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

Vlad BALLY * Gilles PAGÈS † Jacques PRINTEMS‡

July 19, 2002

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 ([22]). An algorithm to compute 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 theoretical orders are probably pessimistic.

Premia 5

Key words: American option pricing, Optimal Stopping, Snell envelope, quantization of random vectors.

2001 AMS classification: 91 B 28, 60G40, (65 C 05, 65 C 20, 93E23, 65 N 50).

1 Introduction and reference model

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:

*Projet MathFi, INRIA. `bally@ccr.jussieu.fr`

†Labo. de Probabilités et Modèles aléatoires, CNRS UMR 7599, Université Paris 6, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5. `gpa@ccr.jussieu.fr`.

‡Centre de Mathématiques & CNRS UMR 8050, Université Paris 12, 61, avenue du Général de Gaulle, F-94010 Créteil. `printems@univ-paris12.fr`

– 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 time-space grid (see *e.g.* [30] 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 [25] and also [27]) in which the Markov chain approximation was first introduced, including its links with the finite difference method. Concerning the consistency of time discretization, see [29].

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 tree into multinomial trees, the difficulty comes from the geometric shape of a tree compatible with all the dimension and correlation constraints.

More recently, 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 [33], the algorithm devised by Longstaff & Schwartz is based on conditional expectation approximation by regression along a finite truncation $(\varphi_i(S_t))_{i\in I}$ of an orthogonal basis $(\varphi_k(S_t))_{k\geq 1}$ of $L^2(\sigma(S_t), \mathbb{P})$. In [37], 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 [19] and [20] Fournié et al. initiated an approach based on Malliavin calculus, to compute conditional expectations and their derivatives with respect to a parameter. This leads to a pure Monte Carlo method. Lions and Régnier in [32] use the same approach to price American options and compute their Greeks.

In this paper, we propose and study a probabilistic method based on grids like in the original finite state Markov chain approximation method. First, we discretize the asset price process at times $t_k := kT/n,\ k = 0, \ldots, n$ (if necessary, we introduce the Euler scheme of the price price diffusion process, still denoted S_{t_k} for a while). The key point is that we will not settle these grids *a priori*: we will use our ability to simulate large samples of $(S_{t_k})_{0\leq k\leq n}$ to produce at each time t_k a grid Γ_k^* which is optimal among all the grids with size N_k in the following sense: 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 $|Z(\Omega)| \leq N_k$. Namely

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

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, $\pi^{\Gamma^*_k}$ or $\pi^{\Gamma^*_k}(S_{t_k})$ are 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. It has been extensively investigated in Signal Processing and Information Theory for more than 50 years (see [23] or more recently [22]). Thus, one knows that it goes to 0 at a $O(N_k^{-\frac{1}{d}})$ rate as $N_k \to \infty$.

Except in some specific 1-dimensional cases of little numerical interest, no closed form is available neither for the optimal grid Γ^*_k, nor for the induced lowest quantization error. In fact little is known on the geometric structure of these grids in higher dimension. However, starting from the integral representation (valid for any grid Γ)

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

and using its regularity properties as an almost everywhere differentiable symmetric function of Γ, one may implement a stochastic gradient algorithm that converges to some (locally) optimal grid. Furthermore, the algorithm yields as by-products the weights ($\mathbb{P}_{S_{t_k}}$*-mass of the Voronoi tessels* of the grid and the quantization error) involved in the pricing of the American option (see subsection 2.2). Thus, Fig.1 illustrates on the bivariate Normal distribution that an optimal grid gets concentrated on heavily weighted areas.

The paper is organized as follows. Section 2 of the paper is devoted to the description of the *quantization tree algorithm for pricing American options*, to the study of its theoretical rate of convergence. Then, its optimization and the algorithmic aspects to achieve this optimization are developed. This section is partially adapted from a general discretization method devised for Reflected Backward Stochastic Differential Equations (*RBSDE*) in [3].

Time discretization (subsection 2.1) amounts to approximating a continuously exercisable American option by its *Bermuda* counterpart to be exercised only at discrete times t_k, $k = 0, \ldots, n$. The theoretical premium of the Bermuda option satisfies a backward dynamic programming formula. The quantization tree algorithm is defined in subsection 2.2: it simply consists in plugging the optimal quantizers $\widehat{S}_{t_k} := \pi^{\Gamma^*_k}(S_{t_k})$ of the S_{t_k}'s in this formula. Some weights appears that are obtained by the stochastic grid optimization procedure mentioned above. In subsection 2.3, the rate of convergence of this algorithm is derived for Lipschitz continuous pay-offs as a function of the time discretization step T/n and of the L^p-quantization errors $\|S_{t_k} - \pi_k^{\Gamma^*}(S_{t_k})\|_p$, $k = 1, \ldots n$. Then a short background on optimal quantization, its asymptotics is provided in subsection 2.4. In subsection 2.5, the algorithm to optimize the grids of the quantization tree is detailed. The last subsection deals with 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).

In Section 3, we design an approximating *quantized hedging strategy* following the ideas by Föllmer & Sondermann on incomplete markets. We are in position to estimate the induced some error bounds, called *local residual risks* of the quantization tree algorithm. To this is the aim of Section 4. To this end, we combine some methods borrowed from *RBSDE* Theory, analytical techniques for p.d.e. and quantization theory. We get a global rate of convergence for the hedging strategy which seems to be the first of that kind.

Section 5 is devoted to the experimental validation of the method. We present extensive 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 that obtained with usual grid methods. The tests are carried out on multi-dimensional American exchange options on (geometric) index in a standard d-dimensional decorrelated Black & Scholes model. This rate, even better than forecast by theory, makes up for the drawback of an irregular approximation. Two settings have been selected for simulation: one "in-the-money" and one "out-of-the-money", both in several dimensions $d = 2, 4, 6, 10$. In the worst case ($d=10$) case, the computed premia remain within $3,5\%$ of the reference price.

Before going into technicalities, one may mention an obvious methodological difference between the quantization tree algorithm and the regression method [33]. The Longstaff-Schwartz approach makes the choice of a *smooth but global* approximation whereas we privilege an *irregular (piecewise constant) but local* approximation. Among the expected advantages of the local feature of quantization approximation, a prominent one is that it may lead to higher order approximations of the price, involving the space derivatives *i.e.* the hedging (see *e.g.* [6] for a first approach in that direction). A second asset, probably the most important for operating applications, is that, once the asset price process has been appropriately quantized, it can almost instantly price all possible American (vanilla) payoffs without any further Monte Carlo simulations. Finally, when the diffusion process is a function of the Brownian motion like in the Black & Scholes model, the quantization tree algorithm becomes completely parameter free: it suffices to call upon some quantization grids of multi-variate Normal distributions, possibly stored on a CD-Rom for ever.

THE REFERENCE MODEL We consider a market on which are traded d risky assets $S^1, \ldots, S^d$ and a deterministic riskless asset $S_t^0 := e^{rt}$, $r \in \mathbb{R}$ 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

$$dS_t^i \;=\; S_t^i(r\,dt + \sum_{1\le j\le q} \sigma_{ij}(e^{-rt}S_t)\,dW_t^j),\; S_0^i := s_0^i > 0,\; 1 \le i \le d \tag{1}$$

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

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

The filtration of interest will be the natural (completed) filtration $\underline{\mathcal{F}} := (\mathcal{F}_t^S)_{t\in[0,T]}$ 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 $p \ge 1$,

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

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

$$d\widetilde{S}_t = c(\widetilde{S}_t).dW_t^j,\; \widetilde{S}_0 := s_0. \tag{3}$$

$\mathbb{P}$ 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*. Under Assumptions (1) and (4), one has

$$\mathbb{E}(\sup_{t\in[0,T]} |h_t|^p) < +\infty \qquad \text{for every } p \ge 1.$$

One shows that – in a complete market – the fair price $\mathcal{V}_t$ at time t for this contract is

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

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, [30]) 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.

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 $d\widetilde{S}_t = c(S_t).dW_t$).

2 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.

2.1 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.*, [3] for a proof). In fact this discrete time markovian setting is the natural framework for the method. In fact, throughout this section, the generic notation $(X_k)_{0\le k\le n}$ will denote indifferently S_{t_k} or $\overline{S}_{t_k}$ (and more generally any L^p-integrable homogeneous $\mathcal{F}_{t_k}$-Markov chain whose transition $P^{(n)}$ satisfies (9). The $\mathcal{F}_{t_k}$-Snell envelope of $h(t_k, X_k)$, denoted by $(V_k)_{0\le k\le n}$, is 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. It $(V_k)_{0\le k\le n}$ satisfies the so-called backward *dynamic programming formula* (see [34]):

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

One derives using the Markov property 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

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

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 appear in [15] for pricing one-dimensional American options. See also [31] 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 $[b]_{Lip}$, $[c]_{Lip}$, $[h]_{Lip}$ and p such that*

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

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

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

2.2 Space discretization: the quantization tree

Our aim is to discretize the random variables X_k by some random variables $\widehat{X}_k$ that can only take a finite number N_k of values. Then, we wish to approximate the dynamic programming formula satisfied by the true Snell envelope $(V_k)_{0\le k\le n}$ with the $\widehat{X}_k$'s.

2.2.1 Abstract quantization of a random vector X, L^p-distortion

Let $X \in L^p_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$. From a probabilistic point of view, L^p-quantization ($p \ge 1$) consists in studying the best L^p-approximation of X by a random vectors $X' = q(X)$ where $q : \mathbb{R}^d \to \mathbb{R}$ is a Borel function taking at most N values $x_1, \cdots, x_{_N} \in \mathbb{R}^d$. 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 $\|X - q(X)\|_p$ reaches its minimum at any q such that $(\{q = x_i\})_{1\le i\le N}$ makes up a *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 Borel partition $C_1(x),\ldots,C_N(x)$ of $\mathbb{R}^d$ is a Voronoi tessellation of the N-tuple x if, for every $i\in\{1,\ldots,N\}$, $C_i(x)$ satisfies*

$$C_i(x) \subset \{y\in \mathbb{R}^d \,/\, |x_i-y| = \min_{1\le j\le N} |y-x_j|\} \quad \textit{where } |\,.\,| \textit{ is for the Euclidean norm on } \mathbb{R}^d.$$

One denotes $\widehat{X}^x = q(X) = \sum_{1\le i\le N} x_i \mathbf{1}_{C_i(x)}(X)$ the corresponding random vector, called a Voronoi quantizer. *One often drops the exponent ^{x} to note $\widehat{X}$.*

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. If the distribution of X weights no hyperplane, then the Voronoi tessellation is $\mathbb{P}_X$- essentially unique and all all the Voronoi quantizers $\widehat{X}$ have the same distribution.

