A quantization tree method for pricing and hedging multi-dimensional American options

Vlad Bally, Gilles Pagès, Jacques Printems

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A quantization tree method for pricing and hedging multi-dimensional American options

Vlad BALLY — Gilles PAGÈS — Jacques PRINTEMS

N° 4465

Mai 2002

THÈME 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4465--FR+ENG

INRIA ROCQUENCOURT

A quantization tree method for pricing and hedging multi-dimensional American options

Vlad BALLY* , Gilles PAGÈS† , Jacques PRINTEMS‡

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Mathfi

Rapport de recherche n° 4465 — Mai 2002 — 63 pages

Abstract: We present here the quantization method which is well-adapted for the pricing and hedging of American options on a basket of assets. Its purpose is to compute a large number of conditional expectations by projection of the diffusion on optimal grid designed to minimize the (square mean) projection error ([25]). An algorithm to computes such grids is described. We provide results concerning the orders of the approximation with respect to the regularity of the pay-off function and the global size of the grids. Numerical tests are performed in dimensions 2, 4, 6, 10 with American style exchange options. They show that our theoretical orders are probably pessimistic.

Key-words: American option pricing, Optimal Stopping, Snell envelope, quantization of random variables

* Univ. du Maine et projet MATHFI, INRIA. E-mail: bally@ccr.jussieu.fr
† Univ. Paris 6. E-mail: gpa@ccr.jussieu.fr.
‡ Univ. Paris 12. E-mail: printems@univ-paris12.fr.

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

Une méthode d'arbre par quantification pour l'évaluation et la couverture d'options américaines sur un panier d'actifs

Résumé : Nous proposons une méthode de quantification adaptée à l'évaluation et à la couverture d'options américaines sur des paniers d'actifs financiers. Elle repose sur le calcul d'un grand nombre d'espérances conditionnelles par projection de trajectoires de la diffusion sur des grilles optimalement disposées pour minimiser l'erreur (quadratique moyenne) ainsi commise ([25]). Un procédé algorithmique pour construire de telles grilles est explicité. Nous fournissons des résultats sur l'ordre d'approximation en fonction de la régularité de l'actif contingent (pay-off) et de la taille globale des grilles. Des tests numériques sont réalisés en dimensions 2, 4, 6, 10 avec des options américaines d'échange. Ces tests montrent que les estimations d'erreur théoriques sont vraissemblablement pessimistes.

Mots-clés : Évaluation d'options américaine, arrêt optimal, enveloppe de Snell, quantification de vecteurs aléatoires

1 Introduction

The aim of this paper is to present, to study and to test a probabilistic method for pricing and hedging American style options on multidimensional baskets of traded assets. The asset dynamics follow a d-dimensional diffusion model between time 0 and a maturity time T. We especially focused on a classical extension of the Black & Scholes model in which the volatility may depend on the asset prices. However, a large part of the algorithmic aspects of this paper can be applied to more general models.

Pricing an American option in a continuous time Markov process $(S_t)_{t\in[0,T]}$ consists in solving the continuous time optimal stopping problem related to an obstacle process. In this paper we are interested in "Markovian" obstacles of the form $h_t = (h(t, S_t)$ which are the most commonly considered on financial markets. Roughly speaking, there are two types of numerical methods for this purpose:

– First, some purely deterministic approaches coming from Numerical Analysis: the solution of the optimal stopping problem admits a representation $v(t, S_t)$ where v satisfies a parabolic variational inequality. So, the various discretizing techniques like finite difference or finite element methods yield an approximation of the function v at discrete points of a regular time-space grid (see *e.g.* [32] for an application to a vanilla put option or [9] for a more comprehensive study).

– Secondly, some probabilistic methods based on the dynamic programming formula or on the approximation of the (lowest) optimal stopping time. In 1-dimension, the most popular approach to American option pricing and hedging remains the implementation of the dynamic programming formula on a Binomial tree, originally initiated by Cox-Ross & Rubinstein as an elementary alternative to continuous time Black & Scholes model. However, let us mention before the massive development of Mathematical Finance, the pioneering work by Kushner in 1977 (see [28] and also [29]) in which the Markov chain approximation was first introduced, including its links with the finite difference method.

These methods are quite efficient to handle vanilla American options on a single asset but they quickly become intractable as the number of the underlying assets increases. Usually, numerical methods become inefficient because the space grids are built regardless of the distributions of the asset prices. The same problem occurs for finite state Markov chain approximation "à la Kushner". For the the extension of binomial treed into multinomial trees, the difficulty comes from the geometric shape of a tree compatible with all the dimension and correlation constraints.

In the past recent years, the problem gave birth to an extensive literature in order to overcome the dimensionality problem. All of them finally lead to some finite state dynamic programming algorithm either in its direct form or through the backward approximation of the (lowest) optimal stopping time. In [8], Barraquant & Martineau a sub-optimal 1-dimensional problem is solved: it amounts to process as if the obstacle process itself had the Markov property. In [36], the algorithm devised by Longstaff & Schwartz is based on conditional expectation approximation by regression, using a finite class $\varphi_i(S_t)$ of random variables derived from a "root" family (φ_i). In [40], Tsitsiklis & Van Roy use a similar idea but for a modified Markov transition. In [11], Braodie & Glassermann generates some random grids at each time step and compute some companion weights using some statistical ideas based on the importance sampling theorem.

Finally in [22] and [23] Fournié et al. initiated a new approach, based on Malliavin calculus, to the computation of conditional expectations and of their derivatives with respect to a parameter. This leads to a pure Monte Carlo method. Lions and Régnier in [36] use the same approach to price and to compute Greeks for American options.

In this paper, we propose a probabilistic method based on grids like in the original finite state Markov chain approximation method. First we start as usual by a time discretization of the asset price process at times $t_k := kT/n,\ k = 0, \ldots, n$ and, if necessary, we introduce the Euler scheme of the diffusion price process (still denoted S_{t_k} for a while). But we will not choose these grid *a priori*. We will use our ability to simulate large samples of the asset price diffusion process – or at least its Euler scheme – to produce at each discretization time t_k a grid Γ_k^* with a given size N_k having the following property: the closest neighbour rule projection $\pi^{\Gamma^*}(S_{t_k})$ of S_{t_k} onto the grid Γ_k^* is the the best least square approximation of S_{t_k} among *all random vectors* Z such that $|Z(\Omega)| = N$. Namely

$$\|S_{t_k} - \pi^{\Gamma_k^*}(\widehat{S}_{t_k})\|_2 = \min\left\{\|S_{t_k} - Z\|_2|,\ Z : \Omega \to \mathbb{R}^d,\ |Z(\Omega)| \le N_k\right\}.$$

In some sense we will produce and then use at each time step the best possible grid of size N_k to approximate the d-dimensional random vector S_{t_k}. For historical reasons coming from Information Theory, $\pi^{\Gamma_k^*}(S_{t_k})$ is often called the *optimal quantizer* of S_{t_k}. The resulting error bound $\|S_{t_k} - \pi^{\Gamma_k^*}(S_{t_k})\|_2$ is called the lowest quadratic mean quantization error and has been extensively investigated in Signal Processing and Information Theory for more than 50 years ([26] or more recently [25]). Namely one knows that it goes to 0 as N_k goes to infinity at a $O(N_k^{-\frac{1}{d}})$ rate.

Except in some specific 1-dimensional cases of little numerical interest, no closed form is available for the grid Γ_k^* that produces the optimal quantizer, nor for the induced lowest quantization error. In fact little is known on the geometric structure of optimal quantizer grids in higher dimension. However, starting from the integral representation

$$\|S_{t_k} - \pi^{\Gamma}(\widehat{S}_{t_k})\|_2^2 = \mathbb{E}\left(\min_{x\in\Gamma}|S_{t_k} - x|^2\right)$$

which is valid for any grid Γ and using its regularity properties as an almost everywhere differentiable (symmetric) function of Γ, one may implement a stochastic gradient descent that converges to some – at least locally – optimal quantizer . Furthermore, the procedure yields as by-products the *Voronoi companion parameters* of the grid which are involved in the pricing of the American option (see below). Simulations (see section) confirm what was expected: the optimal quantizer grid gets concentrated at zones heavily weighted by the diffusion process.

At this stage the time discretization consists in approximating the original American option with payoff $(h(t, S_t))_{0let\le T}$ by its so-called Bermuda counterpart to be exercised exclusively at discrete times t_k, $k = 0, \ldots, n$. It is classical background that then the theoretical premium is the result of a backward dynamic programming formula. Our algorithm simply consists in replacing the random variables S_{t_k} by their optimal quantizers $\widehat{S}_{t_k} := \pi^{\Gamma_k^*}(S_{t_k})$ and then to write down this backward procedure "in distribution". The weights that appears are known from the grid optimization procedure described above.

The second part of the paper is devoted to an extensive study of the rate of convergence of this algorithm a function of the accuracy of the time discretization (T/n) and of the total number $N := 1 + N_1 + \cdots + N_n$ of *elementary quantizers* used to produce the successive optimal quantizer grids on each time layer. We propose an optimal procedure to dispatch *a priori* these N $\mathbb{R}^d$-valued vectors among the layers and we derive some error bounds depending on the mean quantization error when the payoff is Lipschitz continuous. When the quantizer of each layer is optimal we obtain an a priori error bounds of the form $C(n^{-1/2} + n(N/n)^{-\frac{1}{d}})$ which can be improved (1 instead of 1/2 when the payoff is semi-convex).

Then we design an approximating *quantized hedging strategy* following the ideas by Follmer & Sondermann on incomplete markets. We are in position to derive some error bounds, called *local residual risks.* To produce these error bounds, we combine some methods borrowed from Reflected Backward Stochastic Differential Equation Theory, analytical p.d.e. techniques and quantization theory. We get a global rate of convergence for the hedging strategy which seems to be the first of that kind.

The last part of the paper is devoted to the experimental validation of the method on multi-dimensional American exchange options on (geometric) index in a standard d-dimensional Black & Scholes model.

A common fact of most probabilistic methods applied to the American option problem is in the approximation of the conditional expectations implied by the Snell envelope formulation of the problem (76). For example, in [36], the authors computed such conditional expectation by means of least-squares regression on a basis of well suited polynomial functions. Such method has the advantages of the regular approximation methods and the drawbacks of the global approximation methods. From a numerical point of view, our method is a grid method which relies on the approximation of the solution by piecewise constant functions. Its purpose is, as it is common to all probabilistic methods, to compute a large number of conditional expectations along the path of the associated diffusion process. The main differences with the scheme in [36] is that it is a local but irregular approximation scheme.

We will present numerical results which tend to show that when the grid are optimal (in the quadratic quantization sense), the spatial order of convergence is better than in usual grid methods. This rate, better than expected, makes up for the drawback of an irregular approximation. Two settings have been selected for simulation: "in-the-money" and "out-of-the-money", both in several dimensions $d = 2, 4, 6, 10$. The results show that computed prices are more accurately in the first setting (with a maximal error at most of 5% in 10-dimension). Such behaviour seems to be explained by a bigger effect of the numerical incompleteness of the market in the in-the-money setting.

Among the advantages of the local feature of the quantization approximation method, a prominent one is that it naturally leads to higher order approximations of the price involving the space derivatives (the hedging). This is developed in [6]. A second asset is that when the diffusion process is a function of the Brownian motion like in the Black & Scholes model, our method if completely parameter free and all computations can be performed once the optimal quantization of the Brownian motion has been achieved.

2 The investigated model

We consider a market on which are traded d risky assets $S^1, \ldots, S^d$ and a deterministic riskless asset $S^0_t := e^{rt}$, $r \in \mathbb{R}$ between time $t := 0$ and the maturity time $T > 0$. One typical model for the price process of the risky assets is the following diffusion model

$$dS^i_t = S^i_t \left(r\,dt + \sum_{j=1}^{q} \sigma_{ij}(e^{-rt}S_t)\,dW^j_t \right),\ S^i_0 := s^i_0 > 0,\ 1 \le i \le d \qquad (1)$$

where $W := (W^1, \ldots, W^q)$ is a standard q-dimensional Brownian Motion defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and

$$\sigma : \mathbb{R}^d \longrightarrow \mathcal{M}(d \times q) \text{ is a bounded and Lipschitz continuous.} \qquad (2)$$

The filtration of interest will be the natural (completed) filtration $\underline{\mathcal{F}} := (\mathcal{F}^S_t)_{t\in[0,T]}$ of S (which coincides with that of the Brownian motion as soon as $\sigma\sigma^*(x) > 0$ for every x).

For notational convenience, we will denote $c(x) := \mathrm{Diag}(x)\sigma(x)$. Note that c and the drift $b(x) := rx$ are Lipschitz so that a unique strong solution exists for (1) on $(\Omega, \mathcal{A}, \mathbb{P})$. Furthermore, it is classical background that, for every $p \ge 1$,

$$\mathbb{E}_{s_0}\big(\sup_{t\in[0,T]} |S_t|^p\big) < C_p(1 + |s_0|^p).$$

The discounted price process $\widetilde{S}_t := e^{-rt}S_t$ is then a positive $\mathbb{P}$-martingale satisfying

$$d\widetilde{S}_t = c(\widetilde{S}_t).dW^j_t,\ \widetilde{S}_0 := s_0. \qquad (3)$$

$\mathbb{P}$ is the so-called *risk neutral* probability in Mathematical Finance terminology. As long as $q \ne d$, the usual completeness of the market necessarily fails. However, from numerical point of view, this has no influence on the implementation of the quantization method to compute the price of the derivatives: we just compute a $\mathbb{P}$-price. When coming to the problem of hedging these derivatives, then the completeness assumption becomes crucial and will lead us to assume that $q = d$ and that the diffusion coefficient $c(x)$ is invertible everywhere on $(\mathbb{R}^*_+)^d$.

When $q = d$ and $\sigma(x) \equiv \sigma \in \mathcal{M}(d \times d)$, (1) is the usual d-dimensional Black & Scholes model: the risky assets are geometric Brownian motions given by

$$S_t^i = s_0^i \exp\left((r - \frac{1}{2}|\sigma_{i.}|^2)t + \sum_{1\le j\le d} \sigma_{ij} W_t^j\right),\ 1 \le i \le d.$$

An American option related to a payoff process $(h_t)_{t\in[0,T]}$ is a contract that gives the right to receive once and only once the payoff h_t at some time $t \in [0,T]$ where $(h_t)_{t\in[0,T]}$ is a $\underline{\mathcal{F}}$-adapted nonnegative process. In this paper we will always consider the sub-class of payoffs h_t that only depends on (t, S_t) *i.e.* satisfying

$$h_t := h(t, S_t), \quad t \in [0,T] \quad \text{where } h : [0,T] \longrightarrow \mathbb{R}_+ \text{ is a } \textit{Lipschitz} \text{ continuous.} \tag{4}$$

Such payoffs are sometimes called *vanilla* payoffs. Under Assumptions (1) and (4), one has for every $p \ge 1$,

$$\mathbb{E}(\sup_{t\in[0,T]} |h_t|^p) < +\infty$$

One justifies that – in a complete market – the fair price $\mathcal{V}_t$ at time t for such a contract satisfies

$$\mathcal{V}_t := e^{rt}\text{ess sup}\left\{\mathbb{E}(e^{-r\tau} h_\tau / \mathcal{F}_t),\ \tau \in \mathcal{T}_t\right\} \tag{5}$$

where $\mathcal{T}_t := \{\tau : \Omega \to [t,T],\ \underline{\mathcal{F}}\text{-stopping time}\}$. This simply means that the discounted price $\widetilde{\mathcal{V}}_t := e^{-rt}\mathcal{V}_t$ of the option is the *Snell envelope* of the discounted American payoff

$$\widetilde{h}_t := \widetilde{h}(t, \widetilde{S}_t)$$

with $\widetilde{h}(t,x) = e^{-rt}h(t, e^{rt}x)$. This result is based on a hedging argument on which we will come back later on. Note that $\sup_{t\in[0,T]} |\mathcal{V}_t| \le \sup_{t\in[0,T]} |h_t| \in L^p$, $p \ge 1$.

One shows (see [9]) using the Markov property of the diffusion process $(S_t)_{t\in[0,T]}$ that $\mathcal{V}_t := \nu(t, S_t)$ where ν solves the variational inequality

$$\max\left(\frac{\partial \nu}{\partial t} + \mathcal{L}_{r,\sigma}\nu, \nu - h\right) = 0, \quad \nu(T,.) = h(T,.). \tag{6}$$

where $\mathcal{L}_{r,\sigma}$ denotes the infinitesimal generator of the diffusion (1).

Then, it is clear that the approximation problem for $\mathcal{V}_t$ appears as special case of the approximate computation of the Snell envelope of a d-dimensional diffusion with Lipschitz coefficients. To solve this problem in 1-dimension, many methods are available. These methods can be classified in two families: the probabilistic ones based on a weak approximation of the diffusion process (S_t) by a purely discrete dynamics (*e.g.* binomial trees, [**?**]) and the analytic ones based on numerical methods for solving the variational inequality (6) (*e.g.* finite difference or finite element methods). When the dimension d of the market increases, these methods become inefficient.

In [4] (see also [5]) is devised a new method based on quantization which can be implemented in higher dimension. Next section is devoted to a presentation of the method and of the theoretical *a priori* error bounds that have been already derived.

At this stage, *one may assume without loss of generality that the interest rate* r *in (1) is* 0 since Equation (3) for $\widetilde{S}$ appears as a special case of (1) for S since the function $\widetilde{h}(t,x)$ has the same regularity as h. (To derive the "true" formulae when $r \neq 0$ one just has to keep in mind the "original" equation $d\widetilde{S}_t = c(S_t).dW_t$).

3 Pricing an American option using a quantization tree

In that part, the specificity of the martingale diffusion dynamics proposed for the risky assets in (3) (with $r = 0$) has little influence on the results, so it is costless to consider a general drifted Brownian diffusion

$$dS_t = b(S_t)\,dt + c(S_t).dW_t, \qquad S_0 := s_0 \in \mathbb{R}^d. \tag{7}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $c : [0,T] \times \mathbb{R}^d \to \mathcal{M}(d \times q)$ are Lipschitz continuous vector fields and $(W_t)_{t\in[0,T]}$ is q-dimensional Brownian motion.

3.1 Time discretization: the Bermuda options

The exact simulation of a diffusion at time t is usually out of reach (*e.g.* when σ is not constant in the specified model (1)). So one uses a (Markovian) discretization scheme, easy to simulate, *e.g.* the Euler scheme:

$$\overline{S}_{t_{k+1}} = \overline{S}_{t_k} + b(\overline{S}_{t_k})\frac{T}{n} + c(\overline{S}_{t_k}).(W_{t_k} - W_{t_{k-1}}). \tag{8}$$

Then, *the Snell envelope to be approximated by quantization is that of the Euler scheme.*

Sometimes, the diffusion can be simulated simply, essentially because it appears as a closed form $S_t := \varphi(t, W_t)$. This is the case of the regular multi-dimensional Black & Scholes model (set $\sigma(x) := \sigma$ in (1)). Then, it is possible to consider directly the *the Snell envelope of the homogeneous Markov chain* $(S_{t_k})_{0\le k\le n}$ for quantization purpose.

This time discretization corresponds, in the derivative terminology, to approximating the original continuous time American option by a *Bermuda option*, either on $\overline{S}$ or on S itself. By Bermuda option, one means that the set of possible exercise times is finite. Error bounds are available at these exercise times t_k (see Theorem 1 below).

Whatsoever, we want to quantize the Snell envelope of a homogeneous discrete time Markov chain (S_{t_k} or $\overline{S}_{t_k}$) whose transition, denoted $P^{(n)}(x,dy)$, preserves Lipschitz continuity. More precisely, for every Lipschitz continuous $f : \mathbb{R}^d \to \mathbb{R}$

$$[P^{(n)} f]_{Lip} \le (1 + C_{b,\sigma,T}\frac{T}{n})[f]_{Lip}. \tag{9}$$

(see, *e.g.*, [4] for a proof). In fact this discrete time markovian setting is the natural framework for the method. Throughout this section, we will denote by the generic notation $(X_k)_{0\le k\le n}$ any (L^p-integrable) homogeneous $\mathcal{F}_{t_k}$-Markov chain whose transition $P^{(n)}$ satisfies (9) and we will denote by $(V_k)_{0\le k\le n}$ its $\mathcal{F}_{t_k}$-Snell envelope of $h(t_k, X_k)$ defined by

$$V_k := \text{ess sup}\left\{\mathbb{E}\left(h(\theta, X_\theta)/\mathcal{F}_{t_k}\right),\ \theta\in\Theta_k\right\}$$

where Θ_k denotes the set of $\{t_k, \ldots, t_n\}$-valued $\mathcal{F}_{t_\ell}$-stopping times. The Snell envelope V_k satisfies the so-called backward *dynamic programming formula* (see [37]):

$$\left\{\begin{array}{lcl} V_n & := & h(t_n, X_n), \\ V_k & := & \max\left(h(t_k, X_k), \mathbb{E}(V_{k+1}/\mathcal{F}_{t_k})\right), \quad 0\le k\le n-1. \end{array}\right. \tag{10}$$

One derives a dynamic programming formula *in distribution*: $V_k = v_k(X_k),\ k\in\{0,\ldots,n\}$, where the functions v_k are recursively defined by

$$\left\{\begin{array}{l} v_n := h(t_n,\,.), \\ v_k := \max\left\{h(t_k,.),\, P^{(n)}(v_{k+1})\right\},\ 0\le k\le n-1. \end{array}\right. \tag{11}$$

This formula remains intractable for numerical computation since they require to compute at each time step a conditional expectation.

Theorem 1 below gives some L^p-error bounds that hold for $\mathcal{V}_{t_k} - V_{t_k}$ in our original diffusion framework. First we need to introduce some definition about the regularity of h.

Definition 1 *A function* $h:\mathbb{R}^d \longrightarrow \mathbb{R}$ *is* semi-convex *if*

$$\forall\, x,\, y\in\mathbb{R}^d,\ \forall\, t\in\mathbb{R}_+,\qquad h(t,y) - h(t,x) \ge (\delta_h(t,x)|y-x) - \rho|x-y|^2 \tag{12}$$

where δ_h *is a bounded function on* $\mathbb{R}_+\times\mathbb{R}^d$ *and* $\rho\ge 0$.

Remarks: Note that (12) appears as a convex assumption relaxed by $-\rho|x-y|^2$. In most situations, is used in the reverse sense *i.e.* $h(t,x) - h(t,y) \le (\delta_h(t,x)|x-y) + \rho|x-y|^2$. The semi-convexity assumption is fulfilled by a wide class of functions:

– If $h(t,.)$ is $\mathcal{C}^1$ for every $t\in\mathbb{R}_+$ and $\frac{\partial h}{\partial x}(t,x)$ is ρ-Lipschitz in x, uniformly in t, then h is semi-convex (with $\delta_h(t,x) := \frac{\partial h}{\partial x}(t,x)$).

– If $h(t,.)$ is convex for every $t\in\mathbb{R}_+$ with a derivative $\delta_h(t,.)$ (in the distribution sense) which is bounded in (t,x), then h is semi-convex (with $\rho = 0$). Thus, it embodies *most usual pay-off functions used for pricing vanilla and exotic American style options* like $h(t,x) := e^{-rt}(K - \varphi(e^{rt}x))_+$ with φ Lipschitz continuous (on sets $\{\varphi\le L\}$, $L>0$).

The notion of semi-convex function seems to have been introduced in [18] for pricing one-dimensional American options. See also [33] for recent developments in a similar setting.

