Optimal quantization methods and applications to numerical problems in finance

Gilles PAGÈS
Laboratoire de Probabilités et
Modèles Aléatoires
CNRS, UMR 7599
Université Paris 6
gpa@ccr.jussieu.fr

Huyên PHAM
Laboratoire de Probabilités et
Modèles Aléatoires
CNRS, UMR 7599
Université Paris 7
pham@math.jussieu.fr
and CREST

Jacques PRINTEMS
Centre de Mathématiques
CNRS, UMR 8050
Université Paris 12
and INRIA
printems@univ-paris12.fr

April 2003

Abstract

We review optimal quantization methods for numerically solving nonlinear problems in higher dimension associated with Markov processes. Quantization of a Markov process consists in a spatial discretization on finite grids optimally fitted to the dynamics of the process. Two quantization methods are proposed: the first one, called marginal quantization, relies on an optimal approximation of the marginal distributions of the process, while the second one, called Markovian quantization, looks for an optimal approximation of transition probabilities of the Markov process at some points. Optimal grids and their associated weights can be computed by a stochastic gradient descent method based on Monte Carlo simulations. We illustrate this optimal quantization approach with four numerical applications arising in finance: European option pricing, optimal stopping problems and American option pricing, stochastic control problems and mean-variance hedging of options and filtering in stochastic volatility models.

Key words: Quantization, Markov chain, Euler scheme, Numerical integration, Optimal stopping, option pricing, stochastic control, Non linear filtering, stochastic gradient descent.

MSC Classification (2000): 65-00 (90C39, 93E35, 91B28, 65L08, 60G35, 60G40).

1 Introduction

Optimal quantization of random vectors consists in finding the best possible approximation (in L^p) of a $\mathbb{R}^d$-valued random vector X by a measurable function $\varphi(X)$ where φ takes at most N values in $\mathbb{R}^d$. This is a very old story which starts in the early 50's. The idea was to use a finite number N of codes (or "quantizers") to transmit efficiently a continuous stationary signal. Then it became essential to optimize the geometric location of these quantizers for a given distribution of the signal and to evaluate the resulting error. In a more mathematical form, the problem is to find out a measurable function φ^* (if some) such that

$$\|X - \varphi^*(X)\|_p = \inf\left\{\|X - \varphi(X)\|_p,\ \varphi : \mathbb{R}^d \to \mathbb{R}^d,\ |\varphi(\mathbb{R}^d)| \leq N\right\}$$

and then to evaluate $\|X - \varphi^*(X)\|_p$, especially when N goes to infinity. These problems have been extensively investigated in information theory and signal processing (see [12]). However, from a computational point of view, optimal quantization remained essentially limited to one dimensional signals, the optimization process, essentially deterministic, becoming intractable for multi-dimensional signals. The drastic cut down of massive Monte Carlo simulation cost on computers made possible the implementation of alternative procedures based on probabilistic ideas (see the $CLVQ$ algorithm below). This gave birth to many applications and extensions in various fields like automatic classification, data analysis and artificial neural networks. Let us mention *e.g.* the self-organizing maps introduced by Kohonen in the early 80's (see [15]). More recently, this leads to consider optimal quantization as a possible spatial discretization method to solve multi-dimensional (discrete time) problems arising in numerical probability. An important motivation to tackle these questions comes from finance since most problems arising in that field are naturally multi-dimensional.

First, an application to numerical integration in medium dimension ($1 \leq d \leq 4$) was developed and analyzed in [18]. A second step consisted in applying optimal quantization to solve nonlinear problems related to a (discrete time) Markovian dynamics. A first example was provided by discrete time optimal stopping problems (by the way of American option pricing), still in a multi-dimensional setting (see [1], [3] and [4]). From a probabilistic point of view, the nonlinearity usually appears through functionals of conditional expectations that need to be computed. From a computational point of view, the quantization approach leads to some tree algorithms in which, at every time step is associated a grid of quantizers, assumed to be optimal in some sense for the Markov chain. Then, investigating various fields of applications like stochastic control or nonlinear filtering, it turned out that it could be useful to specialize the way one quantizes Markov chains according to the nature of the encountered problem. This gave rise to two variants of the quantization: the *marginal quantization* introduced in [1] that focused on the optimization of the marginal distributions of the Markov chain and the *Markovian quantization* introduced in [19] that enhances the approximation of the conditional distributions at some points. Both approaches are presented here with some applications to finance, along with some further developments (1^{st} order schemes).

The paper is organized as follows: Section 2 is devoted to general background on optimal vector quantization of random vectors. First, the main properties concerning the existence of an optimal quantization and its rate of convergence toward 0 as its size goes to infinity are recalled. Then, numerical methods to get optimal quantizers and their associated weights are described. A first application to numerical integration is presented which points out in a simple setting the main features of this spatial discretization method. In Section 3, we present the two methods used so far to quantize Markov chains, called *marginal* and *Markovian* quantization methods. Both methods are applied to compute expectation of functionals $\phi_0(X_0)\dots\phi(X_n)$ of the Markov chain. Then, the main theoretical and computational features of both methods are discussed. In Section 4, three main applications to finance are described including some numerical illustrations: American option pricing, stochastic control and filtering of stochastic volatility. Finally, in Section 5 we explain on an example how one can design some first order schemes based on optimal quadratic quantization that significantly improve the rate of convergence of the above methods.

Throughout the paper, $|\,\xi\,|$ will denote the usual canonical Euclidean norm of $\xi \in \mathbb{R}^d$.

2 Optimal quantization of a random vector

2.1 Existence and asymptotics of optimal quantization

The basic idea of quantization is to replace a $\mathbb{R}^d$-valued random vector $X \in L^p(\Omega, \mathbb{P})$ by a random vector taking at most N values in order to minimize the induced L^p-error *i.e.* one wishes to solve the minimizing problem error

$$\min\left\{\|X-Y\|_p,\ Y:\Omega\to\mathbb{R}^d,\ \text{measurable}\ ,\ |Y(\Omega)|\le N\right\}.$$

Let $Y:\Omega\to\mathbb{R}^d$ be such a random vector and let $\Gamma = Y(\Omega)$. Then, consider a closest neighbor rule projection $\mathrm{Proj}_\Gamma:\mathbb{R}^d\to\Gamma$ and set,

$$\widehat{X}^\Gamma := \mathrm{Proj}_\Gamma(X).$$

One easily checks that $\|X-\widehat{X}^\Gamma\|_p \le \|X-Y\|_p$. Assume $|\Gamma| = N$ and $\Gamma = \{x^1,\dots,x^N\}$. Closest neighbor rule projections Proj_Γ are in one-to-one correspondence with *Voronoi tessellations* of Γ, that is with Borel partitions $C_1(\Gamma)$, …, $C_N(\Gamma)$ of $\mathbb{R}^d$ satisfying:

$$C_i(\Gamma) \ \subset\ \left\{\xi\in\mathbb{R}^d : |\xi-x^i| = \min_{x^j\in\Gamma}|\xi-x^j|\right\},\quad i=1,\dots,N.$$

Then, one may set $\mathrm{Proj}_\Gamma(\xi) := \sum_{i=1}^N x^i \mathbf{1}_{C_i(\Gamma)}(\xi)$ so that

$$\widehat{X}^\Gamma \ = \ \sum_{i=1}^N x^i \mathbf{1}_{C_i(\Gamma)}(X). \tag{2.1}$$

In the sequel, the exponent $^\Gamma$ in $\widehat{X}^\Gamma$ will be often dropped.

The L^p-error induced by this projection – called L^p-quantization error – is given by $\|X - \widehat{X}^\Gamma\|_p$. It clearly depends on the grid Γ; in fact, one easily derives from (2.1):

$$\|X - \widehat{X}^\Gamma\|_p^p = \mathbb{E}\left[\min_{1\le i\le N} |X - x^i|^p\right], \tag{2.2}$$

for $\Gamma = \{x^1, \ldots, x^N\}$. So, if one identifies a grid Γ of size N with the N-tuple $x := (x^1, \ldots, x^N) \in (\mathbb{R}^d)^N$ or any of its permutation, the p^{th} power of the L^p-quantization error is a symmetric function Q_N^p defined on N-tuples with pairwise distinct components by

$$\begin{aligned} Q_N^p(x^1, \ldots, x^N) &:= \int q_N^p(x, \xi)\, \mathbb{P}_X(d\xi) \qquad (\mathbb{P}_X \text{ is for the distribution of } X) \\ \text{where} \qquad q_N^p(x, \xi) &:= \min_{1\le i\le N} |x^i - \xi|^p, \qquad x = (x^1, \ldots, x^N) \in (\mathbb{R}^d)^N,\ \xi \in \mathbb{R}^d. \end{aligned}$$

(The function $\sqrt[p]{q_N^p}$ is sometimes called *local quantization error*.) The extension of the function Q_N^p on the whole $(\mathbb{R}^d)^N$ is obvious.

Then two questions naturally arise: does this function reach a minimum? how does this minimum behave as N goes to infinity? They have been investigated for a long time as part of quantization theory for probability distributions, first in information theory and signal processing in the 1950's and, more recently in probability for both numerical or theoretical purpose (see [13, 18]). They make up the core of optimal quantization. We will now shortly recall these main results. For a comprehensive, one may consult [13] and the references therein.

First, the size N being settled, the function $\sqrt[p]{Q_N^p}$ is Lipschitz continuous and does reach a minimum (although Q_N^p does not go to infinity as $\max_{1\le i\le N} |x^i| \to \infty$). If $|X(\Omega)| \ge N$, then any N-tuple that achieves the minimum has pairwise distinct components *i.e.* defines a grid Γ^* of size N satisfying

$$\|X - \widehat{X}^{\Gamma^*}\|_p = \min\left\{\|X - Y\|_p,\ Y \text{ random vector in } \mathbb{R}^d,\ |Y(\Omega)| \le N\right\}. \tag{2.3}$$

If $|X(\Omega)|$ is infinite, this minimum (strictly) decreases to 0 as N goes to infinity. Its rate of convergence is ruled by the so-called Zador theorem, completed by several authors: Zador, Bucklew & Wise (see [8]) and finally Graf & Luschgy in [13].

Theorem 2.1 *(see [13]) Assume that $\mathbb{E}|X|^{p+\varepsilon} < +\infty$ for some $\varepsilon > 0$. Then*

$$\lim_N \left(N^{\frac{p}{d}} \min_{|\Gamma|\le N} \|X - \widehat{X}^\Gamma\|_p^p\right) = J_{p,d}\left(\int_{\mathbb{R}^d} g^{\frac{d}{d+p}}(\xi)\, d\xi\right)^{1+\frac{p}{d}} \tag{2.4}$$

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

Remark 2.1 In higher dimension, the true value of $J_{p,d}$ is unknown except in 1 dimension where $J_{p,1} = \frac{1}{2^p(p+1)}$. However, one shows that $J_{2,2} = \frac{5}{18\sqrt{3}}$ and that $J_{p,d} \sim \left(\frac{d}{2\pi e}\right)^{\frac{p}{2}}$ as d goes to infinity (see [13] for some proofs and other results using non Euclidean norms).

This theorem says that $\min_{|\Gamma|\leq N} \|X - \widehat{X}^{\Gamma}\|_p \sim C_{\mathbb{P}_X,p,d} N^{-\frac{1}{d}}$. This is in accordance with the rates $O(N^{-1/d})$ easily obtained with orthogonal lattice grid of size $N = m^d$ for the uniform distribution $U([0,1]^d)$ over the unit hypercube $[0,1]^d$. In fact, even in that very specific setting, these lattice grids are not optimal quantizers for $U([0,1]^d)$ (except if $d = 1$). In fact *optimal quantization provides for every $N \geq 1$ the "best fitting" grid of size N for a given distribution $\mathbb{P}_X$. This grid corresponds to the real constant $C_{\mathbb{P}_X,p,d}$ when N goes to infinity.*

2.2 How to get optimal quantization?

At this stage, the next question clearly is: how to get numerically an optimal N-grid with a minimal L^p-quantization error? Historically, the first attempt to solve this optimization problem – when $p = 2$ and $d = 1$ – is the so-called "Lloyd's methods I". This iterative procedure acts on the grids as follows: let Γ^0 be a grid of size N. Then set by induction

$$\Gamma^{s+1} = \mathbb{E}\left[X \,|\, \mathrm{Proj}_{\Gamma^s}(X)\right](\Omega) \;=\; \left(\mathbb{E}\left[X \,|\, X \in C_i(\Gamma^s)\right]\right)_{1\leq i\leq N}, \quad s \in \mathbb{N}.$$

One shows that $\{\|X - \mathrm{Proj}_{\Gamma^s}(X)\|_2, s \in \mathbb{N}\}$ is a nonincreasing sequence and that, under some appropriate assumptions (see [14]), $\mathrm{Proj}_{\Gamma^s}(X)$ converges toward some random vector $\widehat{X}$ taking N values as s goes to infinity. Moreover, $\hat{X}$ satisfies the *stationary quantizer* property

$$\widehat{X} = \mathbb{E}[X \,|\, \widehat{X}] \tag{2.5}$$

and is the only solution to the original optimization problem $\mathrm{argmin}\,\{\|X - Y\|_2, |Y(\Omega)| \leq N\}$.

When the dimension d is greater than 1, the convergence may fail. When some convergence holds, the limit $\widehat{X}$ is still stationary but has no reason to minimize the quadratic quantization error. In a general setting, this algorithm has two main drawbacks: it is a purely "local" procedure which does not explore the whole state space, and, furthermore, it becomes numerically intractable in its original form since it requires the computation of d-dimensional integrals $\int_C \ldots d\mathbb{P}_X$. When the random vector X *is simulatable*, one can randomize the Lloyd's methods I by using a Monte Carlo simulation to compute the above integrals. This version is sometimes used as a final step of the optimization procedure to "refine" locally the results obtained by other methods like that described below.

We will describe a procedure which partially overcomes these drawbacks, based on another property of the L^p-quantization error function Q_N^p: its smoothness. Let us temporarily identify a grid $\Gamma := \{x^1, \ldots, x^N\}$ of size N with the N-tuple $x = (x^1, \ldots, x^N)$ and let us denote the Voronoi tessel of x^i by $C_i(x)$ instead of $C_i(\Gamma)$.

Proposition 2.1 *([18]) Let $p > 1$. The function Q_N^p is continuously differentiable at any N-tuple $x \in (\mathbb{R}^d)^N$ having pairwise distinct components and a $\mathbb{P}_X$-negligible Voronoi tessellation boundary $\cup_{i=1}^N \partial C_i(x)$. Its gradient $\nabla Q_N^p(x)$ is obtained by formal differentiation:*

$$\nabla Q_N^p(x) \;=\; \mathbb{E}\left[\nabla_x\, q_N^p(x, X)\right], \tag{2.6}$$

$$\textit{where} \quad \nabla_x\, q_N^p(x,\xi) \;=\; \left(\frac{\partial\, q_N^p}{\partial x^i}(x,\xi)\right)_{1\leq i\leq N} := p\left(\frac{x^i - \xi}{|x^i - \xi|}|x^i - \xi|^{p-1}\mathbf{1}_{C_i(x)}(\xi)\right)_{1\leq i\leq N} \tag{2.7}$$

with the convention $\frac{\vec{0}}{|\vec{0}|} = \vec{0}$. *If* $\mathbb{P}_X$ *is continuous the above formula (2.6) still holds for* $p=1$. *(Note that* $\nabla_x q_N^p(x,\xi)$ has exactly one non-zero component $i(x,\xi)$ *defined by* $\xi \in C_{i(x,\xi)}(x)$.*)*

One shows (see [13], p.38) that any N-tuple $x^* \in \text{argmin} Q_N^p$ satisfies the "boundary" assumption of Proposition 2.1 so that $\nabla Q_N^p(x^*) = 0$.

The integral representation (2.6) of ∇Q_N^p strongly suggests, as soon as independent copies of X can be easily simulated on a computer, to implement a *stochastic gradient algorithm* (or *descent*). It is a stochastic procedure recursively defined by

$$\Gamma^{s+1} = \Gamma^s - (\delta_{s+1}/p)\nabla_x\, q_N^p(\Gamma^s, \xi^{s+1}) \tag{2.8}$$

where the initial grid Γ^0 has N pairwise distinct components, $(\xi^s)_{s\geq 1}$ is an i.i.d. sequence of $\mathbb{P}_X$-distributed random vectors, and $(\delta_s)_{s\geq 1}$ a sequence of $(0,1)$-valued step parameters satisfying the usual conditions:

$$\sum_s \delta_s = +\infty \quad \text{and} \quad \sum_s \delta_s^2 < +\infty. \tag{2.9}$$

Note that (2.8) *a.s.* grants by induction that Γ^s has pairwise distinct components. In an abstract framework (see *e.g.* [9] or [17]), under some appropriate assumptions, a stochastic gradient descent associated to the integral representation of a so-called potential function *a.s.* converges toward *a local minimum of this potential function* (Q_N^p in our problem). Although these assumptions are not fulfilled by Q_N^p the encountered theoretical problems can be partially overcome (see [18] for some *a.s.* convergence results in 1-dimension or when $\mathbb{P}_X$ is compactly supported). Practical implementation does provide satisfactory results (a commonly encountered situation with gradient descents). Some estimates of the companion parameters ($\mathbb{P}_X$-weights of the tessels and L^p-quantization errors) can be obtained as by-product of the procedure. This is discussed below.

Stationary quantizers (Back to): When $p = 2$, standard computations show that Equation $\nabla Q_N^2(x) = 0$ is simply the stationary quantizer property: if Γ is the corresponding grid then, $\widehat{X}^\Gamma$ satisfies Equation (2.5). This identity has interesting applications (see the next two paragraphs below). It also implies that, for every $p \in [1,+\infty]$, $\|\widehat{X}^\Gamma\|_p \leq \|X\|_p$.

Note that non optimal quantizers may be stationary: when $\mathbb{P}_X = \mathbb{P}_{X^1} \otimes \cdots \otimes \mathbb{P}_{X^d}$ is a product measure, any "lattice grid" made up with optimal – or even stationary – quantizers of its marginal distributions $\mathbb{P}_{X^i}$ is stationary. It can also be the case of any *local minima* of Q_N^2 which are the natural targets of the above stochastic gradient descent algorithm.

