownian motion. We have immediately

$$\begin{aligned} \mathbb{E}\tilde{f}(\tau, W_\tau) &= \mathbb{E}^\theta \tilde{f}(\tau, W_\tau^\theta) = \mathbb{E}L_T^\theta \tilde{f}(\tau, W_\tau + \theta\tau) \\ &= \mathbb{E}L_\tau^\theta \tilde{f}(\tau, W_\tau + \theta\tau), \end{aligned} \tag{20}$$

2

Definition (stopping time). A random variable $\tau : \Omega \to \mathbb{R}^+ \cup \{0\}$ is a stopping time with respect to a filtration $\mathbb{F} = \mathcal{F}_t)_{0 \le t \le T}$ if and only if $\forall t \in [0, T]$, $\{\tau \le t\}$ is $\mathcal{F}_t$-measurable.

last equality holding because L_t^θ is a martingale.
The second result of proposition is trivial as $\mathcal{T}_{0,M}$ is left unchanged by the choice of a new probability measure $\mathbb{P}^\theta$ and as (18) holds for all stopping time in $\mathcal{T}_{0,M}$.

Let us now introduce the underlying asset process X_t^θ solution of the SDE

$$\begin{cases} \dfrac{dX_t^{\theta,i}}{X_t^{\theta,i}} = \mu^i + (\sigma\theta)^i dt + \sigma dW_t \quad i = 1, \ldots, D \\ X_0^\theta = x \in \mathbb{R}^D \end{cases} \tag{21}$$

and consider an integrable payoff function f; corollary allows us to rewrite pricing problem (1) in the form

$$U_0 = \mathbb{E} f(\tau_0, X_{\tau_0}) = \mathbb{E} L_{\tau_0}^\theta f(\tau_0, X_{\tau_0}^\theta) \doteq U_0^\theta. \tag{22}$$

Remark 4. Note that equation (22) implies that the optimal stopping time τ_0 **does not depend on** θ.

Usual interpretation of eqs. (21)-(22) is that a change in the drift can be *compensated* by a change in the payoff function to give a set of random variables which have the same expectation. Thus, the importance of Proposition 3.1 is in the fact that we now dispose of the family $\{U_0^\theta\}_{\theta \in \mathbb{R}^d}$ of estimators of U_0 which can be used to compute $\mathbb{E} f(\tau_0, X_{\tau_0})$. Having such a family is useful because, as it is often the case, we do have a criterion of *optimality* for estimators and we wish to find the optimal one (whenever it exists).

3.3. **Optimal estimator.** Whenever standard Monte Carlo theory is valid, it is straightworfard to define an optimal estimator. Let us suppose we have to evaluate mathematical expectation $\mathbb{E}Y$ of a square integrable random variable Y and that we dispose of a family of estimators Y_η whose laws **are known**. Importance sampling technique of section 3.1.1 applies then to each estimator Y_η and according to eq.(13) the precision of this procedure depends on the value of $\mathrm{Var}Y_\eta$. Consequently a *natural* definition of optimality is

the optimal (MC) estimator is the one that minimize $\mathit{Var}Y_\eta$.

In the case of the Longstaff-Schwartz algorithm, as $U_0 = \max\{Z_0, \mathbb{E}(Z_{\tau_1})\}$, algorithm convergence is driven by the convergence rate of $1/N \sum_{n=1}^N Z^{(n)}_{\tau_1^{m,N,n}}$ to $\mathbb{E}(Z_{\tau_1})$. We can thus exploit proposition 3.1 obtaining a set of estimators for $\mathbb{E}(Z_{\tau_1})$ and defining as optimal the one which minimise approximation error. Unfortunately, we know that the $Z^{(n)}_{\tau_1^{m,N,n}}$ are not independent and that convergence of $1/N \sum_{n=1}^N Z^{(n)}_{\tau_1^{m,N,n}}$ is not ruled by $\mathrm{Var}(Z_{\tau_1})$ but by an unknown constant. As a consequence we cannot prove by explicit calculus the existence and the unicity of such an optimal estimator for the L-S algorithm.
The main idea of this section is that we can however try to demonstrate the existence of an estimator which is optimal for the unapproximated stopping time problem assuming that the behaviour of the L-S algorithm (which recovers the original one in the limits $m \to \infty$, $N \to \infty$) in terms of existence/unicity will be essentially the same.

Thus, we apply importance sampling (12) to the American option pricing problem (1), using the set of random variables $\{L^\theta_{\tau_0} f(\tau_0, X^\theta_{\tau_0})\}_{\theta\in\mathbb{R}^d}$ as estimators for $f(\tau_0, X_{\tau_0})$ and considering as **optimal** the one which **minimizes** $\mathrm{Var}[L^\theta_{\tau_0} f(\tau_0, X^\theta_{\tau_0})]$.

Before going on with the search for this optimal estimator, we must be sure that expressions like $\mathrm{Var}[L^\theta_i f(t_i, X^\theta_i)]$ make sense and so we require that the $L^\theta_i f(t_i, X^\theta_i)$'s are square integrable variables. As $L^\theta_t \in L^p\ \forall p > 0$, $t \in [0, T]$, $\theta \in \mathbb{R}^d$ to have the required property it is sufficient to make the following hypothesis

(H1) There exists a $\xi > 0$ such that the chosen payoff function f satisfies

$$f(t_i, X_{t_i}) \in L^{2(1+\xi)} \quad \forall i = 1, \dots, M.$$

We will consider hypothesys (H1) as verified throughout the rest of the paper.

At this point we focus on two questions:

i) Has the optimization problem for the unapproximated model got a solution?
ii) Whenever an optimal estimator exists, how could we find or approximate it?

In this section we will answer the first question while the location of the minimum will be investigated in section 5.

One way to proof the existence of a θ minimizing the variance

$$\mathrm{Var}[L^\theta_{\tau_0} f(\tau_0, X^\theta_{\tau_0})] = \mathbb{E}\left(L^\theta_{\tau_0}\tilde{f}(\tau_0, W_{\tau_0} + \theta\tau_0)\right)^2 - U_0^2 \doteq \mathcal{V}(\theta) - U_0^2 \tag{23}$$

consists[3] in showing that function $\mathcal{V}(\theta)$ is strictly convex and tends to infinity as $||\theta|| \to \infty$. These properties are exactly what we state in the following

Proposition 3.2. *Assume that (H1) is verified; assume moreover that* $\forall i = 1, \dots, M$, $\mathbb{P}(f(t_i, X_{t_i}) > 0) \neq 0$. *Then we have that*

$$\begin{cases} \mathcal{V}(\theta) \text{ is twice differentiable} \\ \mathrm{Hess}\mathcal{V}(\theta) \text{ is a positive matrix} \\ \lim_{||\theta||\to+\infty} \mathcal{V}(\theta) = +\infty. \end{cases}$$

Proof. The proof of Proposition 3.2 is achieved by adapting with simple algebra a similar result proved in [A03] for a European option.
First of all we exploit the Girsanov theorem obtaining

$$\begin{aligned} \mathcal{V}(\theta) = \mathbb{E}\left(L^\theta_{\tau_0}\tilde{f}(\tau_0, W_{\tau_0} + \theta\tau_0)\right)^2 &= \mathbb{E}^\theta L^\theta_{\tau_0}\tilde{f}^2(\tau_0, W^\theta_{\tau_0}) \\ &= \mathbb{E}e^{\frac{1}{2}||\theta||^2\tau_0 - \theta\cdot W_{\tau_0}}\tilde{f}^2(\tau_0, W_{\tau_0}) \end{aligned} \tag{24}$$

and then we use the fact that $\tau_0 : \Omega \to \{t_0, t_1, \dots, T\}$ to write[4]

$$\mathcal{V}(\theta) = \sum_{j=1}^{M} \mathbb{E}e^{\frac{1}{2}||\theta||^2 t_j - \theta\cdot W_{tj}}\tilde{f}^2(t_j W_{tj})\mathbb{I}_{\tau_0 = t_j}. \tag{25}$$

[3] Remark that as U_0 is a constant, the variance is driven by $\mathcal{V}(\theta)$.