The problem is then to estimate the L^p*-mean quantization error* $\|X-\widehat{X}\|_p$. Let $\mathbb{P}_X$ denote the distribution of X. Then, the L^p-quantization error is given by

$$\|X-\widehat{X}\|_p^p = \sum_{i=1}^{N} \mathbb{E}\left(\mathbf{1}_{C_i(x)}|X-x_i|^p\right) = \mathbb{E}\left(\min_{1\le i\le N}|X-x_i|^p\right) = \int_{\mathbb{R}^d} \min_{1\le i\le N} |x_i-y|^p \mathbb{P}_X(dy). \quad (15)$$

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

The optimization phase consists in choosing the N-tuple $x := (x_1,\ldots,x_N)$ which achieves *the smallest possible L^p-quantization error* and then *to evaluate how fast it goes to* 0 *as* $N\to\infty$. This will be investigated further on in subsection 2.4, once the way we use quantization and its related error will have been developed.

2.2.2 Quantization tree and quantized pseudo-Snell envelope

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 quantized 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

$$\left\{\begin{array}{lll} \widehat{V}_n & := & h(t_n,\widehat{X}_n),\\ \widehat{V}_k & := & \max\left(h(t_k,\widehat{X}_k), \mathbb{E}(\widehat{V}_{k+1}/\widehat{X}_k)\right), \quad 0\le k\le n-1. \end{array}\right. \quad (16)$$

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 $i\in\{1,\ldots,N_k\}$,*

$$\alpha_i^k := \mathbb{P}(\widehat{X}_k = x_i^k) = \mathbb{P}(X_k\in C_i(x^k)), \quad (17)$$

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

$$\begin{aligned}\pi^k_{ij} &:= \mathbb{P}(\widehat{X}_{k+1} = x^{k+1}_j / \widehat{X}_k = x^k_i) = \mathbb{P}\left(X_{k+1} \in C_j(x^{k+1}) / X_k \in C_i(x^k)\right)\\ &= \frac{\beta^k_{ij}}{\alpha^k_i} \quad \textit{with} \quad \beta^k_{ij} := \mathbb{P}\left(X_{k+1} \in C_j(x^{k+1}),\ X_k \in C_i(x^k)\right).\end{aligned} \tag{18}$$

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

$$\begin{aligned}\widehat{v}_n(x^n_i) &:= h_n(x^n_i), \qquad i \in \{0,\ldots,N_n\}\\ \widehat{v}_k(x^k_i) &:= \max\left(h(t_k, x^k_i), \sum_{j=1}^{N_{k+1}} \pi^k_{ij}\, \widehat{v}_{k+1}(x^{k+1}_j)\right), 1 \le i \le N_k,\ 0 \le k \le, n-1.\end{aligned} \tag{19}$$

Then, $\widehat{V}_k = \widehat{v}_k(\widehat{X}_k)$ *satisfies the above dynamic programming (16) 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 (19) on a computer raises two questions:

– How is it possible to estimate the parameters α^k_i and β^k_{ij} involved in (19) ?

– 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 (17) and (18). An estimate of the L^p-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 2.4 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 (19) 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 always satisfies

$$\text{Complexity} \ge C.\frac{N^2}{n+1}.$$

This purely combinatorial lower bound 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).

2.3 Convergence and rate using L^p-quantization error

The aim of this paragraph is to provide some *a priori* L^p-error bounds for $\|V_k - \widehat{V}_k\|_p$, $0 \le k \le n$, based on the L^p-quantization errors *i.e.* $\|X_k - \widehat{X}_k\|_p$, $0 \le k \le n$ where quantizer $\widehat{X}_k$ is a Voronoi quantizer that takes N_k values $x_1^k, \ldots, x_{N_k}^k$. This error modulus can be obtained as a by-product of a Monte Carlo simulation of $(X_k)_{0\le k\le n}$: it only requires to compute, for every μ_k-distributed simulated random vector, its distance to its closest neighbour in the set $\{x_1^k, \ldots, x_{N_k}^k\}$.

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

$$[Pg]_{Lip} \le K[g]_{Lip} \quad \text{where} \quad [g]_{Lip} := \sup_{x\neq y} \frac{|g(x) - g(y)|}{|x - y|}. \tag{20}$$

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

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

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

Proof: STEP 1: We need to show that the functions v_k defined by (11) are Lipschitz continuous and

$$[v_k]_{Lip} \le (K \vee 1)^{n-k}[h]_{Lip}. \tag{21}$$

Clearly, $[v_n]_{Lip} \le [h]_{Lip}$. Then, one concludes by induction, using that $|\max(a, b) - \max(a', b')| \le \max(|a - a'|, |b - b'|)$: dynamic programming formula (11) yields that

$$[v_k]_{Lip} \le \max\left([h]_{Lip}, [P(v_{k+1})]_{Lip}\right) \le \max\left([h]_{Lip}, K[v_{k+1}]_{Lip}\right)$$

STEP 2: Set $\Phi_k := P(v_{k+1})$ for every $k \in \{0, \ldots, n-1\}$ (and $\Phi_n \equiv 0$). The function Φ_k satisfies $\mathbb{E}(v_{k+1}(X_{k+1})/\mathcal{F}_{t_k}) = \mathbb{E}(v_{k+1}(X_{k+1})/X_k) = \Phi_k(X_k)$. One defines similarly $\widehat{\Phi}_k$ by the equality $\widehat{\mathbb{E}}_k(\widehat{v}_{k+1}(\widehat{X}_{k+1})/\widehat{X}_k) := \widehat{\Phi}_k(\widehat{X}_k)$ (and $\widehat{\Phi}_n \equiv 0$). Then

$$\begin{aligned} |V_k - \widehat{V}_k| &\le |h_k(X_k) - h_k(\widehat{X}_k)| + |\Phi_k(X_k) - \widehat{\Phi}_k(\widehat{X}_k)| \\ &\le [h]_{Lip}|X_k - \widehat{X}_k| + |\Phi_k(X_k) - \widehat{\mathbb{E}}_k(\Phi_k(X_k))| + |\widehat{\mathbb{E}}_k(\Phi_k(X_k)) - \widehat{\Phi}_k(\widehat{X}_k)| \end{aligned} \tag{22}$$

$$\begin{aligned} \text{Now} \quad |\Phi_k(X_k) - \widehat{\mathbb{E}}_k\Phi_k(X_k)| &\le |\Phi_k(X_k) - \Phi_k(\widehat{X}_k)| + \widehat{\mathbb{E}}_k|\Phi_k(X_k) - \Phi_k(\widehat{X}_k)| \\ &\le [\Phi_k]_{Lip}\left(|X_k - \widehat{X}_k| + \widehat{\mathbb{E}}_k|X_k - \widehat{X}_k|\right). \\ \text{Hence,} \quad \|\Phi_k(X_k) - \widehat{\mathbb{E}}\Phi_k(X_k)\|_p &\le 2[\Phi_k]_{Lip}\|X_k - \widehat{X}_k\|_p. \end{aligned}$$

When $p=2$, the very definition of the conditional expectation as a projection in a Hilbert space implies that one may remove the factor 2 in the inequality.

$$\text{Now} \qquad \widehat{\mathbb{E}}_k(\Phi_k(X_k)) - \widehat{\Phi}_k(\widehat{X}_k) = \widehat{\mathbb{E}}_k\left(\mathbb{E}(v_{k+1}(X_{k+1})/X_k)\right) - \widehat{\mathbb{E}}_k\left(\widehat{v}_{k+1}(\widehat{X}_{k+1})\right) = \widehat{\mathbb{E}}_k\left(v_{k+1}(X_{k+1}) - \widehat{v}_{k+1}(\widehat{X}_{k+1})\right)$$

since $\widehat{X}_k$ is $\sigma(X_k)$-measurable. Conditional expectation being a L^p-contraction, it follows

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

Finally, it follows from the above inequalities and (22) that

$$\|V_k - \widehat{V}_k\|_p \le ([h]_{Lip} + c[\Phi_k]_{Lip})\|X_k - \widehat{X}_k\|_p + \|V_{k+1} - \widehat{V}_{k+1}\|_p, \quad k \in \{0,\ldots,n-1\}.$$

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

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

The definition of Φ_i and the K-Lipschitz property of $P(x,dy)$ complete the proof since

$$[\Phi_i]_{Lip} = [P(v_{i+1})]_{Lip} \le K[v_{i+1}]_{Lip}. \quad \diamond$$

2.4 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 some x^*. One may always assume that $x^* \in (\mathcal{H}(\operatorname{supp}\mathbb{P}_X))^N$ (convex hull of $\operatorname{supp}\mathbb{P}_X$). 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 [35] or [22], among others), still lying in $(\mathcal{H}(\operatorname{supp}\mathbb{P}_X))^N$. Furthermore, one shows the following simple facts (see [35] or [22] and references therein):

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

– If supp $\mathbb{P}_X$ is *everywhere dense in its convex hull*, then the N components of an optimal N-tuple x^* all lies in $\mathcal{H}(\operatorname{supp}\mu)$. 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 [23], Bucklew & Wise, see [12] and finally Graf & Luschgy see [22]). It reads as follows

Theorem 3 *(Asymptotics of optimal quantization) Assume that* $\mathbb{E}|X|^{p+\eta} < +\infty$ *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 [23]). Nevertheless some reasonable bounds are available, based on random quantization (see [14]), 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 $U([0,1])$ distribution), optimal quantization provides the best possible "grid method" for a given distribution μ.