Theorem 1 *(a) Let $h : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function and let $p \in [1, +\infty)$. Let V_n denote the Snell envelope of $(\overline{S}_{t_k})_{0\le k\le n}$ or $(S_{t_k})_{0\le k\le n}$. There is some positive real constant C depending on $[b]_{Lip}$, $[c]_{Lip}$ and p such that*

$$\forall\, n \in \mathbb{N}^*,\ \forall\, k \in \{0,\dots,n\}, \qquad \|\mathcal{V}_{t_k} - V_k\|_p \le \frac{e^{CT}(1+|x|)}{\sqrt{n}}. \tag{13}$$

(b) If $X_k = S_{t_k}$, $k = 0,\dots,n$ and if the obstacle h is semi-convex, *then*

$$\forall\, n \in \mathbb{N}^*,\ \forall\, k \in \{0,\dots,n\}, \qquad \|\mathcal{V}_{t_k} - V_k\|_p \le \frac{e^{CT}(1+|x|)}{n} \tag{14}$$

3.2 Space discretization: the quantization tree

3.2.1 Abstract quantization

The starting point of our method is simply to discretize the random variables X_k of the Markov chain by introducing some random variables $\widehat{X}_k$ that can only take a finite number N_k of values. Then, we want to approximate the Snell envelope $(V_k)_{0\le k\le n}$ by a sequence $(\widehat{V}_k)_{0\le k\le n}$ defined by a dynamic programming algorithm quite similar to (10) except that it involves:

– the random variables $(\widehat{X}_k)_{0\le k\le n}$ instead of the $(X_k)_{0\le k\le n}$ and

– the conditional expectation given $\widehat{X}_k$ instead of the regular conditional expectation given the past up to k, *i.e.* $\mathbb{E}_k := \mathbb{E}(\,.\,/\mathcal{F}_k)$.

3.2.2 Elementary abstract quantization, L^p-distortion

Let X be a $\mathbb{R}^d$-valued random vector defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. From a probabilistic point of view, L^p-quantization ($p \ge 1$) consists in studying the best L^p-approximation of X by random vectors X' taking at most N fixed values $x_1, \cdots, x_N \in \mathbb{R}^d$. These random vectors read $X' := \sum_{i=1}^N x_i \mathbf{1}_{A_i}(X)$, $(A_i)_{1\le i\le N}$ Borel partition of $\mathbb{R}^d$. One easily proves that for a fixed N-tuple $x := (x_1, \cdots, x_N) \in (\mathbb{R}^d)^N$, the L^p-mean error $\|X - X'\|_p$ reaches its minimum at any *Voronoi tessellation* $(C_i(x))_{1\le i\le N}$ of x.

Definition 2 *Let $x := (x_1, \cdots, x_N) \in (\mathbb{R}^d)^N$. A partition $C_1(x), \dots, C_N(x)$ of $\mathbb{R}^d$ is a Voronoi tessellation of the N-tuple x if, for every $i \in \{1,\dots,N\}$, $C_i(x)$ is a Borel set satisfying*

$$C_i(x) \subset \left\{ y \in \mathbb{R}^d \,/\, |x_i - y| = \min_{1\le j\le N} |y - x_j| \right\}.$$

where $|\,.\,|$ (usually) denotes the canonical Euclidean norm.

Note that, however the i^{th} tessel $C_i(x)$ always has the same closure and the same boundary, this boundary being included in at most $N-1$ hyperplanes. Furthermore if the distribution of X weights no hyperplane, then the Voronoi tessellation is essentially unique so that all the "x-Voronoi quantizers" $\widehat{X}$ have the same distribution.

The problem is then to estimate the *L^p-mean quantization error* $\|X - \widehat{X}\|_p$ between X and a "Voronoi quantizer"

$$\widehat{X} := \sum_{i=1}^{N} x_i \mathbf{1}_{C_i(x)}(X). \tag{15}$$

Let $\mathbb{P}_X$ denote the distribution of X. Then, the L^p-mean quantization error is given by

$$\|X - \widehat{X}\|_p^p = \sum_{i=1}^{N} \mathbb{E}\left(\mathbf{1}_{C_i(x)}|X - x_i|^p\right) = \mathbb{E}\left(\min_{1\le i\le N}|X - x_i|^p\right) = \int_{\mathbb{R}^d} \min_{1\le i\le N}|x_i - y|^p \mathbb{P}_X(dy). \tag{16}$$

The L^p-mean quantization error only depends upon the distribution $\mathbb{P}_X$ of X.

The next step is to specify the choice of the N-tuple $x := (x_1, \ldots, x_N)$ so as to get the smallest possible L^p-mean quantization error and to evaluate how fast it goes to 0 as N goes to infinity. This will be investigated further on in subsection 3.3.1, once the way we use quantization and it related error will have been developed.

3.2.3 Quantization tree: a pseudo-Snell envelop

We assume from now on that for every $k \in \{0, 1, \ldots, n\}$, we have access some way or another to a Voronoi quantized random vector $\widehat{X}_k$ for X_k, using N_k points $x_1^k, \cdots, x_{N_k}^k$.

The pseudo-dynamic programming formula below is devised by analogy with the original one (10): one simply replaces X_k by its quantized random vector $\widehat{X}_k$ It reads as follows

$$\begin{cases} \widehat{V}_n &:= h(t_n, \widehat{X}_n), \\ \widehat{V}_k &:= \max\left(h(t_k, \widehat{X}_k), \mathbb{E}(\widehat{V}_{k+1}/\widehat{X}_k)\right), \quad 0 \le k \le n-1. \end{cases} \tag{17}$$

NOTATION: for the sake of simplicity, from now on, we will denote $\widehat{\mathbb{E}}_k(\,.\,) := \mathbb{E}(\,.\,/\widehat{X}_k)$.

The main reason for considering conditional expectation with respect to $\widehat{X}_k$ is that the the sequence $(\widehat{X}_k)_{k\in\mathbb{N}}$ is not Markovian. On the other hand, even if the N_k-tuple $x^k := (x_1^k, \ldots, x_{N_k}^k)$ of every term X_k of the chain has been set up *a priori*, this does not make possible to compute explicitly this algorithm. As a matter of fact, one needs to know the coupled distributions $(\widehat{X}_k, \widehat{X}_{k+1})$, $0 \le k \le n-1$. This is enlightened by the easy Proposition below.

Proposition 1 *Let $x^k := (x_1^k, \ldots, x_{N_k}^k)$ denote for every $k \in \{0, \ldots, n\}$ a quantization of the distribution $\mathcal{L}(X_k)$. Set, for every $k \in \{0, \ldots, n\}$ and every $i \in \{1, \ldots, N_k\}$,*

$$\alpha_i^k := \mathbb{P}(\widehat{X}_k = x_i^k) = \mathbb{P}(X_k \in C_i(x^k)), \tag{18}$$

and, for every $k \in \{0,\ldots,n-1\}$, $i \in \{1,\ldots,N_k\}$, $j \in \{1,\ldots,N_{k+1}\}$

$$\begin{aligned}\pi^k_{ij} &:= \mathbb{P}(\widehat{X}_{k+1} = x^{k+1}_j / \widehat{X}_k = x^k_i) = \mathbb{P}\left(X_{k+1} \in C_j(x^{k+1}) / X_k \in C_i(x^k)\right)\\ &= \frac{\beta^k_{ij}}{\alpha^k_i} \quad \textit{with} \quad \beta^k_{ij} := \mathbb{P}\left(X_{k+1} \in C_j(x^{k+1}),\ X_k \in C_i(x^k)\right). \end{aligned} \tag{19}$$

One defines by a backward induction the function $\widehat{v}_k$ *by*

$$\begin{aligned}\widehat{v}_n(x^n_i) &:= h_n(x^n_i), \qquad i \in \{0,\ldots,N_n\}\\ \widehat{v}_k(x^k_i) &:= \max\left(h(t_k,x^k_i), \sum_{j=1}^{N_{k+1}} \pi^k_{ij}\, \widehat{v}_{k+1}(x^{k+1}_j)\right), \ 1 \le i \le N_k,\ 0 \le k \le, n-1. \end{aligned} \tag{20}$$

Then, $\widehat{V}_k = \widehat{v}_k(\widehat{X}_k)$ *satisfies the above dynamic programming (17) of the pseudo-Snell envelop. Thus, if* $\mu_0 := \delta_{x_0}$, *then* $\widehat{v}_0(\widehat{X}_0) = \widehat{v}_0(x_0)$ *is deterministic.*

Simply implementing the algorithm defined by (20) on a computer raises two questions:

– How is it possible to estimate the parameters α^k_i and β^k_{ij} involved in (20) ?

– Is it possible to handle the complexity of such a tree structured algorithm ?

PRELIMINARY ESTIMATION PHASE (FIRST APPROACH): the theoretical tractability of the above algorithm exclusively depends on the parameters α^k_i and β^k_{ij}. Actually, the ability to compute the α^k_i's and the β^k_{ij}'s at a reasonable cost is the key of the whole method presented here for practical implementation. The most elementary solution is simply to process a wide range regular *Monte Carlo simulation of the Markov chain* $(X_k)_{0\le k\le n}$ to estimate the parameters α^k_i and β^k_{ij} of interest defined by (18) and (19). An estimate of the L^p-mean quantization error $\|X_k - \widehat{X}_k\|_p$ can also be computed along the procedure. Actually, this ability to compute these weights and moduli at a reasonable cost is the key of the whole method. When $(X_k)_{0\le k\le n}$ is a Euler scheme (or Black & Scholes diffusion) this makes no problem. More generally, this depends upon the ability to simulate some $P(x,dy)$-distributed random numbers for any $x \in \mathbb{R}^d$.

We will see further on in paragraphs 3.3.1 and 3.3.2 how to choose the N_k-tuples x^k (size and geometric location).

COMPLEXITY OF THE QUANTIZATION TREE : THEORY AND PRACTISE A quick look at the structure of the algorithm (20) shows that going from layer $k+1$ down to layer k needs $C \times N_k.N_{k+1}$ elementary computations (C is a positive real constant). Hence, the cost of a full tree descent in order to get $(\widehat{v}_0(x^0_i))_{1\le i\le N_0}$ approximately is

$$\text{Complexity} = C \times (N_0N_1 + N_1N_2 + \cdots + N_kN_{k+1} + \cdots + N_{n-1}N_n).$$

Setting $N := N_0 + \cdots + N_n$ shows that this complexity is at most equal to

$$\text{Complexity} \leq C.\frac{N^2}{n+1} \qquad \text{whatever the } N_k\text{'s are.}$$

This "blind" combinatorial estimate needs to be tuned. In fact, in most examples the Markov transition $P(x, dy)$ behaves in such a way that, at each layer k, many terms of the "transition matrix" $[\pi^k_{ij}]$ are numerically 0. This means that the estimates of these coefficients will often be 0! Subsequently, the true complexity of the algorithm is more likely close to $O(N)$ instead of the above N^2/n estimation. Thus, the cost of such a "descent" is similar to that of a Cox-Ross-Rubinstein's one dimensional binomial tree with $O(\sqrt{N})$ time discretization instants (such a tree approximately contains $N/2$ points).

3.3 Convergence and rate

The aim of this paragraph is to provide some *a priori* L^p-error bounds for $\|V_k - \widehat{V}_k\|_p$, $0 \leq k \leq n$, based on the L^p-mean quantization errors *i.e.* $\|X_k - \widehat{X}_k\|_p$, $0 \leq k \leq n$ where quantizer $\widehat{X}_k$ is a Voronoi quantizer that takes N_k values $x^k_1, \ldots, x^k_{N_k}$. This error modulus can be obtained as a by-product of a Monte Carlo simulation of $(X_k)_{0 \leq k \leq n}$: it only requires to compute, for every μ_k-distributed simulated random vector, its distance to its closest neighbour in the set $\{x^k_1, \ldots, x^k_{N_k}\}$.

The estimates below can be obtained for any Markov chain having a *Lipschitz* transition $P(x, dy)$ *i.e.* satisfying, for every Lipschitz continuous function $g : \mathbb{R}^d \longrightarrow \mathbb{R}$,

$$[Pg]_{Lip} \leq K[g]_{Lip} \quad \text{where} \quad [f]_{Lip} := \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}. \tag{21}$$

This is the case of the Euler scheme (and the diffusion) having Lipschitz drift and diffusion coefficient as mentioned before see (9).

The theorem below specifies the Lipschitz regularity of the functions u_k defined in (11) and gives the *a priori* error bounds in this Lipschitz setting.

Theorem 2 *Assume that the function h is $[h]_{Lip}$-Lipschitz continuous in x, uniformly time and that the transition P is K-Lipschitz.*

(*a*) *The functions u_k defined by (11) are Lipschitz continuous and*

$$[u_k]_{Lip} \leq (K \vee 1)^{n-k}[h]_{Lip}. \tag{22}$$

(*b*) *For every $k \in \{0, \ldots, n\}$, let $\widehat{X}_k$ denote a (Voronoi) quantizer of X_k. For every $p \geq 1$,*

$$\left\| V_k - \widehat{V}_k \right\|_p \leq \sum_{i=k}^{n} d_i \left\| X_i - \widehat{X}_i \right\|_p$$

with $d_i := [h]_{Lip} + cK[u_{i+1}]_{Lip}$, $0 \leq i \leq n-1$ and $d_n := [h]_{Lip}$. The real constant $c := 2$ if $p \neq 2$ and $c := 1$ if $p = 2$.

Proof: (a) Clearly, $[u_n]_{Lip} \leq [h]_{Lip}$. Then, using that $|\max(a,b) - \max(a',b')| \leq \max(|a - a'|, |b-b'|)$, it follows from the dynamic programming equality (11) that

$$[u_k]_{Lip} \leq \max\left([h]_{Lip}, [P(u_{k+1})]_{Lip}\right) \leq \max\left([h]_{Lip}, K[u_{k+1}]_{Lip}\right)$$

One concludes by induction.

(b) Set $\Phi_k := P(u_{k+1})$ for every $k \in \{0,\ldots,n-1\}$ (and $\Phi_n \equiv 0$). The function Φ_k satisfies $\mathbb{E}(V_{k+1}(X_{k+1})) = \Phi_k(X_k)$. One defines similarly $\widehat{\Phi}_k$ by the equality $\widehat{\mathbb{E}}_k(\widehat{V}_{k+1}/\widehat{X}_{k+1}) := \widehat{\Phi}_k(\widehat{X}_k)$ (and $\widehat{\Phi}_n \equiv 0$). Then

$$\begin{aligned}\left|V_k - \widehat{V}_k\right| &\leq |h_k(X_k) - h_k(\widehat{X}_k)| + \left|\Phi_k(X_k) - \widehat{\Phi}_k(\widehat{X}_k)\right| \\ &\leq [h]_{Lip}|X_k - \widehat{X}_k| + \left|\Phi_k(X_k) - \widehat{\mathbb{E}}_k(\Phi_k(X_k))\right| + \left|\widehat{\mathbb{E}}_k(\Phi_k(X_k)) - \widehat{\Phi}_k(\widehat{X}_k)\right|.\end{aligned}$$

Now
$$\begin{aligned}\left|\Phi_k(X_k) - \widehat{\mathbb{E}}_k\Phi_k(X_k)\right| &\leq \left|\Phi_k(X_k) - \Phi_k(\widehat{X}_k)\right| + \widehat{\mathbb{E}}_k\left|\Phi_k(X_k) - \widehat{\mathbb{E}}_k(\Phi_k(\widehat{X}_k))\right| \\ &\leq [\Phi_k]_{Lip}\left(|X_k - \widehat{X}_k| + \widehat{\mathbb{E}}_k|X_k - \widehat{X}_k|\right).\end{aligned}$$

(Note that $\widehat{X}_k$ is $\mathcal{F}_k$-measurable.) Hence,

$$\|\Phi_k(X_k) - \widehat{\mathbb{E}}\Phi_k(X_k)\|_p \leq c[\Phi_k]_{Lip}\|X_k - \widehat{X}_k\|_p$$

with $c = 2$. When $p = 2$, the very definition of the conditional expectation as a projection in a Hilbert space implies that holds with $c = 1$.

On the other hand, coming back to the definition of $\Phi_k(X_k)$ and $\widehat{\Phi}_k(\widehat{X}_k)$, one gets, using that $\widehat{\mathbb{E}}_k \circ \mathbb{E}_k = \widehat{\mathbb{E}}_k$,

$$\left|\widehat{\mathbb{E}}_k(\Phi_k(X_k)) - \widehat{\Phi}_k(\widehat{X}_k)\right| \leq \widehat{\mathbb{E}}_k\left|V_{k+1} - \widehat{V}_{k+1}\right|.$$

Any conditional expectation being a L^p-contraction, it follows that

$$\left\|\widehat{\mathbb{E}}_k(\Phi_k(X_k)) - \widehat{\Phi}_k(\widehat{X}_k)\right\|_p \leq \left\|V_{k+1} - \widehat{V}_{k+1}\right\|_p.$$

Finally, for every $k \in \{0,\ldots,n-1\}$,

$$\begin{aligned}\left\|V_k - \widehat{V}_k\right\|_p &\leq [h]_{Lip}\left\|X_k - \widehat{X}_k\right\|_p + \left\|\Phi_k(X_k) - \widehat{\mathbb{E}}_k(\Phi_k(X_k))\right\|_p + \left\|\Phi_k(X_k) - \widehat{\Phi}_k(\widehat{X}_k)\right\|_p \\ &\leq \left\|V_{k+1} - \widehat{V}_{k+1}\right\|_p + ([h]_{Lip} + c[\Phi_k]_{Lip})\left\|X_k - \widehat{X}_k\right\|_p.\end{aligned}$$

On the other hand, $\|V_n - \widehat{V}_n\|_p \le [h]_{Lip} \|X_n - \widehat{X}_n\|_p$, so that

$$\left\| V_k - \widehat{V}_k \right\|_p \le \sum_{i=k}^{n} ([h]_{Lip} + c[\Phi_i]_{Lip}) \left\| X_i - \widehat{X}_i \right\|_p$$

The definition of Φ_i and the K-Lipschitz property of the transition $P(x, dy)$ imply that

$$[\Phi_i]_{Lip} = [P(u_{i+1})]_{Lip} \le K[u_{i+1}]_{Lip}$$

and so (a) is proved. $\diamond$

3.3.1 Optimal quantization: existence and asymptotics

The L^p-quantization error has a an attractive specificity among other usual error bounds used in Numerical Integration: it behaves as a regular function of the quantizing N-tuple $x := (x^1, \ldots, x^N)$. More precisely, as a *symmetric* function of the N-tuple x, the L^p-quantization error is 1-Lipschitz continuous. If $\mathbb{P}_X$ has a compact support, it is straightforward that $x \mapsto \|X - \widehat{X}^x\|_p$ reaches a minimum at x^*. One may always assume that $x^* \in \left(\mathcal{H}(\text{supp}\, \mathbb{P}_X)\right)^N$ (convex hull of $\text{supp}\, \mathbb{P}_X$). When $\mathbb{P}_X$ no longer has a compact support, one shows by induction on N that

$$x \mapsto \|X - \widehat{X}^x\|_p \text{ still reaches an absolute minimum on } (\mathbb{R}^d)^N$$

(see [38] or [25], among others), still lying in $(\mathcal{H}(\text{supp}\, \mathbb{P}_X))^N$. Furthermore, one shows the following simple facts (see [38] or [25] and references therein):

– If $\text{supp}\mathbb{P}_X$ has an *infinite* support, any optimal N-tuple x^* has pairwise distinct elements.

– If $\text{supp}\, \mathbb{P}_X$ is *everywhere dense in its convex hull*, then the N components of an optimal N-tuple x^* all lies in $\mathcal{H}(\text{supp}\mu)$. This still holds true for N-tuples corresponding to local minima. In particular, this holds if $\mathbb{P}_X$ has a *positive* density function on $\mathbb{R}^d$.

– The minimal L^p-quantization error goes to zero as $N \to \infty$ *i.e.*

$$\lim_N \min_{x \in (\mathbb{R}^d)^N} \|X - \widehat{X}^x\|_p = 0.$$

As a matter of fact, let $(z_k)_{k\in\mathbb{N}}$ denote an everywhere dense sequence of $\mathbb{R}^d$-valued vectors and set $x_N := \{z_1, \ldots, z_N\}$. It is straightforward that $\|X - \widehat{X}^{x_N}\|_p$ goes to zero by the Lebesgue Dominated Convergence Theorem. Furthermore $0 \le \min_{x\in(\mathbb{R}^d)^N} \|X - \widehat{X}^x\|_p \le \|X - \widehat{X}^{x_N}\|_p$. $\diamond$

At which rate does this convergence to zero hold turns out to be a much more challenging question. The answer was completed by several authors (Zador, see [26], Bucklew & Wise, see [12] and finally Graf & Luschgy see [25]). It reads as follows

Theorem 3 *Assume that $E|X|^{p+\eta} < +\infty$ for some $\eta > 0$. Then*

$$\lim_N \left(N^{\frac{p}{d}} \min_{x\in(\mathbb{R}^d)^N} \|X - \widehat{X}^x\|_p^p \right) = J_{p,d} \|\varphi\|_{\frac{d}{d+p}} \tag{23}$$

where $\mathbb{P}_X(du) = \varphi(u)\,\lambda_d(du) + \nu(du)$, $\nu \perp \lambda_d$ (λ_d Lebesgue measure on $\mathbb{R}^d$) and $\|g\|_q := \left(\int |g|^q(x)\,dx\right)^{\frac{1}{q}}$ for every $q \in \mathbb{R}_+^*$. The constant $J_{p,d}$ corresponds to the case of the uniform distribution on $[0,1]^d$ (or any Borel set of Lebesgue measure 1).

Little is known about the true value of the constant $J_{p,d}$ except in dimension 1 where $J_{p,1} = \frac{1}{2^p(p+1)}$. Some geometric considerations lead to $J_{2,2} = \frac{5}{18\sqrt{3}}$ (see [26]). Nevertheless (see [?] or [17]) some reasonable bounds are available, based on random quantization. (The idea is to upper-bound $\min_{x\in(\mathbb{R}^d)^N} \|X - \widehat{X}^x\|_p^p$ by $\| \min_{1\le i\le N} |X - Z_i|\|_p^p$ where the Z_i's are i.i.d. with an appropriate distribution).

Whatsoever, this theorem says that $\min_{x\in(\mathbb{R}^d)^N} \|X - \widehat{X}^x\|_p \sim C_{X,p,d} N^{\frac{1}{d}}$. This is in accordance with the commonly admitted rates obtained *e.g.* in Numerical Integration by uniform N-tuple methods. In some sense, although optimal quantizers are never uniform square grid (except for the uniform distribution in 1-dimension), optimal quantization provides the best possible “grid method" for a given distribution μ.

3.3.2 Optimal quantization: how to get it ?

When $x = \{x^1, \ldots, x^N\}$, Equation (16) implies that $\|X - \widehat{X}^x\|_p^p = \mathbb{E}\left(\min_{1\le i\le N} |X - x^i|^p\right)$. The induced symmetric function on $(\mathbb{R}^d)^N$ is (Lipschitz) continuous and is denoted D_N^p from now on([1]). One shows (see, *e.g.*, [25] when $p = 2$ or [38]) that, if $p > 1$, D_N^p is continuously differentiable at every N-tuple $y \in (\mathbb{R}^d)^N$ satisfying $\forall\, i \ne j,\ x^i \ne x^j$ and $\mathbb{P}_X\left(\cup_{i=1}^N \partial C_i(y)\right) = 0$. The gradient $\nabla D_N^p(y)$ is obtained by formal differentiation, that is

$$\nabla D_N^p := \left(\mathbb{E}\frac{\partial D_N^p}{\partial x^i}(y,X)\right)_{1\le i\le n} = \left(\int_{\mathbb{R}^d} \frac{\partial D_N^p}{\partial x^i}(y,u)\mathbb{P}_X(du)\right)_{1\le i\le n}$$

$$\text{where}\quad \frac{\partial D_N^p}{\partial x^i}(y,u) := p\frac{u - x^i}{|u - x^i|}|u - x^i|^{p-1}\mathbf{1}C_i(y)(u),\ 1\le i\le n.$$

(The above result still holds when $p = 1$provided that $\mathbb{P}_Y$ is continuous.)