Practical aspects of the optimization, companion parameters: Formula (2.8) can be developed as follows if one sets $\Gamma^s := \{x^{1,s}, \ldots, x^{N,s}\}$,

$$\text{Competitive phase : select } i(s+1) := i(\Gamma^s, \xi^{s+1}) \in \text{argmin}_i |x^{i,s} - \xi^{s+1}| \tag{2.10}$$

$$\text{Learning phase : } \begin{cases} x^{i(s+1),s+1} := x^{i(s+1),s} - \delta_{s+1} \frac{x^{i(s+1),s} - \xi^{s+1}}{|x^{i(s+1),s} - \xi^{s+1}|} |x^{i(s+1),s} - \xi^{s+1}|^{p-1} \\ x^{i,s+1} \quad := x^{i,s}, \qquad i \neq i(s+1). \end{cases} \tag{2.11}$$

The competitive phase (2.10) corresponds to selecting the closest point in Γ^s *i.e.* $i(s+1)$ such that $\xi^{s+1} \in C_{i(s+1)}(\Gamma^s)$. The learning phase (2.11) consists in updating the closest neighbor and leaving still other components of the grid Γ^s.

Furthermore, it is established in [18] that, if $X \in L^{p+\varepsilon}$ ($\varepsilon > 0$), then the sequences $(Q_N^{r,s})_{s\geq 1}$, $0 < r \leq p$, and $(\pi_i^s)_{t\geq 1}$, $1 \leq i \leq N$, of random variables recursively defined by

$$Q_N^{r,s+1} := Q_N^{r,s} - \delta_{s+1}(Q_N^{r,s} - |x^{i(s+1),s} - \xi^{s+1}|^r), \qquad Q_N^{r,0} := 0, \tag{2.12}$$

$$\pi_i^{s+1} := \pi_i^s - \delta_{s+1}(\pi_i^s - \mathbf{1}_{\{i=i(s+1)\}}), \qquad \pi_i^0 := 1/N, \quad 1 \leq i \leq N. \tag{2.13}$$

satisfy on the event $\{\Gamma^s \to \Gamma^*\}$

$$Q_N^{r,s} \xrightarrow{a.s.} Q_N^r(\Gamma^*),\ 0<r\leq p, \quad \text{and} \quad \pi_i^s \xrightarrow{a.s.} \mathbb{P}_X(C_i(\Gamma^*)),\ 1 \leq i \leq N, \ \text{as} \ s \to \infty.$$

These companion procedures are essentially costless since they are steps of the grid optimization procedure itself and they yield the parameters of numerical interest (weights of the Voronoi cells, L^r-quantization errors of Γ^*, $0<r\leq p$) for the grid Γ^*. Note that this holds whatever the limiting grid Γ^* is: this means that the procedure is consistent.

The quadratic case $p = 2$ is the most commonly implemented for applications and is known as the Competitive Learning Vector Quantization (CLVQ) algorithm. Then one considers $(0, 1)$-valued step parameters δ_s so that Γ^{s+1} lives in the convex hull of Γ^s and ξ^{s+1} and the cooperative procedure (2.11) becomes a simple homothety centered at ξ^{s+1} with ratio $1-\delta_{s+1}$. These features have a stabilizing effect on the procedure. One checks on simulation that the CLVQ algorithm does behave better than its non-quadratic counterpart. The numerical aspects of the $CLVQ$ algorithm are deeply investigated in [21] when X is d-dimensional Normal vector.

Figure 1 shows an optimal grid for the bivariate standard Normal distribution with 500 points. It is obtained by the $CLVQ$ procedure described above.

2.3 Application to numerical integration

Consider a simulatable $\mathbb{R}^d$-valued integrable random vector X with probability distribution $\mathbb{P}_X$. The quantization method for numerical integration consists in approximating the probability distribution $\mathbb{P}_X$ by $\mathbb{P}_{\widehat{X}}$, the distribution of (one of) its closest neighbor rule projection(s) $\widehat{X} = \mathrm{Proj}_\Gamma(X)$ on a grid $\Gamma = \{x^1, \ldots, x^N\}$:

$$\mathbb{P}_{\widehat{X}} = \sum_{i=1}^N \hat{p}^i\, \delta_{x^i}.$$

So, $\mathbb{P}_{\widehat{X}}$ is a discrete probability distribution whose weights $\hat{p}^i$ are defined by

$$\hat{p}^i = \mathbb{P}[\widehat{X} = x^i] \ = \ \mathbb{P}_X[C_i(\Gamma)], \qquad 1 \leq i \leq N,$$

where δ_{x^i} is the Dirac mass at x^i and $C_i(\Gamma) = \mathrm{Proj}_\Gamma^{-1}(x^i)$ denotes the Voronoi tessels of $x_i \in \Gamma$. Then, one approximates the expectation of a Lipschitz continuous function ϕ on $\mathbb{R}^d$ w.r.t. $\mathbb{P}_X$, *i.e.*

$$\mathbb{E}[\phi(X)] \ = \ \int_{\mathbb{R}^d} \phi(\xi)\, \mathbb{P}_X(d\xi),$$

by

$$\mathbb{E}[\phi(\widehat{X})] \ = \ \int_{\mathbb{R}^d} \phi(\xi)\, \mathbb{P}_{\widehat{X}}(d\xi) \ = \ \sum_{i=1}^N \hat{p}^i \phi(x^i).$$

THE LIPSCHITZ CASE: When ϕ is simply Lipschitz continuous, the induced error is then simply measured by:

$$|\mathbb{E}[\phi(X)] - \mathbb{E}[\phi(\widehat{X})]| \leq [\phi]_{Lip} \|X - \widehat{X}\|_1, \tag{2.14}$$

$$\leq [\phi]_{Lip} \|X - \widehat{X}\|_p \qquad (p \geq 1). \tag{2.15}$$

Optimal grids (of size N) which minimize the L^1-quantization error then provide a $O\left(N^{-1/d}\right)$ rate. Such a grid, its associated weights $\hat{p}_i$ and the induced L^1-quantization error can be computed by the algorithm described above. It often happens, for stability matter, that one implements the algorithm in the quadratic case ($CLVQ$) and produces an optimal *quadratic* grid Γ^* and its companion parameters (the weights $(\hat{p}_i^*)_{1\leq i\leq n}$ and the L^1-quantization error $\|X - \widehat{X}^{\Gamma^*}\|_1$ as a normalized error bound estimate). Some extensions of (2.14) to locally Lipschitz continuous functions can be found in [11].

THE LIPSCHITZ DERIVATIVE CASE: Assume now that function ϕ is *continuously differentiable with a Lipschitz continuous differential* $D\phi$. Furthermore, assume that the quantization is carried out with an *optimal quadratic* grid Γ. By Taylor's formula, we have

$$|\phi(X) - (\phi(\widehat{X}) + D\phi(\widehat{X}).(X - \widehat{X}))| \leq [D\phi]_{Lip} |X - \widehat{X}|^2$$

$$\text{so that} \quad \left|\mathbb{E}[\phi(X)] - \mathbb{E}[\phi(\widehat{X})] - \mathbb{E}[D\phi(\widehat{X}).(X - \widehat{X})]\right| \leq [D\phi]_{Lip} \|X - \widehat{X}\|_2^2,$$

$$\leq [D\phi]_{Lip} \|X - \widehat{X}\|_p^2 \qquad (p \geq 2).$$

Now, $\widehat{X}$ is in particular a stationary quantizer, hence it satisfies (2.5) so that

$$\mathbb{E}[D\phi(\widehat{X}).(X - \widehat{X})] = \mathbb{E}\left[D\phi(\widehat{X}).\mathbb{E}[X - \widehat{X} \mid \widehat{X}]\right] = 0,$$

$$\text{and} \qquad |\mathbb{E}[\phi(X)] - \mathbb{E}[\phi(\widehat{X})]| \leq [D\phi]_{Lip} \|X - \widehat{X}\|_2^2 = O\left(N^{-2/d}\right). \tag{2.16}$$

THE CONVEX CASE: When ϕ is a convex function and $\widehat{X}$ is a stationary quantizer satisfying $\widehat{X} = \mathbb{E}[X \mid \widehat{X}]$, we have by Jensen's inequality:

$$\mathbb{E}\left[\phi(\widehat{X})\right] = \mathbb{E}\left[\phi\left(\mathbb{E}[X \mid \widehat{X}]\right)\right] \leq \mathbb{E}[\phi(X)], \tag{2.17}$$

so that $\mathbb{E}[\phi(\widehat{X})]$ is always a lower bound for $\mathbb{E}[\phi(X)]$.

2.4 A first numerical Test (European option approximation)

The aim of this section is to test the optimal quantizers that we obtained by the numerical methods described in subsection 2.2 in dimension $d = 4$. Simultaneously, we aim to illustrate the performances of vector quantization for numerical integration. That is why we carry out a short comparison between quantization method and Monte Carlo method on a simple numerical integration problem.

Recall that the Strong Law of Large Numbers implies that, given a sequence $(Z_k)_{k\geq 1}$ of independent copies of a random vector Z with Normal distribution $\mathcal{N}(0; I_d)$,

$$\mathbb{P}(d\omega)\text{-a.s.} \quad \frac{f(Z_1(\omega)) + \cdots + f(Z_N(\omega))}{N} \overset{N\to+\infty}{\longrightarrow} \mathbb{E}(f(Z)) = \int_{\mathbb{R}^d} f(z) \exp\left(-|z|^2/2\right) \frac{dz}{(2\pi)^{d/2}}.$$

for every $f \in L^1(\mathbb{R}^d, \mathbb{P}_Z)$. The Monte Carlo method consists in generating on a computer a path $(Z_k(\omega))_{k\geq 1}$ to compute the above Gaussian integral. The Law of the Iterated Logarithm says that, if $f(Z) \in L^2$, this convergence *a.s.* holds at a

$$\sigma(f(Z))\sqrt{\frac{\log\log N}{N}}$$

rate where $\sigma(f(Z))$ is the standard deviation of $f(Z)$.

When f is twice differentiable, this is to be compared to the error bound provided by (2.16) when using a quadratic optimal N-quantizer $x^* := (x_1^*, \ldots, x_N^*)$, namely

$$[Df]_{Lip} Q_N^2(x^*) \approx \left(J_{2,d}(1+2/d)^{1+d/2}[Df]_{Lip}\right) N^{-2/d}.$$

Consequently, the dimension $d = 4$ appears as the (theoretical) critical dimension for the numerical integration of such functions by quantization for a given computational complexity (quantization formulae involving higher order differentials yield better rates): one assumes that the optimal quantizers have been formerly computed and that the computation times of a (Gaussian) random number and of a weight are both negligible w.r.t. the computation time of a value $f(z)$ of f.

The test is processed in each selected dimension d with five random variables $g_i(Z)$, $i = 0, 1, 2, 3, 4$ where the g_i's are five functions with compact support such that

– g_0 is an indicator function of a (bounded) interval (hence discontinuous),

– g_1 is convex and Lipschitz continuous,

– g_2 is convex and twice differentiable,

– g_3 is difference of two convex functions and Lipschitz continuous,

– g_4 is difference of two convex functions and twice differentiable.

The test functions are borrowed from classical option pricing in mathematical finance: one considers d traded assets $S^1, \ldots, S^d$, following a d-dimensional Black & Scholes dynamics. We assume that these assets are independent (this is not very realistic but corresponds to the most unfavorable case for quantization). We assume as well that $S_0^1 = s_0 > 0$, $i = 1, \ldots, d$ and that the d assets share the same volatility $\sigma^i = \sigma > 0$. At maturity $T > 0$, we then have:

$$S_T^i = s_0 \exp\left((r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}Z^i\right), \qquad i = 1, \ldots, d.$$

One considers, still at time T, the geometric index $I_T = \left(S_T^1 \ldots S_T^d\right)^{\frac{1}{d}}$ satisfying

$$I_T = I_0 \exp\left((r - \frac{\sigma^2}{2d})T + \frac{\sigma\sqrt{T}}{\sqrt{d}}\frac{Z^1 + \cdots + Z^d}{\sqrt{d}}\right) \text{ with } I_0 = s_0 \exp\left(-\frac{\sigma^2(d-1)}{2d}T\right).$$

Then, one specifies the random variables $g_i(Z)$ for $i = 1$ and $i = 3$ as follows

$$g_1(Z) = \qquad e^{-rT}(K_1 - I_T)_+ \qquad \text{(Put}(K_1, T) \text{ payoff)}$$

$$g_3(Z) = \quad e^{-rT}(K_2 - I_T)_+ - e^{-rT}(K_1 - I_T)_+,\ K_1 < K_2, \quad \text{(Put-Spread}(K_1, K_2 T) \text{ payoff)}.$$

The random variables are the payoffs of a Put option with strike price K_1 and a Put-spread option with strike prices $K_1 < K_2$ respectively, both on the geometric index I_T. Some closed forms for the premia $\mathbb{E}[g_1(Z)]$ and $\mathbb{E}[g_2(Z)]$ are given by the Black & Scholes formula:

$$\mathbb{E}[g_1(Z)] \;=\; \pi(I_0, K_1, r, \sigma, T) \quad \text{and} \quad \mathbb{E}[g_3(Z)] \;=\; \psi(I_0, K_1, K_2, r, \sigma, T) \tag{2.18}$$

$$\text{with} \qquad \pi(x, K, r, \sigma, T) \;=\; Ke^{-rT}\mathrm{erf}(-d_2) - I_0\,\mathrm{erf}(-d_1),$$

$$d_1 = \frac{\log(x/K) + (r + \frac{\sigma^2}{2d})T}{\sigma\sqrt{T/d}}, \qquad d_2 = d_1 - \sigma\sqrt{T/d}$$

$$\text{and} \qquad \psi(x, K_1, K_2, r, \sigma, T) \;=\; \pi(x, K_2, r, \sigma, T) - \pi(x, K_1, r, \sigma, T).$$

Then, one sets

$$g_2(Z) = e^{-rT/2}\pi(I_{\frac{T}{2}}, K_1, r, \sigma, T/2) \quad \text{and} \quad g_4(Z) = e^{-rT/2}\psi(I_{\frac{T}{2}}, K_1, K_2, r, \sigma, T/2).$$

The random variables $g_2(Z)$ and $g_4(Z)$ have the distributions of the (discounted) premia at time $T/2$ of the Put(K_1, T) and of the Put-Spread(K_1, K_2, T) respectively. Functions g_2 and g_4 are $\mathcal{C}^\infty$ and using the martingale property of the discounted premia yields

$$\mathbb{E}\, g_2(Z) = \pi(I_0, K_1, r, \sigma, T) \quad \text{and} \quad \mathbb{E}\, g_4(Z) = \psi(I_0, K_1, K_2, r, \sigma, T). \tag{2.19}$$

Finally we specify g_0 as the "hedge function at maturity" of the Put-Spread option:

$$g_0(Z) = -e^{-rT}\mathbf{1}_{\{I_T \in [K_1, K_2]\}}. \tag{2.20}$$

The numerical specifications of the functions g_i's are as follows:

$$s_0 = 100, \qquad K_1 = 98, \qquad K_2 = 102, \qquad r = 5\%, \qquad \sigma = 20\%, \qquad T = 2.$$

• Numerical results in 4-dimension: The comparison with the Monte Carlo estimator

$$\widehat{g_i(Z)}_N = \frac{1}{N}\sum_{k=1}^{N} g_i(Z_k), \qquad Z_k \; i.i.d., \; Z_k \sim \mathcal{N}(0; I_d) \tag{2.21}$$

of $\mathbb{E}[g_i(Z)]$ is presented in the last column on the right: we first computed (a proxy of) the standard deviation $\sigma(\widehat{g_i(Z)}_N)$ of the above estimator (2.21) by a $N = 10\,000$ trial Monte Carlo simulation. Then, in order to measure the error induced by the quantization in the scale of the MC estimator Standard Deviation, we wrote down the *ratio* $\frac{\textit{absolute error}}{\sigma(\widehat{g_i(Z)}_N)}$.

The results in Table 1 illustrate a widely observed phenomenon when integrating functions by quantization: difference of convex functions behave better than convex functions (this is obviously due to (2.17)), and Lipschitz derivative functions behave better than Lipschitz continuous functions (as predicted by (2.16)). The whole tests set suggests that the convexity feature is prominent.

• Graphical comparison in dimensions $d = 3, 4, 5$: We focus her on the convex C^2 function g_2. We wish to emphasize the dimension effect (keeping unchanged the other specifications). So, we depict in Figure 2, in dimension $d = 3$, 4, 5 (in a log-log scale), both the absolute error and the standard deviation $\sigma(\widehat{g_2(Z)}_N)$ of its Monte Carlo estimator as a function of N (the dotted lines are the induced least square regressions)

$d=4$ & $N=6540$	$B\&S$ Reference value	Quantized value	Relative error	$\sigma(\widehat{g_i(Z)}_N)$	$\frac{\text{absolute error}}{\sigma(\widehat{g_i(Z)}_N)}$
$\mathbb{E}\, g_0(Z)$	-0.093	$-0,091$	$2.40\,\%$	0.034	0.064
$\mathbb{E}\, g_1(Z)$	2.077	2.047	$1.44\,\%$	0.054	0.548
$\mathbb{E}\, g_2(Z)$	2.077	2.061	$0.77\,\%$	0.033	0.482
$\mathbb{E}\, g_3(Z)$	1.216	1.213	$0.26\,\%$	0.021	0.015
$\mathbb{E}\, g_4(Z)$	1.216	1.215	$0.08\,\%$	0.012	0.001

Table 1: *Quantization versus Monte Carlo in* 4*-dimension.*

Concerning the dimensionality effect, the theoretical rates for the error bounds ($N^{-1/d}$ in the Lipschitz case and $N^{-2/d}$ for Lipschitz differential case) are confirmed: when $d \leq 4$ quantization overperforms more and more the Monte Carlo method as N increases. When $d > 4$, this is at most true up to a critical number N_d of points (for a given trust level in the MC method). More detailed numerical results are presented in [21].

Remark 2.2 In the above tests, we compared quantization *versus* Monte Carlo for the computation of a *single* integral. If one looks for a uniform error bound over an infinite class of Lipschitz continuous functions, the conclusion can be quite different: thus, with the notations of the former subsection 2.3

$$\begin{aligned}\sup_{f,\,[f]_{Lip}\leq 1}\left|\frac{f(Z_1)+\cdots+f(Z_N)}{N}-\int_{\mathbb{R}^d} f(\zeta)\mathbb{P}_Z(d\zeta)\right| &\geq \int_{\mathbb{R}^d}\min_{1\leq i\leq N}|Z_i-\zeta|\,\mathbb{P}_Z(d\zeta)\\ &\geq Q^1_N(Z_1,\ldots,Z_N)\\ &> \min_{(\mathbb{R}^d)^N} Q^1_N \sim \frac{c_Z}{N^{\frac{1}{d}}},\ (c_Z>0).\end{aligned}$$

This means that for every fixed N the worst behaviour of the Monte Carlo method on 1-Lipschitz functions induces a greater error than that obatined by optimal L^1-quantization. This holds true in any dimension d.