[4] In the expression for $\mathcal{V}$ we omitted the term $\mathbb{E}f^2(t_0, X_{t_0})\mathbb{I}_{\tau_0=t_0}$ which does not depend on θ and will not contribute to gradient and Hessian of $\mathcal{V}$.

Brownian motion is a process with independent and stationary increments and if we introduce M $\mathbb{R}^d$-valued i.i.d. random variables $\epsilon_1, \dots, \epsilon_M$ with $\epsilon_i \sim \mathcal{N}(0, \mathbb{I}_{dxd})$ then we can write for all $i = 1, \dots, M$

$$W_{t_i} = W_{t_{i-1}} + \epsilon_i \sqrt{t_i - t_{i-1}} = W_{t_{i-1}} + \epsilon_i \sqrt{\Delta_i}$$

and introduce a function F such that $\tilde{f}(t_i, W_{t_i}) \doteq F(t_i, \epsilon_1, \dots, \epsilon_i)$. Using this notation eq.(25) rewrites

$$\mathcal{V}(\theta) = \sum_{j=1}^{M} \mathbb{E} \mathrm{e}^{\frac{1}{2}||\theta||^2 t_j - \theta \cdot \sum_{i=1}^{j} \sqrt{\Delta_i}\epsilon_i} F^2(t_j, \epsilon_1, \dots, \epsilon_j) \mathbb{I}_{\tau_0 = j} \doteq \sum_{j=1}^{M} \mathbb{E} v_j(\theta). \tag{26}$$

The next step is to show that it is possible to pass first and second derivatives with respect to $\theta = (\theta_1, \dots, \theta_k, \dots, \theta_d)$ into mathematical expectation and get

$$\begin{cases} \partial_{\theta_k} \mathcal{V}(\theta) = \displaystyle\sum_{j=1}^{M} \mathbb{E} \partial_{\theta_k} v_j(\theta) \\ \partial_{\theta_k} \partial_{\theta_{k'}} \mathcal{V}(\theta) = \displaystyle\sum_{j=1}^{M} \mathbb{E} \partial_{\theta_k} \partial_{\theta_{k'}} v_j(\theta). \end{cases} \tag{27}$$

We can easily prove that this is possible because, as τ_0 is independent from θ (just recall eqs.(18)-(22)) we have

$$|\partial_{\theta_k} v_j(\theta)| \le \left| \sum_{i=1}^{j} (\Delta_i \theta_k - \sqrt{\Delta_i} \epsilon_i^{(k)}) \right| \mathrm{e}^{\frac{1}{2}||\theta||^2 t_j - \theta \cdot \sum_{i=1}^{j} \sqrt{\Delta_i}\epsilon_i} F^2(t_j, \epsilon_1, \dots, \epsilon_j) \tag{28}$$

and

$$\begin{aligned} |\partial_{\theta_k} \partial_{\theta_{k'}} v_j(\theta)| \le \left| \delta_{k,k'} + \left(\sum_{i=1}^{j} (\Delta_i \theta_k - \sqrt{\Delta_i} \epsilon_i^{(k)}) \right) \left(\sum_{i=1}^{j} (\Delta_i \theta_{k'} - \sqrt{\Delta_i} \epsilon_i^{(k')}) \right) \right| \cdot \\ \cdot \mathrm{e}^{\frac{1}{2}||\theta||^2 t_j - \theta \cdot \sum_{i=1}^{j} \sqrt{\Delta_i}\epsilon_i} F^2(t_j, \epsilon_1, \dots, \epsilon_j). \end{aligned} \tag{29}$$

Taking expectation in (28) gives

$$\begin{aligned} &\mathbb{E}|\partial_{\theta_k} v_j(\theta)| \le \\ &\le \mathrm{e}^{\frac{1}{2}||\theta||^2 t_j} \sum_{i=1}^{j} \mathcal{C} \left|\left| (|\Delta_i \theta_k| + \sqrt{\Delta_i} |\epsilon_i^{(k)}|) \mathrm{e}^{||\theta|| \sum_{i=1}^{j} \sqrt{\Delta_i} ||\epsilon_i||} \mathrm{e}^{-\frac{1}{2} \sum_{i=1}^{j} ||\epsilon_i||^2 \frac{\xi}{\xi+1}} \right|\right|_{L^{(\xi+1)/\xi}(\mathbb{R}^{jd})} \cdot \\ &\cdot \left|\left| F^2(t_j, \epsilon_1, \dots, \epsilon_j) \mathrm{e}^{-\frac{1}{2(\xi+1)} \sum_{i=1}^{j} ||\epsilon_i||^2} \right|\right|_{L^{(\xi+1)}(\mathbb{R}^{jd})} < \infty. \end{aligned} \tag{30}$$

where $\mathcal{C}$ is a normalization constant coming from gaussian laws of the ϵ_i's and last inequality holds because $F(t_i, \epsilon_1, \dots, \epsilon_i) \in L^{2(1+\xi)}$.
Eq.(30) is valable for all $j = 1, \dots, M$ and thus first of eqs. (27) holds; the second can be proved with an analogous argument.

Directly from the form of the Hessian and given $\mathbb{P}(f(t_i, X_{t_i}) > 0) \neq 0$, we have

$$u^T \mathrm{Hess}\mathcal{V}(\theta)u =$$
$$= \sum_{j=1}^{M} \mathbb{E}\left(||u||^2 + \left(u \cdot \sum_{i=1}^{j} (\Delta_i \theta - \sqrt{\Delta_i}\epsilon_i) \right)^2 \right) F^2 \mathrm{e}^{\frac{1}{2}||\theta||^2 t_j - \theta \cdot \sum_{i=1}^{j} \sqrt{\Delta_i}\epsilon_i} > 0$$

for all $u \in \mathbb{R}^d - 0$ i.e. the convexity of $\mathcal{V}(\theta)$. Moreover, for each of the v_j's reasonings in [A03] apply and one prove that

$$\lim_{||\theta|| \to +\infty} v_j(\theta) = +\infty$$

which implies the analogous property for $\mathcal{V}(\theta)$.

Remark 5. The results of Proposition 3.2 are important because they ensure us that $\mathcal{V}(\theta)$ has a **unique** minimum and so our search for an optimal drift is not an ill-posed problem. What they clearly **do not** tell is whether the finding of such a minimum will be easy and the gain in term of precision will be worthwile.

4. Construction and convergence of a drift-modified L.-S. algorithm

Thanks to results of section 3 we are finally able to define a modified Longstaff-Schwartz algorithm as follows: first of all we generate Monte Carlo paths $X_{t_0}, X^{\theta}_{t_1}, \ldots, X^{\theta}_{t_M}$ according to (21) and then we use them to solve the approximated dynamic programming problems (2), (9) whith payoff functions $f^{\theta}(t_i, X^{\theta}_{t_i}) \doteq L^{\theta}_i f(t_i, X^{\theta}_{t_i})$ instead of $f(t_i, X_{t_i})$.
Under our hypothesis $f^{\theta} \in L^2$ and we just need to set a few technical hypotesis to prove the following theorems:

Theorem 4.1. *If the sequence $\left(e_k(X_{t_j})\right)_{k\geq 1}$ is total in the space L^2_j for $j = 1$ to M then $\forall\, \theta \in \mathbb{R}^d$,*

$$U_0^{\theta,m} \xrightarrow[m\to\infty]{} U_0^{\theta} \equiv U_0.$$

Theorem 4.2. *Let $\mathbb{P}(\alpha_j \cdot e(X_{t_j}) = Z_{t_j}) = 0$ for $j = 1, \ldots, M$. Then for all $\theta \in \mathbb{R}^d$ and for all $m \in \mathbb{N}$*

$$U_0^{\theta,m,N} \xrightarrow[N\to\infty]{a.s.} U_0^{\theta,m}.$$

Theorem 4.3. *Assume the following hypothesis*

(H2) *For all $\theta \in \mathbb{R}^d$ and for $j = 1$ to M, there exists a neighbourhood V^{θ}_j of α^{θ}_j,$\eta^{\theta}_j > 0$ and $k^{\theta}_j > 0$ such that for $a_j \in V^{\theta}_j$ and for $\epsilon \in [0, \eta^{\theta}_j]$,*

$$\mathbb{E}\Big[\big(1 + \sum_{i=1}^{M} |f^{\theta}(t_i, X^{\theta}_{t_i})| + \sum_{i=1}^{M-1} |e_k(X^{\theta}_{t_i})|\big) \cdot \big(1 + \sum_{i=1}^{M-1} |e_k(X^{\theta}_{t_i})|\big) \cdot$$
$$\cdot \mathbb{I}_{\{|f^{\theta}(t_j, X^{\theta}_{t_j}) - a_j \cdot e_k(X^{\theta}_{t_j})| < \epsilon |e_k(X^{\theta}_{t_j})|\}}\Big] \leq \epsilon k^{\theta}_j.$$

(H3) *For $j = 1$ to M and $\forall\, \theta \in \mathbb{R}^d$, $f^{\theta}(t_j, X^{\theta}_{t_j})$ and $e(X^{\theta}_{t_i})$ are in L^p for all $1 \leq p < +\infty$.*

(H4) *For $j = 1, \ldots, M-1$ and $\forall\, \theta \in \mathbb{R}^d$, $\mathbb{E}[f^{\theta}(\tau^m_j, X^{\theta}_{\tau^m_j})]$ and $\mathbb{E}[f^{\theta}(\tau^m_j, X^{\theta}_{\tau^m_j}) \cdot \cdot e(X^{\theta}_{j-1})]$ thought as function of α are of class $\mathcal{C}^1$ in a neighbourhood of α.*

Then

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\left(f^{\theta}(\tau_j^{m,N,n},X^{\theta,(n)}_{\tau_j^{m,N,n}})-\mathbb{E}f^{\theta}(\tau_j^{m},X^{\theta}_{\tau_j^{m}})\right)$$

converges weakly to a Gaussian random variable as N goes to infinity.

Proof. Theorems 4.1, 4.2, 4.3 are only corollaries of theorems 3.1, 3.2, 4.2 of [CLP02]. In particular the first one is based on the fact that

$$\mathbb{E}[f^{\theta}(\tau_j^{m},X^{\theta}_{\tau_j^{m}})|\mathcal{F}_{t_j}]\xrightarrow[m\to\infty]{L^2}\mathbb{E}[f^{\theta}(\tau_j,X^{\theta}_{\tau_j})|\mathcal{F}_{t_j}]\quad j=1,\dots,M\,,$$

the second rests on the almost sure convergence

$$\frac{1}{N}\sum_{n=1}^{N}f^{\theta}(\tau_j^{m,N,n},X^{\theta,(n)}_{\tau_j^{m,N,n}})\xrightarrow[N\to\infty]{a.s.}\mathbb{E}f^{\theta}(\tau_j^{m},X^{\theta}_{\tau_j^{m}})$$

which holds for all excercise time j and the third is indeed the result of a complicated and technical proof.

At this point of discussion a question arises: if we cannot apply standard MC estimation theory to $1/N\sum_{n=1}^{N}Z^{\theta,(n)}_{\tau_1^{m,N,n}}$, then how could we quantify the convergence rate in order to verify whether our drift-modified algorithm increases precision or not?
Let us look carefully at theorem 4.3: it states that if we introduce a random variable

$$Y^{\theta,N}\doteq\frac{1}{N}\sum_{n=1}^{N}Z^{\theta,(n)}_{\tau_1^{m,N,n}}\,,$$

then when $N\gg 1$, the law of $Y^{\theta,N}$ will be (approximatively) gaussian with mean $\mathbb{E}Z^{\theta}_{\tau_1^{m}}$ and variance $A_N(\theta)/N$. As stressed before, $A_N(\theta)$ is **unknown** and cannot be empirically estimated by means of the $Z^{\theta,(n)}_{\tau_1^{m,N,n}}$'s because they are not independent.
One possible way to quantify $A_N(\theta)$ is to sample $N\cdot N'$ independent paths and to group them in N' *clusters* of N paths each. We then apply L-S algorithm to each cluster obtaining N' independent and identically distributed samples $\{Y^{\theta,N}_{n'}\}_{n'=1,\dots,N'}$ to which *standard Monte Carlo theory* applies; hence the empirical variance $\mathrm{Var}^{e}Y^{\theta,N}$ can be considered as a good measure of the rate of convergence of $\frac{1}{N'}\sum_{n'=1}^{N'}Y^{\theta,N}_{(n')}$.
In particular, we have $A_N(\theta)=N\cdot\mathrm{Var}Y^{\theta,N}\approx N\cdot\mathrm{Var}^{e}Y^{\theta,N}$, and thanks to this procedure we can now estimate the speed of convergence of the algorithm. Obviously, this technique is slow because, in order to have good accuracy, both N' and N must be large. However, as we are firstly interested in testing the algorithm in order to see whether $\mathrm{Var}^{e}Y^{\theta,N}$ presents a minimum or not, this does not constitute a problem.

5. Numerical results

In this section we will give details about numerical results and we will discuss applicability of the introduced algorithm, which was tested in the special case of an American put basket option with $t_0=0$, maturity $T=1$ year and an annual interest rate of 10%. Given a D-dimensional asset model, a strike $K > 0$ and a weight array $w \in [0,1]^D$ such that $\sum_{i=1}^D w_i = 1$ the basket put price is

$$Z_{t_j} = f(t_j, X_{t_j}) \doteq \max(K - \sum_{i=1}^{D} w_i X^i_{t_j}, 0). \tag{31}$$

In order to simplify calculus we chose an uncorrelated model such that

$$\begin{cases} d = D \\ w_i = 1/D \quad i = 1, \dots, d \\ \sigma = \tilde{\sigma}\mathbb{I}_{dxd}, \quad \tilde{\sigma} = 0.2 \end{cases} \tag{32}$$

Let us stress that ***a priori*** the choice of correlated or uncorrelated assets will not affect the presence of a minimum, but only its location. Actually, if for a particular i we chose a weight $w_i >> w_j, \quad j \neq i$, then the basket option price would roughly behave as $\max(K - w_i X_{t_i}, 0)$ and we would expect changes of drift on the other $d-1$ components not to be relevant for variance reduction.

This section is divided in 2 subsections: in the first one we present variance reduction results obtained with different choices of basket dimension, strike and spot vector, emphasizing that, because of some unavoidable numerical effects, some restrictions to the applicability of our technique exist. In the second subsection we will give some rules of thumb on how to locate the requested minimum.

All simulations were performed according to the procedure illustrated in section 4 with 80 clusters of 2000 MC calls each.

5.1. **Variance reduction results.** The reason for which we introduced a drifted model was to increase the precision of the algorithm; consequently, in order to quantify the gain we may have obtained, we must choose two benchmarks to compare our results: one for the price itself and one for Monte Carlo variance. The more natural choice, taking into account what said in section 4, is to consider as reference the estimated pair (price, variance) $(U_0^{0,m,N}, \mathrm{Var}^e Y^{0,N})$ of the original model, i.e. when $\theta \equiv 0$.
We can thus start listing numerical results by considering an unidimensional case and plotting in figure 1 the price ratio $U_0^{\theta,m,N}/U_b$ and the *convergence rate* ratio

$$\mathcal{R}(\theta) \doteq \sqrt{\frac{\mathrm{Var}^e Y^{\theta,N}}{\mathrm{Var}^e Y^{0,N}}} \quad . \tag{33}$$

In this first example, in order to show that our data fit well literature, we used as benchmark for the price U_b, price obtained through a Tree method matching third moment[5].

