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

Vlad BALLY [*] Gilles PAGÈS [†] Jacques PRINTEMS[‡]

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 ([24]). An algorithm to compute such grids is described. We provide results concerning the orders of the approximation with respect to the regularity of the payoff 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.

Key words: American option pricing, Optimal Stopping, Snell envelope, Optimal quantization, local volatility model.

2001 AMS classification: 91B28, 60G40, (65C05, 65C20, 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 focus a classical extension of the Black & Scholes model: the local volatility model. Nevertheless, 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 in financial markets. Roughly speaking, there are two types of numerical methods for this purpose:

– First, some purely deterministic approaches coming from Numerical Analysis: the solution of the optimal stopping problem admits a representation $v(t, S_t)$ where v satisfies a parabolic variational inequality. So, the various discretizing techniques like finite difference or finite element methods yield an approximation of the function v at discrete points of a

[*]Laboratoire d'analyse et de mathématiques appliquées & Projet MathFi (INRIA), UMR 8050, Université de Marne-la-Vallée, Cité Descartes, 5, Bld Descartes, Champs-sur-Marne, F-77454 Marne-la-Vallée Cedex 2, France. bally@math.univ-mlv.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

time-space grid (see *e.g.* [33] for an application to a vanilla put option or [8] 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 the pioneering work by Kushner in 1977 (see [28] and also [29]) in which Markov chain approximation was first introduced, including its links with the finite difference method. This took place before the massive development of Mathematical Finance. Concerning the consistency of time discretization, see [34].

These methods are quite efficient to handle vanilla American options on a single asset but they quickly become intractable as the number of 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". Concerning the extension from binomial to multinomial trees, is seems difficult to design some trees that are both compatible with the dimension/correlation constraints and the probabilistic structure of the dynamics.

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 usual form or based on the backward dynamic approximation of the (lowest) optimal stopping time. In Barraquant & Martineau [7], a sub-optimal 1-dimensional problem is solved: everything is designed as though the obstacle process itself had the Markov property. In [36], the algorithm devised by Longstaff & Schwartz is based on conditional expectation approximation by regression on a finite sub-family $(\varphi_i(S_t))_{i\in I}$ of a basis $(\varphi_k(S_t))_{k\geq 1}$ of $L^2(\sigma(S_t), \mathbb{P})$. The Monte Carlo rate of convergence of this method is deeply analyzed by Clément et al. in [16]. In [41], Tsitsiklis & Van Roy use a similar idea but for a modified Markov transition. In [11], Braodie & Glasserman generate some random grids at each time step and compute some companion weights using some statistical ideas based on the importance sampling theorem.

In [21] and [22] Fournié et al. initiated a Monte Carlo approach based on Malliavin calculus to compute conditional expectations and their derivatives. This leads to a purely probabilistic method. In [35], Lions and Régnier extend this approach to American option pricing (and Greek computation). The crucial step of this method is the variance reduction by localization. Optimal localization is investigated in [27] and [9].

In this paper, we develop a probabilistic method based on grids like in the original finite state Markov chain approximation method (originally described in [5]). 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 diffusion process, still denoted by S_{t_k} for convenience throughout the introduction). The key point is the following: rather than settling 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^* of size N_k which is optimally fitted to S_{t_k} among all grids with size N_k in the following sense: the *closest neighbor rule projection* $q_{\Gamma_k^*}(S_{t_k})$ of S_{t_k} onto the grid Γ_k^* is the best least square approximation of S_{t_k} among *all random vectors* Z such that $|Z(\Omega)| \leq N_k$. Namely

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

In that 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, both the function $q_{\Gamma_k^*}$ and the set $q_{\Gamma_k^*}(\Omega)$ are often called *optimal quantizer* of S_{t_k}. The resulting error bound $\|S_{t_k} - q_{\Gamma_k^*}(S_{t_k})\|_2$ 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 [25] or more recently [24]). 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 mean 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)

$$\|S_{t_k} - q_\Gamma(S_{t_k})\|_2^2 = \mathbb{E}\left(\min_{x_i \in \Gamma} |S_{t_k} - x_i|^2\right)$$

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 distribution of $q_{\Gamma_k^*}(S_{t_k})$, *i.e.* the *weights* $\mathbb{P}(S_{t_k} = x_i^{k,*})$, $x_i^{k,*} \in \Gamma_k^*$ and the induced quantization error. Both are involved in the American option pricing algorithm (see Section 2.2). Thus, Figure 1 illustrates on the bivariate normal distribution how an optimal grid gets concentrated on heavily weighted areas (this grid was obtained by the $CLVQ$ algorithm described in Section 2.4).

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* and to its theoretical rate of convergence. Then, the tree optimization, including the algorithmic aspects, is developed. This section is partially adapted from a general discretization method devised for Reflected Backward Stochastic Differential Equations ($RBSDE$) in [3].

Time discretization (Section 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 Section 2.2: it simply consists in plugging the optimal quantizer $\widehat{S}_{t_k} := q_{\Gamma_k^*}(S_{t_k})$ of S_{t_k} in this formula. Some weights appear that are obtained by the stochastic grid optimization procedure mentioned above. In Section 2.3, the rate of convergence of this algorithm is derived for Lipschitz continuous payoffs as a function of the time discretization step T/n and of the L^p-mean quantization errors $\|S_{t_k} - q_{\Gamma_k^*}(S_{t_k})\|_p$, $k = 1, \ldots n$. Then a short background on optimal quantization is provided in Section 2.4. In Section 2.5, the grid optimization of the quantization tree is addressed, using a stochastic approximation recursive procedure. The last subsection proposes an efficient (analytic) method to design *a priori* the size N_k of the grid at every time t_k is proposed, given that $N := N_0 + N_1 + \cdots + N_n$ elementary quantizers are available. In that case, we obtain some error bounds of the form $C(n^{-1/2} + n(N/n)^{-\frac{1}{d}})$. When the payoff is semi-convex the same holds true with n^{-1} instead of $n^{-1/2}$.

In Section 3, we design an approximating *quantized hedging strategy* following some ideas by Föllmer & Sondermann on incomplete markets. We are in a position to estimate some bounds for the induced hedging default, called *local residual risks* of the quantization tree. 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 obtain different kinds of rates of convergence for the hedging strategy (far from and close to the maturity).

Section 5 is devoted to the experimental validation of the method. We present extensive numerical results which tend to show that when the grids 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 using multi-dimensional American exchange options on (geometric) index in a standard d-dimensional decorrelated Black & Scholes model. This rate, actually better than forecast by theory, compensates for the drawback of an "irregular" approximation (see below). 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.

THE MAIN FEATURES OF THE QUANTIZATION APPROACH. Before going into technicalities, one may mention an obvious methodological difference between the quantization tree algorithm and the regression method [36]. 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 spatial 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 (S_t) is a function of the Brownian motion at time t *i.e.* $S_t = \varphi(t, B_t)$ like in the Black & Scholes model, the quantization tree algorithm may become completely parameter free: it suffices to consider a quantization of the Brownian motion itself which consists of some optimal quantization grids of multi-variate normal distributions with the appropriate sizes. Such optimal grids can be computed systematically in a very accurate way and then kept off line (see [39]). Quadratic optimal N-quantization of the $\mathcal{N}(0; I_d)$ distributions has been carried out systematically for various sizes $N \in \{1, \ldots, 400\}$ and dimensions $d \in \{1, \ldots, 10\}$. Some files of these optimal grids (including their weights) can be downloaded at the URLs:

- `www.proba.jussieu.fr/pageperso/pages.html` or
- `www.univ-paris12.fr/www/labos/cmup/homepages/printems`.

Finally, note that this method of quantization has been implemented in the software PREMIA (see `http://www-rocq.inria.fr/mathfi/Premia/index.html`).

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

$$dS^i_t = S^i_t(r\,dt + \sum_{1\le j\le q} \sigma_{ij}(e^{-rt}S_t)\,dW^j_t),\ S^i_0 := s^i_0 > 0,\ 1 \le i \le d, \qquad (1.1)$$

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

$$\sigma : \mathbb{R}^d \longrightarrow \mathcal{M}(d \times q) := \mathbb{R}^{d\otimes q} \text{ is bounded and Lipschitz continuous.} \qquad (1.2)$$

The filtration of interest will be the natural (completed) filtration $\underline{\mathcal{F}} := (\mathcal{F}^S_t)_{t\in[0,T]}$ of S (which coincides with that of W as soon as $\sigma\sigma^*(\xi) > 0$ for every $\xi \in \mathbb{R}^d$). For notational convenience, we introduce

$$c(\xi) := \mathrm{Diag}(\xi)\sigma(\xi), \qquad \xi := (\xi^1, \ldots, \xi^d) \in \mathbb{R}^d.$$

where $\text{Diag}(\xi)$ denotes the diagonal matrix with diagonal entry ξ^i at row i. The functions $c(\xi)$ and the drift $b(\xi) := r\,\xi$ are Lipschitz continuous so that a unique strong solution exists for (1.1) on $(\Omega, \mathcal{A}, \mathbb{P})$. Furthermore, it is classical background that, for every $p \geq 1$, there exists a constant $C_{p,T} > 0$ such that

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

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

$$d\widetilde{S}_t = c(\widetilde{S}_t)dW_t, \qquad \widetilde{S}_0 := s_0, \tag{1.3}$$

Here $\mathbb{P}$ is the so-called *risk neutral* probability in Mathematical Finance terminology. As long as $q \neq d$, 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 $\sigma(x) \equiv \sigma \in \mathcal{M}(d\times d)$, (1.1) is the usual d-dimensional Black & Scholes model: the risky assets are geometric Brownian motions given by

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

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

$$h_t := h(t, S_t), \quad t \in [0,T] \quad \text{where } h : [0,T] \longrightarrow \mathbb{R}_+ \text{ is a } \textit{Lipschitz} \text{ continuous.} \tag{1.4}$$

Such payoffs are sometimes called *vanilla*. Under Assumptions (1.1) and (1.4), one has

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

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

$$\mathcal{V}_t := e^{rt}\text{ess sup}\left\{\mathbb{E}(e^{-r\tau}h_\tau \mid \mathcal{F}_t),\; \tau\in\mathcal{T}_t\right\} \tag{1.5}$$

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

$$\widetilde{h}_t := \widetilde{h}(t, \widetilde{S}_t) \qquad \text{with} \qquad \widetilde{h}(t,x) := e^{-rt}h(t, e^{rt}x). \tag{1.6}$$

This result is based on a hedging argument on which we will come back further on. Note that $\sup_{t\in[0,T]} |\mathcal{V}_t| \leq \sup_{t\in[0,T]} |h_t| \in L^p$, $p \geq 1$.

One shows (see [8]) 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

$$\max\left(\frac{\partial\nu}{\partial t} + \mathcal{L}_{r,\sigma}\nu, \nu - h\right) = 0, \quad \nu(T,.) = h(T,.). \tag{1.7}$$

where $\mathcal{L}_{r,\sigma}$ denotes the infinitesimal generator of the diffusion (1.1).

Then, it is clear that the approximation problem for $\mathcal{V}_t$ appears as a 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 purely discrete dynamics (*e.g.* binomial trees, [33]) and the analytic ones based on numerical methods for solving the variational inequality (1.7) (*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 (1.1) is* 0: this amounts to assuming that we are in a "discounted world" with $\widetilde{S}$ given by (1.3) and $\widetilde{h}$ given by (1.6) instead of (S_t) and h respectively.

NOTATIONS: • $\mathcal{C}_b^\infty(\mathbb{R}^d)$ denotes the set of functions infinitely differentiable with bounded differentials (so that they have at most linear growth).

• The letters C and K denote positive real constants that may vary from line to line.

• $|\,.\,|$ will denote the Euclidean norm and "$\,.\,$" the inner product on $\mathbb{R}^d = \mathbb{R}^{1\otimes d}$. $\|M\| := \sup_{|x|\le 1}|Mx|$ will denote the operator norm of the matrix $M \in \mathbb{R}^{d\otimes q}$ (d rows, q columns) and M^* its transpose. In particular $x.y = x^*\,y$.

2 Pricing an American option using a quantization tree

In this section, the specificity of the martingale diffusion dynamics proposed for the risky assets in (1.3) (with $r = 0$) has little influence on the results, so it is costless to consider a general drifted Brownian diffusion

$$S_t = S_0 + \int_0^t b(S_s)\,ds + \int_0^t c(S_s)dW_s, \tag{2.1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $c : \mathbb{R}^d \to \mathcal{M}(d \times q)$ 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 (1.1)). So one uses a (Markovian) discretization scheme, easy to simulate, *e.g.* the Euler scheme: set $t_k = kT/n$ and

$$\overline{S}_{t_{k+1}} = \overline{S}_{t_k} + b(\overline{S}_{t_k})\frac{T}{n} + c(\overline{S}_{t_k}).(W_{t_k} - W_{t_{k-1}}), \quad \overline{S}_0 = s_0. \tag{2.2}$$

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 (1.1)). 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).

We want to quantize the Snell envelope of (S_{t_k}) or $(\overline{S}_{t_k})$ or of any family of homogeneous discrete time $\mathcal{F}_{t_k}$-Markov chains $(X_k^{(n)})_{0\le k\le n}$ whose transitions, denoted $P^{(n)}(x,dy)$, preserves Lipschitz continuity in the following sense: for every Lipschitz continuous function $f:\mathbb{R}^d\to\mathbb{R}$

$$[P^{(n)}f]_{\rm Lip}\le (1+C_{b,\sigma,T}T/n)[f]_{\rm Lip} \quad \text{where} \quad [f]_{\rm Lip}:=\sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|} \tag{2.3}$$

(see, *e.g.*, [3] for a proof). In fact this general discrete time markovian setting is the natural framework for the method. To alleviate notations, we drop the dependency in n and keep the notation $(X_k)_{0\le k\le n}$. 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:

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

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

$$\left\{\begin{array}{lcl} V_n & := & h(t_n,X_n),\\ V_k & := & \max\left(h(t_k,X_k),\mathbb{E}(V_{k+1}\mid\mathcal{F}_{t_k})\right),\quad 0\le k\le n-1.\end{array}\right. \tag{2.4}$$

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

$$\left\{\begin{array}{l} v_n:=h(t_n,\,.\,),\\ v_k:=\max\big(h(t_k,.),\,P^{(n)}(v_{k+1})\big),\ 0\le k\le n-1.\end{array}\right. \tag{2.5}$$

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* $h:[0,T]\times\mathbb{R}^d\longrightarrow\mathbb{R}$ *is* semi-convex *if*

$$\forall\,\xi,\,\xi'\in\mathbb{R}^d,\ \forall\,t\in\mathbb{R}_+,\qquad h(t,\xi')-h(t,\xi)\ge(\delta_h(t,\xi)|\xi'-\xi)-\rho|\xi'-\xi|^2 \tag{2.6}$$

where δ_h *is a bounded function on* $[0,T]\times\mathbb{R}^d$ *and* $\rho\ge0$.

Remarks: Note that (2.6) appears as a convex assumption relaxed by $-\rho|\xi'-\xi|^2$. In most situations, is used in the reverse sense *i.e.* $h(t,\xi)-h(t,\xi')\le(\delta_h(t,\xi)|\xi-\xi')+\rho|\xi-\xi'|^2$. The semi-convexity assumption is fulfilled by a wide class of functions:

– If $h(t,.)$ is $\mathcal{C}^1$ for every $t\in[0,T]$ and $\frac{\partial h}{\partial\xi}(t,\xi)$ is bounded, ρ-Lipschitz in ξ, uniformly in t then h is semi-convex (with $\delta_h(t,\xi):=\frac{\partial h}{\partial x}(t,\xi)$).

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

The notion of semi-convex function seems to appear in [14] for pricing one-dimensional American options. See also [32] for recent developments in a similar setting.

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

$$\forall\, n\geq 1,\ \forall\, k\in \{0,\ldots,n\},\quad \|\mathcal{V}_{t_k} - V_k\|_p \leq \frac{e^{CT}(1+|s_0|)}{\sqrt{n}}. \tag{2.7}$$

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

$$\forall\, n\geq 1,\ \forall\, k\in \{0,\ldots,n\},\quad \|\mathcal{V}_{t_k} - V_k\|_p \leq \frac{e^{CT}(1+|s_0|)}{n} \tag{2.8}$$

2.2 Spatial discretization: the quantization tree

The starting point of the method is to discretize the random variables X_k by some $\sigma(X_k)$-random variables $\widehat{X}_k$ taking finitely many values in $\mathbb{R}^d$. Such a random vector $\widehat{X}_k$ is called a *quantization* of X_k. Equivalently, one may define a quantization of X_k by setting $\widehat{X}_k = q_k(X_k)$ where $q_k : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel "quantizing" function such that $|q_k(\mathbb{R}^d)| = |X_k(\Omega)| = N_k < +\infty$. The elements of the set $X_k(\Omega)$ are called *elementary quantizers.* Let $N = N_0 + N_1 + \cdots + N_n$ denote the total number of elementary quantizers used to quantize the whole Markov chain $(X_k)_{0\le k\le n}$.

We aim to approximate the dynamic programming formula (2.4) by a similar dynamic programming formula involving the sequence $(\widehat{X}_k)_{0\le k\le n}$.