3 Optimal quantization of a Markov chain

We consider an $\mathbb{R}^d$-valued $(\mathcal{F}_k)$-Markov chain $(X_k)_{0\leq k\leq n}$, with probability transition $P_k(x,dx')$ (from time $k-1$ to k) and with initial distribution μ for X_0. The joint distribution of $(X_k)_{0\leq k\leq n}$, is then equal to $\mu(dx_0)P_1(x_0,dx_1)\ldots P_n(x_{n-1},dx_n)$.

In this section, we are interested in the quantization of this Markov chain, i.e. an approximation of the distribution of the process (X_k) by the distribution of a process $(\widehat{X}_k)$ valued on finite grids taking into account the probabilistic feature of the process. The naive approach would consist in the quantization of the $\mathbb{R}^{(n+1)d}$-valued random vector $(X_0,\ldots,X_n)$ following the method described in Section 2. However, by Theorem 2.1, for a total number N of points in such a "time-space" grid, the L^p-quantization error would be of order $N^{-\frac{1}{nd}}$. This is of course very slow when n is large.

Instead, we propose an approach based on the fact that a Markov chain is completely characterized by its initial distribution and its transitions probabilities. The idea is then

to "quantize" the initial distribution of X_0 and the conditional probabilities of X_k given X_{k-1}. We propose two different quantization methods which shall provide a better rate of convergence of order $n^{1+1/d}/N^{1/d}$. The first approach, based on a quantization at each time k of the random variable X_k, was introduced in [1] and is called *marginal quantization*. The second one that enhances the preservation of the dynamics, namely the Markov property, was introduced in [19] and is called *Markovian quantization*.

3.1 Marginal quantization

At each time k and given a grid $\Gamma_k = \{x^1, \ldots, x^{N_k}\}$ of N_k points in $\mathbb{R}^d$, associated to a Voronoi tessellation $C_1(\Gamma_k), \ldots, C_{N_k}(\Gamma_k)$, we define:

$$\widehat{X}_k \;=\; \mathrm{Proj}_{\Gamma_k}(X_k), \quad k = 0, \ldots, n. \tag{3.1}$$

Hence, in the marginal approach, the emphasis is put on the accuracy of the distribution approximations: if at every time k, the grid Γ_k is L^p-optimal, then $\widehat{X}_k$ is the best possible L^p-approximation of X_k by a random variable taking $N_k := |\Gamma_k|$ points. Notice that since the projection on the closest neighbor is not injective, the process $(\widehat{X}_k)_k$ constructed in (3.1) is not a Markov chain. However, if we define the probability transition matrices $[\hat{p}_k^{ij}]$ at times $k = 1, \ldots, n$ by:

$$\begin{aligned}
\hat{p}_k^{ij} &= \mathbb{P}\left[\widehat{X}_k = x_k^j \,|\, \widehat{X}_{k-1} = x_k^i\right] = \frac{\hat{\beta}_k^{ij}}{\hat{p}_{k-1}^i}, \quad i = 1, \ldots, N_{k-1},\ j = 1, \ldots, N_k, \\
\text{where} \quad \hat{p}_{k-1}^i &= \mathbb{P}[\widehat{X}_{k-1} = x_{k-1}^i] \;=\; \mathbb{P}[X_{k-1} \in C_i(\Gamma_{k-1})] \\
\hat{\beta}_k^{ij} &= \mathbb{P}[\widehat{X}_{k-1} = x_{k-1}^i,\ \widehat{X}_k = x_k^j] \;=\; \mathbb{P}[X_{k-1} \in C_i(\Gamma_{k-1}),\ X_k \in C_j(\Gamma_k)].
\end{aligned}$$

then it is well-known that there exists a Markov chain $(\widehat{X}_k^c)$ with initial distribution $\hat{p}_0$ and probability transition matrices $[\hat{p}_k^{ij}]$ at times $k = 1, \ldots, n$. The marginal quantization method consists in approximating the distribution of the Markov chain $(X_k)_{0\le k\le n}$ by that of the Markov chain $(\widehat{X}_k^c)_{0\le k\le n}$: by construction, the conditional distribution of $\widehat{X}_{k+1}^c$ given $\widehat{X}_k^c$ is equal to the the conditional distribution of $\widehat{X}_{k+1}$ given $\widehat{X}_k$, and the distribution of $\widehat{X}_0^c$ is equal to the distribution of $\widehat{X}_0$. We will evaluate the rate of approximation (in distribution) of $\widehat{X}^c$ toward X on functions of the form $(x_0, x_1, \ldots, x_n) \mapsto \phi_0(x_0)\phi_1(x_1)\ldots\phi_n(x_n)$, where ϕ_0, …, ϕ_n are bounded Lipschitz continuous functions on $\mathbb{R}^d$. First, notice that both quantities $\mathbb{E}[\phi_0(X_0)\phi_1(X_1)\ldots\phi_n(X_n)]$ and $\mathbb{E}[\phi_0(\widehat{X}_0^c)\phi_1(\widehat{X}_1^c)\ldots\phi_n(\widehat{X}_n^c)]$ follow a dynamic programming formula induced by the Markov property. Namely

$$\mathbb{E}\left[\phi_0(X_0)\ldots\phi_n(X_n)\right] = \mathbb{E}\left[v_0(X_0)\right] \quad \text{and} \quad \mathbb{E}\left[\phi_0(\widehat{X}_0^c)\ldots\phi_n(\widehat{X}_n^c)\right] = \mathbb{E}\left[\widehat{v}_0(\widehat{X}_0^c)\right]$$

where $\widehat{v}_0(\widehat{X}_0^c)$ and $v_0(X_0)$ satisfy

$$v_n(X_n) \;=\; \phi_n(X_n), \qquad v_{k-1}(X_{k-1}) \;=\; \phi_{k-1}(X_{k-1})\mathbb{E}[v_k(X_k)\,|\,X_{k-1}],\ k = 1, \ldots, n \tag{3.2}$$

$$\widehat{v}_n(\widehat{X}_n^c) \;=\; \phi_n(\widehat{X}_n^c), \qquad \widehat{v}_{k-1}(\widehat{X}_{k-1}^c) \;=\; \phi_{k-1}(\widehat{X}_{k-1}^c)\mathbb{E}[\widehat{v}_k(\widehat{X}_k^c)\,|\,\widehat{X}_{k-1}^c],\ k = 1, \ldots, n \tag{3.3}$$

This will be the key to evaluate the error induced by approximating the first expectation term by the second one. Furthermore, the dynamic programming formula for $\widehat{X}^c$, once written "in distribution", provides a simple numerical algorithm to compute $\mathbb{E}\left[\widehat{v}_0(\widehat{X}_0^c)\right]$:

$$\begin{aligned}
\widehat{v}_n(x_n^i) &= \phi_n(x_n^i), \qquad \forall\, x_n^i \in \Gamma_n, \\
\widehat{v}_{k-1}(x_{k-1}^i) &= \phi_{k-1}(x_{k-1}^i)\mathbb{E}[\widehat{v}_k(\widehat{X}_k)\,|\,\widehat{X}_{k-1} = x_{k-1}^i] \\
&= \phi_{k-1}(x_{k-1}^i)\sum_{j=1}^{N_k} \hat{p}_k^{ij}\,\widehat{v}_k(x_k^j), \qquad \forall\, x_{k-1}^i \in \Gamma_{k-1},\ \ k = 1,\dots,n. \qquad (3.4) \\
\mathbb{E}\left[\widehat{v}_0(\widehat{X}_0^c)\right] &= \mathbb{E}\left[\widehat{v}_0(\widehat{X}_0)\right] = \sum_{i=1}^{N_0} \hat{p}_0^i\,\widehat{v}_0(x_0^i).
\end{aligned}$$

We rely on the following Lipschitz assumption on the transitions P_k of the Markov chain (X_k).

(A1) For any $k = 1,\dots,n$, the probability transition P_k is Lipschitz with ratio $[P_k]_{Lip}$, i.e. for any Lipschitz function ϕ on $\mathbb{R}^d$, with ratio $[\phi]_{Lip}$, we have:

$$\left|\int_{\mathbb{R}^d} \phi(x')P_k(x,dx') - \int_{\mathbb{R}^d} \phi(x')P_k(\widehat{x},dx')\right| \ \le\ [P_k]_{Lip}[\phi]_{Lip}|x-\widehat{x}|, \quad \forall x,\widehat{x}\in\mathbb{R}^d.$$

Then we set $[P]_{Lip} = \max_{k=1,\dots,n}[P_k]_{Lip}$. Let

$$BL_1(\mathbb{R}^d) = \left\{\phi : \mathbb{R}^d \to \mathbb{R}, \quad \phi \text{ is bounded by 1 and } \phi \text{ is Lipschitz with ratio } [\phi]_{Lip} \le 1\right\}.$$

Theorem 3.1 *Let $p \ge 1$. Under* **(A1)**, *we have the error estimation in the marginal quantization method: for any functions $\phi_k \in BL_1(\mathbb{R}^d)$, $k = 0,\dots,n$,*

$$\left|\mathbb{E}[\phi_0(X_0)\dots\phi_n(X_n)] - \mathbb{E}[\phi_0(\widehat{X}_0^c)\dots\phi_n(\widehat{X}_n^c)]\right| \le \sum_{k=0}^{n}\left(1+(2-\delta_{2,p})\frac{[P]_{Lip}^{n-k+1}-1}{[P]_{Lip}-1}\right)\|\Delta_k\|_p, \quad (3.5)$$

where $\|\Delta_k\|_p = \|X_k - \widehat{X}_k\|_p$ is the L^p-quantization error at time k of X_k. In (3.5), we make the usual convention that $\frac{1}{u-1}(u^m-1) = m$ if $u = 1$ and $m \in \mathbb{N}$.

Proof. We set $\|\phi\|_{sup} = \max_{k=0,\dots,n}\|\phi_k\|_{sup} \le 1$ and $[\phi]_{Lip} = \max_{k=0,\dots,n}[\phi_k]_{Lip} \le 1$. From (3.2), a standard backward induction shows that

$$\|v_k\|_{sup} \le \|\phi\|_{sup}^{n+1-k} \quad \text{and} \quad [v_k]_{Lip} \le [P]_{Lip}\|\phi\|_{sup}[v_{k+1}]_{Lip} + \|\phi\|_{sup}^{n-k}[\phi]_{Lip}$$

so that
$$[v_k]_{Lip} \le \|\phi\|_{sup}^{n-k}[\phi]_{Lip}\frac{[P]_{Lip}^{n-k+1}-1}{[P]_{Lip}-1}.$$

For any bounded Borel function f on $\mathbb{R}^d$, we set

$$\begin{aligned}
P_k f(x) &= \mathbb{E}[f(X_k)|X_{k-1} = x], \ \ x \in \mathbb{R}^d \\
\widehat{P}_k f(x) &= \mathbb{E}[f(\widehat{X}_k)|\widehat{X}_{k-1} = x], \ \ x \in \Gamma_{k-1},
\end{aligned}$$

for $k = 1, \ldots, n$. Hence, by (3.2) and (3.4), we have

$$\begin{aligned}\left\|v_k(X_k) - \widehat{v}_k(\widehat{X}_k)\right\|_p &\leq \left\|v_k(X_k) - \mathbb{E}[v_k(X_k)|\widehat{X}_k]\right\|_p \\ &+ \left\|\mathbb{E}\left[\left(\phi_k(X_k) - \phi_k(\widehat{X}_k)\right) P_{k+1}v_{k+1}(X_k)\middle|\widehat{X}_k\right]\right\|_p \\ &+ \left\|\phi_k(\widehat{X}_k)\mathbb{E}\left[P_{k+1}v_{k+1}(X_k) - \widehat{P}_{k+1}\widehat{v}_{k+1}(\widehat{X}_k)\middle|\widehat{X}_k\right]\right\|_p \quad (3.6)\end{aligned}$$

On one hand, notice that, for every $p \geq 1$,

$$\begin{aligned}\|v_k(X_k) - \mathbb{E}(v_k(X_k)|\widehat{X}_k)\|_p &\leq \|v_k(X_k) - v_k(\widehat{X}_k)\|_p + \|v_k(\widehat{X}_k) - \mathbb{E}(v_k(X_k)|\widehat{X}_k)\|_p \\ &\leq 2\|v_k(X_k) - v_k(\widehat{X}_k)\|_p \\ &\leq 2[v_k]_{Lip}\|\Delta_k\|_p. \quad (3.7)\end{aligned}$$

When $p = 2$, the very definition of the conditional expectation as an orthogonal projection shows that the above inequality holds without the 2 factor. On the other hand, using that conditional expectation (given $\widehat{X}_k$) is a L^p-contraction and that $\widehat{X}_k$ is $\sigma(X_k)$-measurable yields

$$\left\|\mathbb{E}\left[\left(\phi_k(X_k) - \phi_k(\widehat{X}_k)\right) P_{k+1}v_{k+1}(X_k)\middle|\widehat{X}_k\right]\right\|_p \leq [\phi]_{Lip}\|v_{k+1}\|_{sup}\|\Delta_k\|_p \leq [\phi]_{Lip}\|\phi\|_{sup}^{n-k}\|\Delta_k\|_p \quad (3.8)$$

and

$$\left\|\phi_k(\widehat{X}_k)\left(\mathbb{E}\left[P_{k+1}v_{k+1}(X_k) - \widehat{P}_{k+1}\widehat{v}_{k+1}(\widehat{X}_k)\middle|\widehat{X}_k\right]\right)\right\|_p \leq \|\phi\|_{sup}\|v_{k+1}(X_{k+1}) - \widehat{v}_{k+1}(\widehat{X}_{k+1})\|_p. \quad (3.9)$$

Plugging inequalities (3.7), (3.8) and (3.9) in (3.6) leads to the backward induction formula

$$\begin{aligned}\|v_k(X_k) - \widehat{v}_k(\widehat{X}_k)\|_p &\leq ((2 - \delta_{2,p})[v_k]_{Lip} + [\phi]_{Lip}\|\phi\|_{sup}^{n-k})\|\Delta_k\|_p \\ &+ \|\phi\|_{sup}\|v_{k+1}(X_{k+1}) - \widehat{v}_{k+1}(\widehat{X}_{k+1})\|_p \\ &\leq [\phi]_{Lip}\|\phi\|_{sup}^{n-k}\left(1 + (2 - \delta_{2,p})\frac{[P]_{Lip}^{n-k+1} - 1}{[P]_{Lip} - 1}\right)\|\Delta_k\|_p \\ &+ \|\phi\|_{sup}\|v_{k+1}(X_{k+1}) - \widehat{v}_{k+1}(\widehat{X}_{k+1})\|_p\end{aligned}$$

with $v_n = \widehat{v}_n \equiv \phi_n$. This yields the expected result after some standard computations. □

3.2 Markovian quantization

Here, we suppose that the dynamics of the $(\mathcal{F}_k)$ Markov chain $(X_k)_k$ is given in the form:

$$X_k = F_k(X_{k-1}, \varepsilon_k), \quad k = 1, \ldots, n \quad (3.10)$$

(starting from some initial state X_0), where $(\varepsilon_k)_k$ is a sequence of identically distributed $\mathcal{F}_k$-measurable random variables in $\mathbb{R}^m$, such that ε_k is independent of $\mathcal{F}_{k-1}$, and F_k is some measurable function on $\mathbb{R}^d \times \mathbb{R}^m$. Given a sequence of grids $\Gamma_k = \{x^1, \ldots, x^{N_k}\}$ of

N_k points in $\mathbb{R}^d$, associated to a Voronoi tessellation $C_1(\Gamma_k), \ldots, C_{N_k}(\Gamma_k)$, $k = 0, \ldots, n$, we define the process $(\widehat{X}_k)_k$ by:

$$\widehat{X}_k \;=\; \mathrm{Proj}_{\Gamma_k}(F_k(\widehat{X}_{k-1}, \varepsilon_k)), \quad k = 1, \ldots, n, \tag{3.11}$$

and $\widehat{X}_0 = \mathrm{Proj}_{\Gamma_0}(X_0)$. By construction, the process $(\widehat{X}_k)_k$ is still a Markov chain w.r.t. the same filtration $(\mathcal{F}_k)$. Its probability transition matrix $[\hat{p}_k^{ij}]$ at times $k = 1, \ldots, n$ reads:

$$\hat{p}_k^{ij} \;=\; \mathbb{P}\left[\widehat{X}_k = x_k^j | \widehat{X}_{k-1} = x_k^i\right] = \frac{\hat{\beta}_k^{ij}}{\hat{p}_{k-1}^i}, \quad i = 1, \ldots, N_{k-1},\ j = 1, \ldots, N_k, \tag{3.12}$$

$$\text{where} \quad \hat{p}_{k-1}^i \;=\; \mathbb{P}[\widehat{X}_{k-1} = x_{k-1}^i] \;=\; \begin{cases} \mathbb{P}[F_k(\widehat{X}_{k-2}, \varepsilon_{k-1}) \in C_i(\Gamma_{k-1})], & \text{if} \quad k \geq 2 \\ \mathbb{P}[\widehat{X}_0 \in C_i(\Gamma_0)] & \text{if} \quad k = 1, \end{cases} \tag{3.13}$$

$$\begin{aligned} \text{and} \qquad \hat{\beta}_k^{ij} \;&=\; \mathbb{P}[\widehat{X}_{k-1} = x_{k-1}^i,\ \widehat{X}_k = x_k^j] \\ &=\; \begin{cases} \mathbb{P}[F(\widehat{X}_{k-2}, \varepsilon_{k-1}) \in C_i(\Gamma_{k-1}),\ F_k(\widehat{X}_{k-1}, \varepsilon_k) \in C_j(\Gamma_k)], & \text{if} \quad k \geq 2 \\ \mathbb{P}[\widehat{X}_0 \in C_i(\Gamma_0),\ F(\widehat{X}_0, \varepsilon_1) \in C_j(\Gamma_k)], & \text{if} \quad k = 1. \end{cases} \end{aligned}$$

We still intend to estimate the approximation of (X_k) by the Markov quantized process $(\widehat{X}_k)$ along functions $(x_0, x_1, \ldots, x_n) \mapsto \phi_0(x_0)\phi_1(x_1)\ldots\phi_n(x_n)$, ϕ_0, …, where ϕ_n are bounded Lipschitz continuous functions on $\mathbb{R}^d$. This time, the quantized process $(\widehat{X}_k)$ itself being a Markov chain, one may compute directly $\mathbb{E}[\,\phi_0(\widehat{X}_0)\phi_1(\widehat{X}_1)\ldots\phi_n(\widehat{X}_n)]$. This quantity can be obtained as the final result of a backward dynamic programming formula formally identical to (3.4) but where the coefficient $[\hat{p}_k^{ij}]$ and $\hat{p}_0^i$ are given by (3.12) and (3.13) *i.e.* are based on the Markov chain $(\widehat{X}_k)_{0\leq k\leq n}$ described in (3.11).