2.5 Optimal quantization: how to get it?

Optimal quantization of a single random vector: how to get it? When $x = \{x^1, \ldots, x^N\}$, Equation (15) 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.*, [22] when $p = 2$ or [35]) 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

$$\begin{aligned}
\nabla D_N^p &:= \left(\mathbb{E}\frac{\partial D_N^p}{\partial x^i}(y, X)\right)_{1\le i\le n} = \left(\int_{\mathbb{R}^d} \frac{\partial D_N^p}{\partial x^i}(y,u)\mathbb{P}_X(du)\right)_{1\le i\le n} \\
\text{where}\quad \frac{\partial D_N^p}{\partial x^i}(y,u) &:= p\frac{u - x^i}{|u - x^i|}|u - x^i|^{p-1}\mathbf{1}C_i(y)(u),\ 1 \le i \le n.
\end{aligned}$$

(The above result still holds when $p = 1$ if $\mathbb{P}_Y$ 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 integrals with respect ...to $\mathbb{P}_X$. This stochastic gradient descent is defined as follows: let $(\xi^t)_{t\in\mathbb{N}^*}$ be a sequence of i.i.d. $\mathbb{P}_X$-distributed random variables and let $(\gamma_t)_{t\in\mathbb{N}^*}$ be a sequence of positive steps satisfying

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

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

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

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

(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 [16]) are not fulfilled by D_N^p (see, $e.g.$ [16] or [26] 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 [35] 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 $x^t := \{x^{1,t}, \dots, x^{N,t}\}$,

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

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

Furthermore, it is established in [35] that, if $X \in L^{p+\varepsilon}$ ($\varepsilon > 0$), on the event $\{x^t \to x^*\}$

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

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

These "companion" – hence costless – procedures yield the parameters (weights of the Voronoi cells, L^p-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.

Optimization of the quantization tree: the *CLVQ* algorithm The principle is to modify a Monte Carlo simulation of the chain $(X_k)_{0 \le k \le n}$ by processing a *CLVQ* algorithm at each time step k. One starts from a large scale Monte Carlo simulation of the Markov chain $(X_k)_{0 \le k \le n}$ $i.e.$ independent copies $Z^0 := (Z_0^0, \dots, Z_n^0), Z^1 := (Z_0^1, \dots, Z_n^1), \dots, Z^t := (Z_0^t, \dots, Z_n^t), \dots$ of $(X_k)_{0 \le k \le n}$. Our aim is now to produce for every $k \in \{0, \dots, n\}$ some (almost) optimal N_k-tuple $\Gamma_k^* := (x_1^{k,*}, \dots, x_{N_k}^{k,*})$ with size N_k, their transition kernels $[\pi_{ij}^k]$, their weight vectors $(\alpha_i^k)_{0 \le i \le N_k}$ and the induced quantization errors. Note that, if one set

$$\beta_{ij}^k := \mathbb{P}\left(\{X_{k+1} \in C_j(\Gamma_{k+1})\} \cap \{X_k \in C_i(\Gamma_k)\}\right)$$

then $\pi_{ij}^k = \frac{\beta_{ij}^k}{\alpha_i^k}$. So one can focus on the estimation of the weight vectors β_{ij}^k.

Then the algorithm is as follows

1. *Initialization phase:*

• Initialize the $n+1$ starting N_k-tuples $\Gamma_k^0 := \{x_1^{0,k}, \ldots, x_{N_k}^{0,k}\}$, $0 \le k \le n$, of the $n+1$ CLVQ algorithms that will quantize the distributions $\mathcal{L}(X_k)$.

• Initialize the weight vectors $\alpha_i^{k,0} := 0$, $1 \le i \le N_k$ for every $k \in \{0, \ldots, n\}$.

• Initialize the transitions $\beta_{ij}^{k,0} := 0$, $i \in \{1, \ldots, N_k\}$, $j \in \{1, \ldots, N_{k+1}\}$, $0 \le k \le n-1$.

2. *Updating* $t \rightsquigarrow t+1$*:* At step t, the $n+1$ N_k-tuples Γ_k^t, $0 \le k \le n$, have been obtained. We use now the sample Z^{t+1} to carry on the optimization process *i.e.* building up the Γ_k^{t+1}'s as follows. For every $k = 0$ up to n

• Simulation of Z_k^{t+1} (using Z_{k-1}^{t+1} if $k \ge 1$)

• Selection of the "winner" in the k^{th} CLVQ algorithm *i.e.* the only index $i_k^{t+1} \in \{1, \ldots, N_k\}$ satisfying

$$Z_k^{t+1} \in C_{i_k^{t+1}}(\Gamma_k^t)$$

• Updating of the k^{th} CLVQ algorithm:

$$\Gamma_{k,i}^{t+1} = \Gamma_{k,i}^t - \gamma_{t+1}(\mathbf{1}_{\{i=i_k^{t+1}\}}(\Gamma_{k,i}^t - Z_k^{t+1}))_{1 \le i \le N_k}.$$

• Updating of the k^{th} weight vector $\alpha^{k,t} := (\alpha_i^{k,t})_{1 \le i \le N_k}$:

$$\forall\, i \in \{1, \ldots, N_k\}, \quad \alpha_i^{k,t+1} := \alpha_i^{k,t} + \mathbf{1}_{\{i=i_k^{t+1}\}}.$$

• Updating of the (quadratic) quantization error estimator $D^{k,t}$:

$$D^{k,t+1} := D^{k,t} - \frac{1}{t+1}(|\Gamma_{k,i_k^{t+1}}^t - Z^{t+1}|^2 - D^{k,t})$$

• Updating of the weight vectors $\beta^{k,t} := (\beta_{ij}^{k,t})_{1 \le i \le N_{k-1}, 1 \le j \le N_k}$ $(k \ge 1)$

$$\forall\, i \in \{1, \ldots, N_{k-1}\}, \forall\, j \in \{1, \ldots, N_k\},\ \beta_{ij}^{k-1,t+1} := \beta_{ij}^{k-1,t} + \mathbf{1}_{\{i=i_{k-1}^{t+1},\, j=i_k^{t+1}\}}.$$

• Updating the transition kernels $(\pi_{ij}^{k,t})_{1 \le i \le N_{k-1}, 1 \le j \le N_k}$ $(k \ge 1)$

$$\pi_{ij}^{k,t+1} := \frac{\beta_{ij}^{k,t+1}}{\alpha_i^{k,t+1}} \qquad \text{(possibly only at the end of the simulation process!).}$$

One shows, see [3], that on the event $\{\Gamma_k^t \to \Gamma_k^*\}$, $D^{k,t} \overset{t \to +\infty}{\longrightarrow} D_{N_k}^{\mu_k,2}(\Gamma_k^*)$ and

$$\alpha_k^t \longrightarrow \alpha_k^* = (\mu(C_i(\Gamma_k^*)))_{1 \le i \le N_k} \qquad \text{(since } \mathcal{L}(X_k) \text{ is continuous).}$$

Actually one shows, using the same classical martingale approach, that

$$\beta_{k-1}^t \longrightarrow \beta_{k-1}^* = \big(\mathbb{P}(X_{k-1} \in C_i(\Gamma_{k-1}^*),\ X_k \in C_j(\Gamma_k^*))\big)_{1 \le i \le N_{k-1}, 1 \le j \le N_k} \tag{30}$$

on the event $\{\Gamma_{k-1}^t \longrightarrow \Gamma_{k-1}^*\} \cap \{\Gamma_k^t \longrightarrow \Gamma_k^*\}$.

The main features of this algorithms are essentially those of the regular CLVQ algorithm. Note only that the successive optimizations of the quantization N_k-tuples are not recursive, so there is no deterioration of the process when k increases.

2.6 *A priori* error bounds in time and space

$(\overline{S}_{t_k})_{0\le k\le n}$. Let $(V_k)_{0\le k\le n}$ denote the $\widehat{X}_k$ denote a Voronoi quantizer of Next theorem provides a general error bound for $\|V_k - \widehat{v}_k(\widehat{X}_k)\|_p$ as a function of the quantization errors $\|X_k - \widehat{X}_k\|_p$ (optimality of the Voronoi quantizers $\widehat{X}_k$ is not required). Proposition 2 below is an improvement of Theorem 2 when $X_k = S_{t_k}$ or $\overline{S}_{t_k}$ in which the constant do not depend on n.

Proposition 2 *Assume that the coefficients b and c of the diffusion (7) and the obstacle h are Lipschitz continuous. Let $p\in[1,+\infty)$. There exists $K_{b,\sigma,h,T,p} > 0$ such that*

$$\forall\, n\in\mathbb{N}^*,\ \forall\, k\in\{0,\ldots,n\},\quad \|V_k - \widehat{v}_k(\widehat{X}_k)\|_p \le K_{b,\sigma,h,T,p}\sum_{\ell=k}^{n}\|X_\ell - \widehat{X}_\ell\|_p \tag{31}$$

where $(\widehat{v}_k(\widehat{X}_k))_{0\le k\le n}$ is the pseudo-Snell envelope of $(h(t_k, X_k))_{0\le k\le n}$ defined by (16).

One gets rid of n since the Lipschitz coefficient $K^{(n)}$ of both chains (S_{t_k}) and $(\overline{S}_{t_k})$ satisfy $\limsup_n (K^{(n)})^n < +\infty$ (see [3] for details).

To go further we need a new assumption on the distributions of the X_k's: namely the uniform φ-domination of the quantization errors $\|X_k - \widehat{X}_k\|_p$ in the following sense: there exists a random variable $R\in L^{p+\eta}(\eta>0)$ and a sequence $(\varphi_{k,n})_{0\le k\le n<+\infty}$ such that

$$\forall\, n\ge 1,\ \forall\, k\in\{0,\ldots,n\},\ \forall\, N\in\mathbb{N}^*,\ \min_{x\in(R^d)^N}\|X_k - \widehat{X}_k^x\|_p \le \varphi_{k,n}\min_{x\in(R^d)^N}\|R - \widehat{R}^x\|_p. \tag{32}$$

The point is that the distribution of R may depend on p but *not on N, k or n*. It is shown in [3] (Theorem 3) that uniformly elliptic diffusions with smooth and bounded coefficients satisfy the domination property (32) with $\varphi_{k,n} := c\sqrt{t_k} = c\sqrt{k/n}$. It is shown in the Appendix that, if $q\ge d$ and σ is smooth and uniformly elliptic, then the extended B & S

model (1) is uniformly dominated by $c\sqrt{k/n}$ and $\mathcal{N}(0; I_d)$ in the sense of (32).