We can thus notice that there is minimum located more or less at $\theta^* \approx -2.4$ which gives a ratio $\mathcal{R}(\theta^*) \approx 0.1419$ and a correct value for price. In addition, there exists a *narrow* neighbourhood of θ^* (approximatively $[-3, -1.9]$) on which variance ratio is kept under the threshold of 0.2 and a larger interval $[-3.5, -0.8]$ on which $\mathcal{R}(\theta) \lesssim 0.4$. This is remarkable because it means that our procedure is *good* for it is reliable and leads to a variance reduction of more than a factor 25

[5]We used a routine of ENPC/INRIA software *Premia*. Information about the pricing software Premia can be found at: `http://cermics.enpc.fr/~premia/`

with respect to the case $\theta = 0$. This implies that, if we were in some way able to *predict* the location of minimum, we could have really performing calculus.
In figures 2 and 3 we show results obtained in the case of a 2-dimensional asset with strike $K = 90$ but when we consider respectively a symmetrical ($X_0^1 = X_0^2 = 100$) and an asymmetrical ($X_0^1 = 105$, $X_0^2 = 70$) spot asset. In the symmetrical case, one would expect to have a minimum with $\theta_1^* \sim \theta_2^*$ (remember we chose symmetrical weights too); this revealed to be exactly the case as we found the estimated minimum at $\theta^* \approx (-1.6, -1.6)$ with a corresponding mean square ratio of 0.1533.
On the other hand, in the asymmetrical case, as shown by the two projections of figure 3, we find $\theta^* \approx (-1.48, -1.08)$, i.e. location is asymmetrical, too.

Finally, we performed simulations with $D > 2$ finding similar results. As an exemple we report in table 1 the ratios $U_0^{\theta,m,N}/U_0^{0,m,N}$ and $\mathcal{R}(\theta)$ obtained for some values of θ in a neighbourhood of the minimum in the case $D = 5$.

5.1.1. *Real convergence of the algorithm: a range for θ.* The reader could have noticed that in all the presented pictures we plotted a (relatively) narrow region around the estimated minimum obtaining nice figures. If we did not plot results on a larger range of variation for θ, it was not for propaganda reasons. Actually, even if from a theoretical point of view, once model's parameters are fixed, each estimator constructed according to section 4 is expected to return the same option

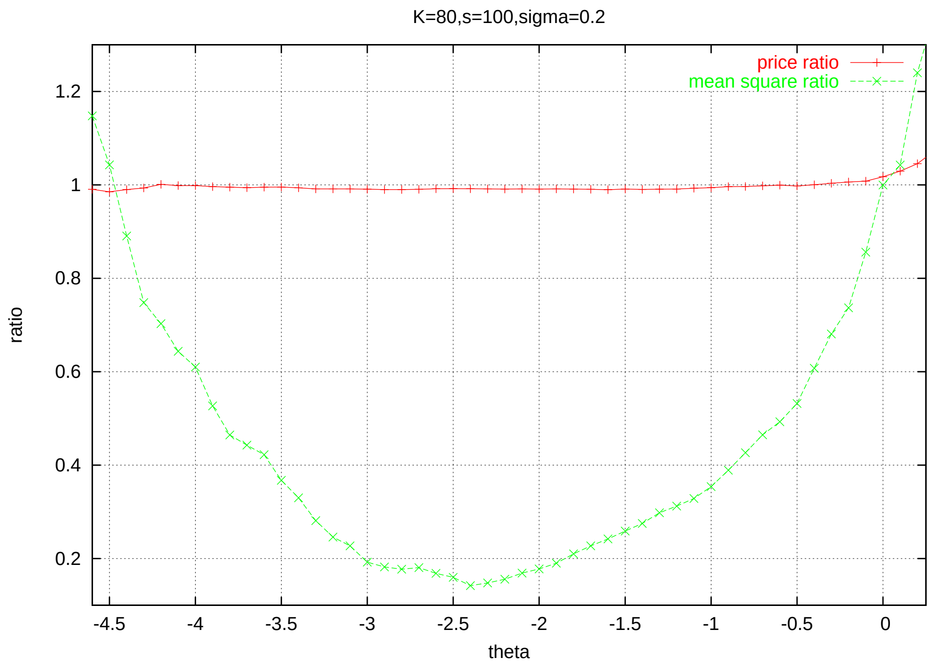


FIGURE 1. $U_0^{\theta,m,N}/U_b$ and $\mathcal{R}(\theta)$ in the unidimensional case. Only region with $U_0^{\theta,m,N} \approx U_b$ and $\mathcal{R}(\theta) \lesssim 1.3$ is plotted.

price, from a numerical point of view things are more delicate. Let us take a look at figure 4-top where we report results obtained in the unidimensional case when $\theta \in [-14, 8]$; we would be induced to state that, except for a region where our datas fit with literature benchmarks[6], prices are sensible to the choice of θ and theory is not verified by practice.

[6]Here again we consider Premia's price as reference.

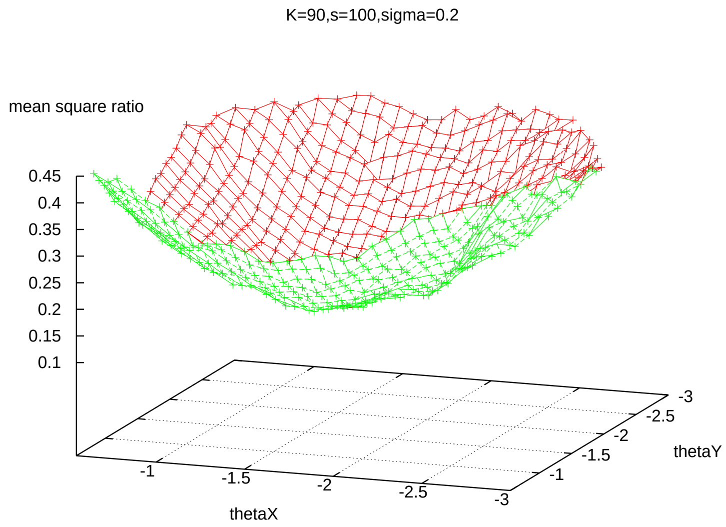
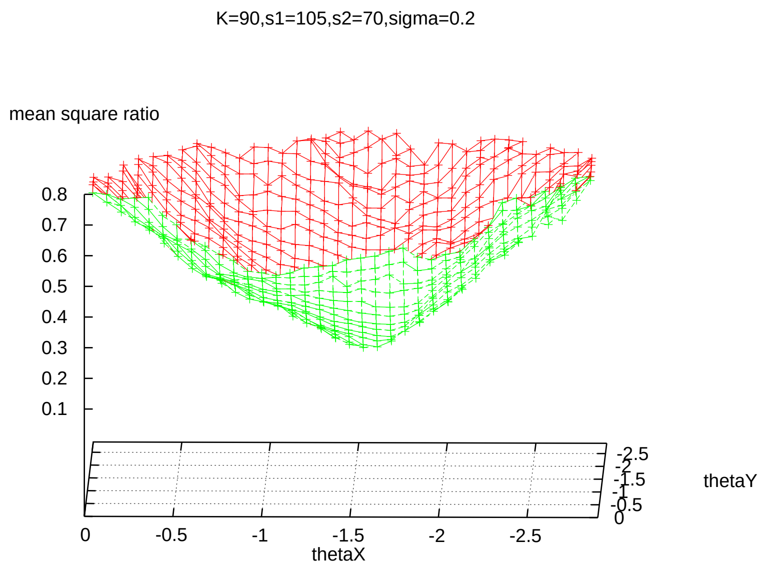


FIGURE 2. $\mathcal{R}(\theta)$ for a symmetrical bidimensional case.