We will rely now on a pathwise Lipschitz assumption on the Markov chain $(X_k)_{0\leq k\leq n}$:

(A1') For any $k = 1, \ldots, n$, there exists some positive constant $[F_k]_{Lip}$ such that:

$$\|F_k(x, \varepsilon_k) - F_k(\widehat{x}, \varepsilon_k)\|_1 \;\leq\; [F_k]_{Lip}|x - \widehat{x}|, \quad \forall x,\, \widehat{x} \in \mathbb{R}^d.$$

We then set $[F]_{Lip} = \max_{k=1,\ldots,n}[F_k]_{Lip}$.

Theorem 3.2 *Under* **(A1')**, *we have the error estimation in the Markov quantization method: for any functions $\phi_k \in BL_1(\mathbb{R}^d)$, $k = 0, \ldots, n$,*

$$\left|\mathbb{E}\left[\phi_0(X_0)\ldots\phi_n(X_n)\right] - \mathbb{E}\left[\phi_0(\widehat{X}_0)\ldots\phi_n(\widehat{X}_n)\right]\right| \;\leq\; \sum_{k=0}^{n} \frac{[F]_{Lip}^{n-k+1} - 1}{[F]_{Lip} - 1}\|\Delta_k\|_1, \tag{3.14}$$

$$\textit{where} \qquad \|\Delta_k\|_1 \;=\; \|F_k(\widehat{X}_{k-1}, \varepsilon_k) - \widehat{X}_k\|_1 \tag{3.15}$$

is the L^1-quantization error at time k of $F_k(\widehat{X}_{k-1}, \varepsilon_k)$. In (3.14), we make the usual convention that $\frac{1}{u-1}(u^m - 1) = m$ if $u = 1$ and $m \in \mathbb{N}$.

Proof. Set $\|\phi\|_{sup} = \max_{k=0,\ldots,n} \|\phi_k\|_{sup} \leq 1$ and $[\phi]_{Lip} = \max_{k=0,\ldots,n}[\phi_k]_{Lip} \leq 1$. We also denote

$$L_k \;=\; \prod_{j=0}^{k} \phi_j(X_j) \quad \text{and} \quad \widehat{L}_k \;=\; \prod_{j=0}^{k} \phi_j(\widehat{X}_j).$$

We then have

$$L_k - \widehat{L}_k \;=\; \Big(\phi_k(X_k) - \phi_k(\widehat{X}_k)\Big)\, L_{k-1} + (L_{k-1} - \widehat{L}_{k-1})\phi_k(\widehat{X}_k).$$

From the boundedness and Lipschitz conditions on ϕ_k, we deduce that

$$\left|L_k - \widehat{L}_k\right| \;\le\; \|\phi\|_{sup}^k [\phi]_{Lip} \left|X_k - \widehat{X}_k\right| + \|\phi\|_{sup} \left|L_{k-1} - \widehat{L}_{k-1}\right|,$$

for all $k = 1, \dots, n$. By a straightforward backward induction, we get

$$|L_n - \widehat{L}_n| \;\le\; \|\phi\|_{sup}^n [\phi]_{Lip} \sum_{k=0}^{n} |X_k - \widehat{X}_k|. \tag{3.16}$$

On the other hand, from the definitions (3.10) and (3.11) of X_k and $\widehat{X}_k$, and (3.15) of Δ_k, we obviously get for any $k \ge 1$:

$$\|X_k - \widehat{X}_k\|_1 \;\le\; \|F_k(X_{k-1}, \varepsilon_k) - F_k(\widehat{X}_{k-1}, \varepsilon_k)\|_1 + \|\Delta_k\|_1.$$

By Assumption **(A1')** and since ε_k is independent of $\mathcal{F}_{k-1}$, we then obtain:

$$\|X_k - \widehat{X}_k\|_1 \;\le\; [F_k]_{Lip} \|X_{k-1} - \widehat{X}_{k-1}\|_1 + \|\Delta_k\|_1.$$

Recalling that $\|X_0 - \widehat{X}_0\|_1 = \|\Delta_0\|_1$, we deduce by backward induction that:

$$\forall\, k \in \{0, \dots, n\}, \qquad \|X_k - \widehat{X}_k\|_1 \;\le\; \sum_{j=0}^{k} [F]_{Lip}^{k-j} \|\Delta_j\|_1 . \tag{3.17}$$

Finally, using (3.16) and (3.17), one completes the proof noting that

$$\begin{aligned}
\left|\mathbb{E}\left[\phi_0(X_0) \dots \phi_n(X_n)\right] - \mathbb{E}\left[\phi_0(\widehat{X}_0) \dots \phi_n(\widehat{X}_n)\right]\right| &\le \|L_n - \widehat{L}_n\|_1 \\
&\le \|\phi\|_{sup}^n [\phi]_{Lip} \sum_{k=0}^{n} \sum_{j=0}^{k} [F]_{Lip}^{k-j} \|\Delta_j\|_1 \\
&\le \|\phi\|_{sup}^n [\phi]_{Lip} \sum_{j=0}^{n} \left(\sum_{k=0}^{n-j} [F]_{Lip}^{k} \right) \|\Delta_j\|_1 . \quad \square
\end{aligned}$$

3.3 Comparison of both methods

THEORETICAL ASPECTS: The marginal and the Markovian quantization processes were assigned two different objectives. The marginal quantization process is originally designed to optimize the marginal distribution approximation at every time step $k = 0, \dots, n$, namely

$$\|\widehat{X}_k - X_k\|_p = \min\left\{\|Y - X_k\|_p,\, |Y(\Omega)| \le N_k\right\}$$

(with in mind some algorithmic stability properties of the grid optimization). Then, at every time k, the conditional distribution $\mathcal{L}(X_{k+1} \,|\, X_k = x) = P(x, dy)$ for a point $x \in C_i(\Gamma_k)$ (*i.e.* x in the tessel of x_k^i) is approximated by

$$\begin{aligned}
\mathcal{L}(X_{k+1} \,|\, X_k = x) &\approx \mathcal{L}(\mathrm{Proj}_{\Gamma_{k+1}}(X_{k+1}) \,|\, \mathrm{Proj}_{\Gamma_k}(X_k) = x) \\
&= \frac{1}{\mathbb{P}_{X_k}(C_i(\Gamma_k))} \int_{C_i(\Gamma_k)} P_{k+1}(x, dy) \circ \mathrm{Proj}_{\Gamma_{k+1}}^{-1} \mathbb{P}_{X_k}(dx).
\end{aligned} \tag{3.18}$$

This induces a loss of the Markov property.

In contrast, the Markovian quantization is designed at every time $k = 0, \ldots, n-1$ to optimize the approximation of the transition $P_{k+1}(x, dy) = F_{k+1}(x, \mathbb{P}_{\varepsilon_{k+1}})(dy)$ of the chain at the points of the quantizing grid $x_k^i \in \Gamma_k$ (in $L^p(\mathbb{P}_{\widehat{X}_k})$), namely

$$\|\widehat{X}_{k+1} - F_{k+1}(\widehat{X}_k, \varepsilon_{k+1})\|_p = \min\{\|Y - F_{k+1}(\widehat{X}_k, \varepsilon_{k+1})\|_p,\ |Y(\Omega)| \le N_{k+1}\}.$$

In this approach, for every $x \in C_i(\Gamma_k)$, one approximates

$$\mathcal{L}(X_{k+1} \,|\, X_k = x) \approx \mathcal{L}(\widehat{X}_{k+1} \,|\, \widehat{X}_k = x_k^i) = P_{k+1}(x_k^i, dy) \circ \mathrm{Proj}_{\Gamma_{k+1}}^{-1} \tag{3.19}$$

Then the Markov property w.r.t. the filtration of $(X_k)_{0 \le k \le n}$ is preserved.

In the marginal quantization, the conditional distributions are not approximated by a specific optimization process, but by averaging the transition w.r.t. the marginal distribution over the Voronoi tessellation of the best possible grid. In the Markovian approach, the conditional distributions are obtained by an optimization procedure which minimizes the error induced at the points of the grid.

One may notice by looking at the *a priori* estimates (3.5) and (3.14) that, provided that Assumption **(A'1)** is satisfied, both approaches lead to quite similar *a priori* error bound structures: they differ by the Lipschitz constants $[P]_{Lip}$ in the marginal quantization and $[F]_{Lip}$ in the Markovian quantization on one hand and by some multiplicative factor (in favor of the Markovian quantization) on the other hand. It is easy to prove that the inequality $[F]_{Lip} \le [P]_{Lip}$ always holds and in many "regular" models (like Lipschitz mixing models, Gaussian Euler schemes of diffusions, etc), the inequality stands as an equality. On the other hand, the multiplicative factor appearing in the marginal quantization is likely to be an artifact of the method of proof. Overall, the assets and drawbacks of both methods essentially annihilate each other.

Then, how to discriminate between the two quantization methods?

One first difference lies in the proofs: the general *a priori* error bounds like (3.14) are significantly easier to get by Markovian approach and so far, provided slightly lower theoretical constants.

When F is the Euler scheme of a diffusion process over $[0, T]$ with Lipschitz coefficients, then $[P]_{Lip} \le [F]_{Lip} = 1 + c_T/n$, where the time step is T/n. Then, in both methods, if one assigns $N/(n+1)$ elementary quantizers to each grid Γ_k and assumes this grid is optimal, inequalities (3.5) and (3.14) lead to the structure of *a priori* global error bound, namely

$$\forall\, n,\, N \ge 1, \qquad \mathrm{Error}(n, N) = O\left(\frac{n}{(N/n)^{1/d}}\right) = O\left(\frac{n^{1+1/d}}{N^{1/d}}\right)$$

In fact, without any further assumption on the probability density functions of the $\mathcal{L}(X_k)$'s, the above bound is only heuristic since it is impossible to control the rates of convergence in the asymptotics of the n optimal quantization errors. So far, this control turned out to be possible with marginal quantization under some *domination-scaling property* (see *e.g.* [1] for American option pricing but has no rigorous counterpart with Markovian quantization (see [19] for such a situation). The preservation of the Markov property maybe

induces a greater degeneracy of the "innovation process": thus, there is more randomness in $\mathrm{Proj}_{\Gamma_k}(X_k)$ where X_k follows (3.10) than in $\mathrm{Proj}_{\Gamma_k}(F(\widehat{X}_{k-1}, \varepsilon_k))$ in (3.11).

So, when the choice is possible, it seems to be essentially motivated by the constraints of the problem: thus, the Markovian quantization, being a Markov chain w.r.t. the filtration of the original chain X_k, seems more appropriate for control problems (for which it was originally designed...) whereas marginal quantization yield more satisfactory results in optimal stopping problems (for which it was originally designed...). But once again, it may be only an artifact.

Let us mention however that the marginal quantization requires only some weak convergence Lipschitz assumption on the chain (namely $[P]_{Lip} < +\infty$) while the Markovian quantization requires some L^p-pathwise Lipschitz assumption (namely $[F]_{Lip} < +\infty$). It may happen that the first approach turns out to be the only available one because

$$[P]_{Lip} < +\infty \;=\; [F]_{Lip}.$$

This is, for example, the case for Markovian dynamics like

$$X_{k+1} = F(X_k, \varepsilon_{k+1}) \quad \text{with} \;\; F(x,\varepsilon) := \mathrm{sign}(x-\varepsilon)G(x,\varepsilon),$$

where $(\varepsilon_k)_k$ is an i.i.d. sequence, $\mathbb{P}_{\varepsilon_1}(du) = g(u)\,\lambda_q(du)$ (λ_q lebesgue measure on $\mathbb{R}^q$) and $(x,\varepsilon) \mapsto G(x,\varepsilon)$ is Lipschitz continuous in x uniformly in ε with ratio $[G]_{Lip}$. Then, one shows that

$$[P]_{Lip} = [G]_{Lip} < +\infty \quad \text{whereas} \quad x \mapsto F(x,\varepsilon_1) = \mathrm{sign}(x-\varepsilon_1)G(x,\varepsilon_1) \quad \text{is not continuous.}$$

Computational aspects: Although, both dynamic programming formulae are formally identical and the fact that, in both cases, the grid optimization phase consists in processing a chain of stochastic gradient descents, one for each time step, the optimization phases are radically different for the marginal and the Markovian quantization processes. Since, these procedures have been extensively described in [1] and [19], we refer to these papers for details of implementation.

We wish to discuss here what make them different. First, they lead to different optimal grids with different transition matrices (using the same set of grids to process the marginal and the Markovian methods would provide two different sets of transition matrices).

In the *marginal quantization*, the optimization consists in two steps

– Computation for every $k = 0, \ldots, n$ of grids Γ_k^* which minimize over all grids of size N_k, the L^p-quantization error $\|\Delta_k\|_p$ *i.e.* solving

$$\mathrm{argmin}_{\Gamma_k}\{\|X_k - \mathrm{Proj}_{\Gamma_k}(X_k)\|_p,\ |\Gamma_k| \le N_k\} \tag{3.20}$$

– Estimation of the companion parameters *i.e.* the resulting transition matrices $[\hat{p}_k^{ij}]$ and the quantization errors $\|X_k - \mathrm{Proj}_{\Gamma_k^*}(X_k)\|_p$.

At every step k, the optimization problem (3.20) *only depends on the distribution of* X_k. The main consequence is that if one looks carefully at the recursive stochastic algorithm described in Section 2.2., the optimization of the grid Γ_k at the k^{th} time step only depends

on the simulation of a large number M of independent copies of X_k. So if one simulates on a computer M independent paths of the whole chain $(X_k)_{0\le k\le n}$, all the grids can be optimized independently by simply implementing procedures (2.10), (2.11).

The estimation of the companion parameters can be carried out "on line" as described in the algorithm of Section 2.2 using (2.12) and (2.13). It may be more efficient to carry on the companion parameter estimation after the grid optimization is achieved: once the optimal grids are settled, the companion parameter estimation procedure becomes a standard Monte Carlo simulation.

At a first glance, in the *Markovian quantization*, the two steps look similar. However, since $\widehat{X}_k = \mathrm{Proj}_{\Gamma_k}(F(\widehat{X}_{k-1}, \varepsilon_k))$, the L^p-optimization problem for the k^{th} grid Γ_k^* reads

$$\mathrm{argmin}_{\Gamma_k}\{\|F(\widehat{X}_{k-1}, \varepsilon_k) - \mathrm{Proj}_{\Gamma_k}(F(\widehat{X}_{k-1}, \varepsilon_k))\|_p,\ |\Gamma_k| \le N_k\}. \tag{3.21}$$

Consequently, *the optimization of the grids Γ_k at time k does depend on the distribution of $\widehat{X}_{k-1}$, i.e.* essentially upon Γ_{k-1}^*. This means that the grid optimization phase of a quantized markov chain is deeply recursive: any optimization default at time k is propagated at times $\ell \ge k$, inducing a great instability of the global optimization process.

This provides an interpretation for a usually observed phenomenon: numerical grid optimization works much better with marginal quantization than Markovian quantization. It is in accordance with the idea that it is more difficult to estimate accurately conditional distributions than marginal ones.

4 Some applications in finance

4.1 Optimal stopping problems and pricing of multi-dimensional American options

We consider a multidimensional diffusion $X = (X^1, \dots, X^d)^*$ governed by:

$$dX_t \;=\; b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0, \tag{4.1}$$

where b, σ are functions on $\mathbb{R}^d$ valued in $\mathbb{R}^d$ and $\mathbb{R}^{d\times m}$, satisfying usual growth and Lipschitz conditions, and W is a m-dimensional standard Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_t, \mathbb{P})$.

Given a reward process $(g(t, X_t))_{t\in[0,T]}$, where g is some continuous function on $[0,T]\times\mathbb{R}^d$, Lipschitz continuous in x, we consider the optimal stopping problem:

$$V_t \;=\; \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}\left[g(\tau, X_\tau)|\mathcal{F}_t\right]. \tag{4.2}$$

Here $\mathcal{T}_{t,T}$ denotes the set of stopping times valued in $[t,T]$ and V is called the Snell envelope of $(g(t, X_t))_{t\in[0,T]}$.

We first approximate this continuous-time optimal stopping problem by a discrete-time optimal stopping problem where the set of possible stopping times is valued in $\{kT/n : k = 0, \dots, n\}$ for n large. When the diffusion X is not simulatable, we approximate it by

a discretization scheme, and we denote by $\overline{X}_k$ this approximation at time $t_k = kT/n$ of X. For example, in the case of an Euler scheme with step T/n, we have:

$$\begin{aligned}\overline{X}_0 = x_0, \qquad \overline{X}_{k+1} &= \overline{X}_k + b(\overline{X}_k)\frac{T}{n} + \sigma(\overline{X}_k)\sqrt{\frac{T}{n}}\varepsilon_{k+1} \\ &=: F(\overline{X}_k, \varepsilon_{k+1}), \quad k = 0, \ldots, n-1,\end{aligned}$$

where $\varepsilon_{k+1} = (W_{t_{k+1}} - W_{t_k})/\sqrt{h}$ is a centered Gaussian random variable in $\mathbb{R}^m$ with variance I_m, independent of $\overline{\mathcal{F}}_k := \mathcal{F}_{t_k}$. The process $(\overline{X}_k)$ is a Markov chain w.r.t. the filtration $(\overline{\mathcal{F}}_k)$. The associated discrete-time optimal stopping problem is:

$$\overline{V}_k = \operatorname*{ess\,sup}_{\tau \in \overline{\mathcal{T}}_{k,n}} \mathbb{E}\left[g(\tau T/n, \overline{X}_\tau)|\overline{\mathcal{F}}_k\right], \tag{4.3}$$

where $\overline{\mathcal{T}}_{k,n}$ denotes the set of stopping times (with respect to the filtration $(\overline{\mathcal{F}}_k)$) valued in $\{j : j = k, \ldots, n\}$.

We have the classical time discretization error estimation:

$$\max_{k=0,\ldots,n} \|V_{kT/n} - \overline{V}_k\|_p \leq \frac{C_{b,\sigma}}{\sqrt{n}}$$

In fact, if g is slightly more regular, namely *semi-convex* and if one replaces the Euler scheme by the diffusion itself sampled at times kT/n, the above bound holds with $\frac{C_{b,\sigma}}{n}$.

It is well-known that the Snell envelope $(\overline{V}_k)_k$ of $(g(t_k, \overline{X}_k))_k$ satisfies $\overline{V}_k = \overline{v}_k(\overline{X}_k)$, $k = 0, \ldots, n$, where the Borel functions $\overline{v}_k$ on $\mathbb{R}^d$ are given by the backward dynamic programming formula:

$$\begin{aligned}\overline{v}_n(x) &= g(T, x), \quad \forall x \in \mathbb{R}^d, \\ \overline{v}_k(x) &= \max\left(g(t_k, x), \mathbb{E}[\overline{v}_{k+1}(\overline{X}_{k+1})|\overline{X}_k = x]\right), \ \forall x \in \mathbb{R}^d, \quad k = 0, \ldots, n-1.\end{aligned}$$

This backward formula remains intractable for numerical computations since it requires to compute at each time step $k = 0, \ldots, n$, conditional expectations of $\overline{X}_{k+1}$ given $\overline{X}_k = x$ at any point $x \in \mathbb{R}^d$ of the state space of $(\overline{X}_k)$.