So, the gradient of D_N^p has an integral representation with respect to the distribution of X this strongly suggests to implement a stochastic gradient descent derived from this representation to approximate some (local) minimum of D_N^p: whenever $d \ge 2$, the implementation of deterministic gradient descent is irrealistic since it would rely on the computation of many

[1]The letter D is a reference to the word *distortion* which used en Information Theory for the L^p-mean quantization error (to the power p)

integrals with respect ... to $\mathbb{P}_X$. This stochastic gradient descent is defined as follows: let $(\xi^t)_{t\in\mathbb{N}^*}$ a sequence of i.i.d. $\mathbb{P}_X$-distributed random variables and let $(\delta_t)_{t\in\mathbb{N}^*}$ be a sequence of positive steps satisfying

$$\sum_t \delta_t = +\infty \quad \text{and} \quad \sum_t \delta_t^2 < +\infty. \tag{24}$$

Then, starting from an initial N-tuple x^0 with N pairwise distinct components, set

$$x^t = x^{t-1} - \delta_t \nabla D_N^p(x^{t-1}, \xi^t) \tag{25}$$

(this formula $a.s.$ grants by induction that x^t has pairwise distinct components). Unfortunately, the usual assumptions that ensure the $a.s.$ convergence of the algorithm (see [19]) are not fulfilled by D_N^p (see, $e.g.$ [19] or [30] for an overview on Stochastic approximation). To be more specific, let us stress that $D_N^p(y)$ does not go to infinity as $|y|$ goes to infinity in $(\mathbb{R}^d)^N$ and ∇D_N^p is clearly not Lipschitz continuous on $(\mathbb{R}^d)^N$. Some $a.s.$ convergence results in the Kushner & Clark sense have been obtained in [38] for compactly supported absolutely continuous distributions $\mathbb{P}_X$, mainly in the quadratic case $p = 2$ (however, regular $a.s.$ convergence is established when $d = 1$). In fact the quadratic case is the most commonly implemented for applications and is known as the Competitive Learning vector Quantization (CLVQ) algorithm.

Formula (25) can be developed as follows if one sets $x^t := \{x^{1,t}, \ldots, x^{N,t}\}$,

$$\text{COMPETITIVE PHASE :} \quad \text{select } \; i(t+1) \in \text{argmin}_i |x^{i,t} - \xi^{t+1}| \tag{26}$$

$$\text{LEARNING PHASE :} \quad \left\{ \begin{array}{lcl} x^{i(t+1),t+1} & := & x^{i(t+1),t} - \delta_{t+1} \frac{x^{i(t+1),t} - \xi^{t+1}}{|x^{i(t+1),t} - \xi^{t+1}|} |x^{i(t+1),t} - \xi^{t+1}|^{p-1} \\ *[.5em] x^{i,t+1} & := & x^{i,t}, \; i \neq i(t+1). \end{array} \right. \tag{27}$$

Furthermore, it is established in [38] that, if $X \in L^{p+\varepsilon}$ ($\varepsilon > 0$), on the event $\{x^t \to x^*\}$

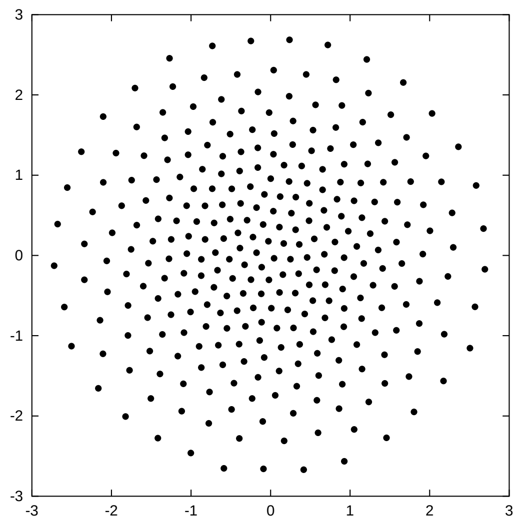
$$D_N^{p,t+1} := D_N^{p,t}(1 - \delta_{t+1}) + \delta_{t+1} \frac{x^{i(t+1),t} - \xi^{t+1}}{|x^{i(t+1),t} - \xi^{t+1}|} |x^{i(t+1),t} - \xi^{t+1}|^{p-1} \xrightarrow{a.s.} \frac{1}{p} D_N^p(x^*) \tag{28}$$

$$\pi^{i,t+1} := \pi^{i,t}(1 - \delta_{t+1}) + \delta_{t+1} \mathbf{1}_{\{i = i(t+1)\}} \xrightarrow{a.s.} \mathbb{P}_X(C_i(x^*)), \; 1 \le i \le N. \tag{29}$$

These "companion" – hence costless – procedures yield the parameters (weights of the Voronoi cells, L^p-mean quantization error of x^*) necessary to exploit the N-tuple x^* for numerical purpose. Note that this holds whatever the limiting N-tuple x^* is: this means that the procedure is consistent.

Concerning practical implementations of the algorithm, it is to be noticed that, when $p = 2$ at each, step the N-tuple x^{t+1} lives in the convex hull of x^t and ξ^{t+1} which has a stabilizing effect on the procedure. One checks on simulation that the CLVQ algorithms does behave better than its non-quadratic counterparts.

Fig.1: Voronoi tessellation of a 300-tuple
with the lowest quadratic quantization error for the Normal distribution



3.3.3 *A priori* error bounds

Next theorem provides a general error bound for $V_k - \widehat{v}_k(\widehat{X}_k)$ as a function of the quantization errors $\|X_k - \widehat{X}_k\|_p$ (for any Voronoi quantizers $\widehat{X}_k$).

Theorem 4 *Assume that the coefficients b and c of the diffusion (7) are Lipschitz continuous as well as the obstacle function h. Let $(X_k)_{0\le k\le n}$ denote either the sampled diffusion $(S_{t_k})_{0\le k\le n}$ or the Euler scheme $(\overline{S}_{t_k})_{0\le k\le n}$. Let $(V_k)_{0\le k\le n}$ denote the Snell envelope of the obstacle $(h(t_k, X_k))_{0\le k\le n}$ and let $\widehat{X}_k$ denote a quantizer of X_k as defined by (15). Then, for every $p\in[1,+\infty)$, there exists some real constant $K_{b,\sigma,T,p} > 0$ such that*

$$\forall\, n\in\mathbb{N}^*,\ \forall\, k\in\{0,\ldots,n\},\quad \|V_k - \widehat{v}_k(\widehat{X}_k)\|_p \le K_{b,\sigma,T,p}\sum_{\ell=k}^{n}\|X_\ell - \widehat{X}_\ell\|_p \qquad (30)$$

where $(\widehat{v}_k(\widehat{X}_k))_{0\le k\le n}$ denotes the pseudo-Snell envelope of $h(t_k, X_k)$ as defined by (17).

The key point in Theorem 4 is that $K_{b,\sigma,T,p}$ does not depend on n One gets rid of n since the Lipschitz coefficient $K^{(n)}$ of the Euler schemes (or the sampled diffusion) satisfies $\limsup_n (K^{(n)})^n < +\infty$ (see [4] for details). Combining the bounds obtained in Theorems 1 and 4 will naturally lead to an as small as possible global error. To this end, we need to settle the parameters of the discretization: the number n of discretization times (or "time layers"), the global number N of elementary quantizers, and the size N_k of the quantizer $\widehat{X}_k$ of the k^{th} layer (time 0 is excluded since it is perfectly quantized by the single quantizer

$X_0 := s_0$). Next theorem provides a theoretic answer which minimize the theoretical error bounds. We proceed as follows: n being fixed, the space discretization error is ruled by the right hand side of (30), namely $\sum_{k=1}^{n} \|X_k - \widehat{X}_k\|_p$. So we need to minimize this quantity for a fixed n and then to make a balance between the resulting space discretization error and the error induced by the time discretization given by Theorem 1.

To be rigorous, we need to state the theorem one further assumption, called *domination* assumption, about the distributions of the X_k's. It is based on the L^p-mean quantization error $\|X_k - \widehat{X}_k\|_p$. Namely, there exists a random variable $R \in L^{p+\eta}(\eta > 0)$ and a sequence $(\varphi_k)_{0\le k\le n}$ such that

$$\forall\, k \in \{0,\dots,n\},\ \forall\, N \in \mathbb{N}^*, \qquad \underline{D}_N^{X_k,p} \le \varphi_k^p\, \underline{D}_N^{R,p}. \tag{31}$$

The point is that the distribution of R may depend on p but *not on N or k*.

It is shown in [4] that uniformly elliptic diffusions with smooth and bounded coefficients and their related Euler schemes satisfy the domination property (31) with $\varphi_k := c\sqrt{t_k}$. It is shown in the Appendix that, if $q \ge d$, σ is smooth and uniformly elliptic, then the extended Black & Scholes model (1) is dominated as well.

Theorem 5 *Assume that all the assumptions of Theorem 4 hold and that the $(X_k)_{0\le k\le n}$ is dominated in the sense of (31). Let n, $N \in \mathbb{N}^*$ and set the size N_k of the quantizer of layer k by*

$$N_k := \left\lceil \frac{t_k^{\frac{d}{d+1}}}{t_1^{\frac{d}{d+1}} + \cdots + t_k^{\frac{d}{d+1}} + \cdots + t_n^{\frac{d}{d+1}}} N \right\rceil,\ 1 \le k \le n \tag{32}$$

(then, $N \le N_1 + \cdots + N_n \le N + n$).

(a) Let $n := \left\lceil \left(\frac{2d}{d+1} C_p(s_0)\right)^{\frac{2}{3d+2}} N^{-\frac{1}{2d+1}} \right\rceil$. Assume furthermore that every quantization $\widehat{X}_k$ is L^p-optimal. Then,

$$\|\mathcal{V}_{t_k} - \widehat{V}_{t_k}^n\|_p \le C'_p(1+|s_0|)e^{C'_pT} \frac{n}{(N/n)^{\frac{1}{d}}} = O\left(N^{-\frac{1}{3d+2}}\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

where $\widehat{V}_k^n$ is the pseudo-Snell envelope of the Euler scheme.

(b) If the obstacle h is semi-convex and if $X_k := S_{t_k}$, then set $n := \left\lceil \left(\frac{2d}{d+1} C_p(s_0)\right)^{\frac{2}{3d+2}} N^{-\frac{1}{2d+1}} \right\rceil$. Assume furthermore that every quantization $\widehat{S}_{t_k}$ is L^p-optimal. Then,

$$\|\mathcal{V}_{t_k} - \widehat{V}_{t_k}^n\|_p \le C'_p(1+|s_0|)e^{C'_pT} \frac{n}{(N/n)^{\frac{1}{d}}} = O\left(N^{-\frac{1}{2d+1}}\right) = O\left(\frac{1}{n}\right).$$

Remark: The real constant $C_p(s_0)$ satisfies $C_p(s_0) \le C_p(1+|s_0|)$ and C_p could be calibrated in using a 1-dimension model.

4 Hedging

Tackling the question of hedging American options needs to go deeper in financial modelling, at least from a heuristic point of view. So, we will shortly recall the principles that govern the pricing and hedging of American options to justify our approach. First, we come back to the original diffusion model (3) which drives the asset price process (S_t) (with $r = 0$). Furthermore, we will assume when necessary that $(q \geq d)$ and

$$\forall\, x \in \mathbb{R}^d,\ \sigma\sigma^*(x) \geq \varepsilon_0 I_d \tag{33}$$

so that
$$\varepsilon_0 \mathrm{Diag}(x_i^2)\, I_d \leq cc^*(x) \leq \|\sigma\sigma^*\|_\infty |x|^2 I_d.$$

4.1 Hedging continuous time American options

First we need to come back shortly to classical European option pricing theory. Let h_T be a European contingent claim that is a nonnegative $\mathcal{F}_T$-measurable $\mathbb{R}^d$-valued random vector. Assume for the sake of simplicity that it lies in $L^2(\mathbb{P}, \mathcal{F}_T)$. The representation theorem for Brownian martingale shows (see [39]) that

$$h_T = \mathbb{E}(h_T) + \int_0^T H_s.dW_s = \mathbb{E}(h_T) + \int_0^T Z_s.dS_s \tag{34}$$

where H is a $d\mathbb{P}dt$-square integrable $\mathcal{F}$-predictable process and $Z_s := [c(S_s)^*]^{-1} H_s$. Hence $M_t := \mathbb{E}(h_T/\mathcal{F}_t)$ satisfies $M_t = M_0 + \int_0^t Z_s.dS_s$.

An analogy with discrete time model shows that the integral $\displaystyle\int_t^T Z_s.dS_s$ represents the (algebraic) gain from time t up to time T provided by the strategy $(Z_s^i)_{1\leq i\leq d}$ (at every time $s \in [t, T]$ the portfolio contains exactly Z_s^i units of asset i). So, at time T, the value of the portfolio invested in risky assets $S^1, \ldots, S^d$ is exactly h_T monetary units: put some way round, the portfolio Z_t *replicates* the payoff h_T; so it is natural to define the (theoretical) *premium* as

$$\mathrm{Premium}_t := \mathbb{E}(h_T/\mathcal{F}_t) = \mathbb{E}(h_T) + \int_0^t Z_s.dS_s. \tag{35}$$

If $h_T := h(T, S_T)$, the Markov property of (S_t) implies that $\mathrm{Premium}_t := p(t, S_t)$. If h is regular enough, then p solves the parabolic P.D.E. $\dfrac{\partial p}{\partial t} + \mathcal{L}_{r,\sigma} p = 0, p(T, .) := h(T, .)$ and a straightforward application of Ito formula shows that $Z_t = \nabla_x p(t, S_t)$.

Let us come back to American option pricing. If one defines the premium process $(\mathcal{V}_t)_{t\in[0,T]}$ of an American option by the $\mathbb{P}$-Snell envelope of its payoff process, then this premium process is a supermartingale that can be decomposed as the difference of a martingale M_t and a nondecreasing path-continuous process K_t *i.e.*, using the representation property of Brownian martingales,

$$\mathcal{V}_t = M_t - K_t = \mathcal{V}_0 + \int_0^t Z_s.dS_s - K_t \qquad (K_0 := 0).$$

So, if a trader replicates the European option related to the (unknown) European payoff M_T using Z_t, he is in position to be the counterpart at every time t of the owner of the option in case of exercise since

$$M_t = \mathcal{V}_t + K_t \geq \mathcal{V}_t \geq h_t.$$

In case of an optimal exercise of his counterpart he will actually have exactly the payoff at time t since all optimal exercise times occur before the process K_t leaves 0.

If the variational inequality (6) admits a regular enough solution $\nu(t,x)$, then $Z_t = \nabla_x \nu(t, S_t)$. In most deterministic numerical methods, the approximation of such a derivative is usually less accurate than that of the function ν itself. So, it is hopeless to implement such methods to this end as soon as the dimension d is greater than 3.

4.2 Hedging Bermuda options

Let $(V^n_{t_k})_{0\leq k\leq n}$ denote the theoretical premium process of the Bermuda option related to $(h(t_k, S_{t_k}))_{0\leq k\leq n}$. It is a $(\mathcal{F}_{t_k})_{0\leq k\leq n}$-supermartingale defined as a Snell envelope by

$$V^n_{t_k} := \text{ess sup}\left\{\mathbb{E}_{t_k}\left(h(\tau, S_\tau)\right),\ \tau \in \Theta^n_k\right\}$$

where Θ^n_k denotes the set of $\{t_k, \ldots, t_n\}$-valued $\underline{\mathcal{F}}$-stopping times.

Then, the $\mathcal{F}_{t_k}$-Doob decomposition of $(V^n_{t_k})$ as a the $(\mathcal{F}_{t_k})$-supermartingale yield:

$$V^n_{t_k} = M^n_k - A^n_k,$$

where $(M^n_{t_k})$ is a $\mathcal{F}_{t_k}$-L^2-martingale and $(A^n_{t_k})$ is a non-decreasing integrable $\mathcal{F}_{t_k}$-predictable process ($A^n_0 := 0$). In fact, the increment of A^n_k can easily be specified since

$$\Delta A^n_k := A^n_k - A^n_{k-1} = V^n_{t_{k-1}} - \mathbb{E}_{t_{k-1}} V^n_{t_k} = \left(h(t_{k-1}, S_{t_{k-1}}) - \mathbb{E}_{t_{k-1}} V^n_{t_k}\right)_+. \tag{36}$$

The representation theorem applied on each time interval $[t_k, t_{k+1}]$, $k = 0, \ldots, n$ then yields a $\mathcal{F}$-progressively measurable process $(Z^n_s)_{s\in[0,T]}$ satisfying

$$M^n_k := \int_0^{t_k} Z^n_s . dS_s, \quad 0 \leq k \leq n, \quad \text{with} \quad \mathbb{E}\int_0^T |c^*(S_s) Z^n_s|^2 ds < +\infty \tag{37}$$

(keep in mind that $< \int_0^{t_k} U_s.dS_s >_t = \int_0^{t_k} |c^*(S_s)U_s|^2 ds$).

Now, in such a setting, continuous time hedging of a Bermuda option is irrealistic since the approximation of an American by a Bermuda option is directly motivated by discrete time hedging (at times t_k). So, it seems natural to look for what a trader can do best when hedging only at times t_k. This leads to consider the closed subspace $\mathcal{P}_n$ of $L^2(c^*(S_.)d\mathbb{P}.dt)$ defined by

$$\mathcal{P}_n = \left\{ (\zeta_s)_{s\in[0,T]},\ \zeta_s := \zeta_{t_k},\ s \in [t_k, t_{k+1}),\ \zeta_{t_k}\ \mathcal{F}_{t_k}\text{-measurable},\ \mathbb{E}\int_0^T |c^*(S_s)\zeta_s|^2 ds < +\infty \right\}. \tag{38}$$

and the induced orthogonal projection proj_n onto $\mathcal{P}_n$ (for notational simplicity a process $\zeta\in\mathcal{P}_n$ will be often referred as $(\zeta_{t_k})_{0\le k\le n}$). In particular, for every $U\in L^2(c^*(S_.)d\mathbb{P}.dt)$

$$\|c^*(S_.)\mathrm{proj}_n(U).\|_{L^2(d\mathbb{P}.dt)}\le \|c^*(S_.)U.\|_{L^2(d\mathbb{P}.dt)}.$$

Doing so, we follow classical ideas introduced by by Follmer & Sonderman ([21]) for hedging purpose in incomplete markets (see also [10]). One checks that $\mathcal{P}_n$ is isometric with the set of square integrable stochastic integrals with respect to $(S_{t_k})_{0\le k\le n}$, namely

$$\left\{\sum_{k=1}^n \zeta_{t_k}.\Delta S_{t_{k+1}},\ (\zeta_{t_k})_{0\le k\le n}\in\mathcal{P}_n\right\}.$$

Computing $\mathrm{proj}_n(Z^n_.)$ amounts to minimizing $\mathbb{E}\left(\sum_{k=1}^n\int_{t_k}^{t_{k+1}}|c^*(S_s)(Z^n_s-\zeta_{t_k})|^2ds\right)$ over $(\zeta_k)_{0\le k\le n}\in\mathcal{P}_n$. Setting $\zeta^n_{t_k}:=\mathrm{proj}_n(Z^n_.)$ and standard computations yield

$$\begin{aligned}\zeta^n_{t_k} &:= \left(\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}cc^*(S_s)\,ds\right)^{-1}\mathbb{E}_{t_k}\left(\int_{t_k}^{t_{k+1}}cc^*(S_s)Z^n_sds\right)\\ &= \left(\mathbb{E}_{t_k}\Delta S_{t_{k+1}}(\Delta S_{t_{k+1}})^*\right)^{-1}\mathbb{E}_{t_k}\left(\Delta M^n_{k+1}\Delta S_{t_{k+1}}\right) \qquad (39)\\ &= \left(\mathbb{E}_{t_k}\Delta S_{t_{k+1}}(\Delta S_{t_{k+1}})^*\right)^{-1}\mathbb{E}_{t_k}\left(\Delta V^n_{t_{k+1}}\Delta S_{t_{k+1}}\right). \qquad (40)\end{aligned}$$

The last equality follows from the fact that A^n_{k-1} is $\mathcal{F}_{t_{k-1}}$-measurable and from the martingale property of (S_{t_k}). The increment

$$\Delta R^n_{t_{k+1}}:=\int_{t_k}^{t_{k+1}}(Z^n_s-\zeta^n_{t_k}).dS_s=\Delta M^n_{k+1}-\zeta^n_{t_k}.\Delta S_{t_{k+1}} \qquad (41)$$

represents the *hedging default* induced by using $\zeta^n_{t_k}$ instead of $Z^n_.$. The sequence $(\Delta R^n_{t_k})_{0\le k\le n}$ is a $\mathcal{F}_{t_k}$-martingale increment process, singular with respect to $(S_{t_k})_{0\le k\le n}$ since $\mathbb{E}_{t_k}(\Delta R_{t_{k+1}}\Delta S_{t_{k+1}})=0$. It is possible to define the *local residual risk* by

$$\mathbb{E}_{t_k}|\Delta R^n_{t_{k+1}}|^2=\mathbb{E}_{t_k}\left(\int_{t_k}^{t_{k+1}}|c^*(S_s)(Z^n_s-\zeta_{t_k})|^2ds\right),\ k\in\{0,\dots,n-1\}. \qquad (42)$$

A little algebra yields the following, more appropriate for quantization purpose:

$$\mathbb{E}_{t_k}|\Delta R^n_{t_{k+1}}|^2=\mathbb{E}_{t_k}|\Delta V^n_{t_{k+1}}-\mathbb{E}_{t_k}\Delta V^n_{t_{k+1}}|^2-(\mathbb{E}_{t_k}\Delta S_{t_{k+1}}\Delta S^*_{t_{k+1}})^{-1}\Big(\mathbb{E}_{t_k}\Delta V^n_{t_{k+1}}\Delta S_{t_{k+1}}\Big)^2. \qquad (43)$$

Formulae (40) or (42), based on S_{t_k} and $V^n_{t_k}$ have natural approximations by quantization. On the other hand, (39) and (42) are more appropriate to produce some *a priori* error bounds (when simulation of the diffusion is possible).