2.2.1 Quantization tree and quantized pseudo-Snell envelope

We assume in that section that for every $k \in \{0, 1,\ldots,n\}$, we have access to a sequence of quantizations $\widehat{X}_k = q_k(X_k)$, $k = 0,\ldots,n$ of the Markov chain $(X_k)_k$. We denote by $\{x^k_1,\cdots,x^k_{N_k}\} = q_k(\mathbb{R}^d)$ the grid of N_k points used to quantize X_k and by $x^k = (x^k_1,\ldots,x^k_{N_k})$ the induced N_k-tuple([1]). The questions related to the optimal choice of x^k and q_k will be addressed in Section 2.4 below. (Note that in our original setting $X_0 = s_0$, so that $\widehat{X}_0 = s_0$ is the best possible L^p-mean quantization of X_0 and $N_0 = 1$).

The quantized dynamic programming formula below is devised by analogy with the original one (2.4): one simply replaces X_k by its quantized random vector $\widehat{X}_k$

$$\left\{\begin{array}{lcl} \widehat{V}_n & := & h(t_n, \widehat{X}_n),\\ \widehat{V}_k & := & \max\Big(h(t_k,\widehat{X}_k), \mathbb{E}(\widehat{V}_{k+1}\mid \widehat{X}_k)\Big), \quad 0\le k\le n-1. \end{array}\right. \tag{2.9}$$

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

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 has been set up *a priori* for every X_k, this does not make the numerical processing of this algorithm possible. As a matter of fact, one needs to know the joint distributions of $(\widehat{X}_k, \widehat{X}_{k+1})$, $k = 0,\ldots,n-1$. This is enlightened by the proposition below whose easy proof is left to the reader.

[1] From now on, for convenience, we will give the priority to the N-tuple notation.

Proposition 1 *(Quantization tree algorithm) For every* $k \in \{0,\dots,n\}$, *let* $x^k := (x_1^k,\dots,x_{N_k}^k)$, $q_k : \mathbb{R}^d \to \{x_1^k,\dots,x_{N_k}^k\}$ *and* $\widehat{X}_k = q_k(X_k)$ *be a quantization of* X_k. *Set, for every* $k \in \{0,\dots,n\}$ *and every* $i \in \{1,\dots,N_k\}$,

$$p_i^k := \mathbb{P}(\widehat{X}_k = x_i^k) = \mathbb{P}(X_k \in C_i(x^k)), \tag{2.10}$$

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

$$\begin{aligned}\pi_{ij}^k &:= \mathbb{P}(\widehat{X}_{k+1} = x_j^{k+1} \mid \widehat{X}_k = x_i^k) = \mathbb{P}\left(X_{k+1} \in C_j(x^{k+1}) \mid X_k \in C_i(x^k)\right)\\ &= \frac{p_{ij}^k}{p_i^k} \quad \text{with} \quad p_{ij}^k := \mathbb{P}\left(X_{k+1} \in C_j(x^{k+1}),\ X_k \in C_i(x^k)\right).\end{aligned} \tag{2.11}$$

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

$$\begin{aligned}\widehat{v}_n(x_i^n) &:= h_n(x_i^n), \qquad i \in \{1,\dots,N_n\}\\ \widehat{v}_k(x_i^k) &:= \max\left(h(t_k,x_i^k), \sum_{j=1}^{N_{k+1}} \pi_{ij}^k\, \widehat{v}_{k+1}(x_j^{k+1})\right), 1 \le i \le N_k,\ 0 \le k \le, n-1.\end{aligned} \tag{2.12}$$

Then, $\widehat{V}_k = \widehat{v}_k(\widehat{X}_k)$ *satisfies the above dynamic programming (2.9) of the pseudo-Snell envelope.*

Remark: If $X_k = S_{t_k}$ or $\overline{S}_{t_k}$, then $X_0 = \widehat{X}_0 = s_0$ and $\widehat{v}_0(\widehat{X}_0) = \widehat{v}_0(s_0)$ is deterministic. In more general settings one approximates $\mathbb{E}\, v_0(X_0)$ by

$$\mathbb{E}\,\widehat{v}_0(\widehat{X}_0) = \sum_{i=1}^{N_0} p_i^0\, \widehat{v}(x_i^0).$$

Implementing the quantization tree algorithm (2.12) on a computer raises two questions:

– How is it possible to estimate the parameters p_i^k and p_{ij}^k involved in (2.12) ?

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

Parameter estimation (A first Monte Carlo approach): the tractability of the above algorithm relies on the parameters $\pi_{ij}^k := p_i^k / p_{ij}^k$. So, the ability to compute them at a reasonable cost is the key of the method. The most elementary solution is to process a wide scale *Monte Carlo simulation of the Markov chain* $(X_k)_{0 \le k \le n}$ to estimate the parameters p_i^k and p_{ij}^k as defined by (2.10) and (2.11). An estimate of the (p^{th} power of the) L^p-mean quantization error $\|X_k - \widehat{X}_k\|_p^p = \mathbb{E}\min_{1\le i\le N} |X_k - x_i^k|_p^p$ can also be computed. 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 sample paths of the chain starting from any $x \in \mathbb{R}^d$.

We will see further on in paragraph 2.4 how to choose the size and the geometric location of the N_k-tuples x^k in an optimal way.

Complexity of the quantization tree: theory and practice A quick look at the structure of the algorithm (2.12) shows that going from layer $k+1$ down to layer k needs $\kappa \times N_k N_{k+1}$ elementary computations (κ is the complexity induced by a connection "$i \to j$"). Hence, the cost of a quantization tree descent is approximately

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

Then an elementary optimization under constraint shows that

$$\kappa \frac{N^2}{n+1} \leq \text{Complexity} \leq \kappa \frac{N^2}{4}.$$

(Lower bound is for $N_k = N/(n+1)$, upper bound for the unrealistic values $N_k = \frac{N}{2}\mathbf{1}_{\{0,1\}}$). This purely combinatorial lower bound needs to be tuned. In fact, in most examples the transition of the Markov chain behaves in such a way that, at each layer k, many terms of the "transition matrix" $[\pi_{ij}^k]$ are negligible because x_i^k and x_j^{k+1} are remote from each other in $\mathbb{R}^d$: the Monte Carlo estimates of these coefficients will be 0. Hence, the complexity of the algorithm is $\nu \times \kappa N$ rather than lower bound $\kappa N^2/(n+1)$, where ν denotes the average *number of active connections* above a regular node i of the tree. Thus, the cost of such a "descent" is similar to that of a one dimensional binomial tree with $\sqrt{\frac{\nu}{2}N}$ time steps (such a tree approximately contains νN points).

2.3 Convergence and rate using L^p-mean quantization error

In this paragraph we provide some *a priori* L^p-error bounds for $\|V_k - \widehat{V}_k\|_p$, $k = 0, \ldots, n$, based on the L^p-mean quantization errors $\|X_k - \widehat{X}_k\|_p$, $k = 0, \ldots, n$, 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 $(X_k)_{0\leq k\leq n}$: it only requires to compute, for every $\mathbb{P}_{X_k}$-distributed simulated random vector, its distance to its closest neighbor in the set $\{x_1^k, \ldots, x_{N_k}^k\}$.

The estimates in Theorem 2 below holds for *any homogeneous Markov chain* $(X_k)_{0\leq k\leq n}$ having a K-*Lipschitz* transition $(P(x, dy))_{x\in\mathbb{R}^d}$ satisfying, for every Lipschitz function g,

$$[Pg]_{\text{Lip}} \leq K[g]_{\text{Lip}}. \tag{2.13}$$

This is the case of a diffusion and of its the Euler scheme with Lipschitz drift and diffusion coefficient as mentioned before, see (2.3). Note that K may be lower than 1: this is, *e.g.*, the case if X_k is the Euler scheme of an Ornstein-Uhlenbeck process with drift $b(x) := -ax$, $a > 0$ (and step $T/n < 1/a$).

Theorem 2 *Assume that the transition $P(x, dy)$ of the chain $(X_k)_{0\leq k\leq n}$ is K-Lipschitz, that h is Lipschitz continuous in x, uniformly in time and set $[h]_{\text{Lip}} := \max_{0\leq k\leq n}[h(t_k, .)]_{\text{Lip}}$. Let $(V_k)_{0\leq k\leq n}$ and $(\widehat{V}_k)_{0\leq k\leq n}$ be like in (2.4) and (2.9) respectively. For every $k \in \{0, \ldots, n\}$, let $\widehat{X}_k$ denote a quantization of X_k. Then, for every $p \geq 1$,*

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

with $d_i^{(n)} := (1 + (2 - \delta_{p,2})(K \vee 1)^{n-i})[h]_{\text{Lip}}$, $0 \leq i \leq n-1$, $d_n^{(n)} := [h]_{\text{Lip}}$ ($\delta_{u,v}$ stands for the Kronecker symbol).

Proof: STEP 1: We first show that the functions v_k recursively defined by (2.5) are Lipschitz continuous with

$$[v_k]_{\text{Lip}} \leq (K \vee 1)^{n-k}[h]_{\text{Lip}}. \tag{2.14}$$

Clearly, $[v_n]_{\text{Lip}} \leq [h]_{\text{Lip}}$ and one concludes by induction, using the inequality

$$|\max(a, b) - \max(a', b')| \leq \max(|a - a'|, |b - b'|).$$

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

$$
\begin{aligned}
|V_k - \widehat{V}_k| &\leq |h_k(X_k) - h_k(\widehat{X}_k)| + |\Phi_k(X_k) - \widehat{\Phi}_k(\widehat{X}_k)| \\
&\leq [h]_{\rm Lip}|X_k - \widehat{X}_k| + |\Phi_k(X_k) - \widehat{\mathbb{E}}_k(\Phi_k(X_k))| + |\widehat{\mathbb{E}}_k(\Phi_k(X_k)) - \widehat{\Phi}_k(\widehat{X}_k)| \quad (2.15)
\end{aligned}
$$

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

Hence,
$$
\|\Phi_k(X_k) - \widehat{\mathbb{E}}\Phi_k(X_k)\|_p \leq 2[\Phi_k]_{\rm Lip}\|X_k - \widehat{X}_k\|_p.
$$

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.

Now
$$
\begin{aligned}
\widehat{\mathbb{E}}_k(\Phi_k(X_k)) - \widehat{\Phi}_k(\widehat{X}_k) &= \widehat{\mathbb{E}}_k\left(\mathbb{E}(v_{k+1}(X_{k+1}) \mid X_k)\right) - \widehat{\mathbb{E}}_k\left(\widehat{v}_{k+1}(\widehat{X}_{k+1})\right) \\
&= \widehat{\mathbb{E}}_k\left(v_{k+1}(X_{k+1}) - \widehat{v}_{k+1}(\widehat{X}_{k+1})\right)
\end{aligned}
$$

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

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

Finally, it follows from the above inequalities and (2.15) that

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

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

$$
\|V_k - \widehat{V}_k\|_p \leq \sum_{i=k}^{n}([h]_{\rm Lip} + (2 - \delta_{p,2})[\Phi_i]_{\rm Lip})\|X_i - \widehat{X}_i\|_p
$$

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

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

2.4 Optimization of the quantization

We begin by a brief introduction to optimal quantization of random vectors (see [24] for on overview), then we address the problem of optimal quantization of Markov chains.

2.4.1 Optimal quantization of a random vector X

Let $X \in L^p_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$. From a probabilistic viewpoint, optimal L^p-mean quantization ($p \geq 1$) consists in studying the best L^p-approximation of X by some random vectors $X' = q(X)$ taking at most N values. Minimizing the L^p-mean quantization error $\|X - q(X)\|_p$ can be decomposed into two successive phases:

– Optimization phase 1. A N-tuple $x = (x_1, \ldots, x_N) \in (\mathbb{R}^d)^N$ being set, find a quantizer $q_x : \mathbb{R}^d \to \{x_1, \ldots, x_N\}$ (if any) such that

$$\|X - q_x(X)\|_p = \inf\Big\{\|X - q(X)\|_p,\ q : \mathbb{R}^d \to \{x_1, \ldots, x_N\},\ \text{Borel function}\Big\}.$$

– Optimization phase 2. Find an N-tuple $x^* \in (\mathbb{R}^d)^N$ (if any) that achieves the infimum of $\|X - q_x(X)\|_p$ over $(\mathbb{R}^d)^N$, *i.e.*

$$\|X - q_{x^*}(X)\|_p = \inf\Big\{\|X - q_x(X)\|_p,\ x \in (\mathbb{R}^d)^N\Big\}.$$

The solution to the first optimization problem is purely geometric: it is the *closest neighbor projections*, denoted q_x, induced by the *Voronoi tessellations* of x as defined below.

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

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

(b) The closest neighbor projection *or* Voronoi quantizer (function) *q_x induced by the Voronoi tessellation $(C_i(x))_{1\le i\le n}$ is defined for every $\xi \in \mathbb{R}^d$, by $q_x(\xi) = \sum_{1\le i\le N} x_i \mathbf{1}_{C_i(x)}(\xi)$.*
(c) The random vector

$$\widehat{X}^x = q_x(X) = \sum_{1\le i\le N} x_i \mathbf{1}_{C_i(x)}(X)$$

is called a Voronoi quantization *of X. The N-tuple x is often called an N-*quantizer.

Notation: From now on, the notation $\widehat{X}^x$ will always denote a *Voronoi quantization* of X. When there is no ambiguity, the exponent ^{x} will often be dropped and we will denote $\widehat{X}$ instead of $\widehat{X}^x$.

Note that, the closure and the boundary of the i^{th} cell $C_i(x)$ are the same for any Voronoi tessellation. This boundary is included into at most $N-1$ hyperplanes. If the distribution $\mathbb{P}_X$ of X weights no hyperplane – that is $\mathbb{P}_X(H) = 0$ for every hyperplane H of $\mathbb{R}^d$ – then all the Voronoi tessellations are $\mathbb{P}_X$- equal and all the Voronoi quantizations $\widehat{X}^x$ have the same distribution.

The second optimization problem consists in minimizing on $(\mathbb{R}^d)^N$ the (symmetric) function $x \mapsto \|X - \widehat{X}^x\|_p$. First, note that the L^p-mean quantization error satisfies

$$\|X - \widehat{X}^x\|_p^p = \sum_{i=1}^N \mathbb{E}\left(\mathbf{1}_{C_i(x)} |X - x_i|^p\right) = \mathbb{E}\left(\min_{1\le i\le N} |X - x_i|^p\right) = \int_{\mathbb{R}^d} \min_{1\le i\le N} |x_i - \xi|^p \mathbb{P}_X(d\xi). \quad (2.16)$$

It follows that the L^p-mean quantization error depends on X through its distribution $\mathbb{P}_X$. The second consequence of (2.16) is an important and attractive feature of the L^p-mean quantization error compared to other usual error bounds: it is a (Lipschitz) continuous function of the N-quantizer $x := (x_1, \ldots, x_N)$.

Hence, as soon as $\mathbb{P}_X$ has a compact support, $x \mapsto \|X - \widehat{X}^x\|_p$ reaches a minimum at some *L^p-optimal N-quantizer* x^*. When $\mathbb{P}_X$ no longer has a compact support, this is still true: one shows by induction on N (see [24] or [38]), that

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

[2] In what follows, we will assume that a partition may contain the empty set: this will happen when $x_i = x_j$ for some $i \neq j$.

Proposition 2 *An L^p-optimal N-quantizer x^* for $X \in L^p(\Omega, \mathbb{P})$ satisfies*

$$\|X - \widehat{X}^{x^*}\|_p = \min\Big\{\|X - Z\|_p,\ Z : \Omega \to \mathbb{R}^d,\ \textit{random vector},\ |Z(\Omega)| \le N\Big\}. \tag{2.17}$$

Proof: Let $Z(\Omega) = \{z_1, \ldots, z_N\}$. Set $z := (z_1, \ldots, z_N)$ (with possibly $z_i = z_j$). Then

$$\|X-\widehat{X}^{x^*}\|_p^p \le \|X-\widehat{X}^z\|_p^p = \int_\Omega \min_i |X(\omega)-z_i|^p \mathbb{P}(d\omega) \le \int_\Omega \min_i |X(\omega)-Z(\omega)|^p \mathbb{P}(d\omega) = \|X-Z\|_p^p. \diamond$$

Moreover, the following simple facts hold true (see [24] or [38] and the references therein):

– If $\mathrm{supp}\,\mathbb{P}_X$ has an *infinite* support, any optimal N-quantizer x^* has pairwise distinct elements, that is $|q_{x^*}(\mathbb{R}^d)| = |\widehat{X}^{x^*}(\Omega)| = N$.

– The closed convex hull $\mathcal{H}_X$ of $\mathrm{supp}\,\mathbb{P}_X$ contains at least an optimal quantizer (obtained as the projection of any optimal quantizer on $\mathcal{H}_X$). Furthermore, if $\mathrm{supp}\,\mathbb{P}_X$ is convex (*i.e.* equal to $\mathcal{H}_X$), then the N distinct components of any optimal N-quantizer x^* all lie in $\overset{\circ}{\mathcal{H}}_X$. This also holds true for $\mathcal{H}_X$-valued locally optimal N-quantizers.

– Rate of convergence: The main function of the L^p-mean quantization error being to be an error bound, it is important to elucidate the behavior of $\|X - \widehat{X}^{x^*}\|_p$ as the *size* N of the optimal N-quantizer x^* go to infinity. The first easy fact is that it goes to 0 as $N \to \infty$ *i.e.*

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

Indeed, let $(z_k)_{k\in\mathbb{N}}$ denote an everywhere dense sequence of $\mathbb{R}^d$-valued vectors and set $x_N := \{z_1, \ldots, z_N\}$. Then $\|X - \widehat{X}^{x_N}\|_p$ goes to zero by the Lebesgue dominated convergence theorem. Furthermore $0 \le \min_{x\in(\mathbb{R}^d)^N} \|X - \widehat{X}^x\|_p \le \|X - \widehat{X}^{x_N}\|_p$. $\diamond$

The rate of this convergence turns out to be a much more challenging problem. Its solution, often referred to as Zador's Theorem, was completed by several authors (Zador, see [25], Bucklew & Wise, see [13] and finally Graf & Luschgy see [24]).