The quantization approach for solving this problem is to first approximate the Markov chain $(\overline{X}_k)$ by a quantized Markov chain as described in Section 3. This means that at each time t_k, $k = 0, \ldots, n$, we are given an (optimal) grid $\Gamma_k = \{x_k^1, \ldots, x_k^{N_k}\}$ of N_k points in $\mathbb{R}^d$, and we approximate the distribution $\overline{X}_0$ by the distribution of $\widehat{X}_0 = \mathrm{Proj}_{\Gamma_0}(\overline{X}_0)$, and the conditional distribution of $\overline{X}_{k+1}$ given $\overline{X}_k$ by the conditional distribution of $\widehat{X}_{k+1}$ given $\widehat{X}_k$: for $k \geq 1$, $\widehat{X}_k$ is defined by $\widehat{X}_k = \mathrm{Proj}_{\Gamma_k}(\overline{X}_k)$ in the marginal quantization method, while $\widehat{X}_k = \mathrm{Proj}_{\Gamma_k}(F(\widehat{X}_{k-1}, \varepsilon_k))$ in the Markovian quantization method. We then approximate the functions $\overline{v}_k$ by the functions $\widehat{v}_k$ defined on Γ_k, $k = 0, \ldots, n$, by the backward dynamic programming formula or *optimal quantization tree descent*:

$$\widehat{v}_n(x_n^i) = g(T, x_n^i), \quad \forall x_n^i \in \Gamma_n \tag{4.4}$$

$$\widehat{v}_k(x_k^i) = \max\left(g(t_k, x_k^i), \mathbb{E}[\widehat{v}_{k+1}(\widehat{X}_{k+1})|\widehat{X}_k = x_k^i]\right) \tag{4.5}$$

$$= \max\left(g(t_k, x_k^i), \sum_{j=1}^{N_{k+1}} \hat{p}_{k+1}^{ij} \widehat{v}_{k+1}(x_{k+1}^j)\right), \ x_k^i \in \Gamma_k, \ k = 0, \ldots, n-1, \tag{4.6}$$

where $\hat{p}_{k+1}^{ij} = \mathbb{P}\left[\widehat{X}_{k+1} = x_{k+1}^j | \widehat{X}_k = x_k^i\right]$.

Then one gets an approximation of the process $(\overline{V}_k)$ by the process $(\widehat{V}_k)_k$, with $\widehat{V}_k = \widehat{v}_k(\widehat{X}_k)$. Namely, if the diffusion is uniformly elliptic, with coefficients b and σ either bounded Lipschitz continuous or $\mathcal{C}_b^\infty(\mathbb{R}^d)$ and if the obstacle function g is Lipschitz over $[0,T]\times\mathbb{R}^d$, then the following error estimation holds for an L^p-optimal marginal quantization (see [1]):

$$\max_{0\le k\le n} \|\overline{V}_k - \widehat{V}_k\|_p \le C_{b,\sigma,T,p}\frac{n^{1+1/d}}{N^{1/d}}, \tag{4.7}$$

where $N = \sum_{k=0}^n N_k$ is the total number of points to be dispatched among all grids Γ_k.

This estimate strongly relies on the sub-Gaussian upper-bound for the probability density of the diffusion density. The same bound holds if one substitutes the diffusion sampled at times t_k, $k = 0,\dots,n$ to its Euler scheme.

Numerical illustration As a numerical illustration, we consider a $2d$-dimensional uncorrelated Black-Scholes model with geometric dividends, *i.e.* for $x = (x^1,\dots,x^{2d})^* \in \mathbb{R}^{2d}$, $b(x) = -(\mu^1 x^1,\dots,\mu^{2d}x^{2d})$, $\sigma(x)$ is a $2d\times 2d$ diagonal matrix with ith diagonal term $\sigma_i\, x^i$, where σ_i, $i = 1,\dots,d$ are constant volatilities. We assume that the short-term interest rate is zero. The American option price at time t of a payoff function $(g(X_t))$ is given by:

$$V_t = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}\left[g(X_\tau)|\mathcal{F}_t\right], \tag{4.8}$$

which is computed by the above algorithm. We consider an American $2d$-dim exchange option characterized by the payoff

$$g(t,x) = \max\left(x^1\dots x^d - x^{d+1}\dots x^{2d}, 0\right)$$

with the following market parameters

$$x_0^1\dots x_0^d = 36, \quad x_0^{d+1}\dots x_0^{2d} = 40, \quad \sigma_i = 20d^{-\frac{1}{2}}\,\%, \quad \mu^1 = 5\,\%,\ \mu^2 = \dots = \mu^{2d} = 0.$$

Our reference price is obtained by a specific difference method devised in [22] for 2-dimension. We reproduce in Figures 3 and 4 for $2d = 4$ and 6 the graphs $\theta \mapsto \widehat{V}_0(\theta)$ where $\widehat{V}_0(\theta)$ denotes the premium at time 0 of the above American option when time to maturity θ runs over $\{k/n,\, k = 1,\dots,n\}$. The numerical parameters are settled as follows: $n = 25$, k time discretization steps when $\theta = k/n$, and

$$N_{25} = 750 \quad \text{if} \quad 2d = 4 \quad \text{and} \quad N_{25} = 1\,000 \quad \text{if} \quad 2d = 6.$$

The sizes N_k of the grid Γ_k is specified following the dispatching rule given in [1].

4.2 A stochastic control problem: mean-variance hedging of options

4.2.1 Error bounds using the Markovian quantization

We consider the following portfolio optimization problem. The dynamics of the controlled process is governed by:

$$\begin{aligned} dX_t &= b(X_t)\,dt + \sigma(X_t)\,dW_t, \\ dY_t &= \alpha_t^* dX_t, \qquad Y_0 = y_0, \end{aligned}$$

where b, σ are functions on $\mathbb{R}^d$ valued in $\mathbb{R}^d$ and $\mathbb{R}^d \times m$, satisfying usual growth and Lipschitz conditions, and W is a m-dimensional standard Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_t, \mathbb{P})$. The control process $\alpha = (\alpha_t)_t$ is an $\mathbb{F}$-adapted process valued in some subset A of $\mathbb{R}^d$. We denote by $\mathcal{A}$ the set of such control processes. Here, $X = (X^1, \ldots, X^d)^*$ represents the dynamics of risky assets and/or volatility, Y is the (self-financed) wealth process of an investor who can trade α_t shares of risky assets at time t, and starting from some initial capital y_0. The set A models the constraints on the portfolio held by the investor. For example, if $A = \mathbb{R} \times \{0\}^{d-1}$, this means that the investor can trade only in the first asset.

We are now given an option written on the risky assets, *i.e.* a payoff function in the form $g(X_T)$, for some Lipschitz continuous function g on $\mathbb{R}^d$, that one wants to hedge with the available risky assets, and according to a quadratic criterion. In other words, one has to solve the stochastic control problem:

$$v(t,x,y) = \inf_{\alpha \in \mathcal{A}} \mathbb{E}\left[(g(X_T) - Y_T)^2 \Big| (X_t, Y_t) = (x,y) \right], \quad (t,x,y) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}. \tag{4.9}$$

We first approximate the continuous-time control problem (4.9) by a discrete-time control problem at dates $t_k = kT/n$, $k = 0, \ldots, n$ for n large. We consider an approximation Euler scheme for (X_t) with step $h = T/n$. The approximation $\overline{X}_k$ of X_{t_k} is then defined by:

$$\begin{array}{rcll} \overline{X}_0 = X_0 \qquad \text{and} \qquad \overline{X}_{k+1} & = & \overline{X}_k + b(\overline{X}_k)\dfrac{T}{n} + \sigma(\overline{X}_k)\sqrt{\dfrac{T}{n}}\,\varepsilon_{k+1} \\ & =: & F(\overline{X}_k, \varepsilon_{k+1}), & k = 0, \ldots, n-1, \end{array}$$

where $\varepsilon_{k+1} = (W_{t_{k+1}} - W_{t_k})/\sqrt{h}$ is a centered Gaussian random variable in $\mathbb{R}^m$ with variance I_m, independent of $\overline{\mathcal{F}}_k := \mathcal{F}_{t_k}$. The process $(\overline{X}_k)$ is a Markov chain w.r.t. the filtration $(\overline{\mathcal{F}}_k)$. We denote by $\overline{\mathcal{A}}$ the set of all $\{\overline{\mathcal{F}}_k, k = 0, \ldots, n-1\}$-adapted processes $\overline{\alpha} = \{\overline{\alpha}_k, k = 0, \ldots, n-1\}$ valued in A. Given $\overline{\alpha} \in \overline{\mathcal{A}}$, we consider the approximation $(\overline{Y}_k)$ of the controlled process (Y_t) at times (t_k), and defined by:

$$\begin{array}{rcll} \overline{Y}_0 = y_0 \qquad \text{and} \qquad \overline{Y}_{k+1} & = & \overline{Y}_k + \overline{\alpha}_k^*(\overline{X}_{k+1} - \overline{X}_k) \\ & =: & G(\overline{Y}_k, \overline{X}_k, \overline{\alpha}_k, \overline{X}_{k+1}), & k = 0, \ldots, n-1. \end{array}$$

We then consider the stochastic control problem in discrete-time:

$$\overline{v}_k(x,y) \;=\; \inf_{\overline{\alpha} \in \overline{\mathcal{A}}} \mathbb{E}\left[\left(g(\overline{X}_n) - \overline{Y}_n\right)^2 \Big| (\overline{X}_k, \overline{Y}_k) = (x,y) \right], \tag{4.10}$$

for all $k = 0, \ldots, n$ and $(x,y) \in \mathbb{R}^d \times \mathbb{R}$. The convergence from the discrete-time control problem to the continuous one may be proved either by probabilistic arguments (see [16]) or by viscosity solutions approach (see [7]):

$$\overline{v}_k(x,y) \;\to\; v(t,x,y),$$

for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}$, as n goes to infinity and $t_k \to t$.

The functions $\overline{v}_k$ satisfy the dynamic programming formula:

$$\begin{aligned}\overline{v}_n(x,y) &= (g(x)-y)^2\,, \quad (x,y)\in\mathbb{R}^d\times\mathbb{R}\\ \overline{v}_k(x,y) &= \inf_{a\in A}\mathbb{E}\left[\overline{v}_{k+1}(\overline{X}_{k+1},\overline{Y}_{k+1})\Big|\,(\overline{X}_k,\overline{Y}_k)=(x,y)\right],\\ &\qquad k=0,\ldots,n-1,\ \ (x,y)\in\mathbb{R}^d\times\mathbb{R}.\end{aligned}$$

From a numerical viewpoint, this backward formula remains intractable since we have to compute at each time step, conditional expectations of $(\overline{X}_{k+1},\overline{Y}_{k+1})$ given $(\overline{X}_k,\overline{Y}_k)=(x,y)$ at every point (x,y) of the state space $\mathbb{R}^d\times\mathbb{R}$. With respect to optimal stopping problems, we have in addition to calculate an infimum of these conditional expectations over the possible values of the control set A.

The starting point in the quantization approach for solving (4.10) is to discretize the controlled $(\overline{\mathcal{F}}_k)$-Markov chain $(\overline{X}_k,\overline{Y}_k)_k$ by a controlled Markov chain $(\widehat{X}_k,\widehat{Y}_k)_k$ valued in a finite state space. Here, recall that $(\overline{X}_k)_k$ is an uncontrolled process while $(\overline{Y}_k)_k$ is one-dimensional controlled process. We shall then consider two different spatial discretizations for $(\overline{X}_k)_k$ and $(\overline{Y}_k)_k$. Moreover, we also want to keep the Markov property of the controlled quantized Markov chain w.r.t. the same filtration $(\overline{\mathcal{F}}_k)$. This means that we wants to approximate the control problem (4.10) by another control problem where the controls are still adapted w.r.t. the filtration $(\overline{\mathcal{F}}_k)$. More precisely, we shall discretize the d-dimensional process $(\overline{X}_k)$ on an optimal grid $\Gamma_k=\{x_k^1,\ldots,x_k^{N_k}\}$ at each time k and define a Markovian quantization of $(\overline{X}_k)$ by:

$$\widehat{X}_0=\mathrm{Proj}_{\Gamma_0}(X_0)\qquad\text{and}\qquad \widehat{X}_{k+1}=\mathrm{Proj}_{\Gamma_{k+1}}\Big(F(\widehat{X}_k,\varepsilon_{k+1})\Big)\,,\quad k=0,\ldots,n-1.$$

The controlled one-dimensional process $(\overline{Y}_k)$ is discretized using a regular orthogonal grid of $\mathbb{R}$, namely $\Gamma^Y=(2\delta)\mathbb{Z}\cup[-R,R]$, and we then define:

$$\widehat{Y}_0=y_0\qquad\text{and}\qquad \widehat{Y}_{k+1} = \mathrm{Proj}_{\Gamma^Y}\Big(G(\widehat{Y}_k,\widehat{X}_k,\overline{\alpha}_k,\widehat{X}_{k+1})\Big)\,,\quad k=0,\ldots,n-1.$$

Therefore, $(\widehat{X}_k,\widehat{Y}_k)$ is a controlled Markov chain w.r.t. $(\overline{\mathcal{F}}_k)$. We then consider the stochastic control problem in discrete-time:

$$\widehat{v}_k(x,y) = \inf_{\overline{\alpha}\in\overline{\mathcal{A}}}\mathbb{E}\left[\Big(g(\widehat{X}_n)-\widehat{Y}_n\Big)^2\,\Big|\,(\widehat{X}_k,\widehat{Y}_k)=(x,y)\right], \tag{4.11}$$

for all $k=0,\ldots,n$ and $(x,y)\in\Gamma_k\times\Gamma^Y$. By the dynamic programming principle, functions $\widehat{v}_k$ are computed recursively by:

$$\begin{aligned}\widehat{v}_n(x,y) &= (g(x)-y)^2\,, \quad (x,y)\in\Gamma_n\times\Gamma^Y\\ \widehat{v}_k(x,y) &= \inf_{a\in A}\mathbb{E}\left[\widehat{v}_{k+1}(\widehat{X}_{k+1},\widehat{Y}_{k+1})\Big|\,(\widehat{X}_k,\widehat{Y}_k)=(x,y)\right],\\ &\qquad k=0,\ldots,n-1,\ \ (x,y)\in\Gamma_k\times\Gamma^Y.\end{aligned}$$

From an algorithmic point of view, this reads:

$$\widehat{v}_n(x_n^i,y) = (g(x_n^i)-y)^2,\quad \forall x_n^i\in\Gamma_n,\ \forall y\in\Gamma^Y$$

$$\widehat{v}_k(x_k^i) = \inf_{a\in A} \sum_{j=1}^{N_{k+1}} \hat{p}_{k+1}^{ij} \widehat{v}_{k+1}\left(x_{k+1}^j, \mathrm{Proj}_{\Gamma^Y}\left(G(y, x_k^i, a, x_{k+1}^j)\right)\right)$$
$$\forall\, x_k^i \in \Gamma_k,\, \forall\, y \in \Gamma^Y, \quad k = 0, \ldots, n-1,$$
$$\text{where} \qquad \hat{p}_{k+1}^{ij} = \mathbb{P}\left[\widehat{X}_{k+1} = x_{k+1}^j | \widehat{X}_k = x_k^i\right].$$

It is proved in [19] that the estimation error for the value functions by this quantization method is measured by:

$$\mathbb{E}\left|\overline{v}_0(\overline{X}_0, y_0) - \widehat{v}_0(\overline{X}_0, y_0)\right| \leq C_1(1+|y_0|)\left(\frac{1}{\sqrt{n}}\sum_{k=0}^{n}(n-k)\|\Delta_k\|_2 + \sum_{k=0}^{n}\|\Delta_k\|_2\right)$$
$$+C_2 n\delta + C_3(1+|y_0|^{\overline{p}})\frac{n}{R^{\overline{p}-1}},$$

for all $\overline{p} > 1$ and $y_0 \in \mathbb{R}$. Here, C_1, C_2 and C_3 are positive constants depending on the coefficients of the diffusion process X and

$$\|\Delta_k\|_2 = \|F(\widehat{X}_{k-1}, \varepsilon_k) - \widehat{X}_k\|_2$$

is the L^2-quantization error at date k in the Markovian approach.

4.2.2 Numerical illustrations

As a numerical illustration, we consider the two following models:

A stochastic volatility model (2-dim X-process) Let $X = (X^1, X^2)$ be governed by:

$$dX_t^1 = X_t^1 X_t^2 dW_t^1$$
$$dX_t^2 = -\eta(X_t^2 - \overline{\sigma})dt + \beta dW_t^2, \quad X_0^2 \sim \mathcal{N}(\overline{\sigma}; \frac{\beta^2}{2\eta})$$

where (W^1, W^2) is a standard two-dimensional Brownian motion. Here X^1 represents the price process of one risky asset and X^2 is the (stationary) stochastic volatility process of the risky asset. The investor trades only in the risky asset X^1, i.e. $A = \mathbb{R} \times \{0\}$, and he wants to hedge a put option on this asset, i.e. $g(x) = (K - x^1)_+$ for $x = (x^1, x^2)$.

By projecting $(K - X_T^1)_+$ on the set of stochastic integrals w.r.t. S, we have by Itô's formula:

$$(K - X_T^1)_+ = \mathbb{E}[(K - X_T^1)_+ \,|\, X_0^2] + \int_0^T \alpha_s^{opt} dX_s^1 + Z_T$$

where $\alpha_t^{opt} = \frac{\partial P}{\partial s}(t, X_t^1, X_t^2)$, $Z_T = \beta\int_0^T \frac{\partial P}{\partial s}(t, X_t^1, X_t^2)dW_t^2$ with $P(t, x^1, x^2) = \mathbb{E}[(K - X_T^1)_+ \,|\, (X_t^1, X_t^2)) = (x^1, x^2)]$.