4.3 Hedging Bermuda option on the Euler scheme

When the diffusion cannot be easily simulated, we substitute the (continuous time) Euler scheme defined by

$$\forall\, t \in [t_k, t_{k+1}), \quad \overline{S}_t = \overline{S}_{t_k} + c(\overline{S}_{t_k})(W_t - W_{t_k}), \quad \overline{S}_0 := s_0 > 0.$$

This process is $\mathbb{P}$-$a.s.$ defined since it is $a.s.$ nonzero (but it may become negative contrarily to the original diffusion). Then, mimicking the above subsection, leads to define some processes $\overline{Z}^n$, $\overline{M}^n$ and $\overline{A}^n$ by

$$\begin{aligned}
\overline{V}^n_{t_k} &:= \overline{M}^n_k - \overline{A}^n_k \qquad \text{(Doob decomposition)}\\
\overline{M}^n_k &:= \int_0^{t_k} \overline{Z}^n_s c(\overline{S}_{\underline{s}})\, d\overline{W}_s = \int_0^{t_k} \overline{Z}^n_s d\overline{S}_s \qquad (\text{with } \underline{s} = t_i,\ s \in [t_i, t_{i+1}))\\
\Delta\overline{A}^n_k &:= \overline{A}^n_k - \overline{A}^n_{k-1} = \overline{V}^n_{t_{k-1}} - \mathbb{E}_{t_{k-1}} \overline{V}^n_{t_k} = \Big(h(t_{k-1}, \overline{S}_{t_{k-1}}) - \mathbb{E}_{t_{k-1}} \overline{V}^n_{t_k}\Big)_+ .
\end{aligned}$$

and $\overline{A}^n_0 := 0$. The (simpler) formulae for the hedging process hold

$$\overline{\zeta}^n_{t_k} := (\mathbb{E}_{t_k} \Delta\overline{S}_{t_{k+1}} \Delta\overline{S}^*_{t_{k+1}})^{-1} \mathbb{E}_{t_k} (\Delta\overline{V}^n_{t_{k+1}} \Delta\overline{S}_{t_{k+1}}) = \frac{1}{\Delta t_{k+1}} \mathbb{E}_{t_k} \int_{t_k}^{t_{k+1}} \overline{Z}^n_s ds. \tag{44}$$

The related hedging default and local residual risk are defined by mimicking (42) and (43):

$$\Delta\overline{R}^n_{t_{k+1}} := \int_{t_k}^{t_{k+1}} (\overline{Z}^n_s - \overline{\zeta}^n_{t_k}).d\overline{S}_s = \Delta M^n_{k+1} - \overline{\zeta}^n_{t_k}.\Delta\overline{S}_{t_{k+1}} \tag{45}$$

$$\mathbb{E}_{t_k} |\Delta\overline{R}^n_{t_{k+1}}|^2 := \mathbb{E}_{t_k} |\Delta\overline{V}^n_{t_{k+1}} - \mathbb{E}_{t_k} \Delta\overline{V}^n_{t_{k+1}}|^2 - (\mathbb{E}_{t_k} \Delta\overline{S}_{t_{k+1}} \Delta\overline{S}^*_{t_{k+1}})^{-1} \Big(\mathbb{E}_{t_k} \Delta\overline{V}^n_{t_{k+1}} \Delta\overline{S}_{t_{k+1}}\Big)^2 \tag{46}$$

4.4 Quantized hedging and local residual risks

The quantized formulae for strategies and residual risks are simply derived from formulae (40) or (44) by replacing S_{t_k} ($\overline{S}_{t_k}$ respectively) by their quantization $\widehat{S}_{t_k}$ ($\widehat{\overline{S}}_{t_k}$ respectively) and $V^n_k := v^n_k(S_{t_k})$ by $\widehat{V}^n_k := \widehat{v}^n_k(\widehat{S}_{t_k})$ ($\widehat{V}^n_k := \widehat{v}^n_k(\widehat{\overline{S}}_{t_k})$ respectively). It follows from section 3 that $V^n_{t_k} := v_k(S_{t_k})$ is approximated by $\widehat{v}^n_k(\widehat{S}_{t_k})$. So, one sets (for the diffusion)

$$\widehat{\zeta}^n_k := \frac{n}{T}\Big(cc^*(\widehat{S}_{t_k})\Big)^{-1} \widehat{\mathbb{E}}_k \Big((\widehat{v}^n_{k+1}(\widehat{S}_{t_{k+1}}) - \widehat{v}^n_k(\widehat{S}_{t_k}))(\widehat{S}_{t_{k+1}} - \widehat{S}_{t_k})\Big). \tag{47}$$

$$|\Delta\widehat{R}^n_{t_{k+1}}|^2 := \mathbb{E}_{t_k} |\Delta\widehat{V}^n_{t_{k+1}} - \mathbb{E}_{t_k} \Delta\widehat{V}^n_{t_{k+1}}|^2 - (\mathbb{E}_{t_k} \Delta\widehat{S}_{t_{k+1}} \Delta\widehat{S}^*_{t_{k+1}})^{-1} \Big(\mathbb{E}_{t_k} \Delta\widehat{V}^n_{t_{k+1}} \Delta\widehat{S}_{t_{k+1}}\Big)^2 \tag{48}$$

One derives their counterparts $\widehat{\widehat{\zeta}}^n_k$, $|\Delta\widehat{\widehat{R}}^n_{t_{k+1}}|^2$ for the Euler scheme by analogy. The point to be noticed is that computing $\widehat{\zeta}^n_{t_k}$ or $\widehat{\widehat{\zeta}}^n_k$ at a given point x^k_i of the k^{th} layer requires to *invert only one matrix* which does not cost much.

5 Convergence of the hedging strategies and rates

This section is devoted to the evaluation of the different errors (quantization, residual risks) induced by space and time discretization.

5.1 From Bermuda to America

First, one extends the definition of V^n_t at any time $t \in [0,T]$ by setting

$$V^n_t := V^n_{t_k} + \int_{t_k}^t Z^n_s dS_s = V^n_{t_{k+1}} - \int_t^{t_{k+1}} Z^n_s dS_s + \Delta A^n_{k+1},\ t \in [t_k, t_{k+1}). \tag{49}$$

This definition implies that, for every $k \in \{0,\dots,n\}$, the left-limit of V^n satisfies

$$V^n_{t_k-} = V^n_{t_k} + \Delta A^n_{k+1}. \tag{50}$$

Proposition 2 *Assume that the payoff process $h_t = h(t,S_t)$ where h is a semi-convex function. Assume that the diffusion coefficient c is Lipschitz continuous.*
(a) For every $k \in \{0,\dots,n\}$, $V^n_{t_k} \le \mathcal{V}_{t_k}$ and for every $t \in (t_k, t_{k+1})$, $\quad (V^n_t - \mathcal{V}_t)_+ \le \Delta A^n_{k+1}$.

Furthermore $\mathbb{P}$-a.s., for every $t \in [0,T]$, $\begin{cases} |V^n_t - \mathcal{V}_t| & \le C_{h,c}\frac{T}{n}(1 + \mathbb{E}_t(\max_{s\ge t}|S_s|^2)), \\ |V^n_t - \overline{V}^n_t| & \le [h]_{Lip}\mathbb{E}_t(\max_{t_k\ge t}|S_{t_k} - \overline{S}_{t_k}|). \end{cases}$

(b) The following bound holds for the hedging strategies (in the "cc metric")*

$$\mathbb{E}\left(\int_0^T |c^*(S_s)(Z_s - Z^n_s)|^2 ds\right) + \mathbb{E}\left(\int_0^T |c^*(S_s)Z^n_s - c^*(\overline{S}_s)\overline{Z}^n_s)|^2 ds\right) \le C_{h,c}\frac{T}{n}. \tag{51}$$

Proof: (a) The inequality between V^n and $\mathcal{V}$ at times t_k is obvious since $\mathcal{V}_t$ is defined as a supremum over a larger set of stopping times than $V^n_{t_k}$. Then, using the supermartingale property of $\mathcal{V}$, equality (49) and Jensen inequality yields

$$(V^n_t - \mathcal{V}_t)_+ \le (\mathbb{E}_t(V^n_{t_{k+1}}) + \Delta A^n_{k+1} - \mathbb{E}_t(\mathcal{V}_{t_{k+1}}))_+ \le \mathbb{E}_t((V^n_{t_{k+1}} - \mathcal{V}_{t_{k+1}} + \Delta A^n_{k+1})_+) \le \Delta A^n_{k+1}.$$

Now, using the expression (36) for ΔA^n_{k+1} and $V^n_{t_k} \ge h(t_{k+1}, S_{t_{k+1}})$ imply

$$\Delta A^n_{k+1} = (h(t_k, S_{t_k}) - \mathbb{E}_{t_k} V^n_{t_{k+1}})_+ \le (h(t_k, S_{t_k}) - \mathbb{E}_{t_k} h(t_{k+1}, S_{t_{k+1}}))_+$$

We need at this stage to use the regularity of h (semi-convex Lipschitz continuous)

$$\begin{aligned} h(t_k, S_{t_k}) - h(t_{k+1}, S_{t_{k+1}}) &= h(t_k, S_{t_{k+1}}) - h(t_{k+1}, S_{t_{k+1}}) + h(t_k, S_{t_k}) - h(t_k, S_{t_{k+1}}) \\ &\le [h]_{Lip}\Delta t_{k+1} - \delta_h(t_k, S_{t_k}).(S_{t_{k+1}} - S_{t_k}) + \rho_h\,(S_{t_{k+1}} - S_{t_k})^2. \end{aligned}$$

Hence

$$\begin{aligned} h(t_k, S_{t_k}) - \mathbb{E}_{t_k} h(t_{k+1}, S_{t_{k+1}}) &\le [h]_{Lip}\Delta t_{k+1} + \rho_h\,\mathbb{E}_{t_k}|S_{t_{k+1}} - S_{t_k}|^2 \\ &\le [h]_{Lip}\Delta t_{k+1} + \rho_h \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} \mathrm{Tr}(cc^*)(S_s)\,ds \\ &\le [h]_{Lip}\Delta t_{k+1} + C\rho_h\Delta t_{k+1}\left(1 + \mathbb{E}_t\,(\max_{s\ge t_k}|S_s|^2)\right) \\ &\le C_{c,h}\frac{T}{n}\left(1 + \mathbb{E}_{t_k}(\max_{s\ge t_k}|S_s|^2)\right) \text{ for some constant } C_{h,c} > 0. \end{aligned}$$

Finally, it yields

$$\Delta A^n_{k+1} \le C_{c,h}\frac{T}{n}\left(1 + \mathbb{E}_{t_k}(\max_{s\ge t_k}|S_s|^2)\right). \tag{52}$$

To complete the inequality for $|\mathcal{V}_t - V^n_t|$, one first notice that, if $t \in [t_k, t_{k+1})$

$$V^n_t = V^n_{t_{k+1}} - \int_t^{t_{k+1}} Z^n_s dS_s + \Delta A^n_{k+1} \le h(t_{k+1}, S_{t_{k+1}}) - \int_t^{t_{k+1}} Z^n_s\,dS_s \tag{53}$$

so that $\quad V^n_t = \mathbb{E}_t(V^n_t) \ge \mathbb{E}_t(h(t_{k+1}, S_{t_{k+1}})) = h(t, S_t) + \mathbb{E}_t(h(t_{k+1}, S_{t_{k+1}}) - h(t, S_t)).$

Using again he semi-convexity property of h at (t, S_t) finally yields that

$$V^n_t + C_{c,h}\frac{T}{n}\left(1 + \mathbb{E}_t(\max_{s\ge t}|S_s|^2)\right) \ge h(t, S_t).$$

As it is a supermartingale as well, it necessarily satisfies

$$\mathbb{P}\text{-}a.s. \quad V^n_t + C_{c,h}\frac{T}{n}\left(1 + \mathbb{E}_t(\max_{s\ge t}|S_s|^2)\right) \ge \mathrm{Snell}(h(t, S_t)) = \mathcal{V}_t$$

which yields the expected result. The second inequality is obvious once noticed

$$|V^n_t - \overline{V}^n_t| \le \max_{t_k\ge t}|h(t_k, S_{t_k}) - h(t_k, \overline{S}_{t_k})| \le [h]_{Lip}\max_{t_k\ge t}|S_{t_k} - \overline{S}_{t_k}|.$$

(b) One considers the *càdlàg* semi-martingale $\mathcal{V}_t - V^n_t = \mathcal{V}_0 - V^n_0 + \int_0^t (Z_s - Z^n_s).dS_s - (K_t - A^n_{\underline{t}})$ where $\underline{t} := k$ on $[t_k, t_{k+1})$. It follows from Itô formula for jump processes that

$$\begin{aligned} \int_0^T |c^*(S_s)(Z_s - Z^n_s)|^2 ds &+ \sum_{t_k\le T}(\Delta A^n_{t_k})^2 + (\mathcal{V}_t - V^n_t)^2 \\ &= -2\int_0^T (\mathcal{V}_s - V^n_{s-})(Z_s - Z^n_s).dS_s + 2\int_t^T (\mathcal{V}_s - V^n_{s-})d(K_s - A^n_{\underline{t}}). \end{aligned}$$

Now $$\int_0^T (\mathcal{V}_s - V^n_{s-})d(K_s - A^n_{\underline{s}}) = \int_0^T (\mathcal{V}_s - V^n_{s-})dK_s + \int_t^T (V^n_{s-} - \mathcal{V}_s)dA^n_{\underline{s}}$$
$$\leq \int_0^T (\mathcal{V}_s - V^n_s)dK_s + \sum_{t_k\leq T}(\Delta A^n_k)^2$$

since $V^n_{t_k-} = V^n_{t_k} + \Delta A^n_k \leq \mathcal{V}_{t_k} + \Delta A^n_k$. This yields, using the inequality obtained in (a) and (52),

$$\int_0^T (\mathcal{V}_s - V^n_{s-})d(K_s - A^n_{\underline{s}}) \leq C_{h,c}\frac{T}{n}\int_0^T (1 + \mathbb{E}_s \max_{u\geq s}|S_u|^2)dK_s + A^n_n \max_{t<t_k\leq T}\Delta A^n_k$$
$$\leq C_{h,c}\frac{T}{n}\left(K_T\left(1 + \sup_{s\in[0,T]}(\mathbb{E}_s \max_{u\geq s}|S_u|^2)\right) + \left(1 + \sup_{s\in[0,T]}(\mathbb{E}_s \max_{u\geq s}|S_u|^2)\right)^2\right).$$

One checks that $\int_0^t (\mathcal{V}_s - V^n_s)(Z_s - Z^n_s)dS_s$ is a true martingale so that

$$\mathbb{E}\left(\int_0^T |c^*(S_s)(Z_s - Z^n_s)|^2 ds\right) \leq C_{h,c}\frac{T}{n}\,(\|K_T\|_2 + 1)\,(1 + \|\max_{s\in[0,T]}|S_s|^2\|_2).$$

Now $K_T \in L^2$ since $0 \leq K_T \leq \mathcal{V}_0 + \int_0^T Z_s dS_s$ which yields the expected result.

The inequality involving the Euler scheme is obtained following the same approach using now $V^n - \overline{V}^n$.

$$\mathbb{E}\int_0^T |c^*(S_s)Z^n_s - c^*(\overline{S}_s)\overline{Z}^n_s|^2 ds \leq 2\,\mathbb{E}\int_0^T (V^n_s - \overline{V}^n_s)d(K^n_s - \overline{K}^n_s) + \mathbb{E}(h(T,S_T) - h(T,\overline{S}_T))^2$$
$$\leq 2[h]_{Lip}\mathbb{E}\int_0^T \mathbb{E}_s\left(\max_{t_k\geq s}|S_{t_k} - \overline{S}_{t_k}|\right)d(K^n_s + \overline{K}^n_s) + [h]^2_{Lip}\|S_T - \overline{S}_T\|^2_2$$
$$\leq C\,\mathbb{E}\left(\sup_{t\in[0,T]}\mathbb{E}_t\left(\max_{t_k\geq t}|S_{t_k} - \overline{S}_{t_k}|\right)(K^n_T + \overline{K}^n_T)\right) + C\|S_T - \overline{S}_T\|^2_2$$
$$\leq C\left\|\sup_{t\in[0,T]}\mathbb{E}_t\max_{t_k\geq t}|S_{t_k} - \overline{S}_{t_k}|\right\|_2\left(\|K^n_T\|_2 + \|\overline{K}^n_T\|_2\right) + C\|S_T - \overline{S}_T\|^2_2$$
$$\leq C_{h,c}\frac{T}{n}\left(\|K^n_T\|_2 + \|\overline{K}^n_T\|_2 + 1\right). \tag{54}$$

Now $\|K^n_T\|_2 \leq \|V^n_0\|_2 + \|\int_0^T (Z_s - Z^n_s)dS_s\|_2 \leq C_1(1 + \|\sup_{s\in[0,T]}|S_s|\,\|_2) + O(1/n)$, hence $\sup_n \|K^n_T\|_2 < +\infty$. Concerning $\overline{K}^n_T$ one has

$$\|K^n_T - \overline{K}^n_T\|_2 \leq \|V^n_0\|_2 + \|\overline{V}^n_0\|_2 + \|\int_0^T Z^n_s dS_s - \int_0^T \overline{Z}^n_s d\overline{S}_s)\|_2 \leq C + O(1/\sqrt{n}) \quad \text{by}(54)$$

so that $\sup_n \|\overline{K}^n_T\|_2 < +\infty$. Plugging this back in (54) completes the proof. $\diamond$

We are now in position to get a first result about the control of residual risks induced by the use of discrete time hedging strategies. It shows that this control is essentially ruled by the path-regularity of the process Z.

Theorem 6 *If h is semi-convex and h and c are Lipschitz continuous, then*

$$\|c^*(S_.)(Z_. - \zeta^n_.)\|_{L^2(d\mathbb{P}\otimes dt)} \leq \|c^*(S_.)(Z_. - \zeta_.)\|_{L^2(d\mathbb{P}\otimes dt)} + \frac{C}{\sqrt{n}} \quad \textit{where } \zeta := \mathrm{proj}_n(Z) \quad (55)$$

is the projection of Z on $\mathcal{P}_n$. Furthermore $\|c^*(S_.)(Z_. - \zeta_.)\|_{L^2(d\mathbb{P}\otimes dt)}$ goes to 0 as n goes to 0. So, this term which depends on the path-regularity of Z_s, rules the rate of convergence.

Proof: Minkowski inequality yields

$$\|c^*(S_.)(Z_s - \zeta^n_.)\|_{L^2(d\mathbb{P}\otimes dt)} \leq \|c^*(S_.)(Z_. - \zeta_.)\|_{L^2(d\mathbb{P}\otimes dt)} + \|c^*(S_.)(\zeta_. - \zeta^n_.)\|_{L^2(d\mathbb{P}\otimes dt)}.$$

Now $\zeta_. - \zeta^n_. = \mathrm{proj}_n(Z_. - Z^n_.)$ so that by Inequality (51) in Proposition 2(b),

$$\|c^*(S_.)(\zeta_. - \zeta^n_.)\|_{L^2(d\mathbb{P}\otimes dt)} \leq \|c^*(S_.)(Z_. - Z^n_.)\|_{L^2(d\mathbb{P}\otimes dt)} \leq \frac{C}{\sqrt{n}}.$$

Now, let F be a bounded adapted continuous-path process. Set $\Phi_s := \frac{n}{T}\int_{t_k}^{t_{k+1}} F_u du$, $s \in [t_k, t_{k+1})$. Using the properties of proj_n, one gets

$$\begin{aligned}\|c^*(S_.)(Z_. - \zeta_.)\|_{L^2(d\mathbb{P}\otimes dt)} &\leq 2\,\|c^*(S_.)(Z_. - F_.)\|_{L^2(d\mathbb{P}\otimes dt)} + \|c^*(S_.)(F_. - \mathrm{proj}_n(F)_.)\|_{L^2(d\mathbb{P}\otimes dt)}\\ &\leq 2\,\|c^*(S_.)(Z_. - F_.)\|_{L^2(d\mathbb{P}\otimes dt)} + \|c^*(S_.)(F_. - \Phi_.)\|_{L^2(d\mathbb{P}\otimes dt)}\\ &\leq 2\,\|c^*(S_.)(Z_. - F_.)\|_{L^2(d\mathbb{P}\otimes dt)} + \left\|\int_0^T \|c(S_s)\|^2 ds (w(F, \frac{T}{n}) \wedge 2\|F\|_\infty)^2\right\|_{L^2(\mathbb{P})}\end{aligned}$$

where $w(F,\delta)$ denotes the uniform continuity modulus of F. One concludes using that the space $L^\infty(c^*(S_t)d\mathbb{P}dt)$ is everywhere dense in $L^2(c^*(S_t)d\mathbb{P}dt)$. $\diamond$

5.2 Hedging error induced by the (quadratic) quantization

We will focus on the error at time $t_0 = 0$.

Proposition 3 *If σ Lipschitz continuous, bounded and uniformly elliptic and if h is semi-convex and Lipschitz continuous, then*

$$|\zeta^n_0 - \widehat{\zeta}^n_0| \leq C(1 + |s_0|)|\frac{n^{\frac{3}{2}}}{(N/n)^{\frac{1}{d}}}.$$

Proof: The hedging vectors ζ_0^n and $\hat{\zeta}_0^n$ satisfy respectively

$$(\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*))\zeta_0^n = \mathbb{E}((V_1^n - V_0^n)\Delta S_{t_1}) \tag{56}$$

$$(\mathbb{E}(\Delta \widehat{S}_{t_1}\widehat{\Delta S}_{t_1}^*))\widehat{\zeta}_0^n = \mathbb{E}\left((\widehat{V}_1^n - \widehat{V}_0^n)\Delta \widehat{S}_{t_1}\right) \tag{57}$$

where $V_1^n = v_1^n(S_{t_1})$ and $V_0^n = v_0^n(s_0)$, etc. The quadratic quantization $\widehat{S}_{t_1}$ of S_{t_1} being optimal and S_0 being deterministic, one has $\mathbb{E}(\Delta S_{t_1}/\Delta\widehat{S}_{t_1}) = \Delta\widehat{S}_{t_1}$. Then a straightforward computation shows that

$$\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*) - \mathbb{E}(\Delta\widehat{S}_{t_1}\Delta\widehat{S}_{t_1}^*) = \mathbb{E}((\Delta S_{t_1} - \Delta\widehat{S}_{t_1})(\Delta S_{t_1} - \Delta\widehat{S}_{t_1})^*)$$

so that
$$\|\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*) - \mathbb{E}(\Delta\widehat{S}_{t_1}\Delta\widehat{S}_{t_1}^*)\| \leq \mathbb{E}\|\Delta S_{t_1} - \Delta\widehat{S}_{t_1}\|_2^2 \leq \left(\frac{C}{N_1^{\frac{1}{d}}}\right)^2.$$

Now
$$\begin{aligned}\left|\mathbb{E}((V_1^n - V_0^n)\Delta S_{t_1}) \right.&- \left.\mathbb{E}\left((\widehat{V}_1^n - \widehat{V}_0^n)\Delta\widehat{S}_{t_1}\right)\right| \\ &\leq \|\Delta\widehat{S}_{t_1}\|_2(\|V_1^n - \widehat{V}_1^n\|_2 + |V_0^n - \widehat{V}_0^n|) + \|V_1^n\|_2\|S_{t_1} - \widehat{S}_{t_1}\|_2 \\ &\leq \|\Delta S_{t_1}\|_2 C(1+|s_0|)\frac{n}{(N/n)^{\frac{1}{d}}} + \frac{C}{N_1^{\frac{1}{d}}} \leq C(1+|s_0|)\frac{\sqrt{n}}{(N/n)^{\frac{1}{d}}}\end{aligned}$$

where we used in the last inequality that $\|\widehat{S}_{t_1}\|_2 \leq \|S_{t_1}\|_2 \leq C\sqrt{\frac{T}{n}}(1+|s_0|)$. One derives from (56) and (57) that

$$\begin{aligned}|\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*)(\zeta_0^n - \widehat{\zeta}_0^n)| &\leq \left|\mathbb{E}((V_1^n - V_0^n)\Delta S_{t_1}) - \mathbb{E}\left((\widehat{V}_1^n - \widehat{V}_0^n)\Delta\widehat{S}_{t_1}\right)\right| \\ &\quad + \|\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*) - \mathbb{E}(\Delta\widehat{S}_{t_1}\Delta\widehat{S}_{t_1}^*)\|\,|\widehat{\zeta}_0| \\ &\leq C(1+|s_0|)\frac{\sqrt{n}}{(N/n)^{\frac{1}{d}}} + \frac{C}{N_1^{\frac{2}{d}}} \leq C(1+|s_0|)\frac{\sqrt{n}}{(N/n)^{\frac{1}{d}}}.\end{aligned}$$

Hence, one obtains the following result by inverting the covariance matrix since

$$|\zeta_0^n - \widehat{\zeta}_0^n| \leq \|(\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*))^{-1}\|C(1+|s_0|)\frac{\sqrt{n}}{(N/n)^{\frac{1}{d}}}.$$

Now, it follows from the obvious $cc^*(x) \geq \varepsilon_0\,\mathrm{Diag}(x_i^2)$ that

$$\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*) \geq \left(\varepsilon_0\int_0^{t_1}\min_{1\leq i\leq d}\mathbb{E}(S_s^i)^2 ds\right)I_d \geq \left(\varepsilon_0\int_0^{t_1}\min_{1\leq i\leq d}(\mathbb{E}S_s^i)^2 ds\right)I_d = (\min_i(s_0^i)^2\frac{\varepsilon_0 T}{n})I_d$$

so that $\|\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*))\| \leq \varepsilon_0^{-2}(\min_i s_0^i)^{-2}\dfrac{n}{T}$ which completes the proof. $\diamond$

5.3 Approximation of the strategy: rate of convergence

In this section we evaluate the "global" residual risk on $[0, T-\delta]$ induced by the use of the time discretization of the diffusion with step T/n *i.e.*

$$\mathbb{E}\int_0^{T-\delta} |c^*(S_s)(Z_s-\zeta_s)|^2\,ds \qquad \text{for some } \delta > n^{-1/3}. \tag{58}$$

where Z_t is defined by (34) and $\zeta_t := \mathrm{proj}_n(Z)$ is the projection on the set $\mathcal{P}_n$ of elementary predictable strategies. Our basic assumption is

$$(\mathcal{H}) \;\equiv\; (i)\ \sigma \in C_b^\infty(\mathbb{R}^d), \quad (ii) \quad \sigma\sigma^* \ge \varepsilon_0\, I_d, \quad (iii) \quad \|\nabla c\|_\infty < +\infty.$$

Note that $\nabla c(x) = \partial\sigma(x)x + \sigma(x)$ where $\partial\sigma = (\partial\sigma_1, \ldots, \partial\sigma_d)$ with $\partial\sigma_i$ the Jacobian matrix of the $i^{'th}$ column of the matrix σ. So ∇c is generally not bounded. However, if $\partial\sigma(x) = O(|x|^{-1})$ when $|x|$ goes to infinity, then $\|\nabla c\|_\infty$ is finite.