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

$$\lim_N \left(N^{\frac{p}{d}} \min_{x\in(\mathbb{R}^d)^N} \|X - \widehat{X}^x\|_p^p\right) = J_{p,d}\left(\int_{\mathbb{R}^d} \varphi(u)^{\frac{d}{d+p}} du\right)^{1+\frac{p}{d}} \tag{2.18}$$

where $\mathbb{P}_X(du) = \varphi(u)\,\lambda_d(du) + \nu(du)$, $\nu \perp \lambda_d$ (λ_d Lebesgue measure on $\mathbb{R}^d$). The constant $J_{p,d}$ corresponds to the case of the uniform distribution on $[0,1]^d$.

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 [25] or [24]). Nevertheless, some upper and lower bounds were established, based on ball packing techniques and on the introduction of random quantizers (see *e.g.* [17] and [24]). It follows that $J_{p,d} \sim (\frac{d}{2\pi e})^{p/2}$ as $d \to +\infty$ (see [24]).

This theorem says that $\min_{x\in(\mathbb{R}^d)^N} \|X - \widehat{X}^x\|_p = C_{X,p,d} N^{-\frac{1}{d}} + o(N^{-\frac{1}{d}})$: this is in accordance with the rates obtained with uniform product lattice grids of size $N = m^d$ for numerical integration with respect to the uniform distribution over $[0,1]^d$. (Even in that very case, no such lattice grid is an optimal quantizer except when $d = 1$). The conclusion is that, for any *distribution* $\mathbb{P}_X$, optimal quantization produces *for every* N the best matching "N-grid" for $\mathbb{P}_X$. Asymptotically, a sequence of optimal quantizers yields the lowest possible constant $C_{X,p,d}$, with an obvious numerical interest.

2.5 How to get optimal quantization using simulation

Optimal quantization of a single random vector: how to get it? In fact the L^p-mean quantization error function is even smoother than Lipschitz continuous. This is at the origin of an important a stochastic optimization method based on simulation. First, we consider for convenience its p^{th} power, denoted $D^p_{_N}$, defined for every $x = (x_1, \ldots, x_{_N}) \in (\mathbb{R}^d)^N$ by

$$D^p_{_N}(x) = \|X - \widehat{X}^x\|^p_{_p} \;=\; \mathbb{E}\left(\min_{1\le i\le N} |X - x_i|^p\right) = \int_{\mathbb{R}^d} d^p_{_N}(x,\xi)\mathbb{P}_{_X}(d\xi)$$

where

$$d^p_{_N}(x,\xi) \;:=\; \min_{1\le i\le N} |x_i - \xi|^p.$$

The letter D refers to the word *distortion* used in Information Theory. The function $d^p_{_N}(x,\xi)$ is often called *local L^p-distortion.*

One shows (see, *e.g.*, [24] or [38]) that, if $p > 1$, $D^p_{_N}$ is continuously differentiable at every $x\in(\mathbb{R}^d)^N$ satisfying the *admissibility* condition

$$\forall\, i \neq j,\; x_i \neq x_j \quad \text{and} \quad \mathbb{P}_X\left(\cup_{i=1}^N \partial C_i(x)\right) = 0. \tag{2.19}$$

Then, its gradient $\nabla D^p_{_N}(x)$ is obtained by formal differentiation, that is

$$\nabla D^p_{_N}(x) \;:=\; \left(\mathbb{E}\,\frac{\partial d^p_{_N}}{\partial x_i}(x,X)\right)_{1\le i\le n} = \left(\int_{\mathbb{R}^d} \frac{\partial d^p_{_N}}{\partial x_i}(x,\xi)\mathbb{P}_{_X}(d\xi)\right)_{1\le i\le n}$$

where

$$\frac{\partial d^p_{_N}}{\partial x_i}(x,\xi) \;:=\; p\,\frac{x_i-\xi}{|x_i-\xi|}|x_i-\xi|^{p-1}\mathbf{1}_{C_i(x)}(\xi), \quad 1\le i\le n,$$

with the convention $\frac{0}{|0|} = 0$. The above differentiability result still holds when $p = 1$ if $\mathbb{P}_{_X}$ is continuous *i.e.* $\mathbb{P}_{_X}(\{\xi\}) = 0$, $\xi\in\mathbb{R}^d$.

One notes that $\nabla D^p_{_N}$ has an integral representation with respect to the distribution of X. When the distribution $\mathbb{P}_{_X}$ is simulatable, this strongly suggests to implement a stochastic gradient descent derived from this representation to approximate some (local) minimum of $D^p_{_N}$: when $d \ge 2$, the implementation of deterministic gradient descent becomes unrealistic 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 $(0,1)$-valued steps satisfying

$$\sum_t \gamma_t = +\infty \quad \text{and} \quad \sum_t \gamma_t^2 < +\infty. \tag{2.20}$$

Set, for every admissible $x\in(\mathbb{R}^d)^N$ in the sense of (2.19), and every $\xi\in\mathbb{R}^d$

$$\nabla_x d^p_{_N}(x,\xi) := \left(\frac{\partial d^p_{_N}}{\partial x_i}(x,\xi)\right)_{1\le i\le N}.$$

Then, starting from a deterministic initial N-tuple $X^0 = x^0$ with N pairwise distinct components, one defines recursively for every $t \ge 1$,

$$X^t = X^{t-1} - \frac{\gamma_t}{p}\nabla_x d^p_{_N}(X^{t-1},\xi^t) \tag{2.21}$$

(this formula *a.s.* grants by induction that x^t has pairwise distinct components).

From a theoretical viewpoint, the main difficulty is that the assumptions usually made that ensure the *a.s.* convergence of such a procedure are not fulfilled by $D^p_{_N}$ (see, *e.g.* [18]

or [30] for an overview on Stochastic approximation). To be more specific, let us stress that $D^p_N(x_1,\ldots,x_N)$ does not go to infinity as $\max_{1\le i\le N}|x_i|$ goes to infinity and ∇D^p_N is clearly not a Lipschitz function. So it is not an appropriate "Lyapunov function". However some (weaker) *conditional* *a.s.* convergence results in the Kushner & Clark sense have been obtained in [38] for compactly supported absolutely continuous distributions $\mathbb{P}_X$ in the case $p=2$. In 1 dimension, regular *a.s.* convergence holds if furthermore the density function of $\mathbb{P}_{X-}$ is bounded.

The quadratic case $p=2$ is the most commonly implemented for applications. It is known in Information Theory literature as as the *Competitive Learning Vector Quantization* (*CLVQ*) algorithm.

The synthetic formula (2.21) can be detailed as follows: set $X^t := (X^t_1,\ldots,X^t_N)$,

$$\textsc{Competitive phase:}\quad \text{select}\quad i(t+1)\in \operatorname{argmin}_i|X^t_i-\xi^{t+1}| \tag{2.22}$$

$$\textsc{Learning phase:}\quad \left\{\begin{array}{rcl} X^{t+1}_{i(t+1)} &:=& X^t_{i(t+1)}-\gamma_{t+1}\frac{X^t_{i(t+1)}-\xi^{t+1}}{|X^t_{i(t+1)}-\xi^{t+1}|}|X^t_{i(t+1)}-\xi^{t+1}|^{p-1}\\ *[.6em]X^{t+1}_i &:=& X^t_i,\qquad i\neq i(t+1).\end{array}\right. \tag{2.23}$$

$\textsc{Companion parameter procedure:}$ Assume that $X\in L^{p(1+\eta)}$ for some $\eta\in(0,1]$ and let $(\widetilde{\gamma}_t)_{t\ge1}$ be a sequence of $(0,1)$-valued steps satisfying

$$\sum_t \widetilde{\gamma}_t=+\infty \qquad \text{and} \qquad \sum_t \widetilde{\gamma}_t^{1+\eta}<+\infty.$$

Then, one defines recursively the following sequences

$$\begin{array}{rrcll} \forall\, t\ge 0, & p^{t+1}_i &:=& p^t_i(1-\widetilde{\gamma}_{t+1})+\widetilde{\gamma}_{t+1}\mathbf{1}_{\{i=i(t+1)\}}, & 1\le i\le N,\\ & p^0_i &:=& 0, & 1\le i\le N,\\ \forall\, t\ge 0, & D^{r,t+1}_N &:=& D^{r,t}_N(1-\widetilde{\gamma}_{t+1})+\widetilde{\gamma}_{t+1}|X^t_{i(t+1)}-\xi^{t+1}|^r, & \\ & D^{r,0}_N &:=& 0 & \end{array}$$

where $r\in[1,p]$. Then, on the event $\{X^t\to x^*\}$,

$$\forall\, i\in\{1,\ldots,N\},\qquad p^t_i \overset{a.s.}{\longrightarrow} \mathbb{P}_X(C_i(x^*)),\qquad \text{as } t\to\infty, \tag{2.24}$$

$$\forall\, r\in[1,p],\qquad D^{r,t}_N \overset{a.s.}{\longrightarrow} D^r_N(x^*)\qquad \text{as } t\to\infty. \tag{2.25}$$

Two natural choices for $(\widetilde{\gamma}_t)_{t\ge1}$ are $\widetilde{\gamma}_t=\gamma_t$ and $\widetilde{\gamma}_t=1/t$ (for some numerical experiments see [39]). The proof of (2.24) and (2.25) relies on some usual martingale techniques coming from Stochastic Approximation (see [38] or [3] for a detailed proof in the second setting). When $\widetilde{\gamma}_t=1/t$, one has a simple synthetic expression for (2.24) and (2.24) which can be attractive for numerical purpose, namely

$$p^t_i=\frac{1}{t}\left|\{s\in\{1,\ldots,t\}\mid \xi^s\in C_i(X^{s-1})\}\right| \quad \text{and} \quad D^{r,t}_N=\frac{1}{t}\sum_{s=1}^t|X^{s-1}_{i(s)}-\xi^s|^r. \tag{2.26}$$

These "companion" procedures are costless since they use some "by-products" of the competitive and learning phases of the procedure. They yield the parameters ($\mathbb{P}_X$-weights of the Voronoi cells $C_i(x^*)$, L^p-mean quantization error $\|X-\widehat{X}^{x^*}\|_p$) needed for a numerical use of the quantizer x^*. The fact that these companion procedures work on the event $\{X^t\to x^*\}$ (whatever the limiting N-tuple x^* is) shows their consistency.

Concerning the practical implementation of the algorithm, it is to be noticed that, in the quadratic case $p = 2$ ($CLVQ$ algorithm), at each step, the N-tuple X^{t+1} remains in the convex hull of X^t and ξ^{t+1}. This induces a stabilizing effect on the procedure which is observed on simulations which explains why the regular $CLVQ$ algorithm is more often implemented than its non-quadratic counterparts.

See [39] for an extensive numerical study of the $CLVQ$ algorithm for Gaussian random vectors. This lead to a large scale quantization of the multivariate normal distributions in dimensions $d = 1$ up to $d = 10$ with a wide range of values of N.

Optimization of the quantization tree: the extended $CLVQ$ 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 $\xi^0 := (\xi^{0,0}, \ldots, \xi^{n,0})$, $\xi^1 := (\xi^{0,1}, \ldots, \xi^{n,1}), \ldots, \xi^t := (\xi^t_0, \ldots, \xi^t_n), \ldots$ of $(X_k)_{0\le k\le n}$. Our aim is now to produce for every $k \in \{0, \ldots, n\}$ a *quadratic* optimal quantizer $X^{k,*} := (x^{k,*}_1, \ldots, x^{k,*}_{N_k})$ with size N_k, with its transition kernel $[\pi^{*,k}_{ij}]$, the distribution $(p^{*,k}_i)_{0\le i\le N_k}$ of $\widehat{X}^{x^*_k}_k$ and the induced mean L^p-quantization errors ($1 \le p \le 2$). Note that, if one sets

$$p^{*,k}_{ij} := \mathbb{P}\left(\{X_{k+1} \in C_j(x^{*,k+1})\} \cap \{X_k \in C_i(x^{*,k})\}\right)$$

then $\pi^{*,k}_{ij} = \frac{p^{*,k}_{ij}}{p^{*,k}_i}$ (and $p^{*,k}_i = \sum_j p^{*,k-1}_{ji}$), $k = 1, \ldots, n$. So one may focus on the estimation of the "joint distribution matrices" $[p^{*,k}_{ij}]$.

In the presentation below of the extended CLVQ algorithm, we assume that the Markov chain starts $X_0 = x_0 \in \mathbb{R}^d$, but other choices are possible. We also assume that

$$\forall\, k \in \{1, \ldots, n\}, \qquad \mathbb{P}_{X_k} \text{ is continuous and } \quad \mathbb{E}|X_k|^{2+\eta} < +\infty \tag{2.27}$$

for some $\eta > 0$. This is not a very demanding assumption when dealing with a diffusion process sampled at discrete times or an Euler scheme. We adopt here the setting in which the companion step sequence is $\widetilde{\gamma}_t = 1/t$ and we rely on the non-recursive expressions like (2.26). We propose to compute the L^r-mean quantization error for a fixed $r \in [1, 2]$ (usually $r = 1$ or 2 in applications). Then the algorithm reads as follows.

1. Initialization phase ($t = 0$):

- Initialize the n starting N_k-tuples $X^{k,0} := \{x^{0,k}_1, \ldots, x^{0,k}_{N_k}\}$, of the n $CLVQ$ algorithms that will quantize the distributions $\mathbb{P}_{X_k}$, $k = 1, \ldots, n$ [set $N_0 = 1$ and $X^{0,0} = \{x_0\}$].
- Initialize the joint distribution counters $\beta^{k,0}_{ij} := 0$, $i \in \{1, \ldots, N_k\}$, $j \in \{1, \ldots, N_k\}$, $k = 0, \ldots, n-1$.
- Initialize the marginal distribution counter $\alpha^{k,0}_i := 0$, $1 \le i \le N_k$, $k = 1, \ldots, n$.
- Initialize the L^r-mean quantization counter $d^{k,0} := 0$, $1 \le i \le N_k$, $k = 1, \ldots, n$.

2. Updating $t \rightsquigarrow t+1$: At step t, the N_k-tuples $X^{k,t}$, $1 \le k \le n$, have been obtained. We use $\xi^{t+1} := (\xi^{0,t+1}, \ldots, \xi^{k,t+1}, \ldots, \xi^{n,t+1})$ to carry on the optimization process at every time step *i.e.* updating the grids $X^{k,t}$ into $X^{k,t+1}$ as follows. For every $k = 1, \ldots, n$:

- Simulate $\xi^{k,t+1}$ (using $\xi^{k-1,t+1}$ if $k \ge 2$ or x_0 if $k = 1$).
- Select the "winner" in the k^{th} CLVQ algorithm *i.e.* the index $i^{k,t+1} \in \{1, \ldots, N_k\}$ satisfying

$$\xi^{k,t+1} \in C_{i^{k,t+1}}(X^{k,t}).$$

• Update the k^{th} $CLVQ$ algorithm:

$$X_i^{k,t+1} = X_i^{k,t} - \gamma_{t+1}\mathbf{1}_{\{i=i^{k,t+1}\}}(X_i^{k,t} - \xi^{k,t+1}), \qquad 1 \le i \le N_k.$$

• Update of the L^r-mean quantization error counter $d^{k,t}$:

$$d^{k,t+1} := d^{k,t} + |X_{i^{k,t+1}}^{k,t} - \xi^{k,t+1}|^p.$$

• Update the distribution counters $\beta^{k-1,t} := (\beta_{ij}^{k-1,t})_{1\le i\le N_{k-1},1\le j\le N_k}$ and $(\alpha_i^{k,t})_{1\le i\le N_k}$, $k = 1,\ldots,n$ (set $\alpha^{0,t+1} = t+1$ and $i^{0,t+1} := 1$):

$$\begin{aligned}
\beta_{ij}^{k-1,t+1} &:= \beta_{ij}^{k-1,t} + \mathbf{1}_{\{i=i^{k-1,t+1},\, j=i^{k,t+1}\}}, \quad 1 \le i \le N_{k-1},\ 1 \le j \le N_k \\
\alpha_i^{k,t+1} &:= \alpha_i^{k,t} + \mathbf{1}_{\{i=i^{k,t+1}\}}, \qquad 1 \le i \le N_k.
\end{aligned}$$

One shows, like for (2.24), that for every $k\in\{1,\ldots,n\}$, on the event $\Big\{X^{k-1,t} \longrightarrow x^{k-1,*}\Big\} \cap \Big\{X^{k,t} \longrightarrow x^{k,*}\Big\}$,

$$\frac{\beta_{ij}^{k-1,t}}{t} \overset{a.s.}{\longrightarrow} p_{ij}^{*,k} = \mathbb{P}(X_{k-1}\in C_i(x^{k-1,*}),\ X_k \in C_j(x^{k,*})), \qquad 1 \le i \le N_{k-1},\ 1 \le j \le N_k, \tag{2.28}$$

$$\frac{\alpha_i^{k,t}}{t} \overset{a.s.}{\longrightarrow} p_i^{*,k} = \mathbb{P}(X_k\in C_i(x^{k,*})), \qquad 1 \le i \le N_k, \tag{2.29}$$

$$\pi_{ij}^{k-1,t} := \frac{\beta_{ij}^{k-1,t}}{\alpha_i^{k-1,t}} \overset{a.s.}{\longrightarrow} \pi_{ij}^{*,k-1} = \mathbb{P}(X_k\in C_j(x^{k,*}) \mid X_{k-1}\in C_i(x^{k-1,*})), \qquad 1 \le i \le N_{k-1},\ 1 \le j \le N_k, \tag{2.30}$$

$$\frac{d^{k,t}}{t} \overset{a.s.}{\longrightarrow} D_{N_k}^{X_k,2}(x^{k,*}) \qquad \text{as } t \to +\infty. \tag{2.31}$$

From a practical viewpoint, this extended version has the same features as the regular $CLVQ$ algorithm as far as convergence is concerned. One important fact is that the optimizations of the quantizers at the successive time steps are processed simultaneously but *independently*: the quantization optimization at time step k does not affect that of time step $k+1$.