Then, the function $y_0 \mapsto \mathbb{E}[v_0(y_0, x_0^1, X_0^2)]$ reaches its minimum at $y_{\min} = \mathbb{E}[(K - X_T^1)_+)]$. So, the optimal control is always α^{opt} regardless of y. Since the volatility process X_t^2 is independent of W^1, we notice by Jensen's inequality that

$$\mathbb{E}[(K - X_T^1)_+] \geq \mathrm{Put}_{\mathrm{B\&S}}(x_0^1, K, T, \overline{\sigma}).$$

Parameters for numerical implementation :

- Model parameters: $T = 1$, $\overline{\sigma} = 20\%$, $\eta = 0.5$, $\beta = 0.05$, $x_0^1 = K = 100$.
- Time discretization: $n = 25$.
- Spatial discretization (quantization grid parameters):
  – Grid Γ^X: $2\delta = \frac{1}{20}$, $n_X = 50$ (*i.e.* $|\Gamma^X| = 2 \times 100 + 1$), centered at $I_0 = 7.96$.
  – Grids Γ_k: Total numbers of points used to produce the $n = 25$ grids that quantize the Euler scheme of (S, σ), $N = 5\,750$ ($N_{25} = 299$).
  – Optimization of the grids using $M = 10^6$ independent trials of the Euler scheme.
- Approximation of the optimal control: dichotomy method on $A = [-1, 0]$.

NUMERICAL RESULTS: Figure 5 below depicts a quantization of $X_T = (X_T^1, X_T^2)$ using $N_{25} := 299$ points obtained as a result of an optimization process described above

Figure 6 and Figure 7 display the computed graph of $y \mapsto \mathbb{E}[\widehat{v}_0(y, x_0^1, \widehat{X}_0^2)]$ and the value of the optimal control α_0^{opt} at $t = 0$ respectively. The global shape of the graph is parabolic and reaches its minimum at $y_{min} = 8.06$. This is to be compared with the premium provided by a direct Monte Carlo simulation, namely 8.00. The optimal control is nearly constant and its value at $y_{min} = 8.06$, $\alpha_0^{mv}(y_{min}) = -0.38$, is satisfactory w.r.t. the theoretical value estimated by Monte Carlo (-0.34).

A put on a Black & Scholes geometric asset (4-dim X-process) We consider a Black & Scholes model $X = (X^1, \ldots, X^4)$ in 4 dimension: $b = 0$ and $\sigma(x)$ is the 4×4 diagonal matrix with i^{th} diagonal term $\sigma_i(x) := \sigma_i x^i$. The payoff function to be hedged is a geometric put option on $J_t = X_t^1 \ldots X_t^d$:

$$g(X_T) = (K - J_T)_+.$$

The investor is allowed to trade only in the first asset X^1 hence $A = \mathbb{R} \times \{0\}^3$.

So, the mean variance hedging problem of the investor at time $t = 0$ is

$$v(0, x_0, y_0) = \min_{\alpha \in \mathcal{A}} \mathbb{E}\left[\left(Y_T^{y_0,\alpha} - (K - J_T)_+\right)^2\right]. \tag{4.12}$$

where x_0 is the initial vector value risky asset. Itô's formula then classically yields

$$(K - J_T)_+ = \mathbb{E}[(K - J_T)_+] + \int_0^T H_t \, dJ_t \quad \text{with } H_t = \frac{\partial P}{\partial x}(T - t, J_t, K, \sigma),\ t \in [0, T),$$

where $P(\theta, x, K, \sigma)$ denotes the price of a one dimensional European Put option with residual maturity θ, asset price x, strike price K, constant volatility σ. It follows that

$$\begin{aligned}\mathbb{E}\left[(Y_T^{y_0,\alpha} - (K - J_T)_+)^2\right] &= \Big(y_0 - P(T, J_0, K, \sigma)\Big)^2 + \sigma_1^2 \int_0^T \mathbb{E}\,(J_s H_s - \alpha_s X_s^1)^2 \, ds \\ &\quad + \sum_{i=2}^d \sigma_i^2 \int_0^T \mathbb{E}\,(J_s H_s)^2 \, ds.\end{aligned}$$

Hence, the solution of (4.12) is given by

$$v_0(x_0, y_0) = \Big(y_0 - P(T, J_0, K, \sigma)\Big)^2 + \sum_{i=2}^d \sigma_i^2 \int_0^T \mathbb{E}\,(J_s H_s)^2 \, ds$$

using the optimal control

$$\alpha_t^{opt} = \frac{J_t H_t}{S_t^1} = X_t^2 \dots X_t^d \frac{\partial P}{\partial x}(T-t, J_t, K, \sigma).$$

In the above model, the non correlation assumption of the assets may look not very realistic but corresponds to the most difficult case to solve for quantization since it corresponds in some way to a "full d-dimensional problem".

Parameters for numerical implementation:

• Model parameters: $d = 4$, $T = 1$, $\sigma = 15\%$, $\sigma_1 = \sigma/\sqrt{2}$ and $\sigma_i = \sigma/\sqrt{2(d-1)}$, $i = 2, \dots, d$, $X_0^i = (100)^{1/d}$, $i = 1, \dots, d$, $K = 100$.

• Spatial discretization (quantization grid parameters):

– Grid Γ^Y: $2\delta := \frac{1}{10}$, $n_Y := 2\,00$ (*i.e.* $|\Gamma^X| = 2 \times 2\,00 + 1$), centered at $I_0 = 5.97$ ($B\&S$ premium of the put option with volatility σ).

– Grids Γ_k: Total numbers of points used to produce the $n = 20$ grids that quantize the geometric Brownian motion S: $N = 22\,656$ ($N_{20} := 1\,540$).

– Optimization of the grids using $M := 10^6$ independent trials of the Euler scheme.

• Approximation of the optimal control: dichotomy method on $A \subset [-2, 0]$.

Numerical results: Figure 8 and Figure 9 below display the computed graphs $x_0 \mapsto v_0(x_0, s_0)$ and $y_0 \mapsto \alpha_0^{opt}(y_0, I_0)$. The global shape of the graph is parabolic, reaches its minimum (equal to 25.82) at $y_{min} = 6.27$ (the true value is 5.97, $B\&S$ premium for the put). The optimal control is satisfactorily constant as expected; its value at y_{min} is $\alpha_0^{opt}(y_{min}) = -0.48$ (true value is -0.47).

4.3 Filtering of stochastic volatility models

We consider the following filtering model. The signal (X_k) is an $\mathbb{R}^d$-valued Markov chain given by:

$$X_k \quad = \quad F(X_{k-1}, \varepsilon_k), \quad k \in \mathbb{N}^*, \tag{4.13}$$

where $(\varepsilon_k)_k$ is a sequence of i.i.d. random variables valued in $\mathbb{R}^m$, and F is some measurable function on $\mathbb{R}^d \times \mathbb{R}^m$. The initial distribution of X_0 is known equal to μ. The observation process valued (Y_k) valued in $\mathbb{R}^q$ takes the form:

$$Y_k \quad = \quad G(X_k, \eta_k), \quad k \in \mathbb{N}, \tag{4.14}$$

where $(\eta_k)_k$ is a sequence of i.i.d. random variables in $\mathbb{R}^l$, independent of $(\varepsilon_k)_k$, and G is a measurable functions on $\mathbb{R}^d \times \mathbb{R}^l$. We assume that for every $x \in \mathbb{R}^d$, the random variable $G(x, \eta_1)$ admits a bounded density $y \mapsto g(x, y)$ w.r.t. the Lebesgue measure on $\mathbb{R}^q$.

We are interested in the computation at some time $n \geq 1$, of the conditional distribution $\Pi_{y,n}$ of the signal X_n given the observations $(Y_0, \dots, Y_n)$ fixed to $y = (y_0, \dots, y_n)$. In other words, we wish to calculate the conditional expectations

$$\Pi_{y,n} f \quad = \quad \mathbb{E}\left[f(X_n) | \, (Y_0, \dots, Y_n) = y \right] \tag{4.15}$$

for all reasonable functions f on $\mathbb{R}^d$. From the Markov property of the pair (X_k, Y_k) and Bayes formula, we have the following expression for $\Pi_{y,n}$:

$$\Pi_{y,n} f = \frac{\pi_{y,n} f}{\pi_{y,n} \mathbf{1}} \qquad \text{where} \qquad \pi_{y,n} f = \mathbb{E}\left[f(X_n) \prod_{k=0}^{n} g(X_k, y_k) \right], \tag{4.16}$$

for any $f \in \mathcal{B}(\mathbb{R}^d)$, the set of bounded measurable functions on $\mathbb{R}^d$. This can be derived by noting that the function $y = (y_0, \ldots, y_n) \in (\mathbb{R}^q)^{n+1} \mapsto \pi_{y,n}\mathbf{1} = \mathbb{E}[\prod_{k=0}^{n} g(X_k, y_k)]$ is actually equal to the density ϕ_{n+1} of $(Y_0, \ldots, Y_n)$ w.r.t. to the Lebesgue measure on $(\mathbb{R}^q)^{n+1}$.

In the sequel, the observations are fixed to $y = (y_0, \ldots, y_n)$ and we write π_n for $\pi_{y,n}$ and $\Pi_n = \Pi_{y,n}$.

The computation of the unnormalized filter $\pi_{y,n}$ is based on the following inductive formula:

$$\pi_k f \quad = \quad \pi_{k-1} H_k f, \quad k = 1, \ldots, n,$$

where H_k is the transition kernel given by:

$$H_k f(x) \quad = \quad \mathbb{E}\left[f(X_k) g(X_k, y_k) \middle| X_{k-1} = x \right].$$

Hence, the inductive formula of the unnormalized filter relies on successive computations of conditional expectations of X_{k+1} given X_k. Notice that with regard to the problems of optimal stopping or stochastic control problems, we have here an infinite-dimensional problem, since we have to calculate these conditional expectations for any Borel bounded functions on $\mathbb{R}^d$. For solving numerically this problem, we are then suggested to approximate the conditional distributions of X_k given X_{k-1} for any $k = 1, \ldots, n$ by a quantization approach as described in Section 3. We are then given, at each time $k = 0, \ldots, n$, an (optimal) grid $\Gamma_k = \{x_k^1, \ldots, x_k^{N_k}\}$ of N_k points in $\mathbb{R}^d$, and we approximate the distribution μ of X_0 by the distribution of $\widehat{X}_0 = \mathrm{Proj}_{\Gamma_0}(X_0)$, and the conditional distribution of X_k given X_{k-1} by the conditional distribution of $\widehat{X}_k$ given $\widehat{X}_{k-1}$: for $k \geq 1$, $\widehat{X}_k$ is defined by $\widehat{X}_k = \mathrm{Proj}_{\Gamma_k}(X_k)$ in the marginal quantization method, while $\widehat{X}_k = \mathrm{Proj}_{\Gamma_k}(F(\widehat{X}_{k-1}, \varepsilon_k))$ in the Markovian quantization method.

We then approximate the transition kernel H_k by the transition matrix $\widehat{H}_k$ defined by:

$$\begin{aligned} \widehat{H}_k^{ij} \quad &= \quad \mathbb{E}\left[\mathbf{1}_{\widehat{X}_k = x_k^j} g(\widehat{X}_k, y_k) \middle| \widehat{X}_{k-1} = x_{k-1}^i \right] \\ &= \quad \hat{p}_k^{ij} g(x_k^j, y_k), \quad i = 1, \ldots, N_{k-1},\ j = 1, \ldots, N_k. \end{aligned}$$

Here, $(\hat{p}_k)_k$ is the probability transition matrix of $(\widehat{X}_k)_k$, i.e.

$$\hat{p}_{k+1}^{ij} \quad = \quad \mathbb{P}\left[\widehat{X}_{k+1} = x_{k+1}^j | \widehat{X}_k = x_k^i \right],$$

and $\hat{p}_0 = (\hat{p}_0^i)_{i=1,\ldots,N_0}$ is the probability distribution of $\widehat{X}_0$, i.e. $\hat{p}_0^i = \mathbb{P}[\widehat{X}_0 = x_0^i]$. The unnormalized filter π_n is then approximated by the discrete probability measure $\hat{\pi}_n$ on Γ_n:

$$\hat{\pi}_n \quad = \quad \sum_{\ell=1}^{N_n} \hat{\pi}_n^\ell \delta_{x_n^\ell},$$

where $(\hat{\pi}_k^\ell)$, $k = 0, \ldots, n$, $\ell = 1, \ldots, N_k$, are computed inductively by:

$$\hat{\pi}_0 = \hat{p}_0 \qquad \text{and} \qquad \hat{\pi}_k^j = \sum_{1 \le i \le N_{k-1}} \widehat{H}_k^{ij} \hat{\pi}_{k-1}^i, \quad k = 1, \ldots, n, \quad j = 1, \ldots, N_k.$$

The normalized filter Π_n is finally approximated by the discrete probability measure $\widehat{\Pi}_n$ on Γ_n:

$$\widehat{\Pi}_n \;=\; \frac{\hat{\pi}_n}{\sum_{\ell=1}^{N_n} \hat{\pi}_n^\ell}.$$

Under the Lipschitz assumption **(A1')** on the scheme F, and assuming also that the function $g(x, y)$ is Lipschitz in x, uniformly in y, with ratio $[g]_{Lip}$, we have the following estimation error for the approximate filter (see [20]): for any $f \in BL_1(\mathbb{R}^d)$,

$$\left|\Pi_{y,n} f - \widehat{\Pi}_{y,n} f\right| \;\le\; \frac{\|g\|_\infty^{n+1}}{\phi_{n+1}(y)} \sum_{k=0}^{n} C_{n,k} \|\Delta_k\|_2,$$

where
$$C_{n,k}(f) \;=\; [F]_{Lip}^{n-k+1} + 2\frac{[g]_{Lip}}{\|g\|_\infty} \left(\frac{[F]_{Lip} + 1}{[F]_{Lip} - 1} ([F]_{Lip}^{n-k+1} - 1) + 1 \right),$$

$\|\Delta_0\|_2 = \|X_0 - \widehat{X}_0\|_2$, $\|\Delta_k\|_2 = \|X_k - \widehat{X}_k\|_2$ in the marginal quantization method, and $\|\Delta_k\|_2 = \|F(\widehat{X}_{k-1}, \varepsilon_k) - \widehat{X}_k\|_2$ in the Markovian quantization, $k = 1, \ldots, n$, are the L^2-quantization errors.

Remark 4.1 • In the marginal quantization method, the constant $[F]_{Lip}$ may be replaced by the constant $[P]_{Lip}$. Note that in regular examples studied here, we have $[P]_{Lip} = [F]_{Lip}$. • The uniform Lipschitz condition in y of $x \mapsto g(x, y)$ may be relaxed into a nonuniform Lipschitz condition in the form: $|g(x, y) - g(x', y)| \le [g]_{Lip}(y)(1 + |x| + |x'|)|x - x'|$, with in this case a more complex estimation error term.

Numerical illustrations : 1. *The (familiar) Kalman-Bucy model*:

$$\begin{aligned} X_{k+1} &= AX_k + \varepsilon_k \quad \in \mathbb{R}^d \\ Y_k &= BX_k + \eta_k \quad \in \mathbb{R}^q \end{aligned}$$

for $k \in \mathbb{N}$, and X_0 is normally distributed with mean $m_0 = 0$ and variance Σ_0. Here A and B are matrices of appropriate dimensions, and $(\varepsilon_k)_k$, $(\eta_k)_k$ are independent centered Gaussian processes, $\varepsilon_k \rightsquigarrow \mathcal{N}(0, I_d)$ and $\eta_k \rightsquigarrow \mathcal{N}(0, \Lambda)$. In this case, we have

$$g(x, y) \;=\; \frac{1}{(2\pi)^{d/2} \det(\Lambda^{\frac{1}{2}})} \exp\left(-\frac{1}{2} \left| \Lambda^{-\frac{1}{2}} (y - Bx) \right|^2 \right).$$

Of course, the filter $\Pi_{y,n}$ is explicitly known, see e.g. [10]: it is a Gaussian distribution of mean m_n and variance Σ_n given by the inductive equations:

$$\begin{aligned} \Sigma_{k+1}^{-1} &= I_d - A\left(A'A + \Sigma_k^{-1}\right)^{-1} A' + B'\Lambda^{-1}B \\ \left(\Sigma_{k+1}^{-1} m_{k+1}\right) &= A\left(A'A + \Sigma_k^{-1}\right)^{-1} \left(\Sigma_k^{-1} m_k\right) + B'\Lambda^{-\frac{1}{2}} y_{k+1}. \end{aligned}$$

We will illustrate the numerical scheme in dimension $d = 3$. Here A and B are

$$A = \begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} \qquad \text{and} \qquad B = I_3.$$

We take $\Gamma = (0.5)^2 I_3$. The variance Σ_0 is such that (X_k) is stationary. In this case, we can work with a single grid (1000 points). In Figure 10 is represented the 1000-optimal quantizer used for $\mathcal{N}(0, I_3)$. Computations are carried out with its Σ_0-rescaled version (which is a non optimal but straightforwardly accessible and quite efficient quantizer for $\mathcal{N}(0, \Sigma_0)$). The number n of observations is equal to 20. We compute the conditional expectations $\mathbb{E}[f(X_n) \mid Y_0, \ldots, Y_n]$ with $f(x) = x$ (the conditional mean) and $f(x) = x \cdot {}^t x$ (the conditional variance). The quantized version of the conditional mean is denoted by $\widehat{m}_n$ and that given by the Kalman filter by m_n. We take the same convention for the conditional variance Σ_n. We represent in Figure 11 the errors $\|m_k - \widehat{m}_k\|$ and $\|\Sigma_k - \widehat{\Sigma}_k\|$ plotted w.r.t. $k \in \{0, \ldots, 20\}$. Finally, Figure 12 depicts the three components of the conditional mean in its Kalman filter version and its quantized version. These figures shows that in this setting the $3d$ Kalman filter is well captured by the quantization method.