Combining the bounds obtained in Theorem 1 (time) and Proposition 2 (space) with the Theorem 3 (asymptotics of optimal quantization) yield an error structure looking like

$$\frac{C_1}{n^\theta} + C_2\sum_{k=1}^{n}\sqrt{t_k}\,N_k^{-\frac{1}{d}} \quad \text{with} \quad N_1+\cdots+N_n = N-1 \tag{33}$$

(the last equality being up to n) N_k denotes the size of the optimal quantizer $\widehat{X}_k$ at the k^{th} layer (time 0 is excluded since $\widehat{X}_0 := s_0$ perfectly quantizes $S_0 = s_0$). Minimizing the right hand of the sum is an easy optimization problem with constraint. Then, in order to minimize (33), one has to make a balance between the time and space discretization errors. The results are detailed in Theorem 4 below.

Theorem 4 *(Optimized quantization tree and resulting error bounds) Assume that all the assumptions of Proposition 2 hold and that the $(S_{t_k})_{0\le k\le n}$ is dominated in the sense of (32) by $\varphi_k := c\sqrt{t_k}$. Let $n\ge 1$, $N\ge n+1$. For every $k\in\{1,\ldots,n\}$, set*

$$N_k := \left\lceil \frac{t_k^{\frac{d}{2(d+1)}}N}{t_1^{\frac{d}{2(d+1)}}+\cdots+t_k^{\frac{d}{2(d+1)}}+\cdots+t_n^{\frac{d}{2(d+1)}}}\right\rceil,\ (N\le N_1+\cdots+N_n\le N+n+1). \tag{34}$$

Assume that, in (a) and (b) below, the Voronoi quantizer $\widehat{X}_k$ has size N_k and is L^p-optimal.

(a) EULER SCHEME:
Then,

$$\|\mathcal{V}_{t_k} - \widehat{v}_k(\widehat{X}_k)\|_p \le C_p(s_0)\left(\frac{1}{\sqrt{n}} + \frac{n^{1+\frac{1}{d}}}{N^{\frac{1}{d}}}\right) \quad \textit{with} \quad C_p(s_0) \le C_p e^{C_p T}(1+|s_0|).$$

where $(v_k(\widehat{X}_k))$ *is the quantized pseudo-Snell envelope of* $(\overline{S}_{t_k})$. *If, furthermore* $n := \left\lceil \left(\frac{d}{2(d+1)}\right)^{\frac{2d}{3d+2}} N^{-\frac{1}{2d+1}} \right\rceil$, *then* $\|\mathcal{V}_{t_k} - \widehat{V}^n_{t_k}\|_p \le C'_p(s_0) N^{-\frac{1}{3d+2}} = O\left(\frac{1}{\sqrt{n}}\right)$.

(b) DIFFUSION: *If the obstacle* h *is semi-convex (and if* $X_k := S_{t_k}$*), then*

$$\|\mathcal{V}_{t_k} - \widehat{v}_k(\widehat{X}_k)\|_p \le C_p(s_0)\left(\frac{1}{n} + \frac{n^{1+\frac{1}{d}}}{N^{\frac{1}{d}}}\right) \quad \textit{with} \quad C_p(s_0) \le C_p e^{C_p T}(1+|s_0|).$$

where $(v_k(\widehat{X}_k))$ *is the quantized pseudo-Snell envelope of* (S_{t_k}). *If furthermore* $n := \left\lceil \left(\frac{d}{2(d+1)}\right)^{\frac{d}{2d+1}} N^{-\frac{1}{2d+1}} \right\rceil$, *then* $\|\mathcal{V}_{t_k} - \widehat{V}^n_{t_k}\|_p \le C'_p(s_0) N^{-\frac{1}{2d+1}} = O\left(\frac{1}{n}\right)$.

3 Hedging

Tackling the question of hedging American options needs to go deeper in financial modeling, 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 $(q \ge d)$ and

$$\forall\, x \in \mathbb{R}^d,\ \sigma\sigma^*(x) \ge \varepsilon_0 I_d \tag{35}$$

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

3.1 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 [36]) that

$$h_T = \mathbb{E}(h_T) + \int_0^T H_s.dW_s = \mathbb{E}(h_T) + \int_0^T Z_s.dS_s \tag{36}$$

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 $(Z^i_s)_{1\le i\le d}$ (at every time $s \in [t,T]$ the portfolio contains exactly Z^i_s 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

$$\text{Premium}_t := \mathbb{E}(h_T/\mathcal{F}_t) = \mathbb{E}(h_T) + \int_0^t Z_s.dS_s. \tag{37}$$

If $h_T := h(T, S_T)$, the Markov property of (S_t) implies that $\text{Premium}_t := p(t, S_t)$. 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 Itô 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 for this purpose as soon as the dimension $d \geq 3$.

3.2 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

$$\Delta A^n_k := A^n_k - A^n_{k-1} = V^n_{t_{k-1}} - \mathbb{E}_{t_{k-1}} V^n_{t_k} = \left(h(t_{k-1}, S_{t_{k-1}}) - \mathbb{E}_{t_{k-1}} V^n_{t_k}\right)_+. \tag{38}$$

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

$$M^n_k := \int_0^{t_k} Z^n_s.dS_s, \quad 0 \leq k \leq n, \quad \text{with} \quad \mathbb{E}\int_0^T |c^*(S_s)Z^n_s|^2 ds < +\infty \tag{39}$$

(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

$$\mathcal{P}_n = \left\{(\zeta_s)_{s\in[0,T]},\ \zeta_s := \zeta_{t_k},\ s\in[t_k,t_{k+1}),\ \zeta_{t_k}\ \mathcal{F}_{t_k}\text{-measurable},\ \mathbb{E}\int_0^T |c^*(S_s)\zeta_s|^2 ds < +\infty\right\}. \tag{40}$$

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 $U\in L^2(c^*(S_.)d\mathbb{P}.dt)$

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

Doing so, we follow classical ideas introduced by by Föllmer & Sondermann ([18]) 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 $\text{proj}_n(Z^n_.)$ amounts to minimizing $\mathbb{E}\left(\sum_{k=1}^n\int_{t_k}^{t_{k+1}}|c^*(S_s)(Z^n_s-\zeta_{t_k})|^2 ds\right)$ over $(\zeta_k)_{0\le k\le n}\in\mathcal{P}_n$. Setting $\zeta^n_{t_k} := \text{proj}_n(Z^n_.)$ and standard computations yield

$$\begin{aligned}\zeta^n_{t_k} &:= \left(\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} cc^*(S_s)\,ds\right)^{-1}\mathbb{E}_{t_k}\left(\int_{t_k}^{t_{k+1}} cc^*(S_s)Z^n_s ds\right)\\ &= \left(\mathbb{E}_{t_k}\Delta S_{t_{k+1}}(\Delta S_{t_{k+1}})^*\right)^{-1}\mathbb{E}_{t_k}(\Delta M^n_{k+1}\Delta S_{t_{k+1}}) \qquad (41)\\ &= \left(\mathbb{E}_{t_k}\Delta S_{t_{k+1}}(\Delta S_{t_{k+1}})^*\right)^{-1}\mathbb{E}_{t_k}\left(\Delta V^n_{t_{k+1}}\Delta S_{t_{k+1}}\right). \qquad (42)\end{aligned}$$

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

$$\Delta R^n_{t_{k+1}} := \int_{t_k}^{t_{k+1}}(Z^n_s-\zeta^n_{t_k}).dS_s = \Delta M^n_{k+1}-\zeta^n_{t_k}.\Delta S_{t_{k+1}} \tag{43}$$

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

$$\mathbb{E}_{t_k}|\Delta R^n_{t_{k+1}}|^2 = \mathbb{E}_{t_k}\left(\int_{t_k}^{t_{k+1}}|c^*(S_s)(Z^n_s-\zeta_{t_k})|^2 ds\right),\ k\in\{0,\dots,n-1\}. \tag{44}$$

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

$$\mathbb{E}_{t_k}|\Delta R^n_{t_{k+1}}|^2 = \mathbb{E}_{t_k}|\Delta V^n_{t_{k+1}}-\mathbb{E}_{t_k}\Delta V^n_{t_{k+1}}|^2-(\mathbb{E}_{t_k}\Delta S_{t_{k+1}}\Delta S^*_{t_{k+1}})^{-1}\left(\mathbb{E}_{t_k}\Delta V^n_{t_{k+1}}\Delta S_{t_{k+1}}\right)^2. \tag{45}$$

Formulae (42) or (44), based on S_{t_k} and $V^n_{t_k}$ have natural approximations by quantization. On the other hand, (41) and (44) are more appropriate to produce some *a priori* error bounds (when simulation of the diffusion is possible).