2.6 *A priori* error bounds in time and space

Proposition 3 below is the application of Theorem 2 to the general diffusion model (2.1) at times $t_k = kT/n$ and its Euler scheme. The error structure is the same except that the real constant does not depend on n (optimality of the quantizers $\widehat{X}_k$ is not required). The main result of this section is Theorem 4 which addresses the last optimization problem: assuming that every quantization $\widehat{X}_k$ is optimal, what is the optimal dispatching of the elementary quantizers among the n time discretization steps.

Proposition 3 *Assume that the coefficients b and c of the diffusion (2.1) and the obstacle function h are Lipschitz continuous. Let $(\widehat{v}_k(\widehat{X}_k))_{0\le k\le n}$ be the pseudo-Snell envelope of $(h(t_k, X_k))_{0\le k\le n}$ defined by (2.9). For every $p \in [1,+\infty)$, there exists a positive real constant $C_{[b]_{\rm Lip},[\sigma]_{\rm Lip},[h]_{\rm Lip},T,p} > 0$ such that*

$$\forall\, n \ge 1,\ \forall\, k\in\{0,\ldots,n\}, \quad \|V_k - \widehat{v}_k(\widehat{X}_k)\|_p \le C_{[b]_{\rm Lip},[\sigma]_{\rm Lip},[h]_{\rm Lip},T,p} \sum_{\ell=k}^{n} \|X_\ell - \widehat{X}_\ell\|_p. \tag{2.32}$$

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 kind of assumption on the marginal distributions of (X_k): we will assume that the L^p-mean quantization errors of the X_k are *φ-dominated* of in the following sense: there exists a random vector $R \in L^{p+\eta}(\mathbb{P})$ $(\eta > 0)$ and a sequence $(\varphi_{k,n})_{0\le k\le n<\infty}$ such that, for every $n \geq 1$, every $k \in \{0, \ldots, n\}$ and every $N \geq 1$,

$$\min_{x\in(\mathbb{R}^d)^N} \|X_k - \widehat{X}_k^x\|_p \leq \varphi_{k,n} \min_{x\in(\mathbb{R}^d)^N} \|R - \widehat{R}^x\|_p. \tag{2.33}$$

The point is that the distribution of R may depend on p but *not on N, k or n*. It is shown in [3] that uniformly elliptic diffusions $(cc^*(x) \geq \varepsilon_0 I_d,\ \varepsilon_0 > 0)$ satisfying either

– $b,\ c \in \mathcal{C}_b^\infty(\mathbb{R}^d)$ (hence with possibly *linear growth*) (following [31])

or

– b and c are *bounded*, b is differentiable, c is twice differentiable and Db, Dc and D^2c are bounded and Lipschitz (following [23], Theorem 5.4, p.148-149),

fulfill the domination property (2.33) with $\varphi_{k,n} := c_{b,\sigma,T}\sqrt{k/n}$. We show here that the local volatility model (1.3) also satisfies this domination property.

Proposition 4 *(Local volatility model) Assume that $q \geq d$ and that $\sigma : (0, +\infty)^d \to \mathbb{R}^{d\otimes q}$ is uniformly elliptic ($\sigma\sigma^*(\xi) \geq \varepsilon_0 I_d$, $\varepsilon_0 > 0$), bounded, three times differentiable and satisfies*

$$\forall\, \ell_1, \ldots, \ell_k \in \{1, \ldots, d\},\ \ \frac{\partial^k \sigma_{ij}}{\partial\xi^{\ell_1}\cdots\partial\xi^{\ell_k}}(\xi^1, \ldots, \xi^d) = O\left(\frac{1}{\xi^{\ell_1}\cdots\xi^{\ell_k}}\right) \quad \text{as} \quad |\xi| \to +\infty \tag{2.34}$$

for every $k = 1, 2, 3$. Then $(S_{t_k})_{0\le k\le n}$ satisfies the φ-domination property (2.33) with

$$\varphi_{k,n} := c_{\sigma,T}|s_0|\sqrt{k/n} \quad (c_{\sigma,T} > 0) \quad \text{and} \quad R := (Z^\ell + e^{Z^\ell})_{1\le\ell\le d},\ Z \sim \mathcal{N}(0; I_d), \tag{2.35}$$

Remark: Assumption (2.34) can be weakened into $\xi \mapsto \sigma\sigma^*(e^{\xi^1}, \ldots, e^{\xi^d})$ is bounded, twice differentiable with bounded Lipschitz first two differentials.

Proof: One starts from the elementary inequality, valid for every $\xi, \xi' \in \mathbb{R}$ and every $\rho > 0$,

$$|e^{\rho\xi'} - e^{\rho\xi}| \leq \rho|\xi' + e^{\xi'} - (\xi + e^\xi)|. \tag{2.36}$$

Let $Y_t := (\ln(S_t^1/s_0), \ldots, \ln(S_t^d/s_0))$ where S denotes a solution of (1.3) (with $r = 0$). Then Y is a diffusion process solution of the SDE

$$dY_t = \delta(Y_t)\,dt + \vartheta(Y_t)\,dW_t, \qquad Y_0 = (1, \ldots, 1),$$

with $\quad \delta(y) := -\dfrac{1}{2}\left(|\sigma_{\ell.}(e^{y^1}, \ldots, e^{y^d})|^2\right)_{1\le\ell\le d} \quad$ and $\quad \vartheta(y) := \sigma(e^{y^1}, \ldots, e^{y^d}).$

It follows from Assumption (2.34) on σ that δ and ϑ are twice differentiable and that δ, $D\delta$ and $D^k(\vartheta\vartheta^*)$, $k = 0, 1, 2$ are Lipschitz continuous and bounded. This implies (see [23], Theorem 5.4, p.148-149) that, for every $t \in (0, T]$, Y_t has an absolutely continuous distribution $\mathbb{P}_{Y_t} = p_t(y)\lambda_d(dy)$ satisfying

$$p_t(y) \leq \alpha\, \pi_{\sqrt{\beta t}Z}(y) \qquad (\alpha,\ \beta > 0)$$

where $\pi_{\sqrt{\beta t}Z}$ denotes the density function of $\sqrt{\beta t}\,Z$, $Z \sim \mathcal{N}(0; I_d)$.

Now let $N \geq 1$ and let $r^* := (r_i^*)_{1\leq i\leq N}$ be an L^p-optimal N-quantizer of the random vector R. One defines for every $k = 1,\ldots,n$, a N-quantizer $x^{k,*} := (x_i^{k,*})_{1\leq i\leq N}$ by

$$(x_i^{k,*})^\ell := s_0^\ell \exp\left(\sqrt{\beta t_k}\,(r_i^*)^\ell\right), \qquad \ell = 1,\ldots,d.$$

Now, coming back to (S_t) which starts now at $S_0 := s_0$, one has for every $k = 1,\ldots,n$,

$$\begin{aligned}
\inf_{x\in(\mathbb{R}_+^d)^N} \|S_{t_k} - \widehat{S}_{t_k}^x\|_p^p &\leq \|S_{t_k} - \widehat{S}_{t_k}^{x^{k,*}}\|_p^p \\
&= \mathbb{E}\left(\min_{1\leq i\leq N} |(s_0^\ell e^{Y_{t_k}^\ell})_{1\leq \ell\leq d} - x_i^{k,*}|^p\right) \\
&\leq \alpha\,\mathbb{E}\left(\min_{1\leq i\leq N}\left|\left(s_0^\ell(e^{\sqrt{\beta\,t_k}\,Z_{t_k}^\ell} - e^{\sqrt{\beta\,t_k}(r_i^*)^\ell}\right)_{1\leq\ell\leq d}\right|^p\right) \\
&\leq \alpha\,(\beta t_k)^{p/2}\max_{1\leq\ell\leq d}|s_0^\ell|^p\,\mathbb{E}\left(\min_{1\leq i\leq N}\left|(Z+e^Z) - \widehat{(Z+e^Z)}^{r^*}\right|^p\right).
\end{aligned}$$

The last inequality follows from (2.36). This completes the proof. $\diamond$

Assume that every quantization $\widehat{X}_k$ is L^p-optimal with size N_k. Then, combining the bounds obtained in Theorem 1 (time discretization error) and Proposition 3 (spatial discretization error) with Zador Theorem (Theorem 3, asymptotics of optimal quantization) yields the following error structure

$$\frac{C_1}{n^\theta} + C_2\sum_{k=1}^n \sqrt{t_k}\,N_k^{-\frac{1}{d}} \qquad \text{with} \qquad N_1+\cdots+N_n = N-1 \tag{2.37}$$

(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 (2.37), one has to make a balance between the time and spatial discretization errors. The results are detailed in Theorem 4 below.

Theorem 4 *(Optimized quantization tree and resulting error bounds) Assume that b, σ and h are Lipschitz continuous, that $(S_{t_k})_{0\leq k\leq n}$ is φ-dominated in the sense of (2.33) by $\varphi_{k,n} := c\sqrt{k/n}$ Let $n \geq 1$, $N \geq n+1$. Set $\widehat{X}_0 = S_0 = s_0$ and assume that, for every $k \in \{1,\ldots,n\}$, $\widehat{X}_k$ is an L^p-optimal (Voronoi) quantization of X_k with size*

$$N_k = |\widehat{X}_k(\Omega)| := \left\lceil \frac{t_k^{\frac{d}{2(d+1)}}(N-1)}{t_1^{\frac{d}{2(d+1)}}+\cdots+t_k^{\frac{d}{2(d+1)}}+\cdots+t_n^{\frac{d}{2(d+1)}}}\right\rceil, \tag{2.38}$$

where $\lceil x\rceil := \min\{k\in\mathbb{N} \mid k\geq x\}$ (then $N_0 = 1$ and $N \leq N_0+\cdots+N_n \leq N+n$). Let $(v_k(\widehat{X}_k))_{0\leq k\leq n}$ be the quantized pseudo-Snell envelope of $(h(t_k,X_k))_{0\leq k\leq n}$.

(a) Diffusion: *If $X_k := S_{t_k}$, $k = 0,\ldots,n$, then*

$$\max_{0\leq k\leq n}\|\mathcal{V}_{t_k} - \widehat{v}_k(\widehat{X}_k)\|_p \leq C_p e^{C_pT}\left(\frac{1+|s_0|}{n^\theta} + \frac{n^{1+\frac{1}{d}}}{N^{\frac{1}{d}}}\right).$$

with $\theta = 1$ if h is semi-convex and $\theta = 1/2$ otherwise.

(b) Euler scheme: *If $X_k := \overline{S}_{t_k}$, $k = 0,\ldots,n$, then*

$$\max_{0\leq k\leq n}\|\mathcal{V}_{t_k} - \widehat{v}_k(\widehat{X}_k)\|_p \leq C_p e^{C_pT}\left(\frac{1+|s_0|}{\sqrt{n}} + \frac{n^{1+\frac{1}{d}}}{N^{\frac{1}{d}}}\right).$$

Remark: If $n, N \to +\infty$ with $n = o(N)$ then $N_k \approx \frac{3d+2}{2(d+1)}\left(\frac{k}{n}\right)^{\frac{d}{2(d+1)}}\frac{N}{n}$ in (2.38).

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 (1.3) which drives the asset price process (S_t) (with $r = 0$). We assume that

$$q = d \qquad \text{and} \qquad \forall\, \xi \in \mathbb{R}^d,\ \sigma\sigma^*(\xi) \geq \varepsilon_0 I_d \tag{3.1}$$

so that $$\varepsilon_0 \text{Diag}((\xi^1)^2, \ldots, (\xi^d)^2) \leq cc^*(\xi) \leq \|\sigma\sigma^*\|_\infty |\xi|^2 I_d$$

where $\|\sigma\sigma^*(\xi)\|_\infty := \sup_{\xi \in \mathbb{R}^d} \|\sigma\sigma^*(\xi)\|$.

NOTATION: For notational convenience we will make the convention throughout this section that if X_t is a continuous time process (and $t_k = kT/n$),

$$\Delta X_{t_{k+1}} := X_{t_{k+1}} - X_{t_k}, \quad k = 0, \ldots, n-1.$$

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 variable. 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 [40]) that

$$h_T = \mathbb{E}(h_T) + \int_0^T H_s.dW_s = \mathbb{E}(h_T) + \int_0^T Z_s.dS_s \tag{3.2}$$

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

An analogy with discrete time model shows that the integral $\displaystyle\int_t^T Z_s.dS_s$ represents the (algebraic) gain from time t up to time T provided by the strategy $(Z_s^\ell)_{1 \leq \ell \leq d, s \in [0,T]}$ (at every time $s \in [t, T]$ the portfolio contains exactly Z_s^ℓ units of asset ℓ). 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

$$\text{Premium}_t := \mathbb{E}(h_T \mid \mathcal{F}_t) = \mathbb{E}(h_T) + \int_0^t Z_s.dS_s. \tag{3.3}$$

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 early 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 (1.7) 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 $(V^n_{t_k})_{0\le k\le n}$ denote the theoretical premium process of the Bermuda option related to $(h(t_k, S_{t_k}))_{0\le k\le n}$. It is a $(\mathcal{F}_{t_k})_{0\le k\le n}$-supermartingale defined as a Snell envelope by

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

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

$$\Delta A^n_k := A^n_k - A^n_{k-1} = V^n_{t_{k-1}} - \mathbb{E}_{t_{k-1}} V^n_{t_k} = \left(h(t_{k-1}, S_{t_{k-1}}) - \mathbb{E}_{t_{k-1}} V^n_{t_k}\right)_+. \tag{3.4}$$

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

$$M^n_k := \int_0^{t_k} Z^n_s . dS_s, \quad 0 \le k \le n, \quad \text{with} \quad \mathbb{E}\int_0^T |c^*(S_s) Z^n_s|^2 ds < +\infty \tag{3.5}$$

(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 unrealistic 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 introduce the closed subspace $\mathcal{P}_n$ of $L^2(c^*(S_.)d\mathbb{P}\otimes dt) := \{(Z)_{s\in[0,T]}$ progressively measurable, $\int_0^T |c^*(S_s)Z_s|^2 ds < +\infty\}$ defined by

$$\mathcal{P}_n = \left\{(\zeta_s)_{s\in[0,T]},\ \zeta_s := \zeta_{t_k},\ s \in [t_k, t_{k+1}),\ \zeta_{t_k}\ \mathcal{F}_{t_k}\text{-measurable},\ \mathbb{E}\int_0^T |c^*(S_s)\zeta_s|^2 ds < +\infty\right\}. \tag{3.6}$$

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

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

Doing so, we follow classical ideas introduced by by Föllmer & Sondermann ([20]) for hedging purpose in incomplete markets (see also [10]). One checks that $\mathcal{P}_n$ is isometric with the set of square integrable stochastic integrals with respect to $(S_{t_k})_{0\le k\le n}$, namely

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

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

$$\begin{aligned}\zeta^n_{t_k} &:= \left(\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}cc^*(S_s)\,ds\right)^{-1}\mathbb{E}_{t_k}\left(\int_{t_k}^{t_{k+1}}cc^*(S_s)Z^n_s ds\right)\\ &= \left(\mathbb{E}_{t_k}\Delta S_{t_{k+1}}(\Delta S_{t_{k+1}})^*\right)^{-1}\mathbb{E}_{t_k}(\Delta M^n_{k+1}\Delta S_{t_{k+1}}) \qquad (3.7)\\ &= \left(\mathbb{E}_{t_k}\Delta S_{t_{k+1}}(\Delta S_{t_{k+1}})^*\right)^{-1}\mathbb{E}_{t_k}\left(\Delta V^n_{t_{k+1}}\Delta S_{t_{k+1}}\right). \qquad (3.8)\end{aligned}$$

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

$$\Delta R^n_{t_{k+1}} := \int_{t_k}^{t_{k+1}}(Z^n_s-\zeta^n_{t_k}).dS_s = \Delta M^n_{k+1}-\zeta^n_{t_k}.\Delta S_{t_{k+1}} \qquad (3.9)$$

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

$$\mathbb{E}_{t_k}|\Delta R^n_{t_{k+1}}|^2 = \mathbb{E}_{t_k}\left(\int_{t_k}^{t_{k+1}}|c^*(S_s)(Z^n_s-\zeta_{t_k})|^2ds\right),\ k\in\{0,\dots,n-1\}. \qquad (3.10)$$

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

$$\mathbb{E}_{t_k}|\Delta R^n_{t_{k+1}}|^2=\mathbb{E}_{t_k}|\Delta V^n_{t_{k+1}}-\mathbb{E}_{t_k}\Delta V^n_{t_{k+1}}|^2-(\mathbb{E}_{t_k}\Delta S_{t_{k+1}}\Delta S^*_{t_{k+1}})^{-1}\left(\mathbb{E}_{t_k}\Delta V^n_{t_{k+1}}\Delta S_{t_{k+1}}\right)^2. \qquad (3.11)$$