Theorem 7 *Assume that $(\mathcal{H})$ holds true. Let $\delta_n := n^{-1/3}$. Then there exists some real constants K and θ (depending on the bounds of c and its first two derivatives) such that*

$$\mathbb{E}\int_0^{T-\delta_n} |c^*(S_s)(Z_s-\zeta_s)|^2\,ds \le \frac{K(1+|s_0|)^q}{a}\frac{1}{n^{\frac{1}{6}-\frac{\theta}{\sqrt{\ln n}}}}. \tag{59}$$

Remarks: Roughly speaking the above result says that on every $[0, T-\delta], \delta > 0$, the speed of convergence in L^2 is of order $\frac{1}{n^{1/6}}$. Let us now comment the true statement.

– The fact that we may take $[0, T-\delta_n], \delta_n = n^{-1/3}$ says that asymptotically we control the whole interval $[0, T)$.

– The fact that $\frac{\theta}{\sqrt{\ln n}}$ comes out is due to the non uniform ellipticity of S: this is the cost of truncation around zero. One may look at that some way round: if we had worked with the uniformly elliptic diffusion $X = \ln S$ instead of S, then the obstacle function becomes $h(t, \exp x)$ and has an exponential growth. So we need to truncate as well and the cost is still $\sqrt{\ln n}$.

– In most financial applications the obstacle h is at most Lipschitz continuous (for example $h(t,x) = e^{-rt}(K - e^{rt}x)_+$ for a put of strike K). However, if the obstacle is more regular, namely $h \in C^{1,2}$, then no regularize is needed and the resulting error is of order $1/n^{1/3}$.

Some technical difficulties arise when evaluating the term in (58) directly, so we first reduce the problem to a simpler one. This is done in two steps.

Lemma 1 (Step 1) *Set $H_s := c^*(S_s)Z_s$ and $\eta_s := \frac{n}{T}\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} H_u du,\ s \in [t_k, t_{k+1})$. Then*

$$\mathbb{E}\int_0^T |c^*(S_s)(Z_s-\zeta_s)|^2\,ds \le \frac{C}{n} + \mathbb{E}\int_0^T |H_s-\eta_s|^2\,ds. \tag{60}$$

Proof: We temporarily define $z_s := \frac{1}{t_{k+1}-t_k}\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} Z_r dr, t_k \le s < t_{k+1}$. Note that z is an adapted process which is piecewise constant. Since ζ is the L^2-projection of Z on the subspace of these type of processes, we have

$$\begin{aligned}\mathbb{E}\int_0^T |c^*(S_s)(Z_s-\zeta_s)|^2\,ds &\le \mathbb{E}\int_0^T |c^*(S_s)(Z_s-z_s)|^2\,ds\\ &\le 2\mathbb{E}\int_0^T |H_s-\eta_s|^2\,ds + 2\mathbb{E}\int_0^T |\eta_s - c^*(S_s)z_s|^2\,ds.\end{aligned}$$

It remains to prove that the second term in the right hand of the above inequality is dominated by C/n. We write this term as

$$\mathbb{E}\sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}}\left|\frac{c^*(S_s)}{\Delta t_{k+1}}\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} Z_u du - \frac{1}{\Delta t_{k+1}}\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} c^*(S_u)Z_u du\right|^2 ds \le 2(I+J)$$

$$\begin{aligned}\text{with } I &:= \mathbb{E}\sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}}\left|\frac{c^*(S_s)-c^*(S_{t_k})}{\Delta t_{k+1}}\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} Z_u du\right|^2 ds,\\ J &:= \mathbb{E}\sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}}\left|\frac{1}{\Delta t_{k+1}}\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} (c^*(S_u)-c^*(S_{t_k}))Z_u du\right|^2 ds.\end{aligned}$$

Let us evaluate J. Set $\underline{s} := t_k$ if $s \in [t_k, t_{k+1})$. Conditional Schwartz's inequality implies that

$$\begin{aligned}\left|\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} (c^*(S_u)-c^*(S_{t_k}))Z_u du\right|^2 &\le \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} \|c^*(S_u)-c^*(S_{t_k}\|^2 du\,\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Z_u|^2 du\\ &\le [c^*]^2_{Lip}\int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k}|S_u - S_{t_k}|^2 du\,\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Z_u|^2 du.\end{aligned}$$

Now, classical results about Euler schemes of diffusions with Lipschitz continuous coefficients yield that, for every $u \in [t_k, t_{k+1})$,

$$\mathbb{E}_{t_k}|S_u - S_{t_k}|^2 \le C\Delta t_{k+1}\mathbb{E}_{t_k}\left((1+\sup_{t\in[0,T]}|S_t|)^2\right).$$

for some positive real constant C. Consequently

$$J \le C\frac{T}{n}\mathbb{E}\left(\sum_{k=0}^{n-1}\mathbb{E}_{t_k}((1+\sup_{t\in[0,T]}|S_t|)^2)\,\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Z_u|^2 du\right)$$

$$
\begin{aligned}
&\leq C\frac{T}{n}\mathbb{E}\left((1+\sup_{t\in[0,T]}|S_t|)^2\sum_{k=0}^{n-1}\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|Z_u|^2du\right)\\
&\leq \frac{C}{n}\left\|(1+\sup_{t\in[0,T]}|S_t|)^2\right\|_2\left\|\sum_{k=0}^{n-1}\mathbb{E}_{t_k}\lambda_{k+1}\right\|_2
\end{aligned}
$$

where $\lambda_{k+1}:=\int_{t_k}^{t_{k+1}}|Z_u|^2\,du$ for every $k\in\{1,\dots,n-1\}$. Since the λ_k's are nonnegative,

$$
\begin{aligned}
\sum_{k=0}^{n-1}\lambda_{k+1}^2 &\leq \left(\sum_{k=0}^{n-1}\lambda_{k+1}\right)^2\\
\text{so that}\quad \mathbb{E}\left(\sum_{k=0}^{n-1}\mathbb{E}_{t_k}\lambda_{k+1}\right)^2 &\leq 2\,\mathbb{E}\left(\sum_{k=0}^{n-1}(\lambda_{k+1}-\mathbb{E}_{t_k}\lambda_{k+1})\right)^2+2\,\mathbb{E}\left(\sum_{k=0}^{n-1}\lambda_{k+1}\right)^2\\
&\leq 2\,\mathbb{E}\sum_{k=0}^{n-1}\left(\lambda_{k+1}-\mathbb{E}_{t_k}\lambda_{k+1}\right)^2+2\,\mathbb{E}\left(\sum_{k=0}^{n-1}\lambda_{k+1}\right)^2\\
&\leq 4\,\mathbb{E}\left(\sum_{k=0}^{n-1}\lambda_{k+1}\right)^2=4\,\mathbb{E}\left(\int_0^T|Z_u|^2\,du\right)^2.
\end{aligned}
$$

$$
\text{Finally}\qquad J\leq\frac{C}{n}\|(1+\sup_{t\in[0,T]}|S_t|)^2\|_2\left\|\int_0^T|Z_u|^2\,du\right\|_2.
$$

It is a standard result on diffusions that $\|(1+\sup_{t\in[0,T]}|S_t|)^2\|_2$ is finite. It remains to prove that the term involving Z is finite. Since $cc^*(S_s)\geq\varepsilon_0\,S_sS_s^*\,I_d$, it follows that $|Z_s|^2\leq\varepsilon_0^{-1}\max_{1\leq i\leq d}(S_s^i)^2\,|H_s|^2$ so that, Schwartz Inequality yields

$$
\mathbb{E}\left(\int_0^T|Z_s|^2\,ds\right)^2\leq\left(\mathbb{E}\sup_{0\leq t\leq T}\left|(S_t^i)^{-1}\right|^8\right)^{1/2}\left(\mathbb{E}\left(\int_0^T|H_s|^2\,ds\right)^4\right)^{1/2}\leq C\left(\mathbb{E}\left(\int_0^T|H_s|^2\,ds\right)^4\right)^{1/2}<+\infty.
$$

As $S_t^{-1}:=(1/S_t^i)$ satisfies an equation similar to (1), its supremum has finite polynomial moments. Finally, the last inequality is a standard fact from $RBSDE$ theory (see [20] or [2]). So we have proved that $J\leq C/n$.

One treats I the same way round. $\diamond$

STEP 2 The second type of difficulty which appears is due to the following two facts:

– The obstacle $h(t,S_t)$ is not sufficiently smooth and so we do not have a nice control on the increasing process K.

– The diffusion process S is not uniformly elliptic (because $c(0)=0$) and so we do not have nice evaluations of the density of S_t.

In order to overcome these difficulties we will replace S by an elliptic diffusion denoted $\underline{S}$ and the obstacle h by a smooth obstacle $\underline{h}$. Namely, let $\varepsilon>0$ and $\lambda>0$. We define:

– A function $\underline{h}\in C^{1,2}(\mathbb{R}_+\times\mathbb{R}^d,\mathbb{R})$ using a regularization by convolution of order ε of h. In particular, since h is Lipschitz continuous, we have

$$\|h-\underline{h}\|_\infty \le C\varepsilon \quad \text{and} \quad \|(\partial_t+\mathcal{L}_c)\underline{h}\|_\infty \le C\varepsilon^{-1} \tag{61}$$

where $\mathcal{L}_c$ is the infinitesimal generator of the diffusion S.

– A function $\varphi_\lambda\in C_b^\infty(\mathbb{R}^d,\mathbb{R}^d)$ satisfying

$$\varphi_\lambda(x) := \begin{cases} x & \text{if } |x|\ge e^{-\lambda} \\ \frac{x}{2|x|}e^{-\lambda} & \text{if } |x|\le \frac{1}{2}e^{-\lambda} \end{cases} \quad \text{and} \quad \sup_{\lambda>0}\|D^\alpha\varphi_\lambda\|_\infty \le C_\alpha \ \text{ for every multi-index } \alpha.$$

Then the approximating diffusion coefficient $c_\lambda := c\circ\varphi_\lambda$ satisfies

$$c_\lambda c_\lambda^*(x) \ge \frac{\varepsilon_0}{4}e^{-2\lambda} \quad \text{and} \quad \|D^\alpha c_\lambda\|_\infty \le C_\alpha \quad \text{for every } \alpha. \tag{62}$$

We consider now the solution $\underline{S}^x$ of the *SDE*

$$d\underline{S}_t = \underline{S}_t(rdt + c_\lambda(\underline{S}_t)dW_t), \quad \underline{S}_0 = x.$$

Sometimes $\underline{S}_t^x$ will denote the solution starting at x. The related Snell envelope

$$\underline{Y}_t = \text{ess sup}_{\tau\in\mathcal{T}_t}\mathbb{E}_t\,\underline{h}(\tau,\underline{S}_\tau),$$

satisfies the *RBSDE*

$$\underline{Y}_t = \underline{h}(T,\underline{S}_T) + \underline{K}_T - \underline{K}_t - \int_t^T \underline{H}_s.dW_s$$

for some non decreasing process $\underline{K}$ and some adapted square integrable process $\underline{H}$. We refer to [20] and [2] for this topic. We also consider the approximation

$$\underline{\eta}_s = \frac{n}{T}\,\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} \underline{H}_s ds,\ t_k\le s<t_{k+1}.$$

Then we have the following lemma.

Lemma 2 *Assume that* $(\mathcal{H})$ *holds*

$$\mathbb{E}\int_0^T |H_s-\eta_s|^2\,ds \le C(e^{-C\lambda^2/T}+\varepsilon^2) + \mathbb{E}\int_0^T \left|\underline{H}_s-\underline{\eta}_s\right|^2 ds \tag{63}$$

Proof: We use the stability property of $RBSDE'$s (see [20] and [2]) in order to obtain

$$\begin{aligned}\mathbb{E}\int_0^T |H_s - \underline{H}_s|^2\, ds &\leq C\,\mathbb{E}\sup_{0\leq s\leq T} |h(s,S_s) - \underline{h}(s,\underline{S}_s)|^2 \\ &\leq C(\varepsilon^2 + \mathbb{E}\sup_{0\leq s\leq T} |h(s,S_s) - h(s,\underline{S}_s)|^2).\end{aligned}$$

Define $\tau := \inf\{t\,/\,|S_t| \leq e^{-\lambda}\}$ and we note that

$$\mathbb{P}(\tau \leq T) = \mathbb{P}(\inf_{0\leq s\leq T} |S_s| \leq e^{-\lambda}) = \mathbb{P}(\sup_{0\leq s\leq T} |\log S_s| \geq \lambda) \leq Ce^{-C\lambda^2/T}$$

the last inequality is a standard large deviation fact (see e.g. [27] although it can be easily checked directly on model (1).

Since $S_t = \underline{S}_t$ for $t \leq \tau$, we obtain

$$\begin{aligned}\mathbb{E}\int_0^T |H_s - \underline{H}_s|^2\, ds &\leq C\left(\varepsilon^2 + \mathbb{E}\left(\sup_{0\leq s\leq T}(|h(s,S_s)|^2 + |h(s,\underline{S}_s)|^2)\mathbf{1}_{\{\tau\leq T\}}\right)\right) \\ &\leq C(\varepsilon^2 + e^{-C\lambda^2/T}).\end{aligned}$$

On the other hand since η and $\underline{\eta}$ are the $L^2(dt\,d\mathbb{P})$-projections of H and $\underline{H}$ respectively on the space $\mathcal{P}_n$ of elementary predictable processes, we have

$$\mathbb{E}\int_0^T \left|\eta_s - \underline{\eta}_s\right|^2 ds \leq \mathbb{E}\int_0^T |H_s - \underline{H}_s|^2\, ds \leq C(\varepsilon^2 + e^{-C\lambda^2/T})$$

and the proof is completed. $\diamond$

We need now some analytical facts that we recall here (see [20] and [2]). First of all we have the representation

$$\underline{Y}_t = u(t,\underline{S}_t), \quad \underline{H}_t = (c_\lambda^*\nabla u)(t,\underline{S}_t)$$

where u is the unique (in some sense not important here, see [2]) solution of the PDE

$$(\partial_t + \mathcal{L}_c)u(t,x) + \underline{F}(t,x,u(t,x)) = 0, \quad u(T,x) = \underline{h}(T,x),$$

with

$$\underline{F}(t,x,u(t,x)) = \alpha(t,x)1_{\{u(t,x)=\underline{h}(t,x)\}}\,((\partial_t + \mathcal{L}_c)\underline{h}(t,x))_+$$

where α is a measurable function such that $0 \leq \alpha \leq 1$. Denote

$$F(t,x) = \underline{F}(t,x,u(t,x))$$

and notice (recall (61)) that

$$\sup_{0\leq t\leq T}\sup_{x\in\mathbb{R}^d} |F(t,x)| \leq \frac{C}{\varepsilon}. \tag{64}$$

With this notation u satisfies

$$(\partial_t + \mathcal{L}_c)u(t,x) + F(t,x) = 0, \quad u(T,x) = \underline{h}(T,x),$$

and consequently u satisfies the mild form of the above PDE

$$u(t,x) = \underline{P}_{T-t}\underline{h}_T(x) + \int_t^T \underline{P}_{s-t}F_s(x)ds$$

where $(\underline{P}_t)_{t\geq 0}$ is the semi-group of the diffusion $\underline{S}_t$, that is $\underline{P}_t f(x) = \mathbb{E}f(\underline{S}_t^x)$. This is the equation that will be used in the sequel.

We turn now to the semi-group. It is well known (see [24] or [31]) that under the hypothesis (62), $\underline{P}_t(x,y) = p_t(x,y)dy$ and for every $k \in \mathbb{N}$ and every multi-index $\alpha = (\alpha_1, \ldots, \alpha_m)$ we have

$$\left|\partial_t^k D_x^\alpha p_t(x,y)\right| \leq \frac{K(1+|x|)^q}{at^{k+\frac{m+1}{2}}} e^{2\lambda} \times e^{-K'\frac{|x-y|^2}{t}} \tag{65}$$

where K, K', q depend on α and on C_α from (62) (but not on λ). Let us pint out some immediate consequences of this evaluation in our framework. Since $|\underline{h}_T(y)| \leq C(1+|y|)$,using (65) we obtain

$$\left|\frac{\partial \underline{P}_\delta \underline{h}_T}{\partial x_k}(x)\right| \leq \int_{\mathbb{R}^d} \left|\frac{\partial p_\delta(x,y)}{\partial x_k}\right| \times C(1+|y|)dy \leq \frac{1}{\sqrt{\delta}} \frac{K(1+|x|)^q}{a} e^{2\lambda} \tag{66}$$

$$\left|\frac{\partial^2 \underline{P}_\delta \underline{h}_T}{\partial x_k \partial x_p}(x)\right| \leq \frac{1}{\delta}\frac{K(1+|x|)^q}{a} e^{2\lambda} \tag{67}$$

$$\left|\frac{\partial}{\partial x_i}\underline{P}_{T-t}\underline{h}_T(x) - \frac{\partial}{\partial x_i}\underline{P}_{T-s}\underline{h}_T(y)\right| \leq \frac{K(1+|x|+|y|)^q}{a\delta^{3/2}} e^{2\lambda}(\sqrt{t-s} + |x-y|). \tag{68}$$

We deal now with the second term in the right hand of (72). Since $\|F\|_\infty \leq C/\varepsilon$, the same computations as above (using (65)) give

$$\left|\frac{\partial \underline{P}_\delta F_s}{\partial x_k}(x)\right| + \left|\frac{\partial^2 \underline{P}_\delta F_s}{\partial x_k \partial x_p}(x)\right| \leq \frac{1}{\delta\varepsilon}\frac{K(1+|x|)^q}{a} e^{2\lambda} \tag{69}$$

and

$$\left|\frac{\partial^2 \underline{P}_\delta F_s}{\partial s \partial x_k}(x)\right| \leq \frac{1}{\delta^{3/2}\varepsilon}\frac{K(1+|x|)^q}{a} e^{2\lambda}. \tag{70}$$

Lemma 3 *Let $v_i = \frac{\partial u}{\partial x_i}$. Under the hypothesis $(\mathcal{H})$ (and consequently under (62)) one has*

$$|v_i(t,x) - v_i(t,y)| \leq \frac{K(1+|x|+|y|)^q}{a} e^{2\lambda} \times \left(\frac{\sqrt{\delta}}{\varepsilon} + \frac{1}{\delta\varepsilon}|x-y|\right) \tag{71}$$

and

$$|v_i(t,x) - v_i(s,x)| \leq \frac{K(1+|x|)^q}{a} e^{2\lambda} \times \left(\frac{\sqrt{\delta}}{\varepsilon} + \frac{1}{\delta\varepsilon}\sqrt{t-s}\right).$$

for every $x, y \in \mathbb{R}^d$ and every $t, s \geq 0$ such that $|t-s| \leq \delta$.

Proof: We take derivatives in the mild equation for u and we obtain, for $t \leq T-\delta$

$$\begin{aligned} v_i(t,x) &= \frac{\partial}{\partial x_i}\underline{P}_{T-t}\underline{h}_T(x) + \int_t^T \frac{\partial}{\partial x_i}\underline{P}_{s-t}F_s(x)ds \\ &= \frac{\partial}{\partial x_i}\underline{P}_{T-(t+\delta)}\underline{P}_\delta \underline{h}_T(x) + \int_{t+\delta}^T \frac{\partial}{\partial x_i}\underline{P}_{s-(t+\delta)}\underline{P}_\delta F_s(x)ds + \int_t^{t+\delta}\frac{\partial}{\partial x_i}\underline{P}_{s-t}F_s(x)ds. \quad (72) \end{aligned}$$

Note that in the first two terms in the above (72) involve $\underline{P}_\delta F$, so one can use the regularization effect of the semi-group which is not the case for the third term. We evaluate first the last term in the right hand of the above equality. Using (65)

$$\begin{aligned} \left|\int_t^{t+\delta}\frac{\partial}{\partial x_i}\underline{P}_{s-t}F_s(x)ds\right| &\leq \|F\|_\infty \int_t^{t+\delta}\int_{\mathbb{R}^d}\left|\frac{\partial}{\partial x_i}p_{s-t}(x,y)\right| dyds \\ &\leq \|F\|_\infty \frac{K(1+|x|)^q}{a} e^{2\lambda}\int_t^{t+\delta}\int_{\mathbb{R}^d}\frac{1}{s-t}e^{-K'\frac{|x-y|^2}{s-t}}dy\,ds \\ &\leq \|F\|_\infty \frac{K(1+|x|)^q}{a} e^{2\lambda}\int_t^{t+\delta}\frac{1}{\sqrt{s-t}}ds \\ &\leq \frac{\sqrt{\delta}}{\varepsilon}\frac{K(1+|x|)^q}{a}e^{2\lambda}, \end{aligned}$$

the last inequality being a consequence of $\|F\|_\infty \leq C/\varepsilon$. We deal now with the first term in the RHS of (72). Using the Feynmann-Kac formula

$$\frac{\partial}{\partial x_i}\underline{P}_{T-(t+\delta)}\underline{P}_\delta\underline{h}_T(x) = \frac{\partial}{\partial x_i}\mathbb{E}\underline{P}_\delta\underline{h}_T(\underline{S}^x_{T-(t+\delta)}) = \sum_{k=1}^d \mathbb{E}\left(\frac{\partial \underline{P}_\delta \underline{h}_T}{\partial x_k}(\underline{S}^x_{T-(t+\delta)})\frac{\partial \underline{S}^{x,k}_{T-(t+\delta)}}{\partial x_i}\right).$$

Using inequalities (66), (67) and (68), one checks that

$$\begin{aligned} &\left|\mathbb{E}\left(\frac{\partial \underline{P}_\delta \underline{h}_T}{\partial x_k}(\underline{S}^x_{T-(t+\delta)})\frac{\partial \underline{S}^{x,k}_{T-(t+\delta)}}{\partial x_i} - \frac{\partial \underline{P}_\delta \underline{h}_T}{\partial x_k}(\underline{S}^y_{T-(t'+\delta)})\frac{\partial \underline{S}^{y,k}_{T-(t'+\delta)}}{\partial x_i}\right)\right| \\ &\qquad \leq \frac{1}{\delta}\frac{K(1+|x|+|y|)^q}{a}e^{2\lambda}(|x-y|+\sqrt{t-t'}). \end{aligned}$$

We turn now to the second term in the right hand of (72). Using (69), one obtains

$$\int_{t+\delta}^T\left|\frac{\partial}{\partial x_i}\underline{P}_{s-(t+\delta)}\underline{P}_\delta F_s(x) - \frac{\partial}{\partial x_i}\underline{P}_{s-(t+\delta)}\underline{P}_\delta F_s(y)\right| \leq \frac{K(1+|x|+|y|)^q}{a\delta\varepsilon}e^{2\lambda}|x-y|.$$

Consider now $t' > t$ and write

$$\left|\int_{t+\delta}^T \frac{\partial}{\partial x_i}\underline{P}_{s-t}F_s(x)ds - \int_{t'+\delta}^T\frac{\partial}{\partial x_i}\underline{P}_{s-t'}F_s(x)ds\right|$$

$$\leq \int_{t+\delta}^{t'+\delta} \left|\frac{\partial}{\partial x_i}\underline{P}_{s-t}F_s(x)\right| ds \quad + \quad \int_{t'+\delta}^{T} \left|\frac{\partial}{\partial x_i}\underline{P}_{s-t}F_s(x)ds - \frac{\partial}{\partial x_i}\underline{P}_{s-t'}F_s(x)\right| ds =: I + J.$$

Using (69) and (70), we obtain
$$\begin{aligned} I &\leq \frac{K(1+|x|)^q}{a\delta\varepsilon} e^{2\lambda} |t-t'| \\ J &\leq \frac{K(1+|x|)^q}{a\delta^{3/2}\varepsilon} e^{2\lambda} |t-t'| \leq \frac{K(1+|x|)^q}{a\delta\varepsilon} e^{2\lambda} \sqrt{|t-t'|} \end{aligned}$$
the last inequality being a consequence of $|t-t'| \leq \delta$. This completes the proof . $\diamond$

The above lemma and the representation $\underline{H}_t^x = (c_\lambda^* \nabla u)(t, \underline{S}_t^x)$ straightforwardly yield

Corollary 1 *For every $s < r < T - \delta$ such that $r - s < 1/n < \delta$,*

$$\left(\mathbb{E}|\underline{H}_r^x - \underline{H}_s^x|^2\right)^{1/2} \leq \frac{K(1+|x|)^q}{a} e^{2\lambda} \times \left(\frac{\sqrt{\delta}}{\varepsilon} + \frac{1}{\delta\varepsilon}\frac{1}{\sqrt{n}}\right). \tag{73}$$

We are now able to prove Theorem 7.