θ_1	θ_2	θ_3	θ_4	θ_5	$U_0^{\theta,m,N}/U_0^{\theta,m,N}$	$\mathcal{R}(\theta)$
-2	-2	-2	-2	-2	0.897768	1.08737
...	...	...	...	...	...	...
-1	-1	-1.5	-0.5	-0.5	1.01046	0.443951
-1	-1	-1	-0.5	-0.5	1.01677	0.345993
-1	-1	-1	-1	-1.5	0.963821	0.266572
-1	-1	-1	-1	-1	1.01038	**0.208331**
-1	-1	-1	-0.5	-1.5	0.956187	0.337243
-1	-1	-1	-0.5	-0.5	1.01677	0.345993
...	...	...	...	...	...	...
-1	-0.5	-2	-0.5	-2	0.920403	0.626063

TABLE 1. D=5, K=100, $X_0 = (100, 100, 100, 100, 100)$; region in the neighbourhood of minimum $\theta^* = (-1, -1, -1, -1, -1)$.



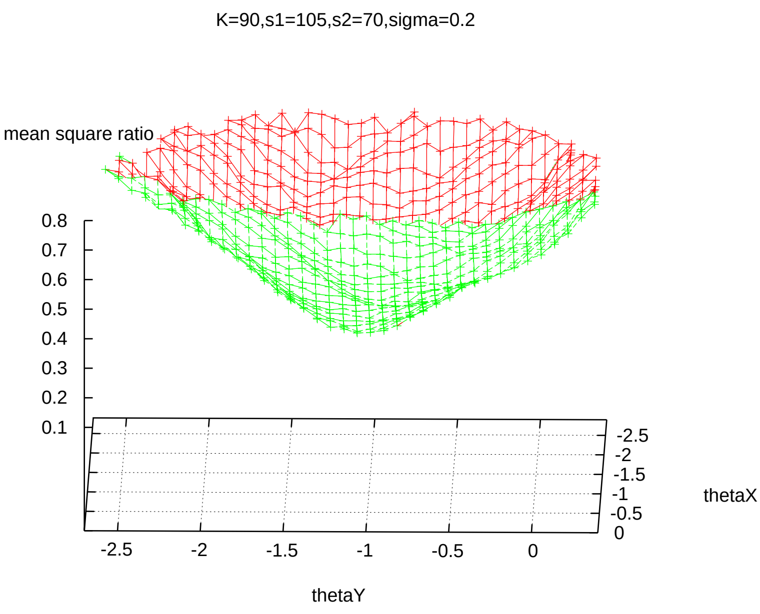


FIGURE 3. $\mathcal{R}(\theta)$ for a bidimensional asymmetrical case. xOz and yOz plane projection.

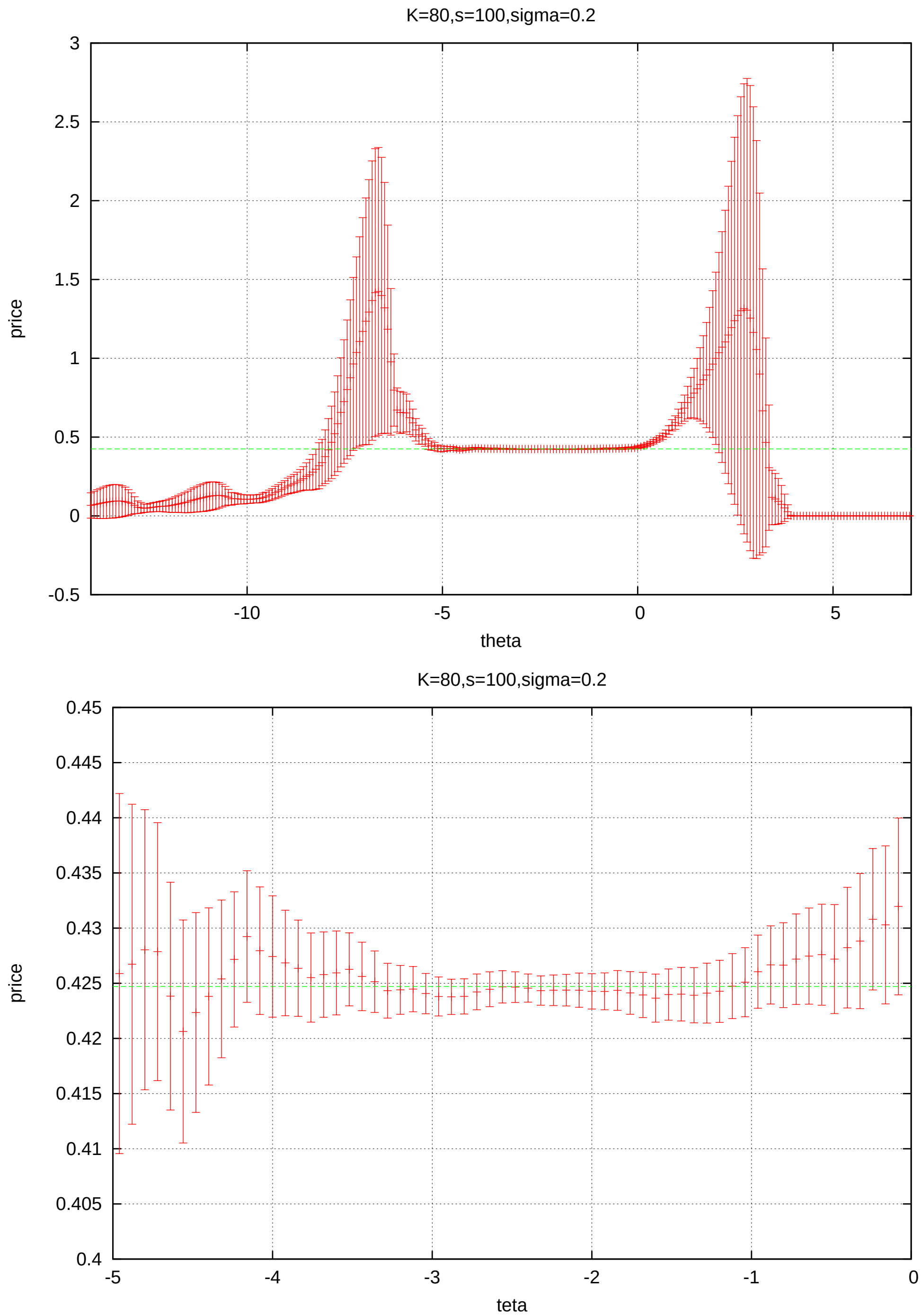


FIGURE 4. 1D model. Prices with errorbars. The dashed orizontal line is the benchmark price. The bottom picture is an enlargement of the top one on the interval [-5,0] .

The point is that nor theory nor our results are wrong. Let us carefully analyse formulae (17) and (31); as payoffs are given by the product of two functions of Brownian motion W_t, namely the Girsanov factor and the put basket payoff, it is possible that one among them presents an *anomalous* behaviour for some values of θ. Let us fix $D = 1$ and let us separate two cases: $\theta \gg 1$ and $\theta \ll -1$. As θ represents the drift, large positive values will force the price to grow while large negative ones will drive it to zero. In our case, we have $X_0 \geq K$ and consequently, when θ is large enough, the simulated trajectories stay completely beyond K, forcing the payoff in case of exercise $f^\theta(t_i, X^\theta_{t_i})$ to be **equal to zero** for all $i = 0$ to M. The result of this artificious behaviour is that, if we performed MC simulation, we would obtain a set of paths on which estimated prices are equal to zero and we would deduce that $U_0^{\theta,m} \sim \mathcal{N}(\approx 0, \approx 0)$. In particular, we would say that our algorithm has infinite precision because $\mathrm{Var}^e Y^{\theta,N} = 0$. Obviously, when $X_0 < K$, we have a similar behaviour: a large value for θ drives the paths beyond the threshold value X_0. Therefore, the optimal stopping time is t_0 and the price of the option is $K - X_0$ on all the paths. Once again, we find that the empirical MC variance is zero.