Formulae (3.8) or (3.10), based on S_{t_k} and $V^n_{t_k}$ have natural approximations by quantization. On the other hand, (3.7) and (3.10) 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 consider 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

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

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

$$\overline{\zeta}^n_{t_k} := (\mathbb{E}_{t_k}\Delta\overline{S}_{t_{k+1}}\Delta\overline{S}^*_{t_{k+1}})^{-1}\mathbb{E}_{t_k}(\Delta\overline{V}^n_{t_{k+1}}\Delta\overline{S}_{t_{k+1}}) = \frac{1}{\Delta t_{k+1}}\,\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}\overline{Z}^n_s ds. \tag{3.12}$$

The related hedging default and local residual risk are defined by mimicking (3.10) and (3.11):

$$\Delta\overline{R}^n_{t_{k+1}} := \int_{t_k}^{t_{k+1}}(\overline{Z}^n_s - \overline{\zeta}^n_{t_k}).d\overline{S}_s = \Delta M^n_{k+1} - \overline{\zeta}^n_{t_k}.\Delta\overline{S}_{t_{k+1}} \tag{3.13}$$

$$\mathbb{E}_{t_k}|\Delta\overline{R}^n_{t_{k+1}}|^2 := \mathbb{E}_{t_k}|\Delta\overline{V}^n_{t_{k+1}} - \mathbb{E}_{t_k}\Delta\overline{V}^n_{t_{k+1}}|^2 - (\mathbb{E}_{t_k}\Delta\overline{S}_{t_{k+1}}\Delta\overline{S}^*_{t_{k+1}})^{-1}\Big(\mathbb{E}_{t_k}\Delta\overline{V}^n_{t_{k+1}}\Delta\overline{S}_{t_{k+1}}\Big)^2. \tag{3.14}$$

3.4 Quantized hedging and local residual risks

The quantized formulae for strategies and residual risks are simply derived from formulae (3.8) or (3.12) 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)

$$\widehat{\zeta}^n_k \;:=\; (\mathbb{E}_{t_k}\Delta\widehat{S}_{t_{k+1}}\Delta\widehat{S}^*_{t_{k+1}})^{-1}\widehat{\mathbb{E}}_k\left((\widehat{v}^n_{k+1}(\widehat{S}_{t_{k+1}}) - \widehat{v}^n_k(\widehat{S}_{t_k}))(\widehat{S}_{t_{k+1}} - \widehat{S}_{t_k})\right). \tag{3.15}$$

$$|\Delta\widehat{R}^n_{t_{k+1}}|^2 \;:=\; \mathbb{E}_{t_k}|\Delta\widehat{V}^n_{t_{k+1}} - \mathbb{E}_{t_k}\Delta\widehat{V}^n_{t_{k+1}}|^2 - (\mathbb{E}_{t_k}\Delta\widehat{S}_{t_{k+1}}\Delta\widehat{S}^*_{t_{k+1}})^{-1}\Big(\mathbb{E}_{t_k}\Delta\widehat{V}^n_{t_{k+1}}\Delta\widehat{S}_{t_{k+1}}\Big)^2. \tag{3.16}$$

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 elementary quantizer 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 time and spatial discretizations.

4.1 From Bermuda to America (time discretization)

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

$$V^n_t := V^n_{t_k} + \int_{t_k}^{t} Z^n_s.dS_s = V^n_{t_{k+1}} - \int_{t}^{t_{k+1}} Z^n_s.dS_s + \Delta A^n_{k+1},\; t \in [t_k, t_{k+1}). \tag{4.1}$$

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

$$V^n_{t_k -} = V^n_{t_k} + \Delta A^n_{k+1}. \tag{4.2}$$

Proposition 5 *Assume that the payoff process $h_t = h(t, S_t)$ where h is a semi-convex function. Assume that the diffusion coefficient c is Lipschitz continuous.*
(a) For every $k \in \{0,\dots,n\}$, $V^n_{t_k} \le \mathcal{V}_{t_k}$ and for every $t \in (t_k, t_{k+1})$, $\quad (V^n_t - \mathcal{V}_t)_+ \le \Delta A^n_{k+1}$.

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

(b) The following bound holds for the hedging strategies (in the "$\sqrt{cc^*}$ metric")

$$\mathbb{E}\left(\int_0^T |c^*(S_s)(Z_s - Z^n_s)|^2 ds\right) + \mathbb{E}\left(\int_0^T |c^*(S_s)Z^n_s - c^*(\overline{S}_s)\overline{Z}^n_s)|^2 ds\right) \le C_{h,c}\frac{T}{n}. \tag{4.3}$$

Proof: (a) The inequality between V^n and $\mathcal{V}$ at times t_k is obvious since $\mathcal{V}_t$ is defined as a supremum over a larger set of stopping times than $V^n_{t_k}$. Then, using the supermartingale property of $\mathcal{V}$, equality (4.1) and Jensen inequality yield

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

Now, using the expression (3.4) for ΔA^n_{k+1} and $V^n_{t_k} \ge h(t_{k+1}, S_{t_{k+1}})$ imply

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

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

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

for some constant $C_{h,c} > 0$. Finally, it yields

$$\Delta A^n_{k+1} \le C_{c,h} \frac{T}{n} \left(1 + \mathbb{E}_{t_k}(\max_{s\in[t_k,T]} |S_s|^2)\right). \tag{4.4}$$

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

$$\begin{aligned}
V^n_t &= V^n_{t_{k+1}} - \int_t^{t_{k+1}} Z^n_s.dS_s + \Delta A^n_{k+1} \le h(t_{k+1}, S_{t_{k+1}}) - \int_t^{t_{k+1}} Z^n_s\,.dS_s \quad (4.5)\\
\text{so that } \quad V^n_t &= \mathbb{E}_t(V^n_t) \ge \mathbb{E}_t(h(t_{k+1}, S_{t_{k+1}})) = h(t, S_t) + \mathbb{E}_t(h(t_{k+1}, S_{t_{k+1}}) - h(t, S_t)).
\end{aligned}$$

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

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

As it is a supermartingale as well, it necessarily satisfies

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

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

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

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

$$\int_0^T |c^*(S_s)(Z_s - Z^n_s)|^2 ds \quad + \quad \sum_{t_k \le T} (\Delta A^n_{t_k})^2 + (\mathcal{V}_t - V^n_t)^2$$

$$= -2\int_0^T(\mathcal{V}_s - V^n_{s-})(Z_s - Z^n_s).dS_s + 2\int_t^T(\mathcal{V}_s - V^n_{s-})d(K_s - A^n_{\underline{s}}).$$

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

since $V^n_{t_k-} = V^n_{t_k} + \Delta A^n_k \leq \mathcal{V}_{t_k} + \Delta A^n_k$. This yields, using the inequality obtained in (a) and (4.4),

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

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

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

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

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

$$\mathbb{E}\int_0^T|c^*(S_s)Z^n_s - c^*(\overline{S}_s)\overline{Z}^n_s|^2ds \leq 2\,\mathbb{E}\int_0^T(V^n_s - \overline{V}^n_s)d(K^n_s - \overline{K}^n_s) + \mathbb{E}(h(T,S_T) - h(T,\overline{S}_T))^2$$
$$\leq 2[h]_{\rm Lip}\mathbb{E}\int_0^T\mathbb{E}_s\left(\max_{t_k\geq s}|S_{t_k} - \overline{S}_{t_k}|\right)d(K^n_s + \overline{K}^n_s) + [h]^2_{\rm Lip}\|S_T - \overline{S}_T\|_2^2$$
$$\leq C\,\mathbb{E}\left(\sup_{t\in[0,T]}\mathbb{E}_t\left(\max_{t_k\geq t}|S_{t_k} - \overline{S}_{t_k}|\right)(K^n_T + \overline{K}^n_T)\right) + C\|S_T - \overline{S}_T\|_2^2$$
$$\leq C\|\sup_{t\in[0,T]}\mathbb{E}_t\max_{t_k\geq t}|S_{t_k} - \overline{S}_{t_k}|\,\|_2\left(\|K^n_T\|_2 + \|\overline{K}^n_T\|_2\right) + C\|S_T - \overline{S}_T\|_2^2$$
$$\leq C_{h,c}\frac{T}{n}\left(\|K^n_T\|_2 + \|\overline{K}^n_T\|_2 + 1\right). \qquad (4.6)$$

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

$$\|K^n_T - \overline{K}^n_T\|_2 \leq \|V^n_0\|_2 + \|\overline{V}^n_0\|_2 + \|\int_0^T Z^n_s.dS_s - \int_0^T\overline{Z}^n_s.d\overline{S}_s)\|_2 \leq C + O(1/\sqrt{n}) \quad \text{by (4.6)}$$

so that $\sup_n\|\overline{K}^n_T\|_2 < +\infty$. Plugging this back in (4.6) 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 and c are Lipschitz continuous and h is semi-convex, then,*

$$\|c^*(S_.)(Z_. - \zeta^n_.)\|_{L^2(d\mathbb{P}\otimes dt)} \le \|c^*(S_.)(Z_. - \mathrm{proj}_n(Z)_.)\|_{L^2(d\mathbb{P}\otimes dt)} + \frac{C}{\sqrt{n}} \tag{4.7}$$

($\mathrm{proj}_n(Z)$ is the projection of Z on $\mathcal{P}_n$). Furthermore $\|c^*(S_.)(Z_. - \mathrm{proj}_n(Z)_.)\|_{L^2(d\mathbb{P}\otimes dt)}$ goes to 0 as $n \to \infty$.

Remark: The term $\|c^*(S_.)(Z_. - \mathrm{proj}_n(Z)_.)\|_{L^2(d\mathbb{P}\otimes dt)}$ which rules the rate of convergence of $\|c^*(S_.)(Z_. - \zeta^n_.)\|_{L^2(d\mathbb{P}\otimes dt)}$ clearly depends on the path-regularity of Z_s. Theorem 6(c) below provides some elements about its own rate of convergence.

Proof: Set for convenience $\zeta := \mathrm{proj}_n(Z)$. 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 (4.3) in Proposition 5(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 $L^\infty(c^*(S_t)\, d\mathbb{P}\otimes dt)$ is everywhere dense in $L^2(c^*(S_t)\, d\mathbb{P}\otimes dt)$. $\diamondsuit$

4.2 Hedging error induced by the (quadratic) quantization

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

Proposition 6 *Assume that σ is Lipschitz continuous, bounded and uniformly elliptic and that h is Lipschitz continuous. Assume that the dispatching rule (2.38) of the N_k applies and that the quadratic quantization of the S_{t_k} are optimal. Assume that N and n go to $+\infty$ so that $\lim_n N/n^{d(1-\frac{1}{2(d+1)})+1} = +\infty$. Then, for every n*

$$|\zeta^n_0 - \widehat{\zeta}^n_0| \le \frac{C(1+|s_0|)}{\varepsilon_0 \min_{1\le\ell\le d}(s^\ell_0)^2}\,\frac{n^{\frac{3}{2}}}{(N/n)^{\frac{1}{d}}}.$$

Proof: The hedging vectors ζ^n_0 and $\hat{\zeta}^n_0$ satisfy respectively

$$(\mathbb{E}(\Delta S_{t_1}\Delta S^*_{t_1}))\zeta^n_0 = \mathbb{E}\left((V^n_1 - V^n_0)\Delta S_{t_1}\right) \tag{4.8}$$

$$(\mathbb{E}(\Delta\widehat{S}_{t_1}(\Delta\widehat{S}_{t_1})^*))\widehat{\zeta}^n_0 = \mathbb{E}\left((\widehat{V}^n_1 - \widehat{V}^n_0)\Delta\widehat{S}_{t_1}\right) \tag{4.9}$$

where $V^n_1 = v^n_1(S_{t_1})$ and $V^n_0 = v^n_0(s_0)$, etc. The quadratic quantization $\widehat{S}_{t_1}$ of S_{t_1} being optimal and $S_0 = \widehat{S}_0 = s_0$ being deterministic, one has $\mathbb{E}(\Delta S_{t_1} \mid \Delta\widehat{S}_{t_1}) = \Delta\widehat{S}_{t_1}$. In particular

$$\mathbb{E}(\Delta S_{t_1}) = \mathbb{E}(\Delta\widehat{S}_{t_1}) \quad \text{and} \quad \|\Delta\widehat{S}_{t_1}\|_2 \le \|\Delta S_{t_1}\|_2 = \|S_{t_1} - s_0\|_2 \le C\sqrt{T/n}(1+|s_0|).$$

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

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

One derives from (4.8) and (4.9) that

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

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

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

so that $\|(\mathbb{E}(\Delta S_{t_1}\Delta S_{t_1}^*))^{-1}\| \le \varepsilon_0^{-1}(\min_{1\le\ell\le d} s_0^\ell)^{-2} n/T$. First, one derives from (4.8) that $|\zeta_0^n| \le \frac{C}{\varepsilon_0}\sqrt{n}$. Hence

$$\begin{aligned}|\zeta_0^n - \widehat{\zeta}_0^n| \;&\le\; \frac{C\,n}{\varepsilon_0\min_{1\le\ell\le d}(s_0^\ell)^2}\left((1+|s_0|)\frac{\sqrt{n}}{(N/n)^{\frac{1}{d}}} + \frac{1}{\sqrt{n}\vee N_1^{\frac{1}{d}}} + \frac{1}{n\vee N_1^{\frac{2}{d}}}(1+|\widehat{\zeta}_0^n|)\right) \\ &\quad \frac{C\,n}{\varepsilon_0\min_{1\le\ell\le d}(s_0^\ell)^2}\left((1+|s_0|)\frac{\sqrt{n}}{(N/n)^{\frac{1}{d}}} + \frac{1}{\sqrt{n}\vee N_1^{\frac{1}{d}}} + \frac{1}{n\vee N_1^{\frac{2}{d}}}(1+|\widehat{\zeta}_0^n - \zeta_0^n| + \frac{1}{\varepsilon_0}\sqrt{n})\right).\end{aligned}$$

The dispatching rule (2.38) implies that $N_1 = C_d N n^{-1-\frac{d}{2(d+1)}} + o(N_1)$, so that, given the above assumption, $\lim_n \frac{N_1^{\frac{1}{d}}}{\sqrt{n}} = +\infty$ *i.e.* $\frac{n}{N_1^{\frac{2}{d}}}$ goes to 0. Consequently

$$|\zeta_0^n - \widehat{\zeta}_0^n| \le \frac{C\,n}{\varepsilon_0\min_{1\le\ell\le d}(s_0^\ell)^2}\left((1+|s_0|)\frac{\sqrt{n}}{(N/n)^{\frac{1}{d}}} + \frac{1}{N_1^{\frac{1}{d}}} + \frac{\sqrt{n}}{\varepsilon_0 N_1^{\frac{2}{d}}}\right).$$

Inspecting the three terms on the righthand side of the inequality completes the proof. $\diamond$

Remark: The above proof points outs the fact that a quantization tree optimized for the premium computation is not optimal at all for the hedging. So, the above error bound could be improved if one adopts another dispatching policy, optimized for the hedging, although it will never reach the performances devoted to the premium computation.

4.3 Approximation of the strategy: rate of convergence

In this section we evaluate the global residual risk on time intervals $[0, T']$, $T' < T$, induced by the use of the time discretization of the diffusion with step T/n, namely

$$\mathbb{E}\int_0^{T-\delta} |c^*(S_s)(Z_s - \zeta_s)|^2\, ds, \tag{4.10}$$

where (Z_t) is defined by (3.2) and $(\zeta_t) := \mathrm{proj}_n(Z)$ is the projection on the set $\mathcal{P}_n$ of elementary predictable strategies. Our basic assumption in this section is

$$(\Sigma)\ \equiv\ (i)\ c\in C_b^\infty((0,+\infty)^d), \qquad (ii)\ \sigma\sigma^*(x) \geq \varepsilon_0\, I_d.$$

Assumption $(\Sigma)(i)$ is fulfilled when $\sigma\in C_b((0,+\infty)^d)\cap C_b^\infty((0,+\infty)^d)$ and

$$\partial^k\sigma_{i.}(x) = O(1/|x^i|) \quad \text{as} \quad |x|\to+\infty,\ \ k\geq 1,\ \ i=1,\ldots,d.$$

Theorem 6 *Assume that* (Σ) *holds, that* h *is Lipschitz continuous and that* $s_0\in(0,+\infty)^d$.
(a) *For every* $T'\in[0,T)$ *there exists some real constants* K *and* θ *and an integer* $q\geq 2$ *(only depending on* T *and on the bounds of* σ *and its first two derivatives) such that*

$$\mathbb{E}\int_0^{T'} |c^*(S_s)(Z_s-\zeta_s)|^2\, ds \leq \left(1+\frac{1}{(T-T')^{3/2}}\right)(1+|s_0|)^q\frac{K}{\varepsilon_0^{5/2}}\frac{e^{\theta\sqrt{\ln n}}}{n^{\frac12}}. \tag{4.11}$$

(b) *Let* $\delta_n := \rho n^{-\frac13}$ $(\rho>0)$*. There exists some real constants* K *and* θ *and an integer* $q\geq 2$ *(depending on* ρ*,* T *and on the bounds of* σ *and its first two derivatives) such that*

$$\mathbb{E}\int_0^{T-\delta_n} |c^*(S_s)(Z_s-\zeta_s)|^2\, ds \leq \frac{K}{\varepsilon_0^{5/2}}(1+|s_0|)^q\frac{e^{\theta\sqrt{\ln n}}}{n^{\frac16}}. \tag{4.12}$$

(c) *If furthermore* h *is semi-convex. Then rates obtained in items* (a) *and* (b) *rule the rate of convergence of* $\|c^*(S_.)(Z_. - \zeta^n_.)\|_{L^2(\Omega\times[0,T_n]d\mathbb{P}\otimes dt)}$ *in Theorem 5 when* $T_n = T' < T$ *or* $T' = T-\delta_n$ *respectively.*

Remarks: • The term $e^{\theta\sqrt{\ln n}}$ 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 $X = \ln(S_t)$ instead of (S_t), then the obstacle function would have become $h(t,\exp x)$, with an exponential growth. So a truncation would have been necessary with a similar cost.