Proof of Theorem 7: Using (73)

$$\begin{aligned} \mathbb{E}\int_0^{T-\delta} \left|\underline{H}_s - \underline{\eta}_s\right|^2 ds &= \sum_{t_k < T-\delta} \mathbb{E}\int_{t_k}^{t_{k+1}} \left|\frac{1}{\Delta t_{k+1}} \int_{t_k}^{t_{k+1}} (\underline{H}_s - \underline{H}_r) dr\right|^2 ds \\ &\leq \sum_{t_k < T-\delta} \int_{t_k}^{t_{k+1}} \frac{1}{\Delta t_{k+1}} \int_{t_k}^{t_{k+1}} \mathbb{E}|\underline{H}_s - \underline{H}_r|^2 dr ds \\ &\leq \frac{K(1+|x|)^{2q}}{a^2} e^{4\lambda} \times \left(\frac{\sqrt{\delta}}{\varepsilon} + \frac{1}{\delta\varepsilon}\frac{1}{\sqrt{n}}\right)^2. \end{aligned}$$

Moreover, as a consequence of the first two lemmas

$$\mathbb{E}\int_0^{T-\delta} |c^*(S_s)(Z_s - \zeta_s)|^2 ds \leq \frac{C}{n} + C(e^{-C'\lambda^2/T} + \varepsilon^2) + \frac{K(1+|x|)^{2q}}{a^2} e^{4\lambda} \times \frac{1}{\varepsilon^2}(\delta + \frac{1}{n\delta^2}).$$

In order to minimize $\delta + \frac{1}{n\delta^2}$ we take $\delta_n = n^{-1/3}$ so that $\delta + \frac{1}{n\delta^2} = Cn^{-1/3}$. Then, in order to minimize $\varepsilon^2 + \varepsilon^{-2}n^{-1/3}$ we take $\varepsilon_n = n^{-1/6}$ so that $\varepsilon^2 + \varepsilon^{-2}n^{-1/3} \sim n^{-1/6}$. Finally we take $\lambda_n = \sqrt{\ln n}$ and to obtain (59). $\diamond$

6 Numerical experiments on American exchange style options

In this section, we present some numerical experiments concerning the pricing and the hedging of American style options in dimensions $d = 2$ up to 10. This study will be divided into two parts. First, we will specify the spatial accuracy in each dimension in order to be able to produce a good choice of time and space discretization. Secondly, we will compute some prices and hedges according to our previous "optimal" discretization.

6.1 The model

Due to the numerical cost of theses simulations (especially in large dimension), we will emphasize on model independent (with respect to the dividends rate, volatilities, ...) computations. For this reason the underlying assets $(S_t)_{t\in[0,T]}$ in $\mathbb{R}^d$ are null correlated and follows here the standard Black and Scholes models

$$\mathrm{d}S_t^i = -\mu_i S_t^i\,\mathrm{d}t + \sigma_i\,\mathrm{d}W_t^i, \quad t\in[0,T], \quad i=1,\dots,d, \tag{74}$$

where $\mu_i > 0$ are divided rates, $\sigma_i > 0$ are the (constant) volatilities and $(W_t)_{t\in[0,T]}$ denotes a d-dimensional standard Brownian motion.

This choice is motivated by the importance of this model for applications. Furthermore, S_t is a closed function of (t, W_t) for every $t\in[0,T]$, namely $S_t^i = s_0^i \exp\left(-(\mu_i + \sigma_i^2/2)t + \sigma_i W_t^i\right)$. Therefore the computation of the quantization tree of (S_t) relies entirely on the one of (W_t). In its turn, because of the null correlation assumption, such quantization is computed from the optimal quantization of the normal law in $\mathbb{R}^d$ and from the edge weights of Proposition 1. We point out that these computations are done once for all.

Concerning the contract, we focus here on American exchange style options with pay-offs

$$h(x) = \max\left(x_1\cdots x_p - x_{p+1}\cdots x_{2p}, 0\right) \qquad \text{(set } d := 2p\text{)}. \tag{75}$$

We know that a fair price $\mathcal{V}_t$ at time t and maturity T of such a contract is given by the Snell envelope of the process $(h(S_t))_{t\in[0,T]}$, namely

$$e^{-rt}\mathcal{V}_t = \sup\left\{\mathbb{E}(e^{-r\tau}h(S_\tau)\mid\mathcal{F}_t), \quad \tau\in[t,T], \quad \tau \text{ stopping time}\right\}, \tag{76}$$

where the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ are those of the Brownian motion (W_t).

In the framework of exchange options, we can show that such a price is independent of the interest rate r, it is why we will consider in this section that $r=0$.

It is important to note that there exists a closed form for the Black & Scholes price of a European contract for (74)–(75) at time t and maturity T given by

$$BS(\theta, x, y, \widetilde{\sigma}, r) = \operatorname{erf}(d_1)\exp(r\theta)\,x - \operatorname{erf}(d_1 - \widetilde{\sigma}\sqrt{\theta})\,y,$$

where $\theta := T-t$, $\widetilde{\sigma} := \sqrt{\sum_{i=1}^d \sigma_i^2}$, $r := \sum_{i=1}^p \mu_i - \sum_{i=p+1}^d \mu_i$, $x := \prod_{i=1}^p S_i(t)$, $\quad y = \prod_{i=p+1}^d S_i(t)$, $d_1 := \dfrac{\ln(x/y) + (\widetilde{\sigma}^2/2 + r)\theta}{\widetilde{\sigma}\sqrt{\theta}}$ and erf denotes the distribution function of the standard normal distribution.

6.2 The numerical scheme

Let us precise now the numerical scheme that we will implement. As mentioned above, our approach to pricing consists first in quantizing the d-dim Brownian motion. More precisely, let $T > 0$ and n, N two integers; set $\Delta t := \frac{T}{n}$ and $t_k := k\Delta t$. Spatial discretization depends on the time t_k. Indeed, we use the partition

$$N = N_0 + N_1 + N_2 + \cdots + N_n,$$

given by (32), which assigns the number of points N_k to te k^{th} time layer. Typically, $N_0 = 1 < N_1 < \cdots < N_n = \rho_d N$ where ρ_d is a constant depending on d. The N_k's are chosen in order to make the smallest possible quantization error for the d-dim Brownian motion. Now, assume that for each $k \in \{0, \ldots, n\}$ the N_k-optimal quantizer of the normal distribution on $\mathbb{R}^d$ has been computed. We deduce the N_k-optimal quantizer $(x_i^k)_{i=1,\ldots,N_k}$ of W_{t_k} by a simple $\sqrt{t_k}$-dilatation.

Finally, introducing an abstract family of control variate variables $(M_i^k)_{1\le k\le n}$, the tested scheme can be written as a slight modification of Algorithm (20)

$$\begin{cases} v_i^n := h_i^n - M_i^n, \quad i = 1, \ldots, N_n, \\ v_i^k := \max\left\{ h_i^k - M_i^k, \displaystyle\sum_{j=1}^{N_{k+1}} \pi_{i,j}^k \, v_j^{k+1} \right\}, \quad i = 1, \ldots, N_k, \; k = 0, \ldots, n-1 \end{cases} \tag{77}$$

where $h_i^k := h(s_{i,1}^k, \ldots, s_{i,d}^k)$ with $s_{i,\ell}^k := s_{0,\ell} \exp\left(-\left(\mu_\ell + \dfrac{\sigma_\ell^2}{2}\right) k\Delta t + \sigma_\ell \, x_i^k \right)$, $\ell = 1, \ldots, d$ and where the weights $\pi_{i,j}^k$ are Monte-Carlo proxies of the theoretical weights

$$\frac{\mathbb{P}(W_{t_{k+1}} \in C_j^{k+1}, W_{t_k} \in C_i^k)}{\mathbb{P}(W_{t_k} \in C_i^k)}.$$

The price at point x_i^k is given by

$$p_i^k = v_i^k + M_i^k. \tag{78}$$

Following (47) the hedging δ_i^k at x_i^k is computed by

$$\delta_{i,\ell}^k := \frac{\displaystyle\sum_{j=1}^{N_{k+1}} \pi_{i,j}^k (v_j^{k+1} - v_i^k)(s_{j,\ell}^{k+1} - s_{i,\ell}^k)}{\displaystyle\sum_{j=1}^{N_{k+1}} \pi_{i,j}^k (s_{j,\ell}^{k+1} - s_{i,\ell}^k)^2}, \quad \ell = 1, \ldots, d. \tag{79}$$

Numerical experiments show that the choice of the so-called "control variables" M_i^k is very important in practice. Note that any "control variables" M_i^k such that

$$\sum_{j=1}^{N_{k+1}} \pi_{i,j}^k \, M_j^{k+1} \approx M_i^k \quad \text{up to the spatial discretization} \tag{80}$$

can be chosen here. A efficient choice is here to take

$$M_i^k = BS\left(T - t_k, \prod_{\ell=1}^{p} s_{i,\ell}^k, \prod_{\ell=p+1}^{d} s_{i,\ell}^k, \widetilde{\sigma}, r\right). \tag{81}$$

We have to note that there exists a natural choice of $\{M_i^k\}$ for which one has equality in (80). It consists in computing in a parallel way the European price approximation by

$$M_i^k = \sum_{j=1}^{N_{k+1}} \pi_{i,j}^k \, M_j^{k+1}, \quad \text{with } M_i^n = h_i^n,$$

But it is not difficult to check that this choice is numerically equivalent to compute directly (v^k) in (77) with $M_i^k = 0$. Numerical computations will be run with the choice (81).

6.3 Accuracy, stability

We will now estimate the rate of convergence of the numerical solution given by (77) towards a reference one. Theorem 5 gives the expected error terms both in time and in space. We recall that for a time discretization parameter n (number of time layers) and space discretization parameter $\overline{N} = N/n$ (average number of points per layer), we have

$$E(n, \overline{N}) = |p_0^0 - P_{th}| \approx \frac{c_1}{n} + c_2 \frac{n}{(\overline{N})^{\alpha}}, \tag{82}$$

where $p(n, \overline{N})$ (resp. P_{th}) is the computed price (resp. theoretical price) at $t = 0$ of american style options.

6.3.1 The one dimensional case

We begin by the case of the 1-dimension. The reference model will be the standard one dimensional Black & Scholes

$$\mathrm{d}S_t = S_t(r \,\mathrm{d}t + \sigma \,\mathrm{d}W_t), \quad t \in [0, T], \quad S_0 = 36,$$

where $r = 0.06$, $\sigma = 0.2$. The american contract will be an American put with the payoff function $x \mapsto \max(K - x, 0)$ with $K = 40$.

We will try to estimate the space convergence rate α in (82). In Table 1, we have computed for n becoming larger ($n \in \{50, \ldots, 100\}$) the error $E(n, \overline{N})$ for $\overline{N} \in \{25, 37, 50, 75\}$. Numerically we observe that for these choices of n the error is almost the spatial error

$$E(n, \overline{N}) \approx c_2 \frac{n}{(\overline{N})^{\alpha}} = C_2\, n.$$

In the last lines of Table 1, we have computed C_2 by means of linear regression. The last line represents the correlation term obtained. Once C_2 computed, we can evaluate the rate α between two values $\overline{N}_1$ and $\overline{N}_2$ by

$$\alpha = \frac{\ln C_2(\overline{N}_2)/C_2(\overline{N}_1)}{\ln(\overline{N}_1)/\ln(\overline{N}_2)}.$$

Values of α are reported in the last line of Table 1. We note that the computed value α is better than the expected value ($\alpha = 1$). In a linear case (European option case) we can show that it is implied by the regularizing properties of the semi-group of the diffusion. Indeed the numerical integration formula applied to $f(X)$ where f admits a second derivative and X is a square integrable random variable can be written as (see [38])

$$\left| \mathbb{E} f(X) - \sum_{i=1}^{N} \pi_i f(x_i) - \sum_{i=1}^{N} f'(x_i) \mathbb{E}((X - x_i)\ \mathbf{1}_{C_i(x)}) \right| \leq \frac{1}{2} L(f'') D_N^{X,2}, \tag{83}$$

where $D_N^{X,2}$ is of the order of $N^{-2/d}$. But the optimality of the grid makes the term $\mathbb{E}((X - x_i)\ \mathbf{1}_{C_i(x)}(X)) = \frac{\partial D_N^{X,2}}{\partial x_i}$ vanishes. It seems difficult to show that this heuristic applies in the American option case although we observed it. This better rate of convergence is a strong argument showing that the optimal quantization is worthwhile.

6.3.2 Higher dimensional case

The determination of spatial order involves long runs (because of the smallness of Δt) with increasing average values of $\overline{N}$. From dimension 4 to 10, the storage of the matrix $\{\pi_{i,j}^k\}$ for such discretization is costly and make the computations intractable. Nevertheless, the previous calculations in one and two dimension conclude to a spatial order of $2/d$ when the grids are optimal. In fact, such optimality becomes harder and harder to get in high dimensions, it is why we can guess that spatial order will be between $1/d$ (the "grid order") and $2/d$.

Table 2 shows

6.4 Numerical experiments

We now present numerical simulations of Eq. (76) using (77), (78), (79), (81). The maximal horizon time T_{max} is chosen to be equal to 1. For a given spatial discretization parameter $\overline{N} = N/n$, we will work with a time discretization $\Delta t = T_{max}/n$ where $n = n(d, \overline{N})$ is given in Theorem 5 in order to ensure stability as shown above. Space discretization is then achieved by the optimal quantization of the Brownian motion in $\mathbb{R}^d$ on a space-time grid $\Gamma^d = \{\Gamma^d_k\}_{0\leq k\leq n(d,\overline{N})}$.

In order to investigate the numerical influence of the free boundary, we will distinguish in each study an in-the-money case and an out-of-the-money case. Furthermore, our numerical price and hedging will be compared to a reference setting which is here the numerical scheme employed by [41] in the 2D exchange option. The model parameters and initial data of the reference scheme are given by Table 3 in the in-the-money case and Table 4 in the out-of-the-money case.

In the following figures are displayed the computed prices and hedges at time $t = 0$ together with the reference ones as a function of the maturity $T \in [0, T_{max}]$.

We first investigate the pricing and hedging of (74)–(76) when $d = 2$. The values of the dividends, volatilities and initial conditions are displayed in Table 3 and 4. We see in Figs. 1–4 that our computations are well fitted with the reference ones in both cases (in and out the money).

In the in-the-money case, when the dimension increases, we can see that our price increases faster than the reference price when the maturity grows (see Figs. 5, 9, 13). The maximal error remains inside 5 % in all the cases. The same phenomenon can be observed on the computed hedges (see Figs. 6, 10, 14).

In the other case (out-of-the-money), very different behaviour are observed on the prices. Indeed whatever the dimension (from 4 to 10) is, the prices seem to be well computed (Figs. 7, 11, 15).

Such a modification of behaviour happens too in Fig. 6.4. Here we show the implied residual risk in the computation of the hedging by least-squares method (see (48)) as a function of the maturity. We can see there that the values of this risk is bigger in the in-the-money case than in the out-of-the-money case. This may suggest an explanation of the antagonism of the Figs. 5/7, 9/11, 13/15. Indeed it seems that numerical incompleteness of the market (“residual risk") has a bigger effect on the price “in-the-money" than “out-of-the-money".

Table 1: Estimation of the spatial rate of convergence α of (82) in the dimension 1 and 2.

	$d=1$				$d=2$	
n	$E(n,25)$	$E(n,37)$	$E(n,50)$	$E(n,75)$	$E(n,300)$	$E(n,600)$
50	0.1871	0.094	0.0573	0.0312		
55	0.2074	0.1043	0.0634	0.0347		
60	0.2213	0.1079	0.0645	0.0328		
65	0.2423	0.1197	0.0707	0.0367		
70	0.2616	0.1302	0.0773	0.0408		
75	0.2822	0.1398	0.0836	0.0442		
80	0.2982	0.147	0.0862	0.0442		
85	0.3186	0.1568	0.0921	0.0474		
90	0.3387	0.167	0.0985	0.0509		
95	0.3555	0.1734	0.102	0.0515		
100	0.3753	0.1854	0.1085			
r	1.00	1.00	0.99	0.97		
C_2	3.77(-3)	1.82(-3)	1.03(-3)	4.79(-4)		
α	1.87	1.90	1.91			

Table 2: Estimation of the spatial rate of convergence α of (82) for dimensions $d = 4, 6, 10$.

	$d=4$		$d=6$		$d=10$	
n	$E(n,500)$	$E(n,750)$	$E(n,500)$	$E(n,1000)$	$E(n,500)$	$E(n,1000)$
75	0.046843					
80	0.049243					
85	0.053543					
90	0.056643					
95	0.060443					
100	0.063443					

i	1	2
μ_i	-5%	0
σ_i	20%	20%
$s_{0,i}$	40	36

Table 3: Values of the parameter model when $d = 2$ in the in-the-money case.

$d = 2, \quad N_{max} = 300, \quad n = 25$

In–the–money case

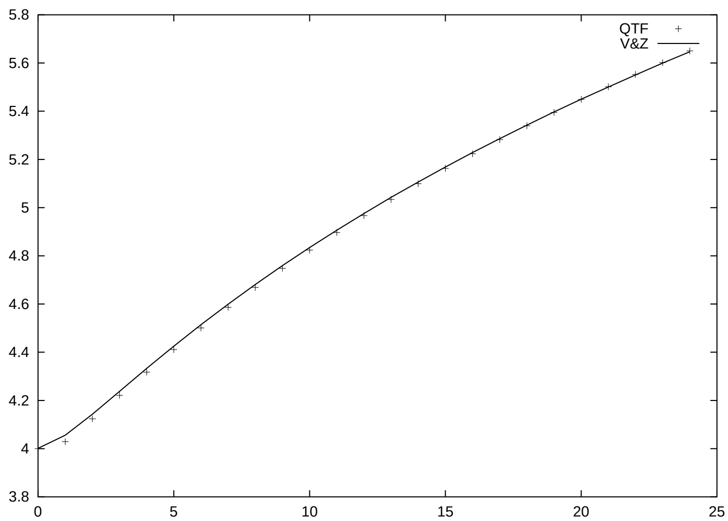
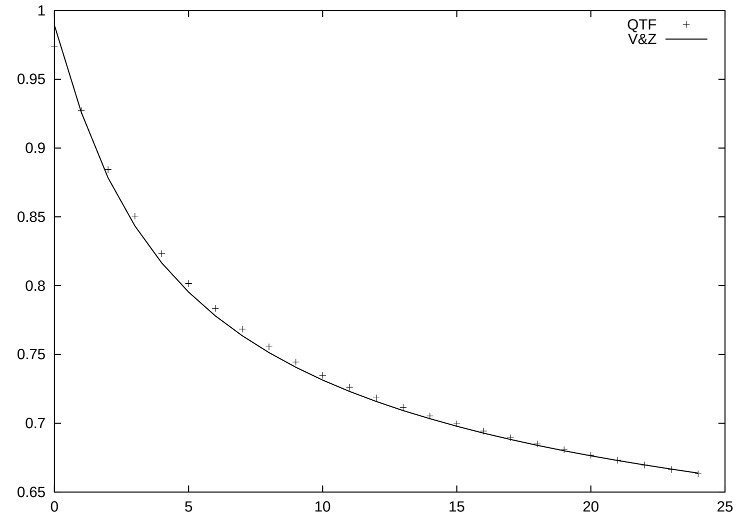


Figure 1: American exchange option price $(S_1 - S_2)_+$ function of the maturity where the number of layers is 25 and the number of points on the top is 300. $+$ depicts the price obtained with the method of quantization and – depicts the reference price (V & Z) (cf. [41]).

a)



b)

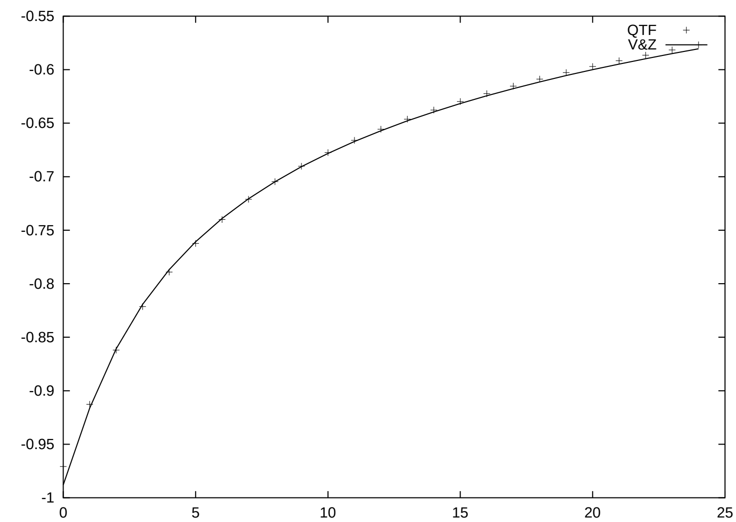


Figure 2: Hedging for the American option of Fig. 1. a) δ_1 (+) fonction of the maturity. b) δ_2 (+) fonction of the maturity. The reference price is denoted by −.