On the other hand, if θ is large and negative, then on almost each path the expected basket payoff will approximatively be equal to $L_i^\theta K$. What now causes our algorithm to fail is the fact that when $\theta << -1$, then[7] $L^\theta_{t_i} K \approx \exp\{-\theta^2 t_i/2\} K \approx 0$ and again the algorithm returns a wrong price.

Let us go back to figure 4 where bottom picture is only an enlargement of the top one: here it is evident that there is a region in which our prices fit the reference and this tells us our results are only in part affected by the defaults described above. In order to obtain reliable results, we are obliged to consider the minimization problem only on a narrowed range for θ, i.e. under the somewhat unnatural constraint that evaluated prices do not differ *so much* from the real one. The point is: **real price is unknown** and it is exactly what we want to evaluate. As the effects described above arise only when $\theta \neq 0$, one possible rule of thumb to obtain coherent data and investigate location of minimum is the following:

(1) Set $\theta \equiv 0$ and price.
(2) Consider $U_0^{0,m,N}$ as reference price.
(3) Choose two real positive numbers $\epsilon \ll 1$, $\epsilon' \ll 1$, price for different values of θ and discard results such that $U_0^{\theta,m,N} \notin [U_0^{0,m,N}(1-\epsilon), U_0^{0,m,N}(1+\epsilon')]$.

This was exactly the way we proceeded to obtain the results presented at the beginning of section 5.1.

5.2. **Empirical approach to minimization problem.** In the previous subsection we gave numerical results and showed that our procedure is able to appreciably reduce the variance of the Longstaff-Schwartz algorithm. Nevertheless our good outcomes are not enough to state that we achieved our goal. Actually, the above discussion was made **after** choosing an $\mathbb{R}^D$ grid as a set of possible values for θ and performing calculus **for each** grid's value. The point is that, when dimension grows, this way of proceeding becomes unbearable as in *real* life we want to be as fast and precise as possible and therefore we do not have time to waste for the

[7] It is sufficient to consider that, as W is a brownian motion, then with probability $p = 0.95$ the value of W_{t_i} will fall into the interval $[-2t_i\tilde{\sigma}, 2t_i\tilde{\sigma}]$ which, in our case ($T = 1$), is included in $[-0.4, 0.4]$. Consequently, $\theta W \in [-0.4\theta, 0.4\theta]$.

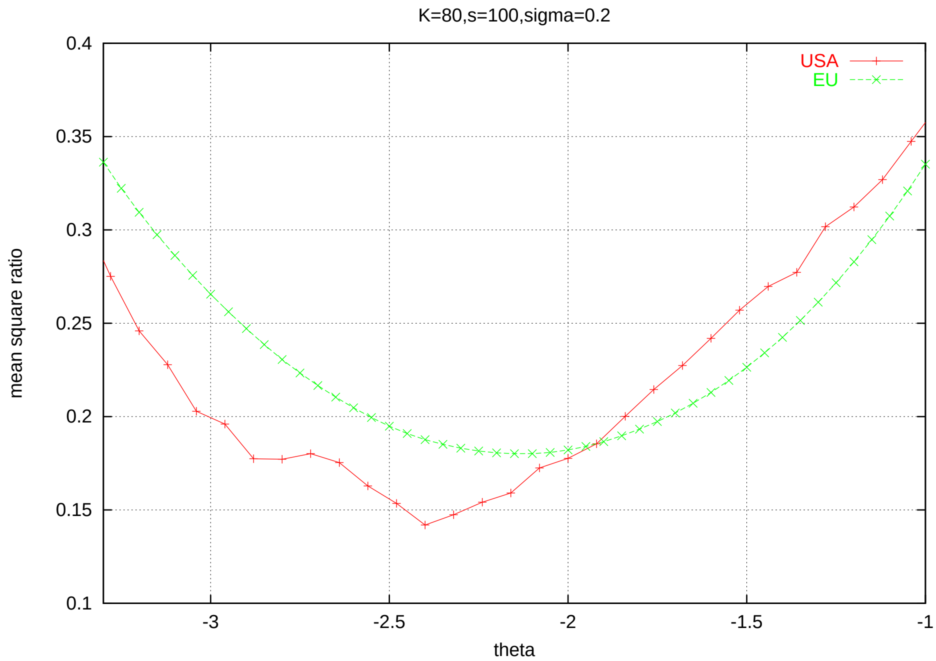


FIGURE 5. Covergence rate ratios for an American and a European basket unidimensional put option.

pricing of the whole grid (note that a five dimensional grid is indeed very *large*!) and the manual search of a minimum.

Thus the challenge is to find a kind of **empirical procedure** which gives us a *sub-optimal* value for θ to be used in pricing.

As dealing with optimal stopping time is complicated, the first thing that comes into mind is to use some **European-derived** criterion which turns out to be *quite reliable* in the sense that it can be considered as sufficently sub-optimal when applied to American case.

We start by analysing the results of the application of our drift-changing procedure to the corresponding European basket option: $M = 1$ and exercise possible only at maturity. As this is only a particular case of the American problem, all results exposed in the paper like the existence of a variance minimum and convergence, still hold.

In figures 5 and 6 we trace European versus American mean square ratios for the unidimensional and bidimensional symmetrical models described in section 5.1 in order to compare goodness of results and location of minima. At a first glance we notice that the European case too has a mean square ratio minimum and we see moreover that it seems to be slightly superior in value than the American one. Concerning its location, we have that for each component $i = 1$ to D, optimal

European $\theta^{*\,EU}$ and optimal American $\theta^{*\,USA}$ verify[8]

$$\theta_i^{*\,USA} \leq \theta_i^{*\,EU};$$

we will come back on this property later.
It is then very useful to estimate from data the location of $\theta^{*\,EU}$ and to compare[9] $\mathcal{R}(\theta^{*\,USA})$ and $\mathcal{R}(\theta^{*\,EU})$. We evaluated these two quantities and we found out that choosing $\theta^{*\,EU}$ to perform American pricing algorithm leads to good sub-optimal results as we obtained a variance reduction of at least a factor 10. Hence, we decided to consider as a quite reliable approximation for the optimal drift the θ which minimize European problem instead of the real minimum $\theta^{*\,USA}$.
Switching from American to European to find a substitute for the minimum presents real advantages because even if we do not dispose of an explicit formula for $\theta^{*\,EU}$, we are able to propose two procedures to approximate it: an empirical and a numerical one.