2. *A stochastic volatility model arising in financial time series*: Let S_k, $k \in \mathbb{N}$, be a positive process describing the stock prices in time, and define $Y_k = \ln S_{k+1} - \ln S_k$, the *log*-returns of the stock prices. A standard stochastic volatility model (SVM) is given by

$$Y_k = \sigma(X_k)\eta_k \quad \in \mathbb{R} \quad \text{with} \quad X_k = \rho\, X_{k-1} + \varepsilon_{k-1} \quad \in \mathbb{R} \tag{4.17}$$

where ρ is a real constant, $\sigma(.)$ is a positive Borel function on $\mathbb{R}$ and $(\varepsilon_k)_k$, $(\eta_k)_k$ are independent Gaussian processes. We consider dynamics (4.17) as a time discretization Euler scheme with step size $\Delta t = 1/n$, of a continuous-time Ornstein-Uhlenbeck stochastic volatility model :

$$dX_t \;=\; -\lambda X_t dt + \tau dW_t, \quad 0 \le t \le 1.$$

We then suppose

$$\rho = 1 - \lambda\, \Delta t, \quad \varepsilon_k \rightsquigarrow \mathcal{N}(0, \tau^2 \Delta t) \quad \text{and} \quad \eta_k \rightsquigarrow \mathcal{N}(0, \Delta t),$$

for some positive parameters λ and τ. Typical examples of SVM are specified with $\sigma(x) = |x| + \gamma$, $\sigma(x) = x^2 + \gamma$, or $\sigma(x) = \exp(x)$ for some positive constant γ. The filtering problem consists in estimating the volatility $\sigma(X_n)$ at step n given the observations of the prices $(Y_0, \ldots, Y_n)$. Here,

$$g(x, y) \;=\; \frac{1}{\sqrt{2\pi\Delta t}\,\sigma(x)} \exp\left(-\frac{y^2}{2\,\sigma^2(x)\Delta t}\right).$$

The values of the parameters in our simulation are for $(\lambda, \tau, \Delta t) = (1, 0.5, 1/250)$. The Gaussian distribution of X_0 is specified so that the sequence $(X_k)_k$ is stationary *i.e.* $X_0 \sim \mathcal{N}(0, \Sigma_0^2)$ with $\Sigma_0 = \tau\sqrt{\Delta t/(1-\rho^2)} \approx \tau/\sqrt{2\,\lambda} = 0.35\ldots$

There are two types of models involved here:

$$(ABS) \equiv \sigma(X_k) = \gamma + |X_k| \quad \text{and} \quad (EXP) \equiv \sigma(X_k) = \overline{\sigma}\exp(X_k),$$

with the values $(\gamma, \overline{\sigma}) = (0.05, 0.2)$.

We represent in Figure 13 the stock price simulation according (EXP) together with the simulation of the volatility σ_n and its mean conditionally to Y_k, $0 \le k \le n = 250$. Idem in Figure 14 with (ABS).

We represent in Figure 15, the conditional variance of the volatility w.r.t. the observations in the two models. Since we are here in a nonlinear model, we cannot compare our results with an explicit filter, but we can see that the filter captures well the dynamic of the stochastic volatility.

5 Toward higher order schemes in quantization methods

The aim of this section is to present in a slightly different setting the first order scheme introduced in [6] and successfully tested on the pricing of Exchange options in a d-dimensional Black& Scholes model ($d = 2$, 4 and 6). One comes back to the expectation computation

$$\mathbb{E}\left[\phi_1(X_1)\dots\phi_n(X_n)\right]$$

along the path of a Markov chain investigated in Section 3. The idea is to try taking advantage of the specificity of the stationary quantizers, like for the numerical integration of smooth functions (see Section 2.3).

We deal here with marginal quantization and the approach is partially heuristic. However, to enhance the quantization aspects we will essentially focus on a smooth setting where the ϕ_k functions are smooth, say C_b^2 (twice differentiable with bounded existing derivatives). We will shortly comment below how it can be somewhat relaxed.

Assume that $(X_k)_{0\le k\le n}$ is a homogeneous Markov chain with a transition $P(x, dy)$ satisfying on $\mathcal{C}_b^2$ functions:

$$\|D(Pf)\|_\infty \le K\|Df\|_\infty \qquad \text{and} \qquad \|D^2(Pf)\|_\infty \le K(\|Df\|_\infty + \|D^2 f\|_\infty) \tag{5.1}$$

for some real constant $K > 0$ (note that then P is Lipschitz with constant K as well). Such an assumption is satisfied *e.g.* by the transition P_t of a (simulatable) diffusion having C_b^2 coefficients b and σ at any time t or by the transition of its Euler scheme. Then, the transition P clearly maps C_b^2 into itself and one shows by induction that the functions v_k defined by (3.2) all lie in C_b^2 and that their first two derivatives can be controlled using K, $\|D\phi_k\|_\infty$ and $\|D^2\phi_k\|_\infty$.

The key result to design a first order scheme is the following proposition.

Proposition 5.1 *Assume that Assumption* (5.1) *and holds and set by induction*

$$\widetilde{v}_n(\widehat{X}_n) = \phi_n(\widehat{X}_n), \tag{5.2}$$

$$\widetilde{v}_{k-1}(\widehat{X}_{k-1}) = \phi_{k-1}(\widehat{X}_{k-1})\mathbb{E}\left(\widetilde{v}_k(\widehat{X}_k) + Dv_k(\widehat{X}_k).(X_k - \widehat{X}_k) \,|\, \widehat{X}_{k-1}\right),\ 1 \le k \le n. \tag{5.3}$$

Then for every $k \in \{0, \dots, n\}$,

$$\left\| \mathbb{E}(v_k(X_k) \,|\, \widehat{X}_k) - \widetilde{v}_k(\widehat{X}_k) \right\|_1 \;\leq\; L^{n+1-k} \sum_{\ell=k}^{n} c_\ell^n(K) \|X_\ell - \widehat{X}_\ell\|_2^2 \tag{5.4}$$

where $$c_\ell^n(K) = \begin{cases} \dfrac{1}{K-1}\left(K^{n-\ell+1}\left(2(n-\ell) - \dfrac{2}{K-1}\right) + \dfrac{1}{2}\dfrac{K+3}{K-1}\right) & \text{if } K \neq 1 \\ 6(n-\ell+1)(n-\ell+3/2) & \text{if } K = 1 \end{cases}$$

and $\max_{1\leq k\leq n} (\|\phi_k\|_\infty, \|D\phi_k\|_\infty, \|D^2\phi_k\|_\infty) \leq L$. *In particular,*

$$|\mathbb{E}\, v_0(X_0) - \mathbb{E}\, \widetilde{v}_0(\widehat{X}_0)| \leq L^{n+1} \sum_{\ell=0}^{n} c_\ell^n(K) \|X_\ell - \widehat{X}_\ell\|_2^2.$$

How to use this result to design a first order scheme? First, one reads (5.3) in distribution *i.e.*

$$\widetilde{v}_{k-1}(x_{k-1}^i) = \phi_{k-1}(x_{k-1}^i) \left(\sum_{j=1}^{N_k} \hat{p}_k^{ij} \widetilde{v}_k(x_k^j) + Dv_k(x_k^j).\hat{\chi}_k^{ij} \right)$$

where the $\mathbb{R}^d$-valued *correcting vectors* $\hat{\chi}_k^{ij}$ are defined by

$$\hat{\chi}_k^{ij} \;:=\; \mathbb{E}\left[(X_k - \widehat{X}_k) \mathbf{1}_{\{\widehat{X}_k = x_k^j\}} \,|\, \widehat{X}_{k-1} = x_{k-1}^i \right]. \tag{5.5}$$

The key point for numerical application is to note that these correcting vectors can easily be estimated like the former companion parameters $\hat{p}_k^{ij}$'s, either on line during the grid optimization phase or using a Monte Carlo simulation once the grids are settled.

The second step is mostly heuristic so far: the weak link in (5.3) is of course that the differential Dv_k is involved in the computation of $\widetilde{v}_{k-1}$ and this function is not numerically accessible since we precisely intend approximating the functions v_k. Note that if Dv_{k-1} had been involved in (5.3), the scheme would have been definitely intractable. In its present form, many approaches are possible.

It often happens, *e.g.* for diffusions or Euler schemes with smooth coefficients, that $D(Pf) = Q(Df)$ where Q is a simulatable integral kernel as well. So one can rely on the backward induction formula

$$Dv_k = D\phi_k\, Pv_{k+1} + \phi_k\, Q(Dv_{k+1})$$

to approximate the differentials Dv_k using quantization. Another approach is to use some approximation by convolution: one approximates Dv_k by $(D\varphi_\varepsilon) * \widetilde{v}_k$ where $(\varphi_\varepsilon)_{\varepsilon>0}$ is *e.g.* a Gaussian unit approximation. The practical task is to tune the band width ε.

When the functions ϕ_k are not smooth enough, one uses the regularizing properties of the Markov semi-group if some. When dealing with diffusions, this is when Malliavin Calculus and the Skorohod integrals come in the game like in [6]. However, some loss in the rate of convergence is to be expected due to the singularity near the maturity n.

We present in Figure 16 a graph that emphasizes the improvement provided by this first order *quantization tree descent* versus the original one for pricing 6-dimensional American

exchange options, as described in paragraph 4.1. We consider an "in-the-money" case consisting in setting $x_0^1 \dots x_0^3 = 40$ and $x_0^4 \dots x_0^6 = 36$, all other parameters being unchanged. For more details we refer to [6].

ACKNOWLEDGMENT : We thank Harald Luschgy for fruitful comments.

References

[1] V. Bally, Pagès G. (2000): A quantization algorithm for solving discrete time multi-dimensional optimal stopping problems, forthcoming in *Bernoulli*,

[2] Bally V., Pagès G. (2001): Error analysis of the quantization algorithm for obstacle problems, forthcoming in *Stochastic Processes and their Applications*,

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

[4] Bally V., Pagès G., Printems J. (2001): A quantization method for pricing and hedging multi-dimensional American style options, Preprint $n^0$753, Laboratoire de Probabilités & modèles aléatoires, Université Paris VI (France).

[5] Bally V. (2002): The Central Limit Theorem for a non linear algorithm based on quantization, forthcoming in *Proceedings of the Royal Society.*

[6] Bally V., Pagès G., Printems J. (2003): First order schemes in the numerical quantization method, *Mathematical Finance*, **13**, n^0 1, pp.1–16.

[7] Barles G., Souganidis P. (1991): Convergence of approximation schemes for fully nonlinear second-order equations, *Asymptotics Analysis*, **4**, pp.271-283.

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

[9] Duflo, M. (1997): *Random Iterative Models*, Coll. Applications of Mathematics, **34**, Springer-Verlag, Berlin, 1997, 385p.

[10] Elliott R., Aggoun L. and J. Moore (1995): *Hidden Markov Models, Estimation and Control*, Springer Verlag, 361 pp.

[11] Fort J.C., Pagès G. (2002): Asymptotics of optimal quantizers for some scalar distributions, *Journal of Comput. and Applied Mathematics*, **146**, pp.253-275.

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

[13] Graf S., Luschgy H. (2000): *Foundations of Quantization for Probability Distributions*, Lecture Notes in Mathematics n^{0}1730, Springer, Berlin, 230 pp.

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

[15] Kohonen T. (1982): Analysis of simple self-organizing process, *Biological Cybernetics*, **44**, pp. 135–140.

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

[17] Kushner H.J., Yin G.G. (1997): *Stochastic Approximation Algorithms and Applications*, Springer, New York.

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

[19] Pagès G., Pham H. (2001): A quantization algorithm for multidimensional stochastic control problems, Preprint PMA-697, Laboratoire de Probabilités et modèles aléatoires, Universités Paris 6 & 7 (France).

[20] Pagès G., Pham H. (2002): Optimal quantization methods for nonlinear filtering with discrete-time observations, Preprint PMA-778, Laboratoire de Probabilités et modèles aléatoires, Universités Paris 6 & 7 (France).

[21] Pagès G., Printems J.,(2002): Optimal quadratic quantization for numerics: the Gaussian case, Preprint PMA-809, Laboratoire de Probabilités et modèles aléatoires, Universités Paris 6 & 7 (France).

[22] Villeneuve S., Zanette A. (2002) Parabolic A.D.I. methods for pricing american option on two stocks, *Mathematics of Operation Research*, **27**, $n^0 1$, pp.121-149.

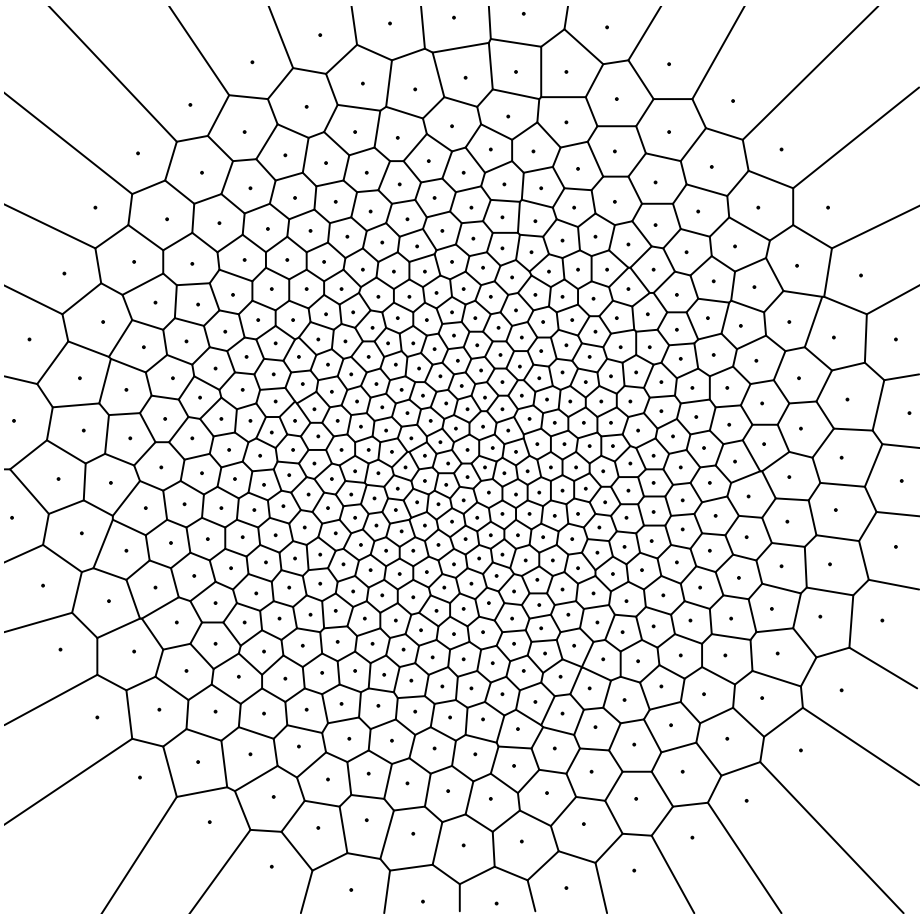


Figure 1: A L^2-optimal 500-quantization grid for the Normal distribution $\mathcal{N}(0; I_2)$ with its Voronoi tessellation.

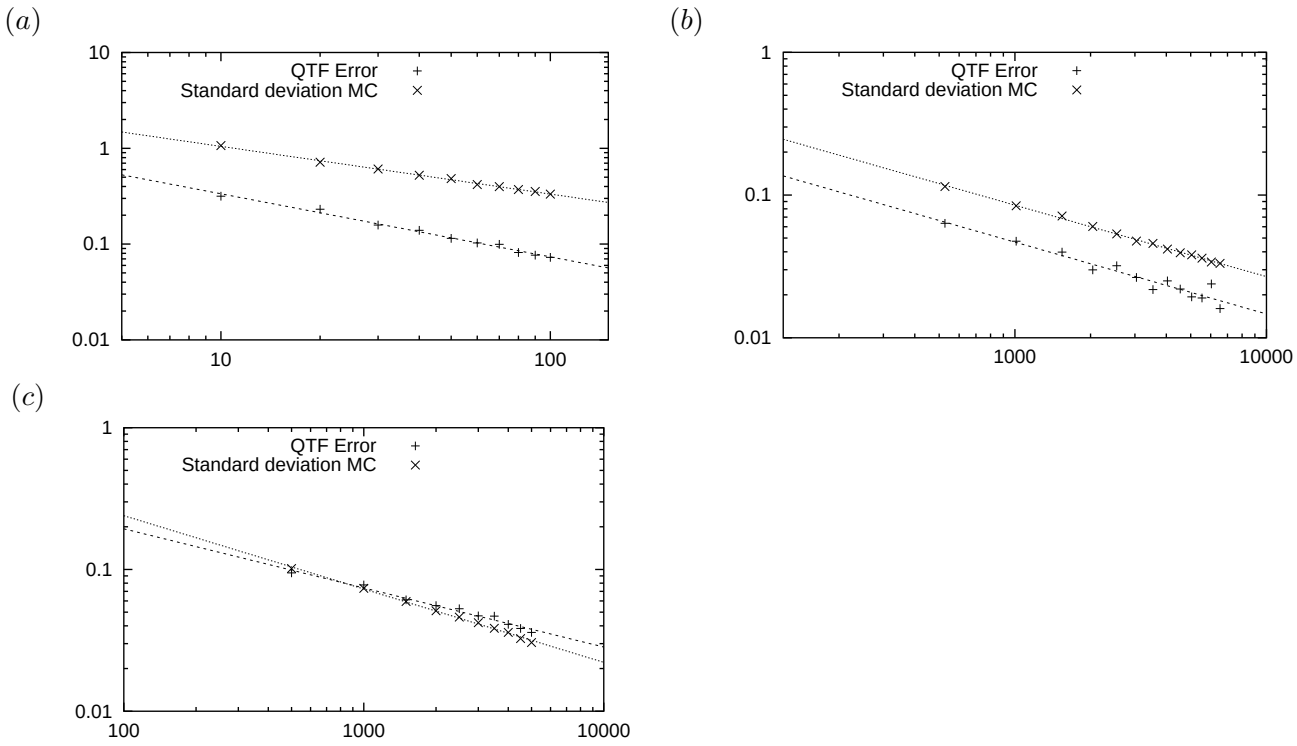


Figure 2: Linear regression in log-log scale of $N \mapsto |\mathbb{E}g_2(\widehat{Z}) - \widehat{\mathbb{E}g_2(Z)}_N|$. In a) $d = 3$; b) $d = 4$; c) $d = 5$.

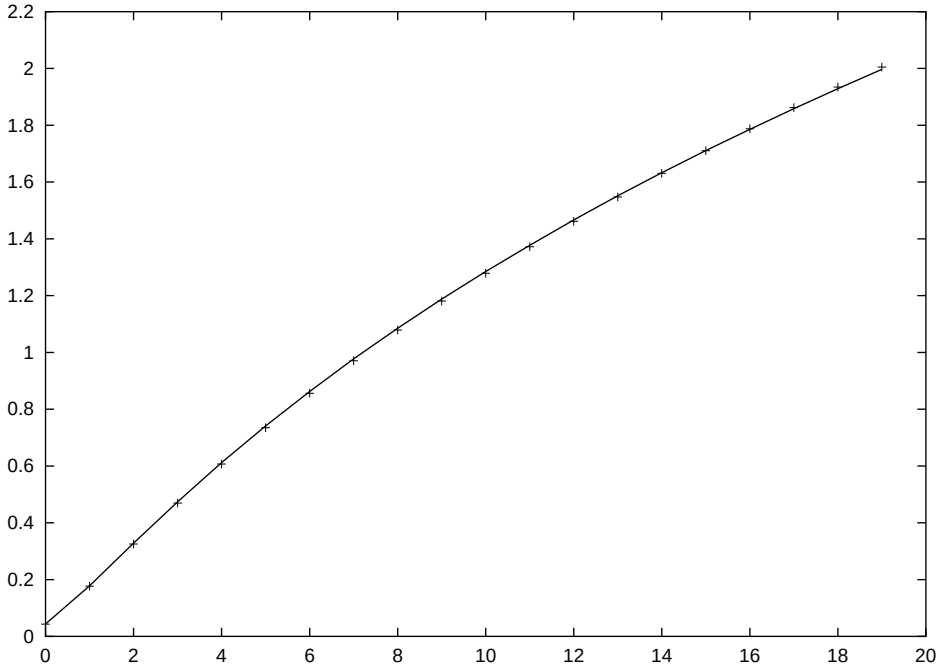


Figure 3: American exchange option in dimension 4 (out-of-the-money case). The reference price is depicted by a line and the quantized price by the cross line.