• 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 is $O(e^{\theta\sqrt{\ln n}}/n)$ and $O(e^{\theta\sqrt{\ln n}}/n^{1/3})$ in claims (a) and (b) of Theorem 6 respectively. Finally, in case of a *uniformly elliptic diffusion* our method of proof would lead to $O(1/n)$ and $O(1/n^{1/3})$ rates respectively.

Some technical difficulties arise when evaluating the term in (4.10) 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* $\eta_s := \dfrac{n}{T}\mathbb{E}_{t_k}\displaystyle\int_{t_k}^{t_{k+1}} H_u du$, $s\in[t_k,t_{k+1})$. *Then, under the assumptions of Theorem 6,*

$$\mathbb{E}\int_0^T |c^*(S_s)(Z_s-\zeta_s)|^2\, ds \leq \frac{C}{n} + 2\,\mathbb{E}\int_0^T |H_s-\eta_s|^2\, ds. \tag{4.13}$$

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

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

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

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

with

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

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

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

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

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

for some positive real constant C. Consequently

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

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,

$$\sum_{k=0}^{n-1}\lambda_{k+1}^2 \le \left(\sum_{k=0}^{n-1}\lambda_{k+1}\right)^2$$

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

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

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

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

It is a standard result on diffusions that $\|(1+\sup_{t\in[0,T]}|S_t|)^2\|_2$ is finite. It remains to prove that the term involving Z is finite. Since $cc^*(S_s) \geq \varepsilon_0\,\mathrm{Diag}((S_s^1)^2,\ldots,(S_s^d)^2)$, 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, by Schwartz's Inequality,

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

where $(S^{-1})_t := (1/S_t^1,\ldots,1/S_t^d)$ satisfies an SDE with bounded coefficients, so that its supremum has finite polynomial moments. Finally, the last inequality is a standard fact from *RBSDE* theory (see [19] or [2]). So we have proved that $J \leq C/n$.

The term I can be treated 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_t).

– The diffusion process (S_t) 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, when necessary, the obstacle h by a smoother obstacle $\underline{h}$. Namely, let $\varepsilon\in(0,1]$ and $\lambda>0$. We consider:

– 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

$$\|h-\underline{h}\|_\infty \leq C_h\,\varepsilon \quad\text{and}\quad \|(\partial_t+\mathcal{L}_c)\underline{h}\|_\infty \leq C_h\,\varepsilon^{-1} \tag{4.14}$$

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

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

$$\varphi_\lambda(\xi) := \xi \quad\text{if}\ \xi\geq e^{-\lambda}, \qquad \varphi_\lambda(\xi) := \frac{e^{-\lambda}}{2} \quad\text{if}\ \xi\leq \frac{1}{2}e^{-\lambda} \tag{4.15}$$

and

$$\forall\, m\geq 1, \qquad \left\|\varphi_\lambda^{(m)}\right\|_\infty \leq C_m\,e^{C_m\lambda} \tag{4.16}$$

where C_m is a real constant (not depending upon λ). Then the approximating diffusion coefficient c_λ defined for every $x=(x^1,\ldots,x^d)\in\mathbb{R}^d$ by

$$c_\lambda(x) := c(\varphi_\lambda(x^1),\ldots,\varphi_\lambda(x^d))$$

satisfies

$$c_\lambda \in \mathcal{C}_b^\infty(\mathbb{R}^d) \quad \text{ and } \quad c_\lambda c_\lambda^*(x) \geq \frac{\varepsilon_0}{4} e^{-2\lambda} I_d. \tag{4.17}$$

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

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

Let $\underline{P}_t(x,dy)$ denote its markov semi-group defined by $\underline{P}_t f(x) = \mathbb{E} f(\underline{S}_t^x)$. We will denote by $\underline{S}_t$ the solution $\underline{S}^{s_0}$ starting at $s_0 \in (0,+\infty)^d$. 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 progressively measurable $d\mathbb{P} \otimes dt$-square integrable process $\underline{H}$ (see [19] and [2] for this topic). We also consider the approximation

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

Lemma 2 *Assume that* (Σ) *holds. Then*

$$\mathbb{E} \int_0^T |H_s - \eta_s|^2\, ds \leq C \left(\varepsilon^2 + (1+|s_0|)^2 e^{-C\lambda^2/T} \right) + \mathbb{E} \int_0^T \left| \underline{H}_s - \underline{\eta}_s \right|^2 ds. \tag{4.18}$$

Proof: We rely on the stability property of $RBSDE$ (see [19] and [2]).

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

Let $\tau := \inf\{t > 0 \mid S_t \leq e^{-\lambda}\}$. One checks directly on model (1.1) that

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

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

$$\begin{aligned} \mathbb{E} \int_0^T |H_s - \underline{H}_s|^2\, ds &\leq C \left(\|h - \underline{h}\|_\infty^2 + \mathbb{E} \left(\sup_{0 \leq s \leq T} (|h(s, S_s)|^2 + |h(s, \underline{S}_s)|^2) \mathbf{1}_{\{\tau \leq T\}} \right) \right) \\ &\leq C \left(\varepsilon^2 + (1+|s_0|)^2 \sqrt{\mathbb{P}(\tau \leq T)} \right) \\ &\leq C(\varepsilon^2 + (1+|s_0|)^2 e^{-C\lambda^2/T}). \end{aligned}$$

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

$$\mathbb{E} \int_0^T \left| \eta_s - \underline{\eta}_s \right|^2 ds \leq \mathbb{E} \int_0^T |H_s - \underline{H}_s|^2\, ds. \qquad \diamond$$

We need now some analytical facts that we briefly recall here (see [19] and [2]). First of all we have the representation

$$\underline{Y}_t = u(t, \underline{S}_t), \quad \underline{H}_t = (c_\lambda^* \nabla_x u)(t, \underline{S}_t) \tag{4.19}$$

where u is the unique solution in a variational sense (see [2]) of the PDE

$$\begin{aligned}(\partial_t + \mathcal{L}_{c_\lambda})u(t,x) + \underline{F}(t,x,u(t,x)) &= 0, \; u(T,x) = \underline{h}(T,x), \quad (4.20)\\ \text{with} \qquad \underline{F}(t,x,u(t,x)) &= \vartheta(t,x)\mathbf{1}_{\{u(t,x)=\underline{h}(t,x)\}}\left((\partial_t + \mathcal{L}_{c_\lambda})\underline{h}(t,x)\right)_+\end{aligned}$$

where ϑ is a measurable function such that $0 \le \vartheta \le 1$. Set $F_t(x) := \underline{F}(t,x,u(t,x))$. It follows from (4.14) that $\sup\limits_{0\le t\le T}\sup\limits_{x\in\mathbb{R}^d} |F_t(x)| \le C_h/\varepsilon$ (where C_h is the real constant introduced in (4.14)). With this notation (4.20) becomes

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

in a variational sense. Then, it is a standard fact that u satisfies the mild form of the above PDE

$$u(t,x) = \underline{P}_{T-t}(\underline{h}_T)(x) + \int_t^T \underline{P}_{s-t}(F_s)(x)ds. \tag{4.21}$$

We focus now on the semi-group $\underline{P}_t$. It is well known (see [31]) that under assumption (4.17), $\underline{P}_t(x,dy) = \underline{p}_t(x,y)dy$ and for every $k \in \mathbb{N}$ and every multi-index $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}^m$ we have

$$\forall\, x,\, y \in \mathbb{R}^d,\ \forall\, t \in [0,T], \qquad \left|\partial_t^k D_x^\alpha \underline{p}_t(x,y)\right| \le K_{\alpha,k}(1+|x|)^{q_{\alpha,k}} \frac{e^{K_{\alpha,k}\lambda}}{\varepsilon_0^{1+k+|\alpha|/2}} \frac{e^{-C\frac{|x-y|^2}{t}}}{t^{k+\frac{|\alpha|+d}{2}}} \tag{4.22}$$

where $D_x^\alpha := \frac{\partial^{\alpha_1+\cdots+\alpha_m}}{\partial(x^1)^{\alpha_1}\cdots\partial(x^d)^{\alpha_d}}$, $|\alpha| = \alpha^1 + \cdots + \alpha_d$ and $K_{\alpha,k}$ and $q_{\alpha,k}$ are real constants depending on α, k and $C_{|\alpha|}$ (but not on λ).

One derives some straightforward consequences from this evaluation. First, using that $\left|\underline{h}_T(y)\right| \le C(1+|y|)$, it follows from (4.22) that there exists some constants K and q such that, for every $t \in (0,T]$, $x = (x^1, \ldots, x^d) \in \mathbb{R}^d$,

$$\left|\frac{\partial \underline{P}_t(\underline{h}_T)}{\partial x^k}(x)\right| \le C\int_{\mathbb{R}^d}\left|\frac{\partial \underline{p}_t(x,y)}{\partial x^k}\right|(1+|y|)\,dy \le \frac{K}{\varepsilon_0^{3/2}}(1+|x|)^q\frac{e^{K\lambda}}{\sqrt{t}} \tag{4.23}$$

$$\left|\frac{\partial^2 \underline{P}_t(\underline{h}_T)}{\partial x^k \partial x^\ell}(x)\right| \le \frac{K}{\varepsilon_0^2}(1+|x|)^q\frac{e^{K\lambda}}{t}, \quad \left|\frac{\partial^2 \underline{P}_t(\underline{h}_T)}{\partial t \partial x^k}(x)\right| \le \frac{K}{\varepsilon_0^{5/2}}(1+|x|)^q\frac{e^{K\lambda}}{t^{3/2}}. \tag{4.24}$$

Similarly, using that $\|F\|_\infty \le C_h/\varepsilon$ (and changing K for $K(C_h \vee 1)$), one gets for $F_s(x)$,

$$\left|\frac{\partial \underline{P}_t(F_s)}{\partial x^k}(x)\right| \le \frac{K}{\varepsilon_0^{3/2}}(1+|x|)^q\frac{e^{K\lambda}}{\varepsilon}\frac{1}{\sqrt{t}}, \quad \left|\frac{\partial^2 \underline{P}_t(F_s)}{\partial x^k \partial x^\ell}(x)\right| \le \frac{K}{\varepsilon_0^2}(1+|x|)^q\frac{e^{K\lambda}}{\varepsilon}\frac{1}{t} \tag{4.25}$$

and

$$\left|\frac{\partial^2 \underline{P}_t(F_s)}{\partial t \partial x^k}(x)\right| \le \frac{K}{\varepsilon_0^{5/2}}(1+|x|)^q\frac{e^{K\lambda}}{\varepsilon}\frac{1}{t^{3/2}}. \tag{4.26}$$

Lemma 3 *Set for every $t \in [0,T)$ and every $x \in \mathbb{R}^d$, $v_k(t,x) := \frac{\partial u}{\partial x_k}(t,x)$, $k = 1,\dots,d$. Assume that (4.17) holds.*

(*i*) *For every $t \in (0,T)$,*

$$|v_k(t,x)| \;\leq\; \frac{K}{\varepsilon_0^{3/2}}(1+|x|)^q e^{K\lambda}\left(\frac{1}{\varepsilon}+\frac{1}{\sqrt{T-t}}\right). \tag{4.27}$$

(*ii*) *For every $t, t' \in [0,T)$,*

$$\left|v_k(t,x)-v_k(t',x)\right| \;\leq\; \frac{K}{\varepsilon_0^{5/2}}(1+|x|)^q \frac{e^{K\lambda}}{\varepsilon}\left(\sqrt{|t-t'|}+\frac{|t-t'|}{(T-t\vee t')^{3/2}})\right). \tag{4.28}$$

(*iii*) *Let $\delta \in (0,T)$, for every $t \in (0,T-\delta)$,*

$$|v_k(t,x)-v_k(t,x')| \;\leq\; \frac{K}{\varepsilon_0^{2}}(1+|x|+|x'|)^q \frac{e^{K\lambda}}{\varepsilon}\left(\left(\frac{\varepsilon}{T-t}+\ln\left(\frac{T}{\delta}\right)\right)|x-x'|+\sqrt{\delta}\right). \tag{4.29}$$

Proof: We take derivatives in the mild equation (4.21) for u and we obtain

$$v_k(t,x) = \frac{\partial \underline{P}_{T-t}(\underline{h}_T)}{\partial x^k}(x) + \int_t^T \frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x)ds.$$

Let us begin by (*iii*). For every $t \in [0,T)$ and every $x, x' \in \mathbb{R}^d$,

$$v_k(t,x)-v_k(t,x') = \frac{\partial \underline{P}_{T-t}(\underline{h}_T)}{\partial x^k}(x) - \frac{\partial \underline{P}_{T-t}(\underline{h}_T)}{\partial x^k}(x') + \int_t^T \left(\frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x) - \frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x')\right)ds.$$

Hence, if $t \in [0,T-\delta)$, one derives using (4.24) and (4.25) that

$$\begin{aligned}
|v_k(t,x)-v_k(t,x')| \;&\leq\; \left|\frac{\partial \underline{P}_{T-t}(\underline{h}_T)}{\partial x^k}(x) - \frac{\partial \underline{P}_{T-t}(\underline{h}_T)}{\partial x^k}(x')\right| + \int_{t+\delta}^T \left|\frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x) - \frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x')\right|ds \\
&\quad + \int_t^{t+\delta}\left|\frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x)\right|ds + \int_t^{t+\delta}\left|\frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x')\right|ds \\
&\leq\; \frac{K}{\varepsilon_0^2}(1+|x|+|x'|)^q e^{K\lambda}\left(\left(\frac{1}{T-t}+\frac{1}{\varepsilon}\ln\left(\frac{T-t}{\delta}\right)\right)|x-x'|+\frac{\sqrt{\delta}}{\varepsilon}\right).
\end{aligned}$$

which yields the second inequality. Claim (*i*) follows similarly. Let us come to claim (*ii*). Assume without loss of generality that $t < t'$.

$$\begin{aligned}
v_k(t',x)-v_k(t,x) \;&=\; \frac{\partial \underline{P}_{T-t'}(\underline{h}_T)}{\partial x^k}(x) - \frac{\partial \underline{P}_{T-t}(\underline{h}_T)}{\partial x^k}(x) + \int_{t'}^T \left(\frac{\partial \underline{P}_{s-t'}(F_s)}{\partial x^k}(x) - \frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x)\right)ds \\
&\quad - \int_t^{t'} \frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x)ds
\end{aligned}$$

so that

$$\begin{aligned}
|v_k(t',x)-v_k(t,x)| \;&\leq\; \left|\frac{\partial \underline{P}_{T-t}(\underline{h}_T)}{\partial x^k}(x) - \frac{\partial \underline{P}_{T-t'}(\underline{h}_T)}{\partial x^k}(x)\right| + \int_{t'}^T \left|\frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x) - \frac{\partial \underline{P}_{s-t'}(F_s)}{\partial x^k}(x)\right|ds \\
&\quad + \int_t^{t'}\left|\frac{\partial \underline{P}_{s-t}(F_s)}{\partial x^k}(x)\right|ds.
\end{aligned}$$

Hence, one derives using (4.24) and (4.26) that

$$|v_k(t,x)-v_k(t',x)| \;\leq\; \frac{K}{\varepsilon_0^{5/2}}(1+|x|)^q\frac{e^{K\lambda}}{\varepsilon}\left(\frac{|t-t'|}{(T-t')^{3/2}}+\frac{|t-t'|}{(T-t')}+\sqrt{|t-t'|}\right)$$

which completes the proof. $\diamond$

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

Lemma 4 *(a) Let* $T' \in [0,T]$ *and* $\delta \in (0, T-T']$. *For every* $s, t \in [0,T']$,

$$\left(\mathbb{E}\,|\underline{H}_s^x - \underline{H}_t^x|^2\right)^{1/2} \le \frac{K(1+|x|)^{q+1}}{\varepsilon_0^{5/2}(T-T')^{\frac{3}{2}}}\frac{e^{K\lambda}}{\varepsilon}\left(\left(2+\sqrt{T}+\ln\!\left(\frac{T}{\delta}\right)\right)\sqrt{|t-s|}+\sqrt{\delta}\right). \tag{4.30}$$

(b) Let $\delta \in (0,T)$. *For every* $s,\ t \in [0, T-\delta)$, $|t-s| \le \delta$,

$$\left(\mathbb{E}\,|\underline{H}_s^x - \underline{H}_t^x|^2\right)^{1/2} \le \frac{K}{\varepsilon_0^{5/2}}(1+|x|)^{q+1}\frac{e^{K\lambda}}{\varepsilon}\left(\frac{T+2}{\delta}\sqrt{|t-s|}+\sqrt{\delta}\right). \tag{4.31}$$