5.2.1. *An empirical estimation of* $\theta^{*\,EU}$. Let us describe the empirical (and less precise) one first.
From the results of section 3.3 and from the equation (28) we have

$$\partial_{\theta_k}\mathcal{V}(\theta) = \mathbb{E}\partial_{\theta_k} v_1(\theta) = \mathbb{E}\left[(T\theta_k - \sqrt{T}\epsilon_1^{(k)})\mathrm{e}^{\frac{1}{2}||\theta||^2 T - \theta\cdot\epsilon_1\sqrt{T}} F^2(T,\epsilon_1)\right] \tag{34}$$

which, given (31) and (32), has the explicit form

$$\partial_{\theta_k}\mathcal{V}(\theta) = \tag{35}$$
$$= \mathbb{E}\left[(T\theta_k - \sqrt{T}\epsilon_1^{(k)})\mathrm{e}^{\frac{1}{2}||\theta||^2 T - \theta\cdot\epsilon_1\sqrt{T}} \left(K - \sum_{i=1}^{D}\lambda_i \mathrm{e}^{\sigma\sqrt{T}\epsilon_1^{(i)}}\right)^2 \mathbb{I}_{\{K \geq \sum_1^D \lambda_i \mathrm{e}^{\sigma\sqrt{T}\epsilon_1^{(i)}}\}}\right]$$

where by definition $\lambda_i \doteq w_i X_0^i \mathrm{e}^{T(r-\sigma^2/2)}$.
The main idea is to study the behaviour of $\mathcal{V}(\theta)$ when θ varies on a particular direction of $\mathbb{R}^D$. In this case we are able to prove the following

Proposition 5.1. *Set* $\theta = \lambda\sigma\alpha$, $\alpha \in \mathbb{R}$. *Then for all* $\alpha \geq \frac{K - \sum_1^D \lambda_i}{T\sigma^2 \sum_{i=1}^D \lambda_i^2} \doteq \bar{\alpha}$, $\partial_\alpha \mathcal{V}(\alpha) \geq 0$.
Assume moreover $K \leq \sum_{i=1}^D \lambda_i$; *then*

$$\mathcal{V}(\alpha) \leq \mathcal{V}(0) \quad \forall\, \bar{\alpha} \leq \alpha \leq 0$$

Proof. It is very simple to see that when we set $\theta = \alpha c$, c constant vector in $\mathbb{R}^D$, the derivative of $\mathcal{V}$ with respect to α becomes

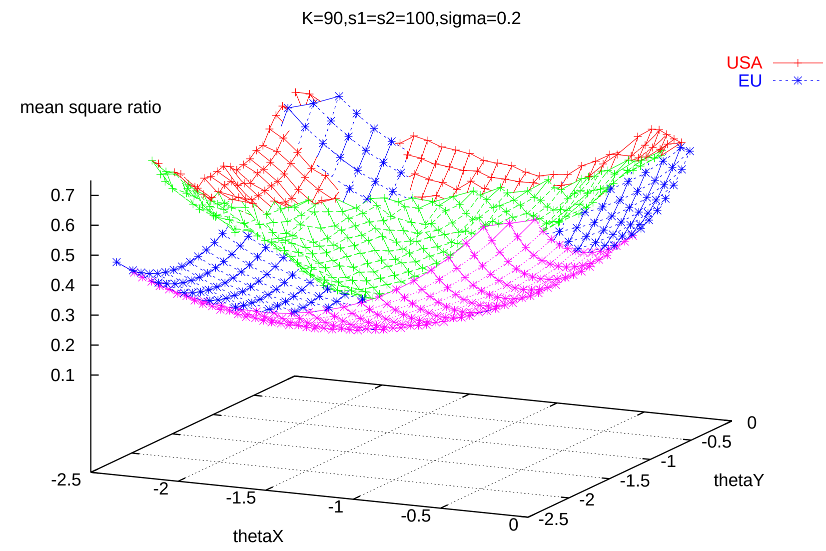
$$\partial_\alpha \mathcal{V}(\theta(\alpha)) = \mathbb{E}\left[g(\epsilon_1,\alpha)\mathbb{I}_{\{K \geq \sum_1^D \lambda_i \mathrm{e}^{\sigma\sqrt{T}\epsilon_1^{(i)}}\}} \sum_{i=1}^{D}(Tc_i^2\alpha - c_i\sqrt{T}\epsilon_1^{(i)})\right] \tag{36}$$

with g non negative function.
On the integration domain we have $K \geq \sum_{i=1}^D \lambda_i \mathrm{e}^{\sqrt{T}\sigma\epsilon_1^{(i)}}$ which implies

$$K - \sum_{i=1}^{D}\lambda_i \geq \sum_{i=1}^{D}\lambda_i\sigma\sqrt{T}\epsilon_1^{(i)}.$$

[8]Empirically we have $0.7 \lesssim |\theta_i^{*\,EU}/\theta_i^{*\,USA}| \lesssim 0.9$ but do not think of it as to a general result because *a priori* the ratio depends on the maturity T.

[9]By definition $\mathcal{R}$ is the mean square ratio of American problem $M > 1$ defined in eq.(33).



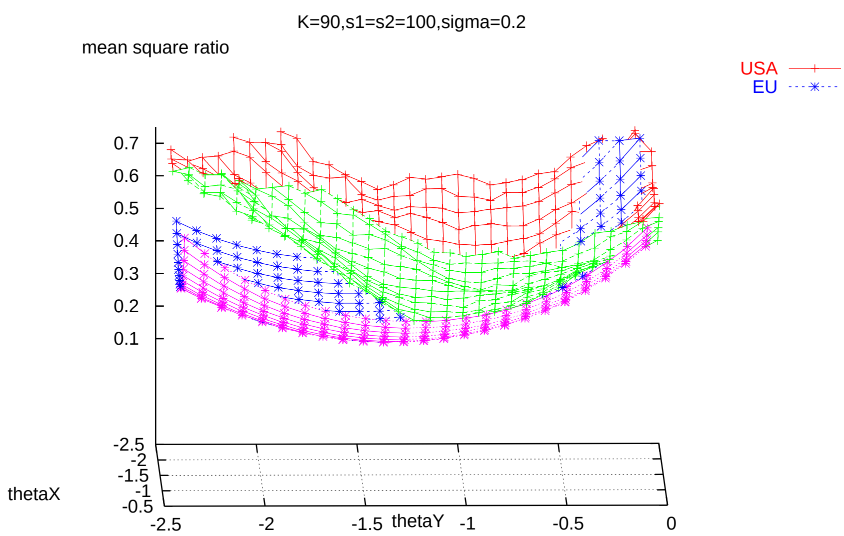


FIGURE 6. Convergence rate ratios for an American and for a European bidimensional symmetrical model.

Consequently, it sufficient to choose $c_i = \lambda_i \sigma$ to find the first result. The second is trivial as, under the additional assumption, $\bar{\alpha} \le 0$.

At this point of discussion we have time for three remarks

Remark 6. Proposition 5.1 reveals to be useful because, when $K \leq \sum_{i=1}^{D} \lambda_i$, as we are sure to reduce the European variance up to $\bar{\alpha}$, we can roughly approximate $\theta^{*\,EU}$ with $\bar{\alpha}\lambda\sigma \doteq \theta^{*\,e}$ and then price with $\theta^{*\,e}$ instead of $\theta^{*\,USA}$.

Remark 7. To prove the second part of proposition 5.1, we needed condition $K \leq \sum_{i=1}^{D} \lambda_i$. Let us stress that $\sum_{i=1}^{D} \lambda_i$ represents the value of the basket at the time T in the case $\epsilon_1 \equiv 0$ and is related to the expected basket return. This means that, whenever the basket tends to a value superior to K (which corresponds to a European price equal to zero), we can reduce variance by driving the price process towards K. On the other side, when $K \geq \sum_{i=1}^{D} \lambda_i$, we have two possible scenarios: $\sum_{i=1}^{D} \lambda_i > \sum_{i=1}^{D} w_i X_0^i$ or $\sum_{i=1}^{D} \lambda_i \leq \sum_{i=1}^{D} w_i X_0^i$. In the first case, which is the more typical one, we expect the optimal stopping time to be t_0 and therefore, price is $\approx K - \sum_{i=1}^{D} w_i X_0^i$. In the second case, our euristic approach cannot be used and we must switch to the numerical approximation of $\theta^{*\,e}$.