Table 4: Value of the parameter model when $d = 2$ in the out-of-the-money case.

i	1	2
μ_i	-5%	0
σ_i	20%	20%
$s_{0,i}$	36	40

$$d = 2, \quad N_{max} = 300, \quad n = 25$$

Out–of–the–money case

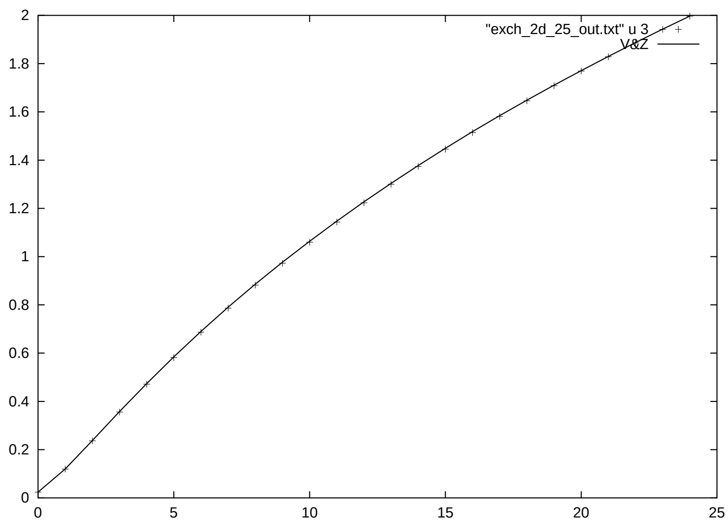
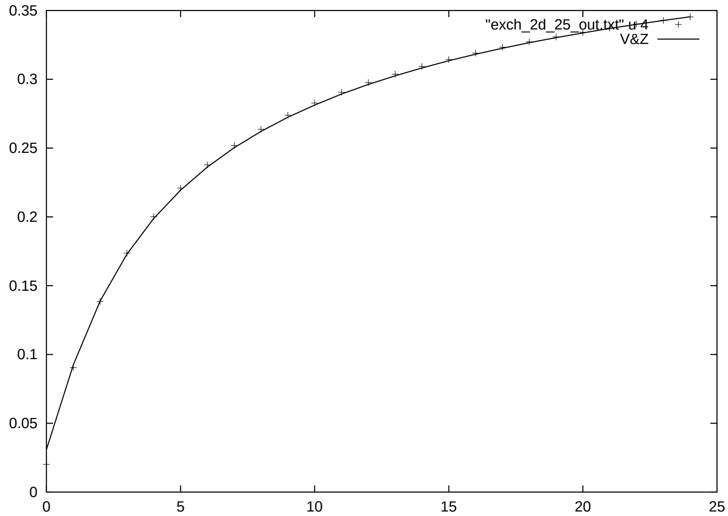


Figure 3: American exchange option price $(S_1 - S_2)_+$ fonction of the maturity where the number of layers is 25 and the number of points on the top is 300. + depicts the price obtained with the method of quantization and – depicts the reference price (V & Z) (cf. [41]).

a)



b)

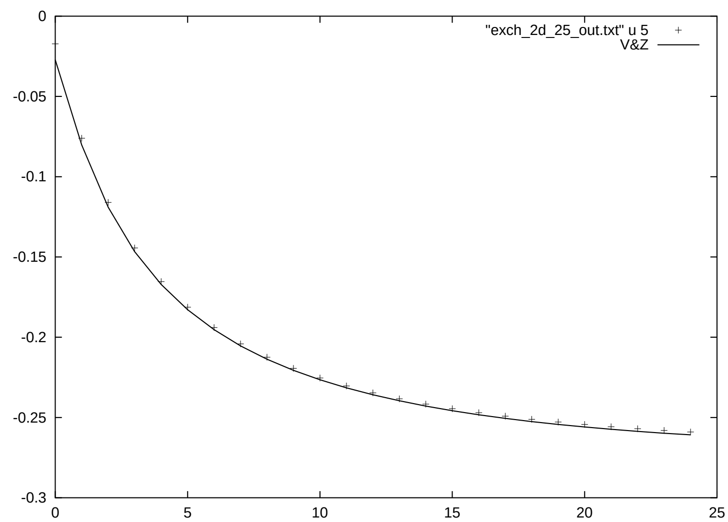


Figure 4: Hedging for the American option of Fig. 3. a) δ_1 (+) fonction of the maturity. b) δ_2 (+) fonction of the maturity. The reference price is denoted by $-$.

Table 5: Value of the model parameter when $d = 4$ in the in-the-money case.

i	1	2	3	4
μ_i	-5 %	0	0	0
σ_i	14.14 %	14.14 %	14.14 %	14.14 %
$s_{0,i}$	6.32	6.32	6.00	6.00

$$d = 4, \quad N_{max} = 750, \quad n = 20$$

In–the–money case

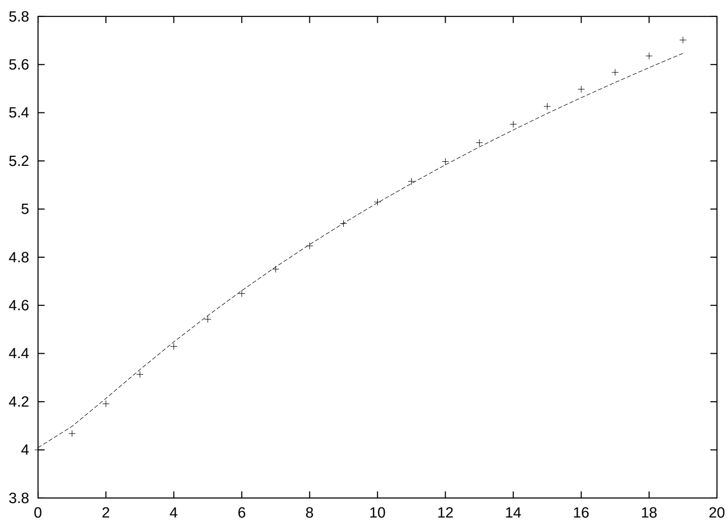
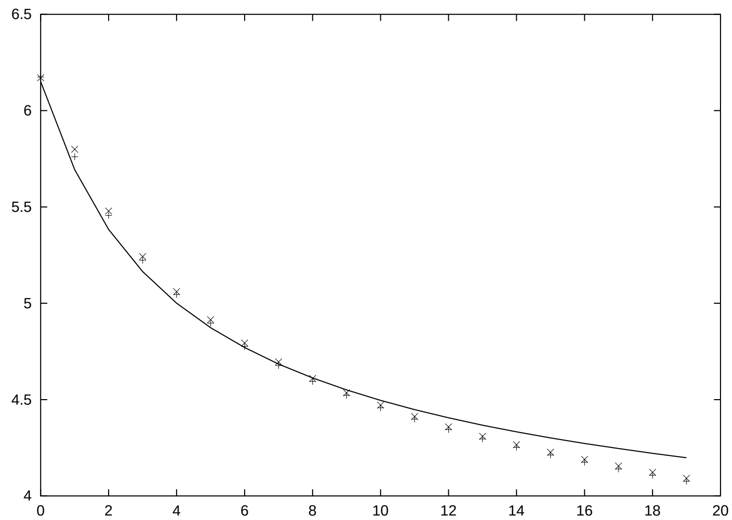


Figure 5: American exchange option price $(S_1S_2 - S_3S_4)_+$ in–the–money fonction of the maturity where the number of layers is 20 and the number of points on the top is 750. $+$ depicts the price obtained with the method of quantization and – depicts the reference price (V & Z) (cf. [41]).

a) δ_1 (+), δ_2 ($\times$), Reference hedging (–)



b) δ_3 (+), δ_4 ($\times$), Reference hedging (–)

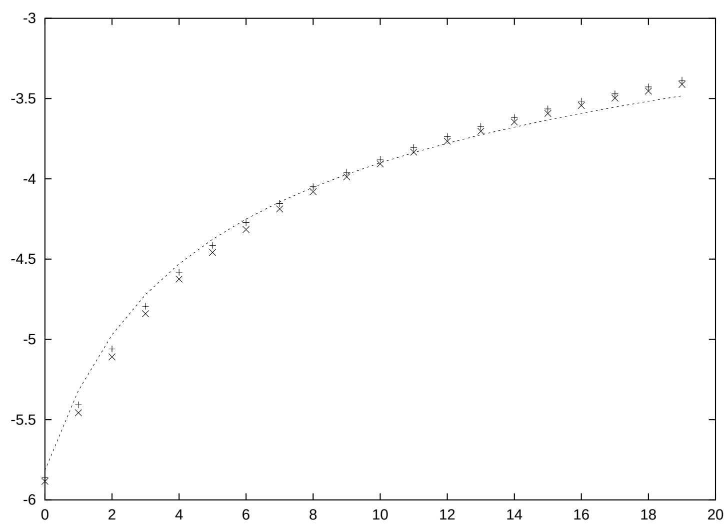


Figure 6: Hedging for the American option of Fig. 5. a) δ_1 (+), δ_2 ($\times$) fonction of the maturity. b) δ_3 (+), δ_4 ($\times$) fonction of the maturity. The reference price is denoted by –.

Table 6: Value of the model parameter when $d = 4$ in the out-of-the-money case.

i	1	2	3	4
μ_i	-5 %	0	0	0
σ_i	14.14 %	14.14 %	14.14 %	14.14 %
$s_{0,i}$	6.00	6.00	6.32	6.32

$d = 4, \quad N_{max} = 750, \quad n = 20$

Out–of–the–money case

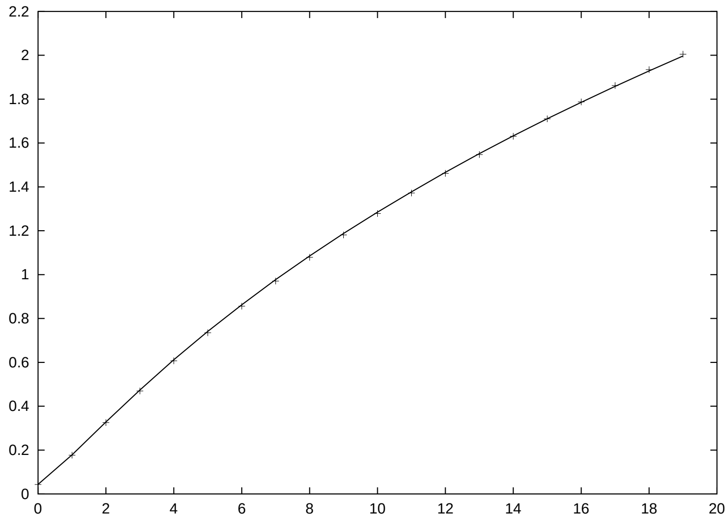
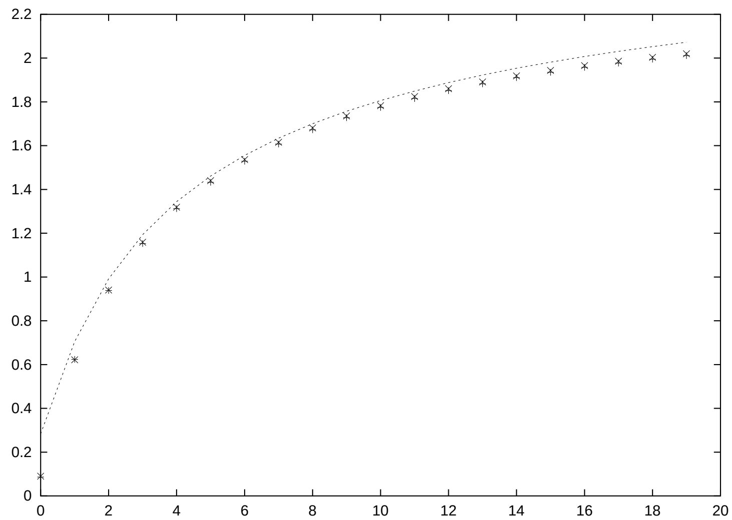


Figure 7: American exchange option price $(S_1S_2 - S_3S_4)_+$ out–the–money fonction of the maturity where the number of layers is 20 and the number of points on the top is 750. $+$ depicts the price obtained with the method of quantization and – depicts the reference price (V & Z) (cf. [41]).

a) δ_1 (+), δ_2 (×), Reference hedging (–)



b) δ_3 (+), δ_4 (×), Reference hedging (–)

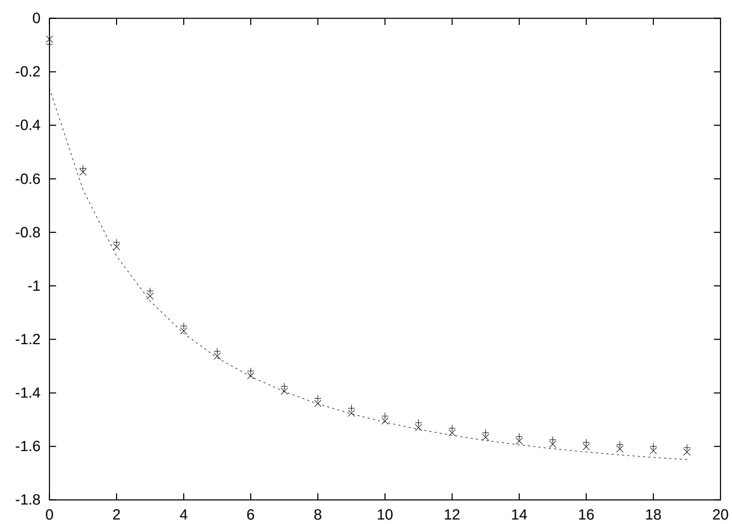


Figure 8: Hedging for the American option of Fig. 7. a) δ_1 (+), δ_2 (×) fonction of the maturity. b) δ_3 (+), δ_4 (×) fonction of the maturity. The reference price is denoted by –.

Table 7: Value of the model parameter when $d = 6$ in the in-the-money case.

i	1	2	3	4	5	6
μ_i	-5 %	0	0	0	0	0
σ_i	11.55 %	11.55 %	11.55 %	11.55 %	11.55 %	11.55 %
$s_{0,i}$	3.42	3.42	3.42	3.30	3.30	3.30

$d = 6, \quad N_{max} = 1000, \quad n = 25$

In–the–money case

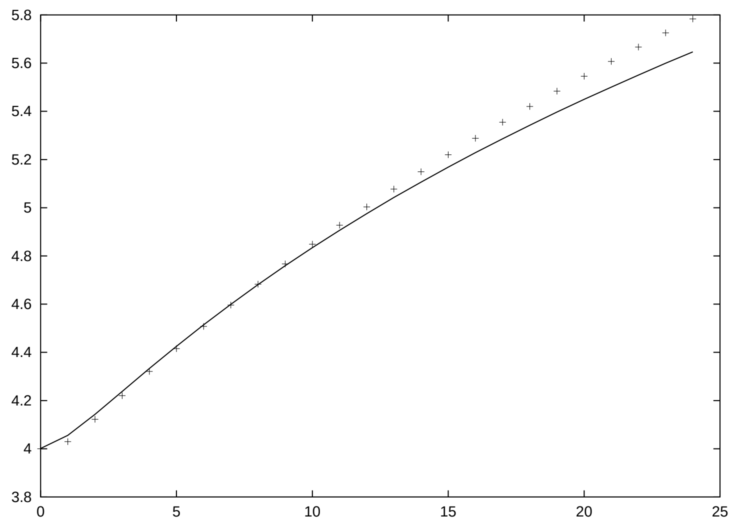
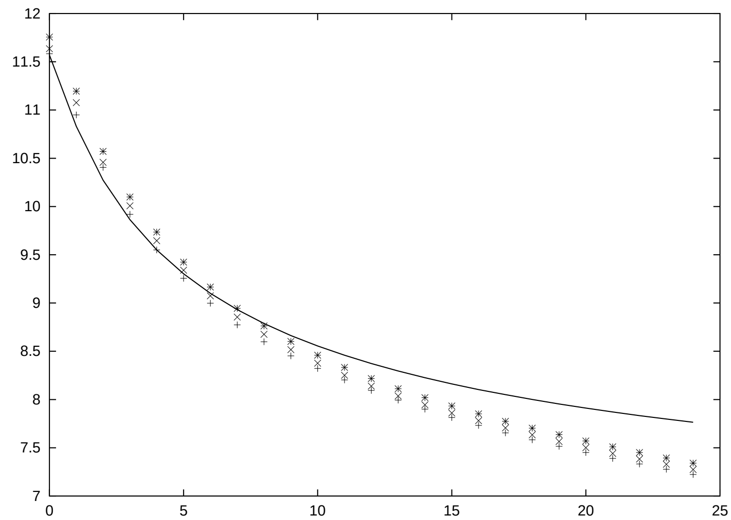


Figure 9: American exchange option price $(S_1S_2S_3 - S_3S_4S_5)_+$ fonction of the maturity where the number of layers is 25 and the number of points on the top is 1000. + depicts the price obtained with the method of quantization and – depicts the reference price (V & Z) (cf. [41]).

a)



b)

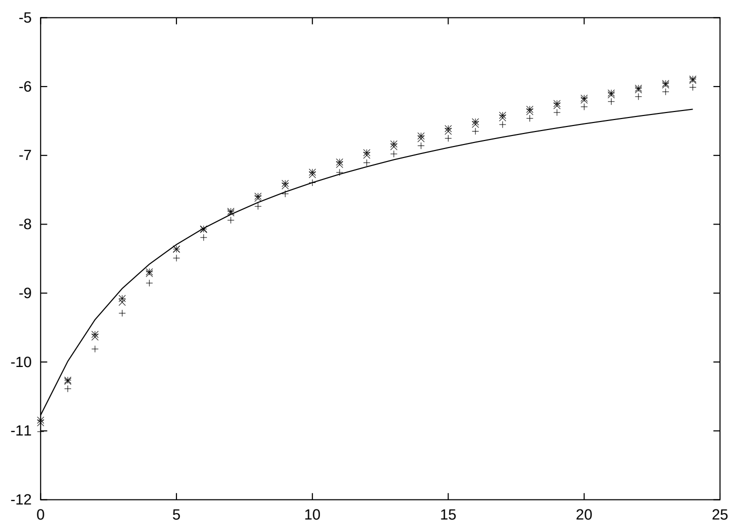


Figure 10: Hedging for the American option of Fig. 9. a) δ_1 ($+$), δ_2 ($\times$), δ_3 ($\ast$) fonction of the maturity. b) δ_4 ($+$), δ_5 ($\times$), δ_6 ($\pm$) fonction of the maturity. The reference price is denoted by $-$.

Table 8: Value of the model parameter when $d = 6$ in the out-of-the-money case.

i	1	2	3	4	5	6
μ_i	-5 %	0	0	0	0	0
σ_i	11.55 %	11.55 %	11.55 %	11.55 %	11.55 %	11.55 %
$s_{0,i}$	3.30	3.30	3.30	3.42	3.42	3.42

$d = 6, \quad N_{max} = 1000, \quad n = 25$

Out–of–the–money case

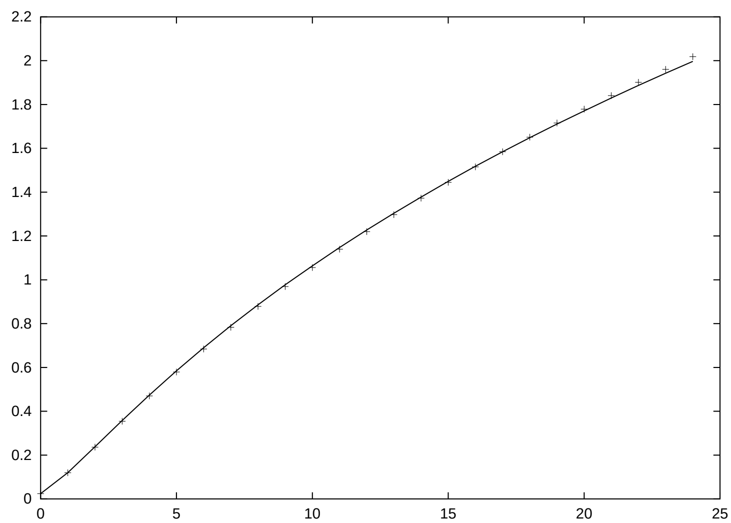
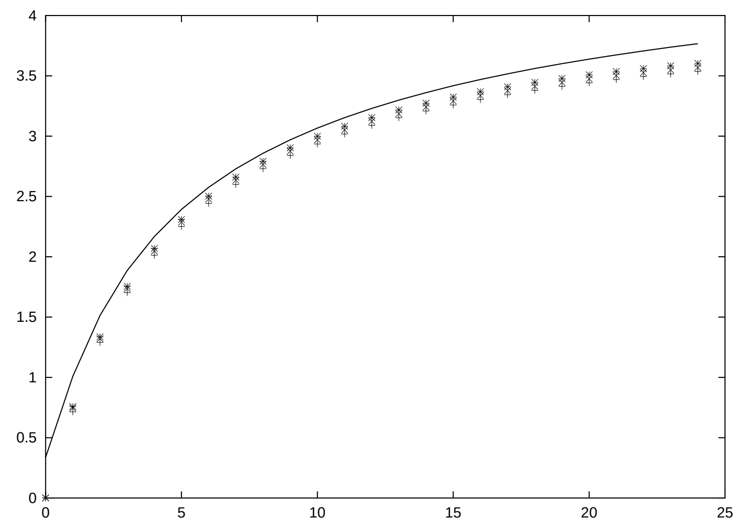


Figure 11: American exchange option price $(S_1S_2S_3 - S_3S_4S_5)_+$ fonction of the maturity where the number of layers is 25 and the number of points on the top is 1000. + depicts the price obtained with the method of quantization and – depicts the reference price (V & Z) (cf. [41]).

a)



b)

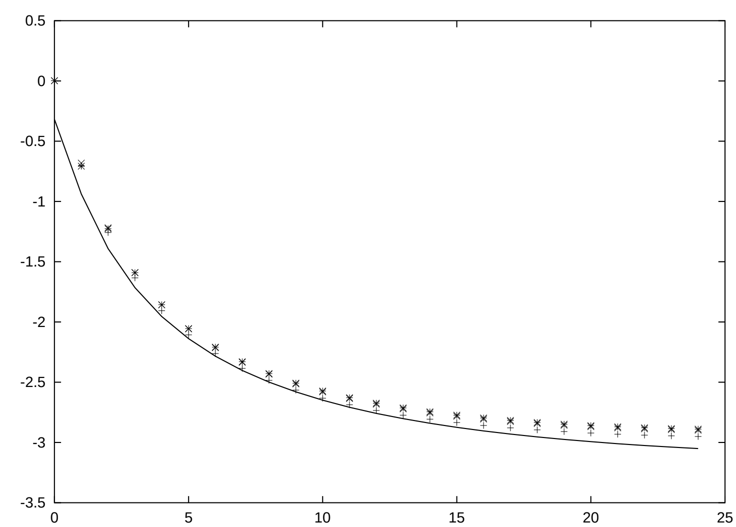


Figure 12: Hedging for the American option of Fig. 11. a) δ_1 $(+)$, δ_2 $(\times)$, δ_3 $(\ast)$ fonction of the maturity. b) δ_4 $(+)$, δ_5 $(\times)$, δ_6 $(\pm)$ fonction of the maturity. The reference price is denoted by $-$.