3.3 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 adverse 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

$$\overline{\zeta}^n_{t_k} := (\mathbb{E}_{t_k}\Delta\overline{S}_{t_{k+1}}\Delta\overline{S}^*_{t_{k+1}})^{-1}\mathbb{E}_{t_k}(\Delta\overline{V}^n_{t_{k+1}}\Delta\overline{S}_{t_{k+1}}) = \frac{1}{\Delta t_{k+1}}\,\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} \overline{Z}^n_s ds. \tag{46}$$

The related hedging default and local residual risk are defined by mimicking (44) and (45):

$$\Delta\overline{R}^n_{t_{k+1}} := \int_{t_k}^{t_{k+1}} (\overline{Z}^n_s - \overline{\zeta}^n_{t_k}).d\overline{S}_s = \Delta M^n_{k+1} - \overline{\zeta}^n_{t_k}.\Delta\overline{S}_{t_{k+1}} \tag{47}$$

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

3.4 Quantized hedging and local residual risks

The quantized formulae for strategies and residual risks are simply derived from formulae (42) or (46) 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 2 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)

$$\widehat{\zeta}^n_k := \frac{n}{T}\Big(cc^*(\widehat{S}_{t_k})\Big)^{-1}\widehat{\mathbb{E}}_k\Big((\widehat{v}^n_{k+1}(\widehat{S}_{t_{k+1}}) - \widehat{v}^n_k(\widehat{S}_{t_k}))(\widehat{S}_{t_{k+1}} - \widehat{S}_{t_k})\Big). \tag{49}$$

$$|\Delta\widehat{R}^n_{t_{k+1}}|^2 := \mathbb{E}_{t_k}|\Delta\widehat{V}^n_{t_{k+1}} - \mathbb{E}_{t_k}\Delta\widehat{V}^n_{t_{k+1}}|^2 - (\mathbb{E}_{t_k}\Delta\widehat{S}_{t_{k+1}}\Delta\widehat{S}^*_{t_{k+1}})^{-1}\Big(\mathbb{E}_{t_k}\Delta\widehat{V}^n_{t_{k+1}}\Delta\widehat{S}_{t_{k+1}}\Big)^2 \tag{50}$$

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

4 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.

4.1 From Bermuda to America

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

$$V_t^n := V_{t_k}^n + \int_{t_k}^t Z_s^n dS_s = V_{t_{k+1}}^n - \int_t^{t_{k+1}} Z_s^n dS_s + \Delta A_{k+1}^n,\ t\in[t_k,t_{k+1}). \tag{51}$$

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

$$V_{t_k-}^n = V_{t_k}^n + \Delta A_{k+1}^n. \tag{52}$$

Proposition 3 *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 $k\in\{0,\dots,n\}$, $V_{t_k}^n \le \mathcal{V}_{t_k}$ and for every $t\in(t_k,t_{k+1})$, $\quad (V_t^n-\mathcal{V}_t)_+ \le \Delta A_{k+1}^n$.

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

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

$$\mathbb{E}\left(\int_0^T |c^*(S_s)(Z_s-Z_s^n)|^2 ds\right) + \mathbb{E}\left(\int_0^T |c^*(S_s)Z_s^n - c^*(\overline{S}_s)\overline{Z}_s^n)|^2 ds\right) \le C_{h,c}\frac{T}{n}. \tag{53}$$

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 $V_{t_k}^n$. Then, using the supermartingale property of $\mathcal{V}$, equality (51) and Jensen inequality yield

$$(V_t^n-\mathcal{V}_t)_+ \le (\mathbb{E}_t(V_{t_{k+1}}^n)+\Delta A_{k+1}^n - \mathbb{E}_t(\mathcal{V}_{t_{k+1}}))_+ \le \mathbb{E}_t((V_{t_{k+1}}^n - \mathcal{V}_{t_{k+1}} + \Delta A_{k+1}^n)_+) \le \Delta A_{k+1}^n.$$

Now, using the expression (38) for ΔA_{k+1}^n and $V_{t_k}^n \ge h(t_{k+1},S_{t_{k+1}})$ imply

$$\Delta A_{k+1}^n = (h(t_k,S_{t_k}) - \mathbb{E}_{t_k}V_{t_{k+1}}^n)_+ \le (h(t_k,S_{t_k}) - \mathbb{E}_{t_k}h(t_{k+1},S_{t_{k+1}}))_+$$

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

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

Finally, it yields

$$\Delta A_{k+1}^n \le C_{c,h}\frac{T}{n}\left(1+\mathbb{E}_{t_k}(\max_{s\ge t_k}|S_s|^2)\right). \tag{54}$$

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

$$V_t^n = V_{t_{k+1}}^n - \int_t^{t_{k+1}} Z_s^n dS_s + \Delta A_{k+1}^n \leq h(t_{k+1}, S_{t_{k+1}}) - \int_t^{t_{k+1}} Z_s^n\, dS_s \quad (55)$$

$$\text{so that} \quad V_t^n = \mathbb{E}_t(V_t^n) \geq \mathbb{E}_t(h(t_{k+1}, S_{t_{k+1}})) = h(t, S_t) + \mathbb{E}_t(h(t_{k+1}, S_{t_{k+1}}) - h(t, S_t)).$$

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

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

As it is a supermartingale as well, it necessarily satisfies

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

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

$$|V_t^n - \overline{V}_t^n| \leq \max_{t_k\geq t}|h(t_k, S_{t_k}) - h(t_k, \overline{S}_{t_k})| \leq [h]_{Lip} \max_{t_k\geq t}|S_{t_k} - \overline{S}_{t_k}|.$$

(b) One considers the *càdlàg* semi-martingale $\mathcal{V}_t - V_t^n = \mathcal{V}_0 - V_0^n + \int_0^t (Z_s - Z_s^n).dS_s - (K_t - A_{\underline{t}}^n)$ where $\underline{t} := k$ on $[t_k, t_{k+1})$. It follows from Itô formula for jump processes that

$$\begin{aligned}\int_0^T |c^*(S_s)(Z_s - Z_s^n)|^2 ds \; &+ \sum_{t_k\leq T}(\Delta A_{t_k}^n)^2 + (\mathcal{V}_t - V_t^n)^2 \\ &= -2\int_0^T (\mathcal{V}_s - V_{s-}^n)(Z_s - Z_s^n).dS_s + 2\int_t^T (\mathcal{V}_s - V_{s-}^n)d(K_s - A_{\underline{t}}^n).\end{aligned}$$

$$\begin{aligned}\text{Now} \quad \int_0^T (\mathcal{V}_s - V_{s-}^n)d(K_s - A_{\underline{s}}^n) &= \int_0^T (\mathcal{V}_s - V_{s-}^n)dK_s + \int_t^T (V_{s-}^n - \mathcal{V}_s)dA_{\underline{s}}^n \\ &\leq \int_0^T (\mathcal{V}_s - V_s^n)dK_s + \sum_{t_k\leq T}(\Delta A_k^n)^2\end{aligned}$$

since $V_{t_k-}^n = V_{t_k}^n + \Delta A_k^n \leq \mathcal{V}_{t_k} + \Delta A_k^n$. This yields, using the inequality obtained in (a) and (54),

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

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

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

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 $V^n - \overline{V}^n$.

$$\begin{aligned}
\mathbb{E}\int_0^T |c^*(S_s)Z^n_s - c^*(\overline{S}_s)\overline{Z}^n_s|^2 ds &\le 2\,\mathbb{E}\int_0^T (V^n_s - \overline{V}^n_s)d(K^n_s - \overline{K}^n_s) + \mathbb{E}(h(T,S_T) - h(T,\overline{S}_T))^2\\
&\le 2[h]_{Lip}\mathbb{E}\int_0^T \mathbb{E}_s\Big(\max_{t_k\ge s}|S_{t_k} - \overline{S}_{t_k}|\Big)\, d(K^n_s + \overline{K}^n_s) + [h]^2_{Lip}\|S_T - \overline{S}_T\|_2^2\\
&\le C\,\mathbb{E}\left(\sup_{t\in[0,T]}\mathbb{E}_t\Big(\max_{t_k\ge t}|S_{t_k} - \overline{S}_{t_k}|\Big)(K^n_T + \overline{K}^n_T)\right) + C\|S_T - \overline{S}_T\|_2^2\\
&\le C\|\sup_{t\in[0,T]}\mathbb{E}_t\max_{t_k\ge t}|S_{t_k} - \overline{S}_{t_k}|\,\|_2\left(\|K^n_T\|_2 + \|\overline{K}^n_T\|_2\right) + C\|S_T - \overline{S}_T\|_2^2\\
&\le C_{h,c}\frac{T}{n}\left(\|K^n_T\|_2 + \|\overline{K}^n_T\|_2 + 1\right). \qquad (56)
\end{aligned}$$

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

$$\|K^n_T - \overline{K}^n_T\|_2 \le \|V^n_0\|_2 + \|\overline{V}^n_0\|_2 + \|\int_0^T Z^n_s dS_s - \int_0^T \overline{Z}^n_s d\overline{S}_s)\|_2 \le C + O(1/\sqrt{n}) \quad \text{by (56)}$$

so that $\sup_n \|\overline{K}^n_T\|_2 < +\infty$. Plugging this back in (56) 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 5 *If h is semi-convex and h and c are Lipschitz continuous, then*

$$\|c^*(S_.)(Z_. - \zeta^n_.)\|_{L^2(d\mathbb{P}\otimes dt)} \le \|c^*(S_.)(Z_. - \zeta_.)\|_{L^2(d\mathbb{P}\otimes dt)} + \frac{C}{\sqrt{n}} \quad \textit{where } \zeta := \mathrm{proj}_n(Z) \quad (57)$$

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

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

Now $\zeta_. - \zeta^n_. = \mathrm{proj}_n(Z_. - Z^n_.)$ so that by Inequality (53) in Proposition 3(b),

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

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

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

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 $L^2(c^*(S_t)d\mathbb{P}dt)$. $\diamond$

4.2 Hedging error induced by the (quadratic) quantization

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

Proposition 4 *If σ Lipschitz continuous, bounded and uniformly elliptic and if h is semi-convex and Lipschitz continuous, then*

$$|\zeta_0^n - \widehat{\zeta}_0^n| \le C(1+|s_0|)|\frac{n^{\frac{3}{2}}}{(N/n)^{\frac{1}{d}}}.$$

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

$$(\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*))\zeta_0^n = \mathbb{E}\left((V_1^n - V_0^n)\Delta S_{t_1}\right) \tag{58}$$

$$(\mathbb{E}(\Delta \widehat{S}_{t_1}\widehat{\Delta S}_{t_1}^*))\widehat{\zeta}_0^n = \mathbb{E}\left((\widehat{V}_1^n - \widehat{V}_0^n)\Delta \widehat{S}_{t_1}\right) \tag{59}$$