Remark 8. When $D = 1$ the integration domain is $\epsilon_1 < (\sqrt{T}\sigma)^{-1} \ln(K/\lambda)$ and, as $\ln(K/\lambda) \leq (K/\lambda) - 1 \leq 0$, we find out that variance is indeed reduced when α goes down to $\alpha^* \doteq (1/T\sigma^2\lambda)\ln(K/\lambda)$ i.e. even beyond $\bar{\alpha}$. This is only a consequence of the fact that domain $K - \sum_{1}^{D} \lambda_i \geq \sum_{1}^{D} \lambda_i \sqrt{T}\sigma\epsilon_1^{(i)}$ is **included** in the integration domain $K > \sum_{i=1}^{D} \lambda_i e^{\sqrt{T}\sigma\epsilon_1^{(i)}}$ for all $D \geq 1$. As there are reasons why when $D > 1$ things should be different, we expect to have always (when $K \leq \sum_{i=1}^{D} \lambda_i$, of course) the existence of an α^* such that $\theta^{*\,EU} \lesssim \lambda\sigma\alpha^* \leq \bar{\alpha}\lambda\sigma$.
Finally, let us spend a word about the signification of $\bar{\alpha}$ and α^*. Obviously, as we want to set $\theta \propto \alpha$, α plays a role in the drift-change of the modified model. In dimension 1 it easy to see that when $\theta = \lambda\sigma\alpha^*$, then

$$\sum_{i=1}^{D\equiv 1} w_i X_T^{\lambda\sigma\alpha^*,i}|_{\epsilon_i \equiv 0} = X_T^{\lambda\sigma\alpha^*,1}|_{\epsilon_1 \equiv 0} \equiv K$$

that is variance is reduced at least **as far as θ drifts the basket price on K.** As $\alpha^* \leq \bar{\alpha}$, we believe that even when $D > 1$, $\theta = \lambda\sigma\bar{\alpha}$ leads the spot price with $\epsilon_i \equiv 0$ towards a value strictly superior to K and that driving the basket to K on a direction which keeps memory of λ and σ is still a good way of proceeding. Let us notice that, if we had tried very simply to choose θ so that *each component* X_T^i had been drifted *separately* to K, then we would have found very bad variance results. This strenghten the idea that our empirical proposal, even if simple, captures the essential behaviour of the model.

To conclude this subsection, let us just stress that we have different possibilities if we want to increase precision of this rough estimate of $\theta^{*\,EU}$: we can for instance keep θ in the form $\sigma\lambda\alpha$ and try to find a better approximation of integration domain or alternatively we can choose another parametrization for theta, simple enough to allow us to perform explicit calculus.

5.2.2. *A numerical estimation of $\theta^{*\,EU}$.* The proposal is simple: as we find in literature some very fast algorithms (see [A03] for a Robbins-Monro-based example) which allows us to estimate with a good precision $\theta^{*\,EU}$, we just have use one of them, obtain a $\theta^{*\,RM} \approx \theta^{*\,EU}$ and then do price the American option with it.

Model	$\mathcal{R}(\theta^{*\,USA})$	$\mathcal{R}(\theta^{*\,RM})$	$\mathcal{R}(\theta^{*\,e})$
$1D$, $K = 80$, $X_0 = 100$	0.14	0.16	0.30
$2D$, $K = 90$, $X_0 = (100, 100)$	0.15	0.17	0.35
$2D$, $K = 90$, $X_0 = (105, 70)$	0.23	0.37	0.74
$3D$, $K = 95$, $X_0 = (100, 100, 100)$	0.200	0.31	0.49
$5D$, $K = 100$,$X_0 = (100, 100, 100, 100, 100)$	0.21	0.31	0.67

TABLE 2. Comparison between American mean square ratios evaluated at the American minimum, at the Robbins Monro minimum, at the euristic $\theta^{*\,e}$.

In table 2 we reported values of $\mathcal{R}(\theta)$ at the estimated $\theta^{*\,USA}$, at the Robbins-Monro $\theta^{*\,RM}$ and at the empirical $\theta^{*\,e} \doteq \lambda\sigma\bar{\alpha}$. We didn't reported results when $\theta = \theta^{*\,EU}$ because $\theta^{*\,EU}$ essentially coincides with $\theta^{*\,RM}$. The outcome is that, exactly as we expected, when $\theta = \theta^{*\,RM}$, variance reduction is still remarkable, while in the case $\theta = \theta^{*\,e}$ the reduction vary between a factor 10 and a factor 2. If we want to have really good results, it is then worthwile to use a reliable (numerical) estimation of $\theta^{*\,EU}$, while the empirical procedure will be precise enough if we are satisfied by a variance reduction of one half .

6. Conclusions and perspectives

In this paper we showed how, thanks to a simple application of Girsanov theorem [GHS99] and importance sampling, it is possible to develop a variance reduction technique which works for American option pricing, thus extending some similar results already proved in the European case [A03]. The construction of a suitable set of estimators and the existence among them of an optimal one minimising the variance was demonstrated under quite general hypothesis on the payoff function. We then restricted ourselves to the Longstaff-Schwartz pricing algorithm of which we introduced a modified version obtained by changing the drift of the underlying diffusion processs. Almost sure convergence of the modified algorithm was proved. Finally, we chose a Black-Scholes uncorrelated multidimensional model and we numerically tested the case of a put basket option. We showed how at the minimum variance could be reduced of more than a factor 25 with respect to the original algorithm. As we cannot predict the location of the minimum and as the search for it can be slow, we suggested two ways of proceeding in order to approximate it by the optimal estimator of the corresponding European option pricing problem. We gave moreover some hints on possible improvements of these approximations.

It is worth to stress that the coordinates of the drift which realise American minimum appear to be larger in absolute value than the corresponding European ones, all of them being negative in the relevant case; the optimal drift seems to drive the basket towards a neighbourhood of the strike. As for any given time the price of an American option is superior or equal to the corresponding European one, there are good reasons to believe that if, we disposed of some reliable approximation of critical price $\bar{S}$, then we could improve empirical criterion of section 5.2.1 togheter with remark 7 and price with the drift which drives the basket toward $\bar{S}$.

References

[A03] B. Arouna, *Robbins-Monro algorithms and variance reduction in finance*, forthcoming (accepted for publication in "The Journal of Computational Finance").

[BG97] M. Broadie, P. Glasserman, *Pricing American-style securities using simulation*, Journal of Economics Dynamics and Control **21**, pp.1323-1352, (1997).

[BPP02] V. Bally, G. Pagès, J. Printems *A quantization method for pricing and hedging multidimensional American style options*, preprint. LPMA n.753,(2002).

[CLP02] E. Clément, D. Lamberton, P. Protter, *An analysis of a least squares regression method for American option pricing*, Finance and Stochastic **6**, 449-471 (2002)

[LL97] D. Lamberton, B. Lapeyre, *Introduction au calcul stochastique appliqué à la finance*, Ellipse, (1997).

[LST] B. Lapeyre, A. Sulem, D. Talay, *Understanding Numerical Analysis for Financial Models*, forthcoming.

[LS01] F.A. Longstaff, E.S. Schwartz, *Valuing American options by simulation: a simple least squares approach*, Rev. Financial Stud. **14**, 113-148 (2001)

[GHS99] P. Glasserman, P. Heidelberger, P. Shahabuddin, *Asymptotically Optimal Importance sampling and Stratification for Pricing Path Dependent Options*, Mathematical Finance, Vol.9, No.2, 117-152, (April 1999).

[PTVF92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, 2nd edition*, Cambridge University Press, New York, 1992.

[TVR01] J.N. Tsitsiklis, B. Van Roy, *Regression Methods for Pricing Complex American-Style Options*, IEEE Transactions on Neural Networks, Vol. 12, No. 4 (special issue on computational finance), pp. 694-703, (July 2001).

Centre d'Enseignement et de Recherche en Mathématiques, Informatique et Calcul Scientifique, ENPC, 6 et 8 avenue Blaise Pascal, Cité Descartes, Champs sur Marne -77455 Marne la Vallée Cedex 2, France

Dipartimento di Fisica nucleare e Teorica, Università degli Studi di Pavia, Via U.Bassi, 6 - 27100 Pavia, Italia

E-mail address: moreni@cermics.enpc.fr