Proof: (a) The functions c_λ are Lipschitz continuous with $[c_\lambda]_{\text{Lip}} \le Ce^{C\lambda}$ and satisfy $\|c_\lambda(x)\| \le C(1+|x|)$ where the real constant C does not depend on λ, consequently

$$|c_\lambda^*(x)\nabla_x u(t,x) - c_\lambda^*(x')\nabla_x u(t',x')| \le Ce^{C\lambda}|\nabla_x u(t,x)|\,|x-x'| + C(1+|x'|)|\nabla_x u(t,x) - \nabla_x u(t',x')|.$$

Combining the bounds obtained in Lemma 3 for the functions $v_k(t,x)$ leads to

$$\begin{aligned}
&|c_\lambda^*(x)\nabla_x u(t,x) - c_\lambda^*(x')\nabla_x u(t',x')| \\
&\le \frac{K}{\varepsilon_0^{5/2}}(1+|x|+|x'|)^q\frac{e^{K\lambda}}{\varepsilon}\left(\left(\frac{\varepsilon}{T-T'}+\frac{1}{\sqrt{T-T'}}+\ln\!\left(\frac{T}{\delta}\right)\right)|x-x'|+\sqrt{\delta}+\frac{|t-t'|}{(T-T')^{3/2}}\right) \\
&\le \frac{1}{(T-T')^{3/2}}\frac{K}{\varepsilon_0^{5/2}}(1+|x|+|x'|)^q\frac{e^{K\lambda}}{\varepsilon}\left(\left(2+\ln\!\left(\frac{T}{\delta}\right)\right)|x-x'|+\sqrt{\delta}+|t-t'|\right).
\end{aligned}$$

Consequently, using Holder Inequality and the 1/2-Holder regularity of $t \mapsto \underline{S}_t^x$ from $[0,T]$ into $L^4(\mathbb{P})$ (uniformly with respect to λ), one has for every $s,\ t \in [0,T]$,

$$\begin{aligned}
&\|\underline{H}_s^x - \underline{H}_t^x\|_2 \\
&\le \frac{1}{(T-T')^{3/2}}\frac{K}{\varepsilon_0^{5/2}}\|(1+|\underline{S}_s^x|+|\underline{S}_t^x|)^q\|_4\frac{e^{K\lambda}}{\varepsilon}\left(\left(2+\ln\!\left(\frac{T}{\delta}\right)\right)\|\underline{S}_s^x-\underline{S}_t^x\|_4+\sqrt{\delta}+|t-s|\right) \\
&\le \frac{1}{(T-T')^{3/2}}\frac{K}{\varepsilon_0^{5/2}}(1+|x|)^q\frac{e^{K\lambda}}{\varepsilon}\left(\left(2+\sqrt{T}+\ln\!\left(\frac{T}{\delta}\right)\right)(1+|x|)\sqrt{|t-s|}+\sqrt{\delta}\right).
\end{aligned}$$

(b) Still using the estimates Lemma 3 and, this time, $\ln(u) \le u$ and $\frac{|t-t'|}{(T-t\vee t')^{3/2}} \le \frac{\sqrt{|t-t'|}}{\delta}$

yields $$\begin{aligned}|c_\lambda^*(x)\nabla_x u(t,x) \;-\; &c_\lambda^*(x')\nabla_x u(t',x')| \\
\le\; &\frac{K}{\varepsilon_0^{5/2}}(1+|x|+|x'|)^q\frac{e^{K\lambda}}{\varepsilon}\left(\left(\frac{T+2}{\delta}\right)|x-x'|+2\sqrt{\delta}\right).\end{aligned}$$

One concludes the same way round. $\diamond$

Proof of Theorem 6: (a) Using (4.30) (still using the notation q instead of $q+1$)

$$\begin{aligned}
\mathbb{E}\int_0^{T'}\left|\underline{H}_s - \underline{\eta}_s\right|^2 ds &= \sum_{t_k<T'}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|\frac{1}{\Delta t_{k+1}}\int_{t_k}^{t_{k+1}}(\underline{H}_s-\underline{H}_r)dr\right|^2 ds \\
&\le \sum_{t_k<T'}\int_{t_k}^{t_{k+1}}\frac{1}{\Delta t_{k+1}}\int_{t_k}^{t_{k+1}}\mathbb{E}\,|\underline{H}_s-\underline{H}_r|^2\,drds \\
&\le \frac{K^2}{\varepsilon_0^5(T-T')^3}(1+|s_0|)^{2q}\frac{e^{2K\lambda}}{\varepsilon^2}\left(\varepsilon+\ln\!\left(\frac{T}{\delta}\right)\frac{1}{\sqrt{n}}+\sqrt{\delta}\right)^2.
\end{aligned}$$

Moreover, as a consequence of the first two lemmas,

$$\mathbb{E}\int_0^{T'}|c^*(S_s).(Z_s-\zeta_s)|^2\,ds \;\le\; \frac{C}{n}+C\,(1+|s_0|)^2e^{-C\lambda^2/T}+C\,\varepsilon^2$$
$$\frac{K^2(1+|s_0|)^{2q}}{\varepsilon_0^5(T-T')^3}\frac{e^{2K\lambda}}{\varepsilon^2}\left(\left(2+\sqrt{T}+\ln\!\left(\frac{T}{\delta}\right)\right)\frac{1}{\sqrt{n}}+\sqrt{\delta}\right)^2.$$

At this stage, we choose our parameters λ, ε and δ, depending on n. We set $\lambda_n := \sqrt{\frac{T}{2C}\ln n}$, $\delta_n := 4/n$ so that,

$$\left(2+\sqrt{T}+\ln\!\left(\frac{T}{\delta_n}\right)\right)\frac{1}{\sqrt{n}}+\sqrt{\delta_n}\le\frac{4+\sqrt{T}+\ln n+\ln(T/4)}{\sqrt{n}}.$$

Then, set $A(T,T') := \frac{K}{\varepsilon_0^{5/2}(T-T')^{3/2}}(1+|s_0|)^q$ and take the regularization parameter $\varepsilon := \varepsilon_n$ such that

$$\varepsilon_n^2 := \frac{A(T,T')}{\sqrt{C}}\frac{4+\sqrt{T}+\ln n+\ln(T/4)}{\sqrt{n}}e^{2K\lambda_n}.$$

Consequently

$$\begin{aligned}\mathbb{E}\int_0^{T'}|c^*(S_s)(Z_s-\zeta_s)|^2\,ds &\le \frac{C}{n}+C\,(1+|s_0|)^2e^{-C\lambda_n^2/T}+2\sqrt{C}A(T,T')\frac{4+\ln n+\ln(T/4)}{\sqrt{n}}e^{2K\lambda_n}\\ &\le \frac{C}{n}+\frac{C(1+|s_0|)^2}{\sqrt{n}}+CA(T,T')\frac{\ln n}{\sqrt{n}}e^{2K\frac{T}{C}\sqrt{\ln n}}\\ &\le C(1+|s_0|)^{q\vee 2}\frac{K}{\varepsilon_0^{5/2}}(1+(T-T')^{-\frac{3}{2}})\frac{e^{2K\frac{T}{C}\sqrt{\ln n}}}{\sqrt{n}}.\end{aligned}$$

(b) One carries out a similar optimization process, based this time on (4.31). One sets, for large enough n,

$$\delta_n := \rho n^{-1/3},\ \varepsilon_n^2 := \frac{K}{\varepsilon_0^5}(1+|s_0|)^q e^{2K\lambda_n}((T+2)/\rho+\sqrt{\rho})n^{-1/6},\ \lambda_n := \sqrt{\frac{T}{6C}\ln n}. \qquad \diamond$$

5 Numerical results on American style options

In this section, we present some numerical experiments concerning the pricing and the hedging of American style options in dimensions $d = 2$ up to 10. This study will be divided in two parts. First, we will show how to numerically estimate the spatial accuracy in each dimension in order to be able to produce a good choice of time and spatial discretization. Secondly, we will compute some prices and hedges following this choices.

5.1 The model

We specify the underlying asset model (1.1) into a d-dimensional Black & Scholes ($B\&S$) model, *i.e.* constant volatilities σ_ℓ with constant dividend rates μ_ℓ, $\ell = 1,\dots,d$:

$$dS_t^\ell = (r-\mu_\ell)S_t^\ell\,\mathrm{d}t+\sigma_\ell S_t^\ell\,\mathrm{d}W_t^\ell,\quad t\in[0,T],\quad \ell=1,\dots,d, \tag{5.32}$$

where $(W_t)_{t\in[0,T]}$ denotes a d-dimensional standard Brownian motion. The traded assets vector are $(e^{\mu_\ell t}S_t^\ell)$, $\ell = 1,\dots,d$, so that the discounted price satisfies (5.32) 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 to carry out convincing numerical experiments.

Beyond its importance for applications, in the of $B\&S$ model S_t is a *closed function* of (t, W_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 payoffs 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 achieved very accurately once for all and then stored off line. Indeed, the Brownian quantization is made of optimal quantizations of the d-dim standard Normal distributions by appropriate dilatations (see Figure 1) which are actually stored with all their companion parameters for a wide range of sizes (see [39]).

We focus on American style "geometric" exchange options which payoffs read

$$h(\xi) = \max\left(\xi^1 \cdots \xi^p - \xi^{p+1} \cdots \xi^{2p}, 0\right) \qquad \text{with} \qquad d := 2p. \tag{5.33}$$

It follows from the pricing formula (1.5) 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

$$\begin{aligned} Ex_{BS}(\theta, \xi, \xi', \widetilde{\sigma}, \mu) &:= \operatorname{erf}(d_1)\exp(\mu\theta)\,\xi - \operatorname{erf}(d_1 - \widetilde{\sigma}\sqrt{\theta})\,\xi', \\ d_1(\xi, \xi', \widetilde{\sigma}, \theta, \mu) &:= \frac{\ln(\xi/\xi') + (\widetilde{\sigma}^2/2 + \mu)\theta}{\widetilde{\sigma}\sqrt{\theta}} \qquad \text{and} \qquad \operatorname{erf}(\xi) := \int_{-\infty}^{\xi} e^{-\frac{u^2}{2}} \frac{du}{\sqrt{2\pi}} \end{aligned}$$

$$\text{with } \theta := T - t,\ \widetilde{\sigma} := \left(\sum_{\ell=1}^{d} \sigma_\ell^2\right)^{1/2},\ \mu := \sum_{\ell=1}^{p} \mu_\ell - \sum_{\ell=p+1}^{d} \mu_\ell,\ \xi := \prod_{\ell=1}^{p} S_t^\ell,\ \xi' = \prod_{\ell=p+1}^{d} S_t^\ell. \tag{5.34}$$

We will also use some American geometric put payoffs:

$$h(\xi^1, \ldots, \xi^d) := \left(K - \left(\xi^1 \cdots \xi^d\right)^{1/d}\right)_+.$$

In this case, the explicit formulæ for the European Put with strike K and maturity T at time t (with $\mu_i = 0$ and $\sigma_i = \sigma$, $i = 1, \ldots, d$) reads

$$\begin{aligned} P_{BS}(\theta, K, \xi, \sigma, r) &:= \operatorname{erf}(-d_2 + \sigma\sqrt{\theta/d})\exp(-r\theta)K - \operatorname{erf}(-d_2)\,\xi, \qquad (5.35) \\ d_2(K, \xi, \sigma, \theta, r) &:= \frac{\ln(\xi/K) + (r + \sigma^2/(2d))\theta}{\sigma\sqrt{\theta/d}}, \end{aligned}$$

where $\theta = T - t$ and $\xi = \left(\prod_{i=1}^{d} S_t^i\right)^{1/d} \exp\left(-\frac{\sigma^2(d-1)}{2d}\right)$.

5.2 Specification of the numerical scheme

Let us specify now the implemented numerical scheme. As mentioned above, our approach to pricing consists first in quantizing the d-dim Brownian motion W. 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 (2.38) "size" to the N_k-quantizer of time t_k so that $N_0 = 1$, $N \le 1 + N_1 + N_2 + \cdots + N_n \le N + n$. First, we compute for every $k \in \{1, \dots, n\}$ an optimal (quadratic) N_k-quantizer of $\mathcal{N}(0; I_d)$ by processing a $CLVQ$ algorithm (2.21) (the final converging phase is refined using a randomized version of the so-called Lloyd I fixed point procedure, see *e.g.* [26]). For further details about the implementation, see [39]. As a second step, we get the optimal N_k-quantizer $(x_i^k)_{i=1,\dots,N_k}$ of W_{t_k} by a $\sqrt{t_k}$-dilatation. All the companion parameters (weights p_i^k, p_{ij}^k, L^2-quantization errors) are then estimated by a standard Monte Carlo simulation. Note that all these quantities are universal objects that can be kept off line, once computed accurately enough.

In this very particular but important case, we only need the original CLVQ algorithm defined by (2.22) and (2.23), not its extended version developed for general diffusions.

Finally, the *quantization tree algorithm* (2.12) reads

$$\begin{cases} v_i^n := h_i^n, \quad i = 1, \dots, N_n, \\ v_i^k := \max\left(h_i^k, \sum_{1 \le j \le N_{k+1}} \pi_{ij}^k \, v_j^{k+1}\right), \quad i = 1, \dots, N_k, \ k = 0, \dots, n-1 \end{cases} \tag{5.36}$$

where the *obstacle* is given by

$$h_i^k := h(s_i^{k,1}, \dots, s_i^{k,d}) \ \text{ with } \ s_i^{k,\ell} := s_0^\ell \exp\left(-\left(\mu_\ell + \frac{\sigma_\ell^2}{2}\right) k\Delta t + \sigma_\ell \, x_i^k\right), \ \ell = 1, \dots, d,$$

and the weights π_{ij}^k are Monte-Carlo proxies of the theoretical weights *i.e.*

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

(About the error induced by the Monte Carlo approximation, see [4] and [1]). Following (3.15) the hedging δ_i^k at x_i^k is then computed by

$$\delta_i^{k,\ell} := \frac{\displaystyle\sum_{j=1}^{N_{k+1}} \pi_{ij}^k (v_j^{k+1} - v_i^k)(e^{\mu_\ell t_{k+1}} s_{j,\ell}^{k+1} - e^{\mu_\ell t_k} s_i^k)}{\displaystyle\sum_{j=1}^{N_{k+1}} \pi_{ij}^k (e^{\mu_\ell t_{k+1}} s_j^{k+1,\ell} - e^{\mu_\ell t_k} s_i^{k,\ell})^2}, \quad \ell = 1, \dots, d. \tag{5.37}$$

In practice, we often need to introduce in the quantization tree algorithm a sequence of “control variate variables”. This is usually achieved by considering a $\mathcal{F}_{t_k}^S$-martingale

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

Then one sets $M_k^i := m(t_k, s_i^k)$ so that the (explicit) sequence $(M_i^k)_{1 \le i \le N_k, 1 \le k \le n}$, *i.e.* approximately satisfies:

$$\sum_{j=1}^{N_{k+1}} \pi_{ij}^k \, M_j^{k+1} \approx M_i^k. \tag{5.38}$$

The approximation comes from the spatial discretization by quantization (in fact if the equality did hold it would be of no numerical interest). Here, an efficient choice is to take

$$M_i^k = Ex_{BS}(T - t_k, \prod_{\ell=1}^{p} s_i^{k,\ell}, \prod_{\ell=p+1}^{d} s_i^{k,\ell}, \widetilde{\sigma}, \mu). \tag{5.39}$$

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

$$\text{Premium}^h(t_k, s_i^k) := m(t_k, s_i^k) + v_i^{h-m,k} \tag{5.40}$$

where $(v_i^{h-m,k})_{1\le k\le n}$ is obtained by the scheme (5.36) with the obstacle $(h_k^i - m(t_k, s_i^k))_{1\le k\le n}$.

Let us emphasize that "control variate variables" (M_i^k) such that (5.38) holds exactly is useless in practice since in this case it is not difficult to see that

$$\forall\, i, k, \quad v_i^{h-m,k} = v_i^{h,k} - M_i^k.$$

5.3 Numerical accuracy, stability

We will now estimate numerically the rate of convergence (at time $t = 0$) of the numerical premium $p(n, \bar{N}) := \text{Premium}^h(0, s_0)$ given by (5.36) using (5.40) towards a reference p_{th} as a function of $(n, \bar{N})$ where $\bar{N} := N/n$ (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 [42]). The error terms both in time and in space given by Theorem 4 are

$$E(n, \bar{N}) = |p(n, \bar{N}) - p_{th}| \approx \frac{c_1}{n} + c_2\,\frac{n}{\bar{N}^{\alpha}} \quad \text{with} \quad \alpha = 1/d \qquad (5.41)$$

for semi-convex payoffs. Two questions are raised by this error bound:

– are these rates optimal?

– Is it possible to compute an optimal number n_{opt} of time steps to minimize the global error?

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

We begin by the 1 and 2-dimension settings. The specifications of the reference model (5.32) 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

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

Note first that c_1 does not depend upon $\bar{N}$: this confirms the above global error structure (5.41). 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 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: let X be a square integrable random variable, x an optimal quadratic N-quantizer; if f admits a Lipschitz continuous differential Df, then (see [38])

$$|\mathbb{E}f(X) - \sum_{1\le i\le N} \mathbb{P}(\widehat{X}^x = x_i) f(x_i) - \sum_{1\le i\le N} Df(x_i).\underbrace{\mathbb{E}((X - x_i)\mathbf{1}_{C_i(x)})}_{=0 \text{ since } x \text{ is optimal}}| \le [Df]_{\text{Lip}} \|X - \widehat{X}^x\|_2^2, \quad (5.42)$$

where $\|X - \widehat{X}^x\|_2^2 = c_X N^{-2/d} + o(N^{-2/d})$ as $N \to \infty$. The optimality of x 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 favor of optimal quantization.