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

$$\begin{aligned}
\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*) - \mathbb{E}(\Delta\widehat{S}_{t_1}\Delta\widehat{S}_{t_1}^*) &= \mathbb{E}((\Delta S_{t_1} - \Delta\widehat{S}_{t_1})(\Delta S_{t_1} - \Delta\widehat{S}_{t_1})^*) \\
\text{so that}\qquad \|\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*) - \mathbb{E}(\Delta\widehat{S}_{t_1}\Delta\widehat{S}_{t_1}^*)\| &\le \mathbb{E}\|\Delta S_{t_1} - \Delta\widehat{S}_{t_1}\|_2^2 \le C N_1^{-\frac{2}{d}}.
\end{aligned}$$

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

where we used in the last inequality that $\|\widehat{S}_{t_1}\|_2 \le \|S_{t_1}\|_2 \le C\sqrt{\frac{T}{n}}(1+|s_0|)$. One derives from (58) and (59) that

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

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

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

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

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

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

4.3 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.*

$$\mathbb{E}\int_0^{T-\delta} |c^*(S_s)(Z_s-\zeta_s)|^2\,ds \qquad \text{for some } \delta > n^{-1/3}. \tag{60}$$

where Z_t is defined by (36) and $\zeta_t := \mathrm{proj}_n(Z)$ is the projection on the set $\mathcal{P}_n$ of elementary predictable strategies. Our basic assumption is

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

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 6 *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*

$$\mathbb{E}\int_0^{T-\delta_n} |c^*(S_s)(Z_s-\zeta_s)|^2\,ds \leq \frac{K(1+|s_0|)^q}{a}\frac{1}{n^{\frac{1}{6}-\frac{\theta}{\sqrt{\ln n}}}}. \tag{61}$$

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 regularization is needed and the resulting error $O(1/n^{1/3})$.

Some technical difficulties arise when evaluating the term in (60) 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*

$$\mathbb{E}\int_0^T |c^*(S_s)(Z_s-\zeta_s)|^2\,ds \leq \frac{C}{n} + \mathbb{E}\int_0^T |H_s - \eta_s|^2\,ds. \tag{62}$$

Proof: We temporarily define $z_s := \dfrac{1}{t_{k+1}-t_k}\mathbb{E}_{t_k}\displaystyle\int_{t_k}^{t_{k+1}} Z_r dr, t_k \leq s < t_{k+1}$. 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

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

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

$$\begin{aligned}\sum_{k=0}^{n-1}\lambda^2_{k+1} &\le \left(\sum_{k=0}^{n-1}\lambda_{k+1}\right)^2\\ \text{so that}\quad \mathbb{E}\left(\sum_{k=0}^{n-1}\mathbb{E}_{t_k}\lambda_{k+1}\right)^2 &\le 2\,\mathbb{E}\left(\sum_{k=0}^{n-1}(\lambda_{k+1}-\mathbb{E}_{t_k}\lambda_{k+1})\right)^2+2\,\mathbb{E}\left(\sum_{k=0}^{n-1}\lambda_{k+1}\right)^2\end{aligned}$$

$$\leq \quad 2\,\mathbb{E}\sum_{k=0}^{n-1}(\lambda_{k+1}-\mathbb{E}_{t_k}\lambda_{k+1})^2+2\,\mathbb{E}\left(\sum_{k=0}^{n-1}\lambda_{k+1}\right)^2$$

$$\leq \quad 4\,\mathbb{E}\left(\sum_{k=0}^{n-1}\lambda_{k+1}\right)^2=4\,\mathbb{E}\left(\int_0^T|Z_u|^2\,du\right)^2.$$

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

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 [17] 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

$$\|h-\underline{h}\|_\infty\leq C\varepsilon\quad\text{ and }\quad\|(\partial_t+\mathcal{L}_c)\underline{h}\|_\infty\leq C\varepsilon^{-1}\tag{63}$$

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

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

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

$$c_\lambda c_\lambda^*(x)\geq\frac{\varepsilon_0}{4}e^{-2\lambda}\quad\text{ and }\quad\|D^\alpha c_\lambda\|_\infty\leq C_\alpha\quad\text{ for every }\alpha.\tag{64}$$

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 [17] and [2] for this topic. We also consider the approximation

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

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

$$\mathbb{E}\int_0^T |H_s - \eta_s|^2\, ds \le C(e^{-C\lambda^2/T} + \varepsilon^2) + \mathbb{E}\int_0^T \left|\underline{H}_s - \underline{\eta}_s\right|^2 ds \tag{65}$$

Proof: We use the stability property of $RBSDE'$s (see [17] and [2]) in order to obtain

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

Let $\tau := \inf\{t \,/\, |S_t| \le e^{-\lambda}\}$. Note that

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

the last inequality is a standard large deviation fact (although it can be easily checked directly on model (1)). Since $S_t = \underline{S}_t$ for $t \le \tau$, we obtain

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

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 complete the proof by noting that

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

We need now some analytical facts that we recall here (see [17] 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

$$\begin{aligned}(\partial_t + \mathcal{L}_c)u(t,x) + \underline{F}(t,x,u(t,x)) &= 0, u(T,x) = \underline{h}(T,x),\\ \text{with} \quad \underline{F}(t,x,u(t,x)) &= \alpha(t,x)\mathbf{1}_{\{u(t,x)=\underline{h}(t,x)\}}\,((\partial_t + \mathcal{L}_c)\underline{h}(t,x))_+\end{aligned}$$

where α is a measurable function such that $0 \le \alpha \le 1$. Denote $F(t,x) := \underline{F}(t,x,u(t,x))$ and notice (recall (63)) that $\sup_{0\le t\le T}\sup_{x\in\mathbb{R}^d} |F(t,x)| \le C/\varepsilon$. 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 $\underline{P}_tf(x)=\mathbb{E}f(\underline{S}_t^x)$. This is the equation that will be used in the sequel.

We turn now to the semi-group. It is well known (see [21] or [28]) that under the hypothesis (64), $\underline{P}_t(x,y)=p_t(x,y)dy$ and for every $k\in\mathbb{N}$ and every multi-index $\alpha=(\alpha_1,\dots,\alpha_m)$ we have

$$\left|\partial_t^k D_x^\alpha p_t(x,y)\right|\leq\frac{K(1+|x|)^q}{at^{k+\frac{m+1}{2}}}e^{2\lambda}\times e^{-K'\frac{|x-y|^2}{t}} \tag{66}$$

where K,K',q depend on α and on C_α from (64) (but not on λ). Let us point out some immediate consequences of this evaluation in our framework. Since $|\underline{h}_T(y)|\leq C(1+|y|)$,using (66) we obtain

$$\left|\frac{\partial\underline{P}_\delta\underline{h}_T}{\partial x_k}(x)\right| \leq \int_{\mathbb{R}^d}\left|\frac{\partial p_\delta(x,y)}{\partial x_k}\right|\times C(1+|y|)dy\leq\frac{1}{\sqrt{\delta}}\frac{K(1+|x|)^q}{a}e^{2\lambda} \tag{67}$$

$$\left|\frac{\partial^2\underline{P}_\delta\underline{h}_T}{\partial x_k\partial x_p}(x)\right| \leq \frac{1}{\delta}\frac{K(1+|x|)^q}{a}e^{2\lambda} \tag{68}$$

$$\left|\frac{\partial}{\partial x_i}\underline{P}_{T-t}\underline{h}_T(x) - \frac{\partial}{\partial x_i}\underline{P}_{T-s}\underline{h}_T(y)\right|\leq\frac{K(1+|x|+|y|)^q}{a\delta^{3/2}}e^{2\lambda}(\sqrt{t-s}+|x-y|). \tag{69}$$

We deal now with the second term in the right hand of (73). Since $\|F\|_\infty\leq C/\varepsilon$, the same computations as above (using (66)) give

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

and

$$\left|\frac{\partial^2\underline{P}_\delta F_s}{\partial s\partial x_k}(x)\right| \leq \frac{1}{\delta^{3/2}\varepsilon}\frac{K(1+|x|)^q}{a}e^{2\lambda}. \tag{71}$$

Lemma 3 *Let $v_i=\frac{\partial u}{\partial x_i}$. Under the hypothesis $(\mathcal{H})$ (and consequently under (64)) one has*

$$|v_i(t,x)-v_i(t,y)| \leq \frac{K(1+|x|+|y|)^q}{a}e^{2\lambda}\times\left(\frac{\sqrt{\delta}}{\varepsilon}+\frac{1}{\delta\varepsilon}|x-y|\right) \tag{72}$$

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

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 $t\leq T-\delta$

$$\begin{aligned}v_i(t,x)&=\frac{\partial}{\partial x_i}\underline{P}_{T-t}\underline{h}_T(x)+\int_t^T\frac{\partial}{\partial x_i}\underline{P}_{s-t}F_s(x)ds\\&=\frac{\partial}{\partial x_i}\underline{P}_{T-(t+\delta)}\underline{P}_\delta\underline{h}_T(x)+\int_{t+\delta}^T\frac{\partial}{\partial x_i}\underline{P}_{s-(t+\delta)}\underline{P}_\delta F_s(x)ds+\int_t^{t+\delta}\frac{\partial}{\partial x_i}\underline{P}_{s-t}F_s(x)ds.\end{aligned} \tag{73}$$

Note that in the first two terms in the above (73) 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 (66)

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

the last inequality being a consequence of $\|F\|_\infty \le C/\varepsilon$. We deal now with the first term in the RHS of (73). Using the Feynman-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 (67), (68) and (69), one checks that

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

We turn now to the second term in the right hand of (73). Using (70), one obtains

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

Consider now $t' > t$ and write

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

Using (70) and (71), we obtain
$$\begin{aligned}I &\le \frac{K(1+|x|)^q}{a\delta\varepsilon}e^{2\lambda}\,|t-t'|\\ J &\le \frac{K(1+|x|)^q}{a\delta^{3/2}\varepsilon}e^{2\lambda}\,|t-t'| \le \frac{K(1+|x|)^q}{a\delta\varepsilon}e^{2\lambda}\sqrt{|t-t'|}\end{aligned}$$

the last inequality being a consequence of $|t-t'|\le\delta$. This completes the proof . $\diamond$

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

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

$$\left(\mathbb{E}\,|\underline{H}^x_r - \underline{H}^x_s|^2\right)^{1/2} \le \frac{K(1+|x|)^q}{a}e^{2\lambda}\times\left(\frac{\sqrt{\delta}}{\varepsilon}+\frac{1}{\delta\varepsilon}\frac{1}{\sqrt{n}}\right). \qquad (74)$$

Proof of Theorem 6: Using (74)

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

Moreover, as a consequence of the first two lemmas

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

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 (61). $\diamond$

5 Numerical results on American exchange options

In this section, we present some numerical experiments concerning the pricing and the hedging of American style options in (even) dimensions $d=2$ up to 10. This study will be divided in two parts. First, we will numerically estimate 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 following this choices.