Table 9: Value of the model parameter when $d = 10$ in the in-the-money case.

i	1	2	3	4	5	6	7	8	9	10
μ_i	-5 %	0	0	0	0	0	0	0	0	0
σ_i	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %
$s_{0,i}$	2.09	2.09	2.09	2.09	2.09	2.04	2.04	2.04	2.04	2.04

$d = 10, \quad N_{max} = 1000, \quad n = 50$

In–the–money case

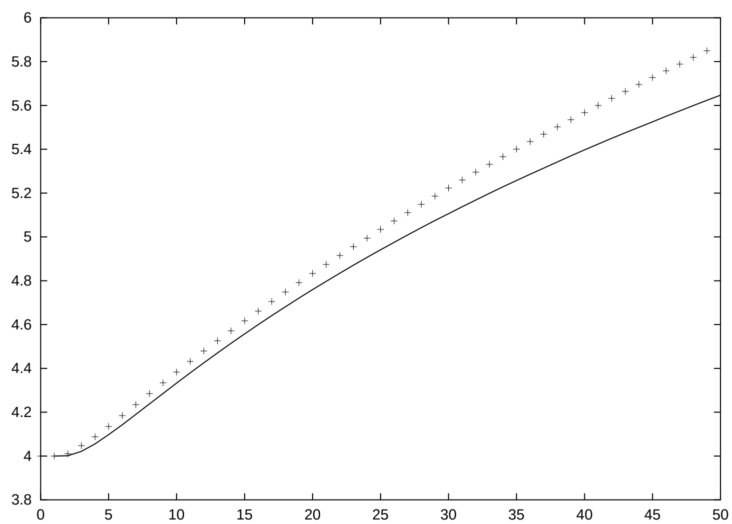
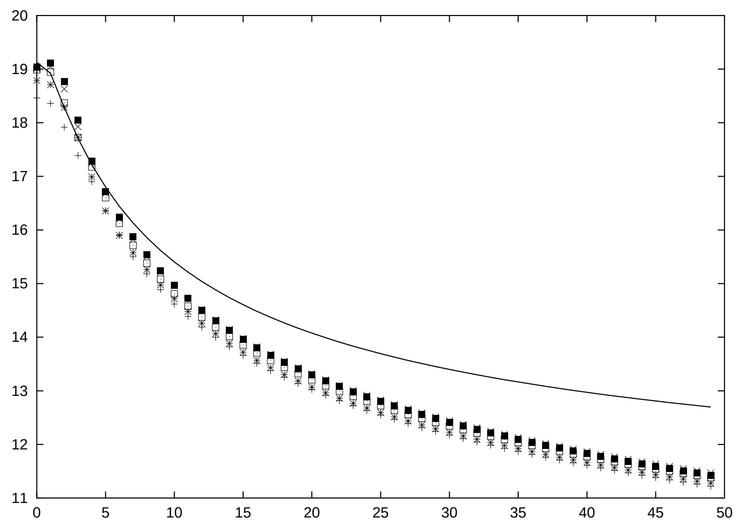


Figure 13: American exchange option price $(S_1S_2S_3S_4S_5 - S_6S_7S_8S_9S_{10})_+$ fonction of the maturity where the number of layers is 50 and the number of points on the top is 1000. + depicts the price obtained with the method of quantization and – depicts the reference price (V & Z) (cf. [41]).

a)



b)

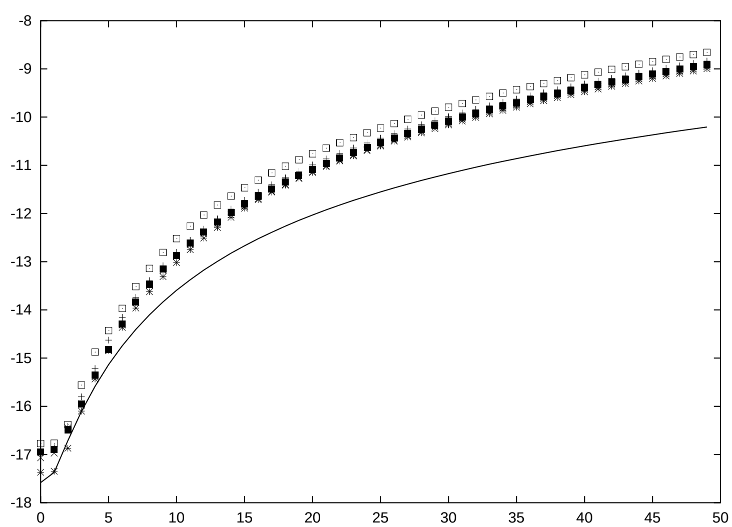


Figure 14: Hedging for the American option of Fig. 13. a) δ_1 (+), δ_2 ($\times$), δ_3 ($\ast$), δ_4 ($\blacksquare$), δ_5 ($\square$) fonction of the maturity. b) δ_6 (+), δ_7 ($\times$), δ_8 ($\ast$), δ_9 ($\blacksquare$), δ_{10} ($\square$) fonction of the maturity. The reference price is denoted by –.

Table 10: Value of the model parameter when $d = 10$ in the out-of-the-money case.

i	1	2	3	4	5	6	7	8	9	10
μ_i	-5 %	0	0	0	0	0	0	0	0	0
σ_i	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %	8.94 %
$s_{0,i}$	2.04	2.04	2.04	2.04	2.04	2.09	2.09	2.09	2.09	2.09

$d = 10, \quad N_{max} = 1000, \quad n = 50$

Out–of–the–money case

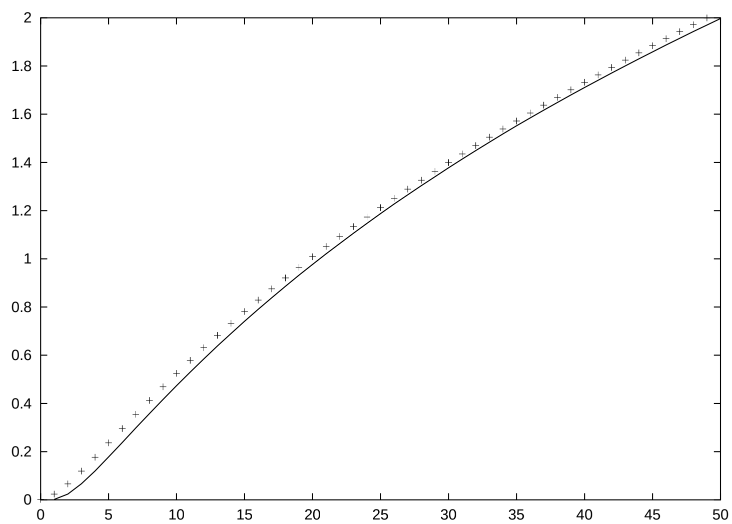
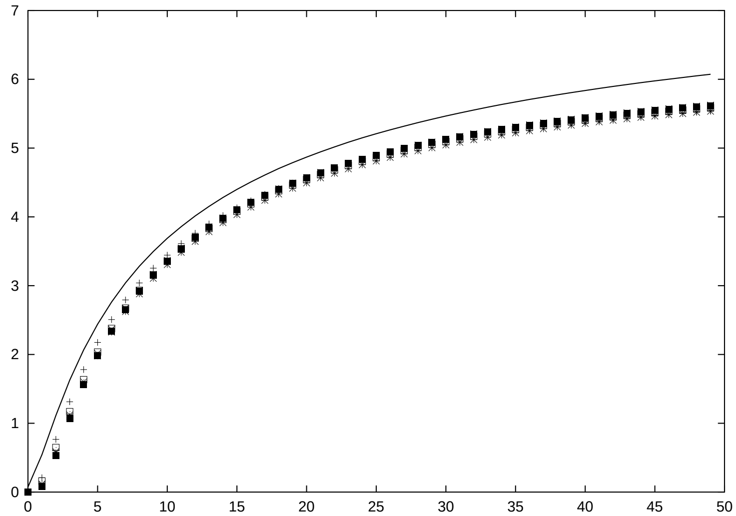


Figure 15: American exchange option price $(S_1S_2S_3S_4S_5 - S_6S_7S_8S_9S_{10})_+$ fonction of the maturity where the number of layers is 50 and the number of points on the top is 1000. $+$ depicts the price obtained with the method of quantization and – depicts the reference price (V & Z) (cf. [41]).

a)



b)

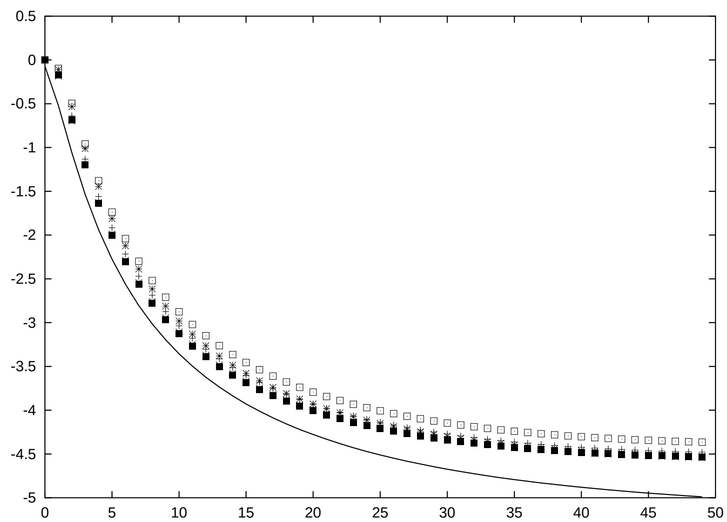


Figure 16: Hedging for the American option of Fig. 15. a) δ_1 ($+$), δ_2 ($\times$), δ_3 ($\ast$), δ_4 ($\blacksquare$), δ_5 ($\square$) fonction of the maturity. b) δ_6 ($+$), δ_7 ($\times$), δ_8 ($\ast$), δ_9 ($\blacksquare$), δ_{10} ($\square$) fonction of the maturity. The reference price is denoted by $-$.

References

[1] D.G. Aronson, Bounds for the fundamental solution of a parabolic equation, *Bull. of the A.M.S.*, **73**, 1967, pp.890-903.

[2] V. Bally, M.E. Caballero, B. Fernandez, N. El Karoui, Reflected BSDE's, PDE's and Variational inequalities. Pré-print INRIA, 2002.

[3] V. Bally, D. Talay, The law of the Euler scheme for stochastic differential equations. I: Convergence rate of the distribution function, *Probab. Theory Relat. Fields*, **104**, No.1, pp.43-60, 1996.

[4] V. Bally, G. Pagès, A quantization algorithm for solving multi-dimensional Optimal Stopping problems, technical report n°628, Laboratoire de Probabilités & modèles aléatoires, Université Paris VI (France), 2000.

[5] V. Bally, G. Pagès, J. Printems, A stochastic quantization method for non linear problems, *Monte Carlo Methods and Applications*, **7**, n^{0}1-2, pp.21-34, 2001.

[6] V. Bally, G. Pagès, J. Printems, First order schemes in the numerical quantization method, forthcoming in (selected papers of the colloquium *Application of Malliavin calculus to Finance*).

[7] V. Bally, G. Pagès, Error analysis of a quantization algorithm for obstacle problems, forthcoming in *Stoch. Proc. and their Appl.*, 2002.

[8] J. Barraquand , D. Martineau, Numerical valuation of high dimensional multivariate American securities, *Journal of Finance and Quantitative Analysis*, **30**, 1995.

[9] A. Bensousan, J.L.Lions, *Applications of the Variational Inequalities in Stochastic Control*, North Holland, 1982, or *Applications des inéquations variationnelles en contrôle stochastique*, Dunod, Paris, 1978.

[10] N. Bouleau, D. Lamberton, Residual risks and hedging strategies in Markovian markets, *Stoch. Proc. and their Appl.*, **33**, pp.131-150, 1989.

[11] M. Broadie, P. Glasserman, Pricing American-Style Securities Using Simulation, *Journal of Economic Dynamics and Control*, **21**, $n^0$8-9, pp.1323-1352, 1997.

[12] J. Bucklew, G. Wise, Multidimensional Asymptotic Quantization Theory with r^{th} Power distortion Measures, *IEEE on Information Theory, Special issue on Quantization*, **28**, n^0 2, pp.239-247, 1982.

[13] D. Chevance, Thèse de l'Université de Provence & INRIA, Septembre 1997, 140p.

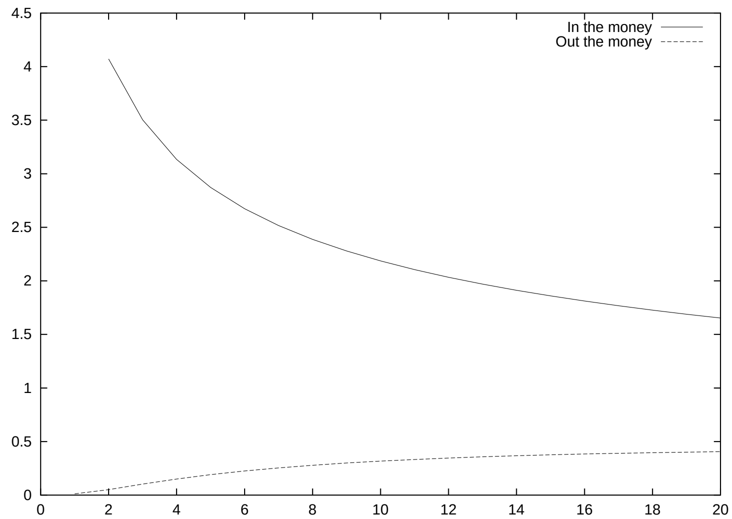
[14] D. Chevance, Numerical methods for backward stochastic differential equations, *Numerical Methods in Finance*, L. Rogers and D. Talay eds., Publications of the Newton Institute series, Cambridge University Press, 1997.

[15] P. Cohort, A geometric approach for the uniqueness of a locally optimal quantizer, *Technical Report*, Labo. Probabilités, Univ. Paris 6 (France), 1997, submitted for publication.

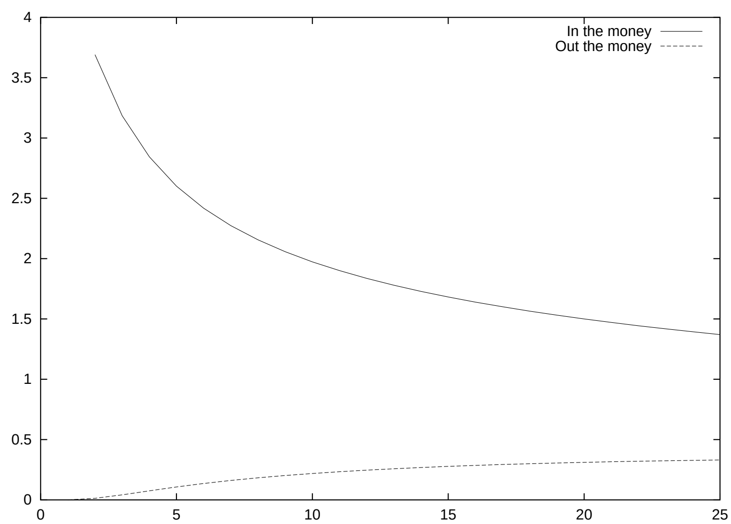
[16] P. Cohort, Strong Law of large Number for the random distortion, *Technical Report*, Centre de Mathématiques de l'Université Paris 12, submitted for publication.

[17] P. Cohort, *Sur quelques problèmes de quantification*, Thèse de l'Université Pierre et Marie Curie, Janvier 2000.

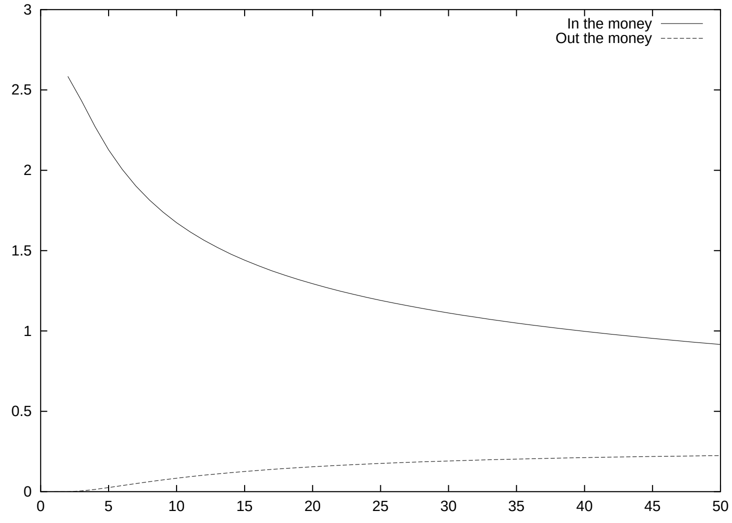
a) Residual risk in $d = 4$.



b) Residual risk in $d = 6$.



c)



[18] A.P. Caverhill, N. Webber, American options: theory and numerical analysis, in *Options: recent advances in theory and practise*, Manchester University press, 1990.

[19] M. Duflo, *Random Iterative systems*, Berlin, Springer, 1998.

[20] N. El Karoui, C. Kapoudjan, E. Pardoux, S. Peng, M.C. Quenez, Reflected solutions of Backward Stochastic Differential Equations and related obstacle problems for PDE's, *The Ann. of Proba.*, **25**, No 2, pp.702-737, 1997.

[21] H. Föllmer, D. Sondermann, Hedging of non redundant contingent claims, *Contributions to Mathematical Economics*, pp.205-223, North-Holland, Amsterdam, 1986.

[22] É. Fournié, J.M. Lasry, J. Lebouchoux, P.L. Lions, N. Touzi, Aplications of Malliavin calculus to Monte Carlo methods in Finance, *Finance & Stochastics*, **3**, pp.391-412, 1999.

[23] É. Fournié, J.M. Lasry, J. Lebouchoux, P.L. Lions, Aplications of Malliavin calculus to Monte Carlo methods in Finance II, *Finance & Stochastics*, **5**, pp.201-236, 2001.

[24] A. Friedmann, *Partial Differential Equations and Applications*, Academic Press, **1**, 1975.

[25] S. Graf, H. Luschgy, *Foundations of quantization for random vectors*, Lecture Notes in Mathematics $n^0$1730, Springer, 2000, 230p.

[26] A. Gersho, R. Gray (eds.), *IEEE on Information Theory, Special issue on Quantization*, **28**, 1982.

[27] N. Ikeda and S. Watanabe *Stochastic Differential Equations and Diffusion processes, second edition*, North Holland 1989.

[28] Kushner H.J., *Approximation and weak convergence methods for random processes, with applications to stochastic systems theory*, MIT Press Series in Signal Processing, Optimization, and Control, **6**, MIT Press, Cambridge, MA, (1977), 1984, 269 pp.

[29] Kushner H.J., Dupuis P., *Numerical methods for stochastic control problems in continuous time*, 2^{nd} edition, Applications of Mathematics, **24**, Stochastic Modelling and Applied Probability, Springer-Verlag, New York, 2001, 475 pp.

[30] H.J. Kushner, G.G. Yin, *Stochastic Approximations Algorithms and Applications*, Springer, New York, 1997.

[31] S. Kusuoka and D. Stroock, Application of the Malliavin calculus II, J. *Fac. Sci. Univ. Tokyo*, Sect IA Math., **32**, pp.1-76, 1985.

[32] D. Lamberton, B. Lapeyre, *Introduction to stochastic calculus applied to Finance*, Chapman & Hall, London, 1996, 185 pp.

[33] D. Lamberton, Brownian optimal stopping and random walks, technical report Univ. Marne-la-Vallée, 1998 (revised 1999).

[34] D. Lamberton D., B. Lapeyre, Introduction to Stochastic Calculus applied to Finance, Chapman & Hall, 1996, 185p.

[35] P.L. Lions, H. Régnier, Calcul des prix et des sensibilités d'une option américaine par une méthode de Monte Carlo, working paper.

[36] F.A. Longstaff, E.S. Schwartz, Valuing American options by simulation: a simple least-squares approach, *Review of Financial Studies*, **14**, 113-148, 2001.

[37] J. Neveu, *Martingales à temps discret*, Masson, Paris, 1971, 215p.

[38] G. Pagès, A space vector quantization method for numerical integration, *Journal of Applied and Computational Mathematics*, **89**, pp.1-38, 1997.

[39] D. Revuz, M. Yor, *Continuous Martingales and Brownian Motion*, Springer-Verlag, 2^{nd} edition, Berlin-Heidelberg, 1991, 560p.

[40] J.N. Tsitsiklis, B. Van Roy, Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives, *IEEE Trans. Automat. Control*, **44**, $n^0$10, 1999, pp.1840-1851.

[41] S. Villeneuve, A. Zanette (2002), Parabolic A.D.I. methods for pricing american option on two stocks, forthcoming in *Mathematics of Operation Research*.

ANNEX

Proposition 4 *(a) If b, $c \in \mathcal{C}_b^\infty(\mathbb{R}^d)$ and c is uniformly elliptic, then both $(S_{t_k})_{0\le k\le n}$ and $(\overline{S}_{t_k})_{0\le k\le n}$ satisfy the domination property (31) with*

$$\varphi_k := c_{b,\sigma}\sqrt{t_k} \quad (c_{b,\sigma}>0) \qquad \textit{and} \qquad R := Z \sim \mathcal{N}(0; I_d).$$

(b) In the extended Black & Scholes model (1), if $\sigma \in \mathcal{C}_b^\infty(\mathbb{R}^d)$ is uniformly elliptic, then both $(S_{t_k})_{0\le k\le n}$ and $(\overline{S}_{t_k})_{0\le k\le n}$ satisfy the domination property (31) with

$$\varphi_k := c_{b,\sigma}\sqrt{\frac{t_k}{T}} \quad (c_{b,\sigma}>0) \qquad \textit{and} \qquad R := \left(s_0^i\,\psi(\sqrt{T}Z^i)\right)_{1\le i\le d},\ Z \sim \mathcal{N}(0; I_d),$$

where $\psi(u) := \left(u^i + e^{u^i}\right)_{1\le i\le d}$, $u = (u^1, \ldots, u^d) \in \mathbb{R}^d$.

Proof: (a) cf. [4], Theorem 4.

(b) One starts from the obvious inequality, valid for every u, $v \in \mathbb{R}$ and every $\rho > 0$,

$$|e^{\rho v} - e^{\rho u}| \le \rho|v + e^v - (u + e^u)|. \tag{84}$$

The diffusion $Y_t := \ln S_t$ starting at 0 is clearly a diffusion with diffusion coefficient $\sigma(S_t)$, hence $\ln S_t$ is uniformly elliptic. It follows from item (a) that the density function $\pi_{\ln S_{t_k}}$ satisfies

$$\pi_{\ln S_{t_k}}(y) \le \alpha\,\pi_{\sqrt{\beta t_k}\,Z}(y), \qquad (\alpha,\,\beta>0).$$

Consequently, if $X_k := S_{t_k}$ starting now at $X_0 := s_0 > 0$, one has for every N-tuple $x \in (\mathbb{R}_+^d)^N$

$$D_N^{X_k,p}(x) = \mathbb{E}\left(\min_{1\le i\le N} |(s_0^\ell e^{Y_{t_k}^\ell})_{1\le \ell\le d} - x_i|^p\right) \le \alpha\,\mathbb{E}\left(\min_{1\le i\le N} |(s_0^\ell e^{\beta\, t_k Z_{t_k}^\ell})_{1\le \ell\le d} - x_i|^p\right).$$

Now, one easily derives (with obvious notations) that

$$\underline{D}_N^{X_k,p} \le \alpha \inf_{y\in(\mathbb{R}^d)^N} \mathbb{E}\left(\min_{1\le i\le N}\left|\left(s_0^\ell\,(e^{\beta\, t_k Z_{t_k}^\ell} - e^{\beta\, t_k y_i^\ell})\right)_{1\le \ell\le d}\right|^p\right).$$

For every $i \in \{1, \ldots, n\}$, Inequality (84) yields

$$\sum_{\ell=1}^d (s_0^\ell)^2\left(e^{\sqrt{\beta\, t_k}Z^\ell} - e^{\sqrt{\beta\, t_k}y_i^\ell}\right)^2 \le \left(\sqrt{\frac{t_k}{T}}\right)^p \sum_{\ell=1}^d (s_0^\ell)^2\left(\psi(\sqrt{T}Z^\ell) - \psi(\sqrt{T}y_i^\ell)\right)^2$$

which finally yields the expected result since $u \mapsto s_0^\ell\psi(\sqrt{T}u)$ is a bijective from $\mathbb{R}$ onto $\mathbb{R}_+^*$. $\diamond$

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

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

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