From dimension 4 to 10, the storage of the matrix $[\pi_{ij}^k]$ for increasing values of $\bar{N}$ 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, true optimal quantizers 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 $\bar{N}$ and n (rather than N and n). For a given value of $\bar{N}$, one proceeds as above a nonlinear regression that yields numerical values for c_1 and $C_2 := c_2\bar{N}^{-1/d}$. Finally set

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

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

5.4 Numerical results for American style options

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

We first settle the value of $\bar{N}$ and then read on Table 2 the optimal number $n = n_{opt}(d,\bar{N})$ of time steps. Space discretization is the one used for the above numerical experiments. The model parameters and initial data are settled so that μ and $\widetilde{\sigma}$ remain constant, equal to -5% and 20% respectively in (5.34):

$$\begin{aligned}
&\mu_1 := -5,\ \mu_i := 0,\ i = 2,\ldots,d, \quad \sigma_i := 20/\sqrt{d},\ i = 1,\ldots,d,.\\
&\quad s_0^i := 40^{2/d},\ i = 1,\ldots,d/2, \qquad s_0^i := 36^{2/d},\ i = d/2+1,\ldots,d \quad \text{(in-the-money)},\\
&\quad s_0^i := 36^{2/d},\ i = 1,\ldots,d/2, \qquad s_0^i := 40^{2/d},\ i = d/2+1,\ldots,d \quad \text{(out-of-the-money)}.
\end{aligned}$$

In Figure 2 are displayed the computed premia a) and hedges b) in 2-dimension at time $t = 0$ together with the reference ones as a function of the maturity $T \in [0, T_{max}]$ for $T_{max} = 1$. Figure 2 emphasizes that both premia and hedges in 2-dimension are very well fitted with the reference premium. It also holds true in the Out-of-the-money case (not depicted here).

In general, in the In-the-money case, we can see on Figure 3(a) and Table 3 that the computed premium tends to overestimate the reference one when the maturity grows. This phenomenon grows also when the dimension d increases. 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). In the Out-of-the-money setting, we can see on Figure 3(b) that very different behaviors are observed on the premia. Indeed whatever the dimension is (from 4 to 10), the premia seem to be well computed (dimension other than 4 are not depicted here). Figure 4 depicts the quantized residual risk (at $t = 0$) as a function of the maturity. It suggests that numerical incompleteness of the market has a bigger impact "in-the-money" than "out-of-the-money".

We will now test the influence of the European premium when used as a "control variate variable" in the simulations. To this aim, we will price American puts on a geometrical index in dimension $d = 5$. The model parameters and initial data are

$$\mu_i = 0,\ s_0^i = 100,\ \sigma_i = 20\%, \quad i = 1,\ldots,d,$$

and

$$r = \ln(1.1), \quad K = 100.$$

This choice is motivated by the fact that then the European premium is significantly lower the American premium. The reference prices and hedges are computed using a BBSR algorithm (see [12]) with 1000 time steps in dimension 1 with

$$s_{0,eq} = 100,\ \sigma_{eq} = \sigma_1/\sqrt{d},\ \delta_{eq} = \frac{\sigma_1^2(d-1)}{2d},$$

where $s_{0,eq}$, σ_{eq} and δ_{eq} are the "1d-equivalent"s spot, volatility and dividend rate. The quantized prices are still computed using (5.40) and algorithm (5.36) where the "control variate variable" is known by (5.35) and the hedges are computed using (5.37).

Table 4 shows the price and hedges computed for $(n, N_{max}) = (10, 2800)$. We can see that the price error is 0.5% and the sum of the hedge errors of each components is 0.8%.

Now, Figure 5 shows the influence of the European "control variate variable" (5.35). We have plotted the American premium computed following (5.35), (5.36) and (5.40) for an "optimal" time and space discretization found in Table 4, namely $(n, N_{max}) = (10, 2800)$. We can see that the European premium counts for a little part in the American one. Here we can see that the quantization is able to capture by itself a significant part of the price as the maturity T varies in $[0, 1]$.

References

[1] V. Bally, The central limit theorem for a nonlinear algorithm based on quantization. Stochastic analysis with applications to mathematical finance., *Proc. R. Soc. Lond.*, **460**, No 2041, 221-241, 2004.

[2] V. Bally, M.E. Caballero, B. Fernandez, N. El Karoui, Reflected BSDE's, PDE's and Variational inequalities. Pre-print RR-4455 INRIA, 2002.

[3] V. Bally, G. Pagès, A quantization algorithm for solving discrete time multi-dimensional discrete time Optimal Stopping problems, *Bernoulli*, **9**, 1003-1049, 2003.

[4] V. Bally, G. Pagès, Error analysis of a quantization algorithm for obstacle problems, *Stoch. Proc. and their Appl.*, **106**, 1-40, 2003.

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

[6] V. Bally, G. Pagès, J. Printems, First order schemes in the numerical quantization method, *Mathematical Finance*, **13**, No 1, 1-16, 2002.

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

[8] 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.

[9] B. Bouchard, N. Touzi, Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations, *Stoch. Proc. and their Appl.*, **111**, No 2, 175-206, 2004.

[10] N. Bouleau, D. Lamberton, Residual risks and hedging strategies in Markovian markets, *Stoch. Proc. and their Appl.*, **33**, 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, 1323-1352, 1997.

[12] M. Broadie, J. Detemple, American option valuation: new bounds, approximations and a comparison with an existing method, *Review of Financial Studies*, **9**, $n^0$4, 1211–1250, 1996.

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

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

[15] 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.

[16] É. Clément, P. Protter, D. Lamberton, An analysis of a least squares regression method for American option pricing, *Finance & Stochastics*, **6**, n^0 2, 449-471, 2002.

[17] P. Cohort, Limit Theorems for the Random Normalized Distortion, *The Ann. of Appl. Probab.*, **14**, No 1, 118-143, 2004.

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

[19] 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 Probab.*, **25**, No 2, 702-737, 1997.

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

[21] 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**, 391-412, 1999.

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

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

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

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

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

[27] A. Kohatsu-Higa, R. Pettersson, Variance reduction methods for simulation of densities on Wiener space, *SIAM J. Numer. Anal.*, **40** No. 2, 431-450, 2002.

[28] 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

[29] 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

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

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

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

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

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

[35] P.L. Lions, H. Régnier, Calcul des prix et des sensibilités d'une option américaine par une méthode de Monte Carlo, working paper, 2002.

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

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

[38] G. Pagès, A space vector quantization method for numerical integration, *Journal of Comput. Appl. Math.*, **89**, 1-38, 1997.

[39] G. Pagès, J. Printems, Optimal quadratic quantization for numerics: the Gaussian case, *Monte Carlo Meth. and Appl.*, **9**, $\mathrm{n}^0 2$, 135-168, 2003.

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

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

[42] S. Villeneuve, A. Zanette (2002), Parabolic A.D.I. methods for pricing american option on two stocks, *Mathematics of Operation Research*, **27**, $\mathrm{n}^o 1$, 121-149, 2002.

Table 1: Estimation of the spatial convergence exponent α of (5.41) in dimensions $d = 1, 2$.

	$d=1$				$d=2$	
$\bar{N}_i$	$\bar{N}_1 = 20$	$\bar{N}_2 = 30$	$\bar{N}_3 = 40$	$\bar{N}_4 = 60$	$\bar{N}_1 = 235$	$\bar{N}_2 = 455$
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 2: Estimation of the optimal number of time steps for $d = 1, 2, 4, 6, 10$.

	$d=1$	$d=2$	$d=4,\ \bar{N}=750$	$d=6,\ \bar{N}=1000$	$d=10,\ \bar{N}=1000$
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}	$0.63\,\bar{N}$	$1.31\,\bar{N}^{1/2}$	19	24	50

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

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 4: Value of an American put at time $t = 0$ and the hedging strategy on a geometrical index in dimension 5 for maturity $T = 1$, $\sigma_i = 0.2$, $r = \ln(1.1)$, $s_0^i = 100 = K$, $i = 1, \ldots, 5$.

n	N_{max}	AM Qtf.	BBSR	δ_i Qtf.					BBSR
10	2800	1.576	1.584	-0.0739	-0.0779	-0.0750	-0.0751	-0.0789	-0.0756

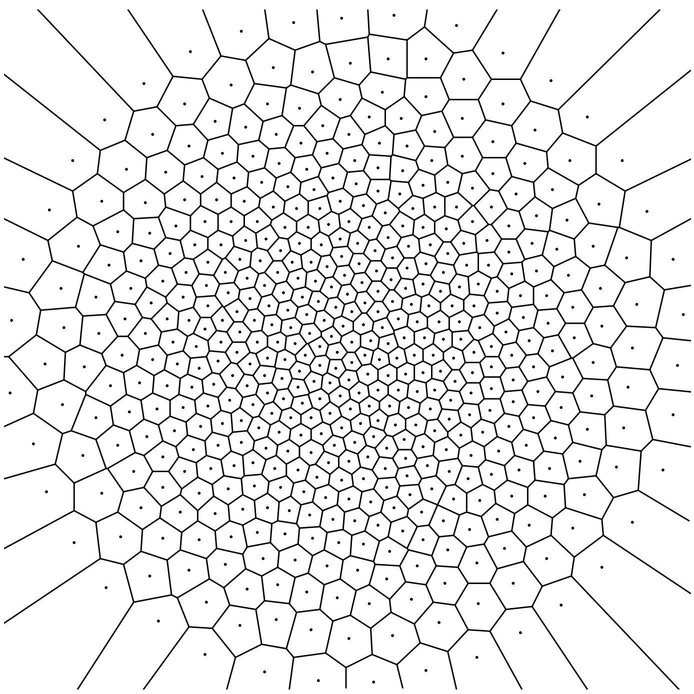

Figure 1: *A 500-tuple with its Voronoi tessellation with the lowest quadratic quantization error for the bi-variate normal distribution.*

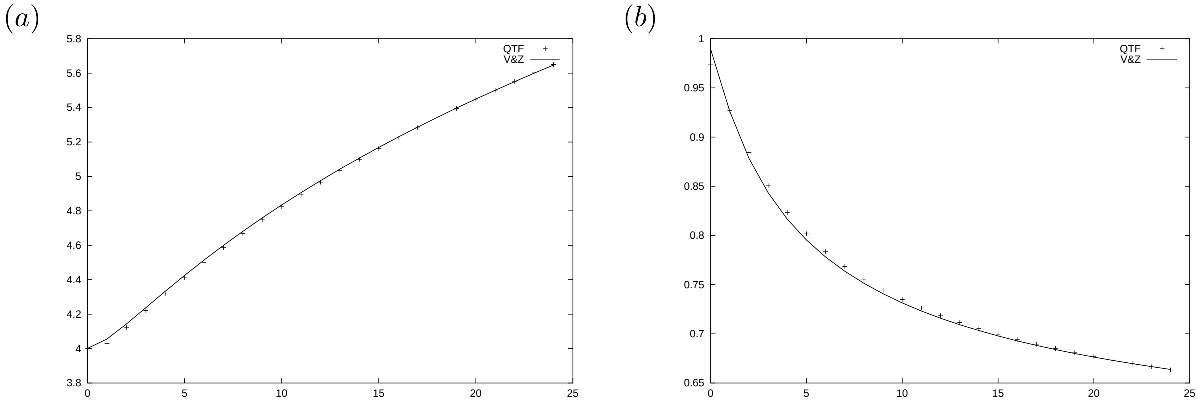

Figure 2: *$d = 2$, $n = 25$ and $\bar{N} = 300$. American premium as a function of the maturity: a); Hedging strategy on the first asset: b). The cross + depicts the premium obtained with the method of quantization and – depicts the reference premium (V & Z) (cf. [42]).*

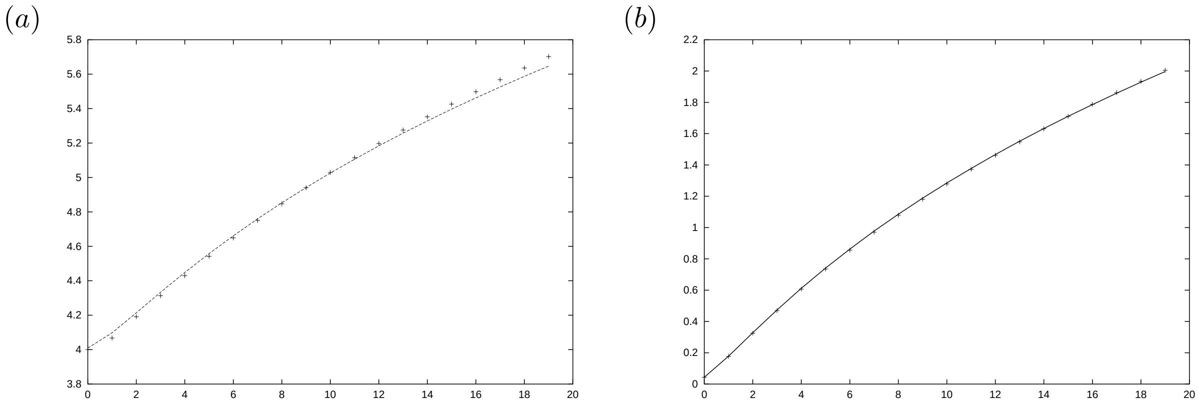

Figure 3: *$d = 4$. In-the-money: (a); Out-of-the-money: (b). American premium as a function of the maturity. + depicts the premium obtained with the method of quantization and – depicts the reference premium (V & Z) (cf. [42]).*

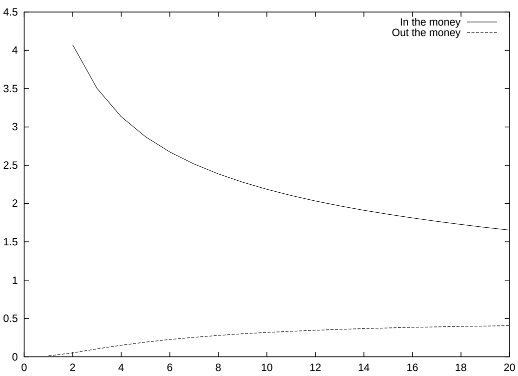

Figure 4: *Quantized local residual risk $|\Delta\widehat{R}^n_{t_1}|^2$ as a function of the maturity in 4-dimension with* $n = 20$, $\overline{N} = 750$ *(see Table 2) (see the definition of local residual risk in (3.10) computed owing to (3.16) in the "In-the-money" case (solid line) and "Out-of-the-money" case (dash line)*

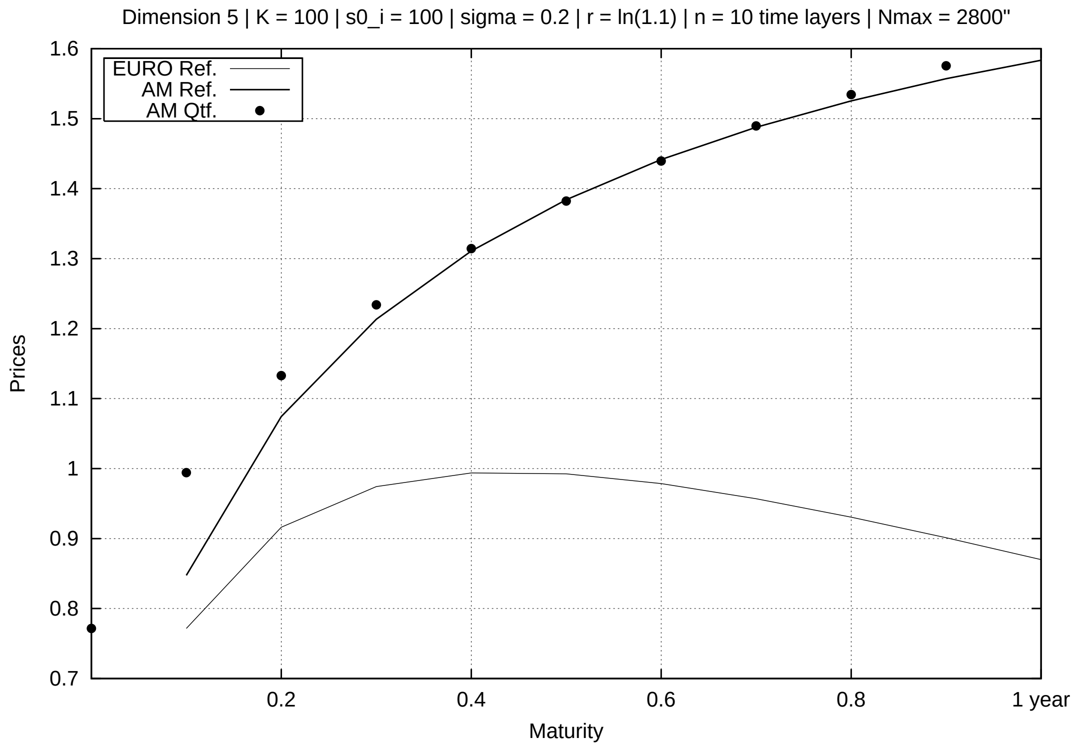

Figure 5: *American Put premium on a geometrical index in dimension 5 as a function of maturity. Here,* $s_0^i = 100$*,* $\sigma_i = 0.2$*,* $r = \ln(1.1)$ *and* $K = 100$*. Time and space discretization are* $(n, N_{max}) = (10, 2800)$*. The bold line depicts the reference price computed by a BBSR 1d–algorithm, the thin line depicts the European premium (i.e. the "control variate variable") and the points depict the quantized American Premium at each time step.*