5.1 The model

dimension),

We specify the underlying asset model (1) into a d-dimensional Black & Scholes ($B\&S$) model (constant volatility *sigma*) with constant dividend rates μ_ℓ, $\ell=1,\ldots,d$ *i.e.*

$$dS_t^\ell=(r-\mu_\ell)S_t^\ell\,\mathrm{d}t+\sigma_\ell S_t^\ell\,\mathrm{d}W_t^\ell,\quad t\in[0,T],\quad \ell=1,\ldots,d \tag{75}$$

where $(W_t)_{t\in[0,T]}$ denotes a d-dimensional standard Brownian motion. The traded assets vector are $(e^{\mu_\ell t}S_t^\ell)$, $\ell=1,\ldots,d)$ so that the discounted price satisfies (75) with $r=0$. The assets are assumed to be independent for technical reasons: it turns out to be the worst setting for quantization, so the most appropriate for experiments.

Beyond the importance of $B\&S$ for applications, S_t is then a closed function of (t,W_t) for every $t\in[0,T]$ since $S_t^\ell=s_0^\ell\exp\left((r-(\mu_\ell+\sigma_\ell^2/2))t+\sigma_\ell W_t^\ell\right)$. Therefore, one can either implement a quantization tree for $(S_t)_{t\in[0,T]}$ or for $(W_t)_{t\in[0,T]}$. Although the pay-offs functions are *stricto sensu*, no longer Lipschitz continuous as functions of W, we chose the second approach because of its universality: an optimal quantization of the Brownian motion can be done once for all and can be derived from optimal quantizations of the d-dim standard Normal distributions by appropriate dilatations (see Fig.1).

We focus on American style "geometric" exchange options which pay-offs read

$$h(x)=\max\left(x_1\cdots x_p-x_{p+1}\cdots x_{2p},0\right)\qquad \text{with } d:=2p. \tag{76}$$

It follows from the pricing formula (5) that the European and American premia for exchange options do not depend upon the interest rate r so we can set $r = 0$ w.l.g. An important remark is that there exists a closed form for the Black & Scholes premium of a European exchange option with maturity T at time t given by

$$Ex_{BS}(\theta, x, y, \widetilde{\sigma}, \mu) \;:=\; \mathrm{erf}(d_1)\exp(\mu\theta)\, x - \mathrm{erf}(d_1 - \widetilde{\sigma}\sqrt{\theta})\, y,$$

$$d_1(x, y, \widetilde{\sigma}, \theta, \mu) \;:=\; \frac{\ln(x/y) + (\widetilde{\sigma}^2/2 + \mu)\theta}{\widetilde{\sigma}\sqrt{\theta}} \quad \text{and} \quad \mathrm{erf}(x) := \int_{-\infty}^{x} e^{-\frac{u^2}{2}} \frac{du}{\sqrt{2\pi}}$$

$$\text{where } \theta := T - t,\ \widetilde{\sigma} := \left(\sum_{\ell=1}^{d} \sigma_\ell^2\right)^{1/2},\ \mu := \sum_{\ell=1}^{p} \mu_\ell - \sum_{\ell=p+1}^{d} \mu_\ell,\ x := \prod_{\ell=1}^{p} S_t^\ell,\ y = \prod_{\ell=p+1}^{d} S_t^\ell. \quad (77)$$

We will also use some American geometric put pay-offs $h(x_, \ldots, x_d) := (K - x_1 \cdots x_d)_+$.

5.2 Specification of 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. We use the optimized dispatching rule (34) that assigns N_k points to the grid Γ_k of time t_k so that $N \leq N_0 + N_1 + N_2 + \cdots + N_n \leq N + n$, (typically, $N_0 = 1 < N_1 < \cdots < N_n$). Now, we compute for every $k \in \{0, \ldots, n\}$ a N_k-optimal quantizer of $\mathcal{N}(0; I_d)$, from which we derive the N_k-optimal quantizer $(x_\ell^k)_{\ell=1,\ldots,N_k}$ of W_{t_k} by a $\sqrt{t_k}$-dilatation. The optimal grids of the Normal distributions are obtained by processing the $CLVQ$ algorithm (25). The final converging phase is refined using a randomized version of the so-called Lloyd I fixed point procedure (see *e.g.* [24]). For further details about the implementation, see [7]. Then, all the companion parameters (weights α_i^k, β_{ij}^k and quantization errors) are then estimated by a standard Monte Carlo simulation. In this very particular but important case, we do need the extended CLVQ procedure proposed for general diffusions (see [3]).

Finally, the Quantization tree algorithm (19) reads

$$\begin{cases} v_i^n := h_i^n, \quad i = 1, \ldots, N_n, \\ v_i^k := \max\left(h_i^k, \sum_{1 \leq j \leq N_{k+1}} \pi_{i,j}^k\, v_j^{k+1}\right), \quad i = 1, \ldots, N_k,\ k = 0, \ldots, n-1 \end{cases} \tag{78}$$

where the *obstacle* $h_i^k := h(s_{i,1}^k, \ldots, s_{i,d}^k)$ 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 *i.e.*

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

where $C_i^k = C_i(x^k)$. Following (49) the hedging δ_i^k at x_i^k is computed by

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

In practise, we need to introduce a kind of control variate variables $(M_i^k)_{1\le i\le N_k, 1\le k\le n}$ for the quantization tree algorithm. That means some explicitly known variables satisfying (ideally):

$$\sum_{j=1}^{N_{k+1}} \pi_{i,j}^k \, M_j^{k+1} = M_i^k. \tag{80}$$

($M^k$is a martingale with respect to the natural filtration of $(\widehat{W}_{t_k})_{0\le k\le n}$). Of course this can only be achieved up to the spatial discretization by considering a $\mathcal{F}_{t_k}^S$-martingale

$$M_{t_k} := m(t_k, S_{t_k}) \qquad \text{where } m \text{ is explicitly known}$$

An efficient choice is here to take

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

Then, we use the following proxy for the premium of the American pay-off $(h(t_k, S_{t_k}))_{0\le k\le n}$

$$\text{Premium}^h(t_k, s_i^k) := m(t_k, s_i^k) + v_i^{h-m,k} \tag{82}$$

where $(v_i^{h-m,k})_{1\le k\le n}$ is obtained by the scheme (78) with the obstacle $h_k^i - m(t_k, s_i^k)_{1\le k\le n}$.

5.3 Numerical accuracy, stability

We will now estimate numerically the rate of convergence (at time $t = 0$) of the numerical premium $p(n, \overline{N}) := \text{Premium}^h(0, s_0)$ given by (78) using (82) towards a reference p_{th} as a function of $(n, \overline{N})$ where $\overline{N} := N/n$ (average number of points per layer). The reference premium p_{th} is obtained by a finite difference method for vanilla American put options in 1-dimension and derived from a 2-dimensional difference method due to Villeneuve & Zanette in higher dimensions (see [38]). The error terms both in time and in space given by Theorem 4 are

$$E(n, \overline{N}) = |p(n, \overline{N}) - p_{th}| \approx \frac{c_1}{n} + c_2 \frac{n}{\overline{N}^{\alpha}} \qquad \text{with} \qquad \alpha = 1/d. \tag{83}$$

for semi-convex pay-offs. Two questions are raised by this error bound: are these rates optimal? Is it possible to compute an optimal number n_{opt} of time layers to minimize the global error?

We are able to answer the first one: we compute by c_1 and $C_2 := c_2\overline{N}^{-\alpha}$ by nonlinear regression of the function $n \mapsto E(n, \overline{N})$ for several fixed values of $\overline{N}$ and n.

We begin by the 1 and 2-dimension settings. The specifications of the reference model (75) are ($d = 1$, *vanilla put*, $r = 0.06$, $\sigma = 0.2$, $S_0 = 36$, $K = 40$) and (*exchange*, $d = 2$, $\sigma = 0.2$, $\mu = -0.05$, $S_0^1 = \sqrt{40}$, $S_0^2 = \sqrt{36}$).

In Table 1 are displayed numerical approximations of c_1, C_2 and

$$\alpha_i := \frac{\ln(C_2(\overline{N}_{i+1})/C_2(\overline{N}_i))}{\ln(\overline{N}_i/\overline{N}_{i+1})}, \; i = 1, 2, 3.$$

Note first that c_1 does not depend upon $\overline{N}$: this confirms the above global error structure (83). These empirical values for α are closer to $2/d$ than the theoretical $1/d$ and strongly suggests that $\alpha = 2/d$ is the true order. This can be explained by the

	$d=1$				$d=2$	
$\overline{N}_i$	$\overline{N}_1=20$	$\overline{N}_2=30$	$\overline{N}_3=40$	$\overline{N}_4=60$	$\overline{N}_1=235$	$\overline{N}_2=455$
c_1	0.47	0.45	0.45	0.46	3.54(-1)	3.41(-1)
C_2	3.77(-3)	1.82(-3)	1.03(-3)	4.79(-4)	6.61(-4)	3.55(-4)
α_i	1.87	1.90	1.91	$\times$	0.89	$\times$

Table 1: Estimation of the spatial convergence exponent α of (83) in dimensions $d=1,2$.

following heuristics: in the linear case (*e.g.* a European option computed by a descent of the quantization tree algorithm), the semi-group of the diffusion quickly regularizes the premium. Then, the second order numerical integration formula by quantization applies: if f admits a Lipschitz continuous derivative and X is a square integrable random variable, then (see [35])

$$|\mathbb{E}f(X)-\sum_{1\le i\le N}\mathbb{P}(\widehat{X}^x=x_i)f(x_i)-\sum_{1\le i\le N}f'(x_i).\underbrace{\mathbb{E}((X-x_i)\mathbf{1}_{C_i(x)})}_{=0\text{ if }x\text{ optimal}}|\le 1/2[f']_{Lip}\|X-\widehat{X}^x\|_2^2, \quad (84)$$

where $\|X-\widehat{X}^x\|_2^2$ is $O(N^{-2/d})$. The optimality of the grid makes the term $\mathbb{E}((X-x_i)\,\mathbf{1}_{C_i(x)}(X))=\frac{1}{2}\frac{\partial\|X-\widehat{X}^x\|_2^2}{\partial x_i}$ vanish. Applying rigorously this idea to American option pricing remains an open question (however see [6]). Whatsoever this better rate of convergence is a strong argument in favour of optimal quantization.

From dimension 4 to 10, the storage of the matrix $\{\pi^k_{i,j}\}$ for increasing values of $\overline{N}$ and large n is costly and make the computations intractable. The above computations suggest a spatial order of $2/d$ when the grids are optimal. In fact, truly optimal grids become harder and harder to obtain in higher dimensions, that is why we verify that spatial order becomes closer and closer to $1/d$ rather than $2/d$.

Several answers to the second question are possible according to the variables used in the error bound. Here, we chose to compute n_{opt} as a function of $\overline{N}$ and n. (rather than N and n). For a given value of $\overline{N}$, one proceeds as above a non linear regression that yields numerical values for c_1 and $C_2:=c_2\overline{N}^{-1/d}$. Finally set

$$n_{opt}(d,\overline{N}):=\sqrt{\frac{c_1}{C_2}}.$$

In lower dimension ($d\le 3$), the order α can be estimated and one may set directly for every $\overline{N}$, $n_{opt}(d,\overline{N})=\sqrt{\frac{c_1}{c_2}}\,\overline{N}^{1/d}$. In Table 2 are displayed the numerical values.

	$d=1$	$d=2$	$d=4$, $\overline{N}=750$	$d=6$, $\overline{N}=1000$	$d=10$, $\overline{N}=1000$
c_1	0.45	0.35	8.84(-1)	1.46	2.10
c_2	1.12	2.05(-1)	$\times$	$\times$	$\times$
C_2	$\times$	$\times$	2.62(-3)	2.57(-3)	8.75(-4)
n_{opt}	$0.63\,\overline{N}$	$1.31\,\overline{N}^{1/2}$	19	24	50

Table 2: Estimation of the optimal number of time layers for $d=1,2,4,6,10$.

5.4 Numerical results for American exchange options

We now present numerical computations for American geometric exchange functions based on the model described in section 5.1. Namely, we present the premium of in- and out-of-the money options as a function of the maturity T (expressed in year), $T\in\{\frac{k}{n},\,0\le k\le n\}$. This distinction gives an insight about the numerical influence of the free boundary.

Maturity	3 months		6 months		9 months		12 months	
AM_{ref}	4.4110		4.8969		5.2823		5.6501	
	Price	Error (%)	Price	Error (%)	Price	Error (%)	Price	Error (%)
$d = 2$	4.4111	0.0023	4.8971	0.0041	5.2826	0.0057	5.6505	0.0071
$d = 4$	4.4076	0.08	4.9169	0.34	5.3284	0.82	5.7366	1.39
$d = 6$	4.4156	0.1	4.9276	0.63	5.3550	1.38	5.7834	2.20
$d = 10$	4.4317	0.47	4.9945	2.00	5.4350	2.89	5.8496	3.53

Table 3: American premium & relative error for different maturities and dimensions.

We first settle the value of $\overline{N}$ and then read on Table 2 the optimal number $n = n_{opt}(d, \overline{N})$ of time layers. Space discretization is that used for the above numerical experiments. The model parameters and initial data settled so that μ and $\widetilde{\sigma}$ remain constant, equal to -5% and 20% respectively in (77):

$$\mu_1 := -5\,\%, \qquad \mu_i := 0,\ 2 \leq i \leq d, \quad \sigma_i := 20/\sqrt{d}\,\%.$$

$$s_0^i := 40^{2/d},\ 1 \leq i \leq d/2, \quad s_0^i := 36^{2/d},\ d/2+1 \leq i \leq d \text{ (in-of-the-money)},$$

$$s_0^i := 36^{2/d},\ 1 \leq i \leq d/2, \quad s_0^i := 40^{2/d},\ d/2+1 \leq i \leq d \text{ (out-of-the-money)}.$$

In Fig.2 and 3 are displayed the computed premia (and hedges in 2-dim) at time $t = 0$ together with the reference ones as a function of the maturity $T \in [0, T_{max}]$. Fig.2 emphasizes that both premia and hedges in 2-dimension are very well fitted with the reference premium.

In in-the-money case, we can see on Fig.3(a)-(c)-(e) that the computed premium tends to overestimate the reference one when the dimension d increases and when the maturity grows. However, the maximal error remains within 3,5 % in all the cases as displayed in Table 3. The same phenomenon occurs for the computed hedges, within a similar range (hedges are not depicted here). A piecewise constant approximation scheme is usually not efficient to compute derivatives. Furthermore the parameters of the quantization tree have not been settled to minimize the local residual risks (48). In the out-of-the-money setting, very different behaviours are observed on the premia. Indeed whatever the dimension is (from 4 to 10), the premia seem to be well computed (Figs.3(b)-(d)-(f)). Fig. 4 depicts the quantized residual risk (at $t = 0$) as a function of the maturity. It suggests that that numerical incompleteness of the market has a bigger impact on the premium “in-the-money” than “out-of-the-money”.

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, G. Pagès, A quantization algorithm for solving discrete time multi-dimensional discrete time Optimal Stopping problems, technical report $n°628$, Laboratoire de Probabilités & modèles aléatoires, Université Paris 6 (France), 2000, accepted for publication in *Bernoulli*.

[4] V. Bally, G. Pagès, Error analysis of a quantization algorithm for obstacle problems, Laboratoire de Probabilités & modèles aléatoires, Université Paris 6 (France), 2000, technical report $n°642$, forthcoming in *Stoch. Proc. and their Appl.*

[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 *Mathematical Finance* (selected papers of the colloquium *Applications of Malliavin calculus to Finance*), 2002.

[7] V. Bally, G. Pagès, J. Printems, Optimal quadratic quantization for numerics: the Gaussian case, in progress, 2002.

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

[9] A. Bensoussan, 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. INRIA, Septembre 1997, 140p.

[13] 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.

[14] P. Cohort, Limit Theorems for the Random Normalized Distortion, forthcoming in *Ann. of Applied Probability.*

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

[16] M. Duflo, *Random Iterative Systems*, Berlin, Springer, 1998.

[17] N. El Karoui, C. Kapoudjan, É. 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.

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

[19] 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.

[20] 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.

[21] A. Friedmann, *Stochastic Differential Equations and Applications*, Academic Press, **1**, 1975.

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

[23] A. Gersho, R. Gray (eds.), *IEEE on Infor. Theory, Special issue on Quantization*, **28**, 1982. Equations and

[24] J. Kiefer, Exponential rate of Convergence for the Lloyd's Method I, *IEEE on Information Theory, Special issue on Quantization*, **28**, n^0 2, pp.205-210, 1982.

[25] H.J. Kushner, *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.

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

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

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

[29] D. Lamberton, G. Pagès, Sur l'approximation des réduites, *Ann. Inst. Poincaré*, **26**, $n^0 2$, 1990, pp.331-355.

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

[31] D. Lamberton, Brownian optimal stopping and random walks, *Applied Mathematics and Optimization*, **45**, 2002, pp.283-324.

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

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

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

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

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

[37] 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**, $\mathrm{n}^0 10$, 1999, pp.1840-1851.

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

25, 26 2, 5, 13, 14, 30 3, 32 30 1 1, 4 17 1 10 11 6 12 25, 26 17 1 1 27 0, 2, 10, 11 2, 10, 11 30 1 12 1 27 1 1, 4 6 1 1, 3 5 10, 11, 12, 32 15 1 31, 38, 39

ANNEX: Uniform domination of the log-normal quantization error

Proposition *In the extended Black & Scholes model (1), if $\sigma \in \mathcal{C}_b^\infty(\mathbb{R}^d)$ is uniformly elliptic, then the (minimal) quantization errror $(S_{t_k})_{0\le k\le n}$ satisfies the uniform domination property (32) with*

$$\varphi_k := c_{\sigma,T}\sqrt{t_k} \quad (c_{\sigma,T} > 0) \qquad \textit{and} \qquad R := (s_0^i\,\psi(\sqrt{T}Z^i))_{1\le i\le d},\; Z \sim \mathcal{N}(0; I_d), \tag{85}$$

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

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

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

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

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

For every $i\in\{1,\ldots,n\}$, Inequality (85) yields

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

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

Figure 1: *A* 300-*tuple with the lowest quadratic quantization error for the Normal distribution*

(a) (b)

(c) (d)

(e) (f)

Figure 2: $d = 2$, $n = 25$ *and* $\overline{N} = 300$. *In-the-money : a), c), e) ; Out-of-the-money : b), d), f). American premium fonction of the maturity : a), b) ; Hedging strategy on the first asset : c), d) ; Hedging strategy on the second asset : e), f). The cross + depicts the premium obtained with the method of quantization and – depicts the reference premium (V & Z) (cf. [38]).*

(a) (b)

(c) (d)

(e) (f)

Figure 3: $d = 4, 6, 10$. *In-the-money : (a), (c), (e) ; Out-of-the-money : (b), (d), (f). American premium fonction of the maturity. + depicts the premium obtained with the method of quantization and – depicts the reference premium (V & Z) (cf. [38]).*

Figure 4: *Example of residual risk as a function of the maturity in* 4*-dmension*