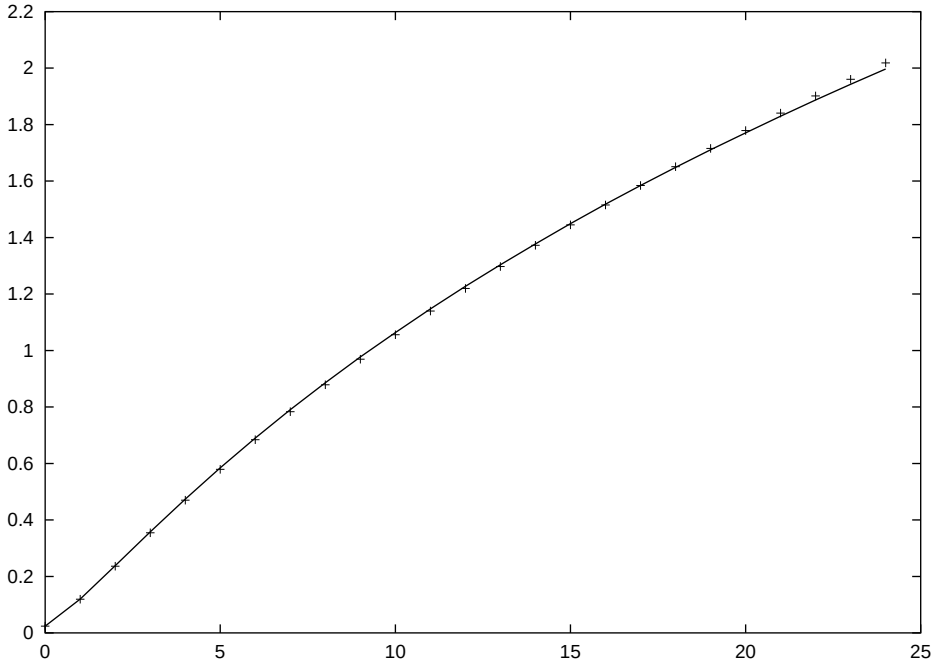


Figure 4: American exchange option in dimension 6 (out-of-the-money case). The reference price is depicted by a line and the quantized price by the cross line.

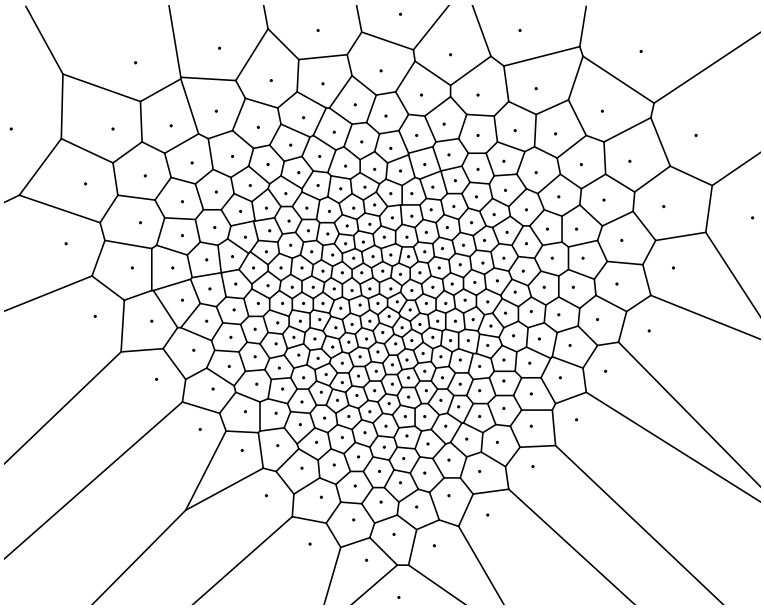


Figure 5: L^2-optimal 300-quantizer of $(\log(X_T^1), X_T^2)$ with its Voronoi tessellation.

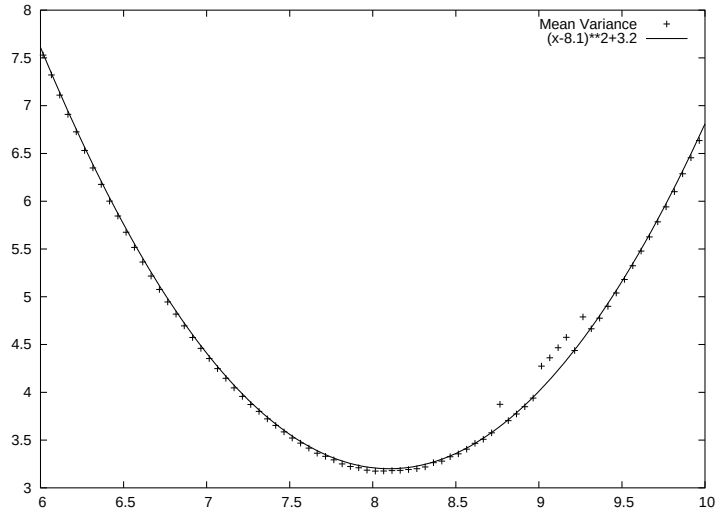


Figure 6: Graph of $y \mapsto \mathbb{E}[\widehat{v}_0(y, x_0^1, \widehat{X}_0^2)]$ (dot line), closest parabola (thick line).

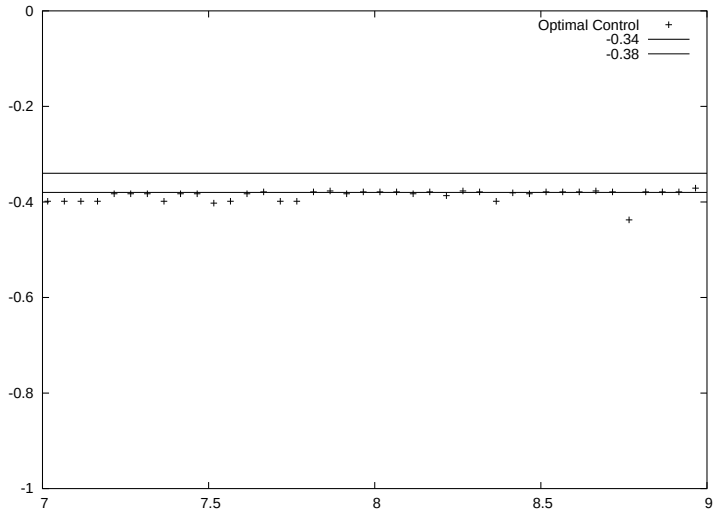


Figure 7: Quantized optimal control (dot line), Optimal control computed by a Monte Carlo simulation.

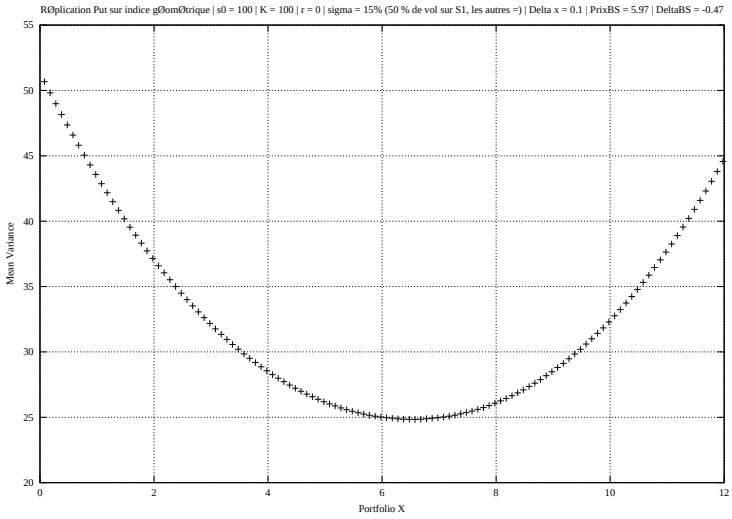


Figure 8: Graph of $x_0 \mapsto v_0(x_0, s_0)$ (dot line), closest parabola (thick line).

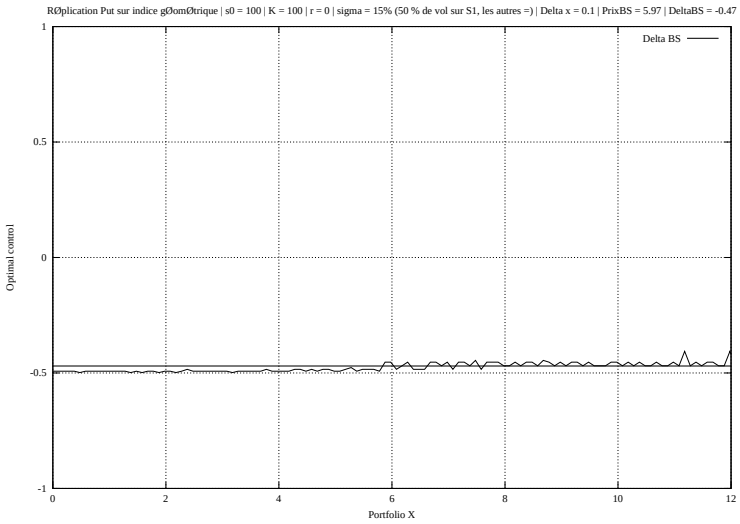


Figure 9: Quantized optimal control with theoretical optimal control (straight line).

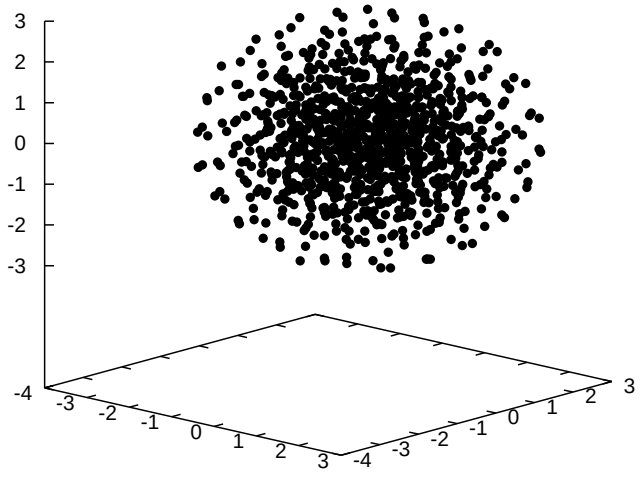


Figure 10: A L^2 optimal 1 000-quantizer of $\mathcal{N}(0, I_3)$. L^2-quantization error is equal to 0.233.

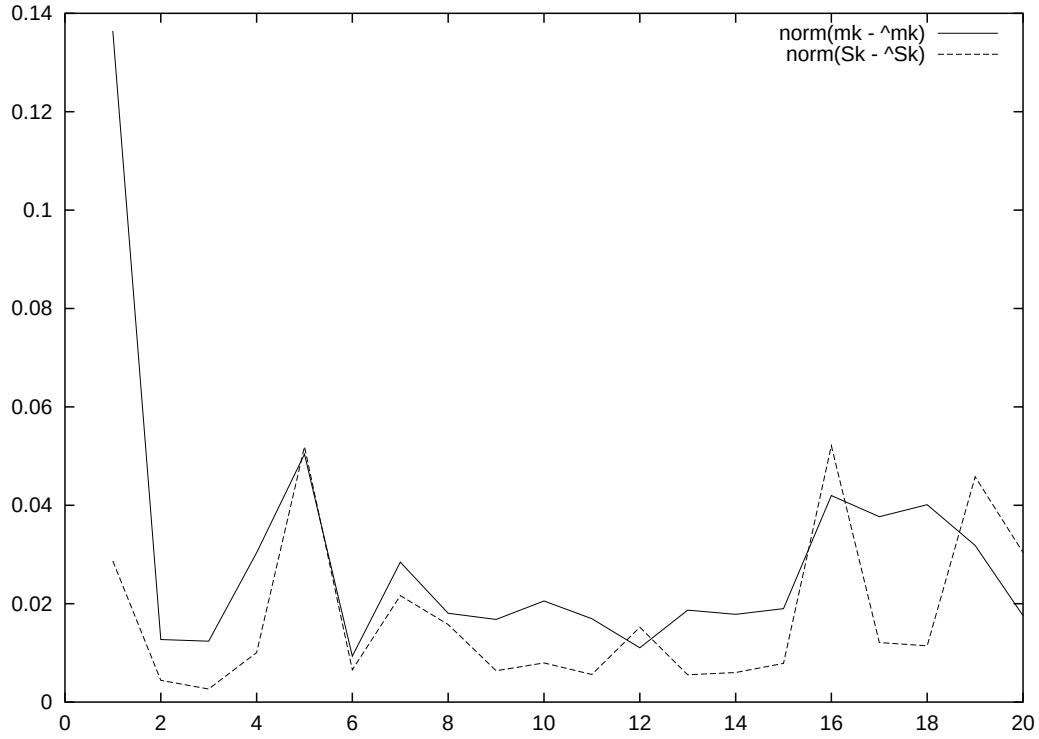
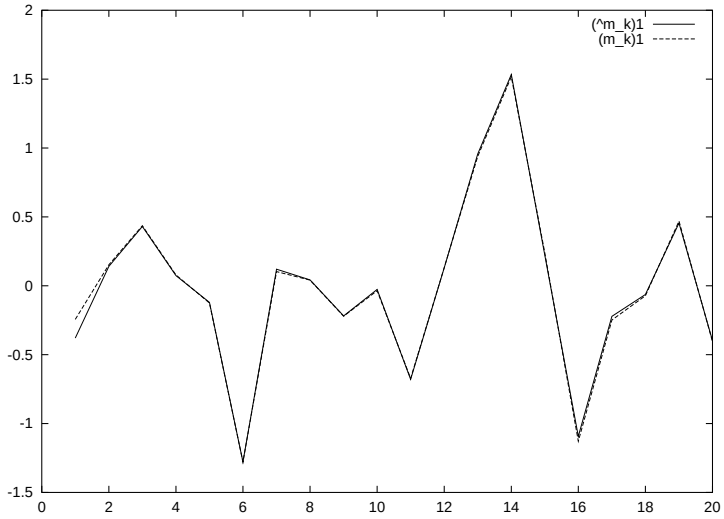
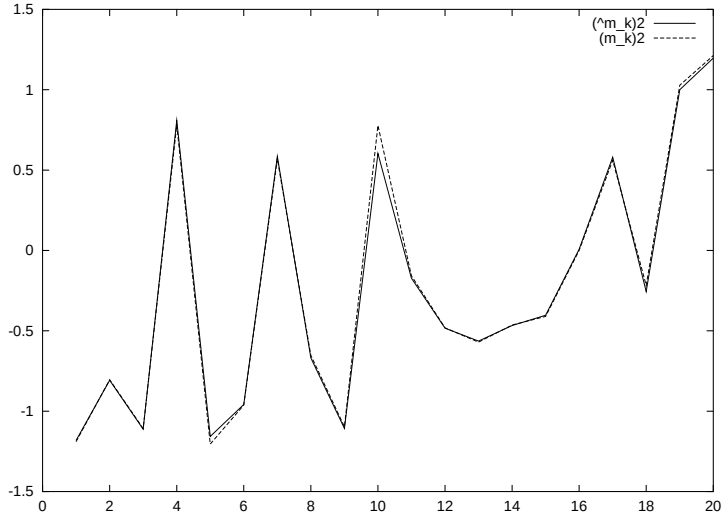


Figure 11: Errors $\|m_n - \widehat{m_n}\|$ (line) and $\|\Sigma_n - \widehat{\Sigma_n}\|$ (dot line) plotted w.r.t. n.

(a)



(b)



(c)

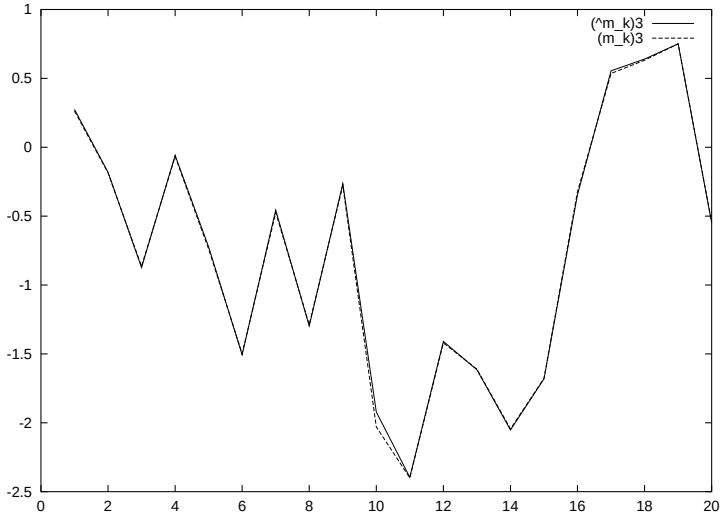
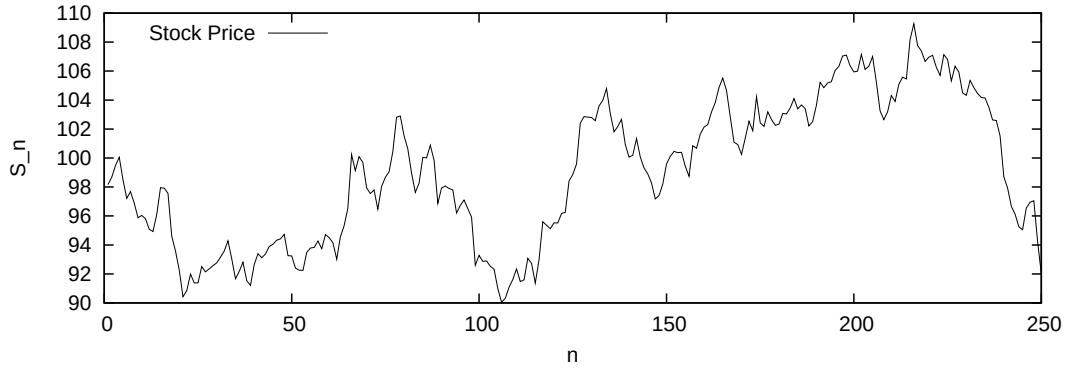


Figure 12: Components of $\widehat{m_k}$ (Quantized filter) (line)nd of m_k (Kalman filter) (dot line). (a) x, (b) y and (c) z.

a)



b)

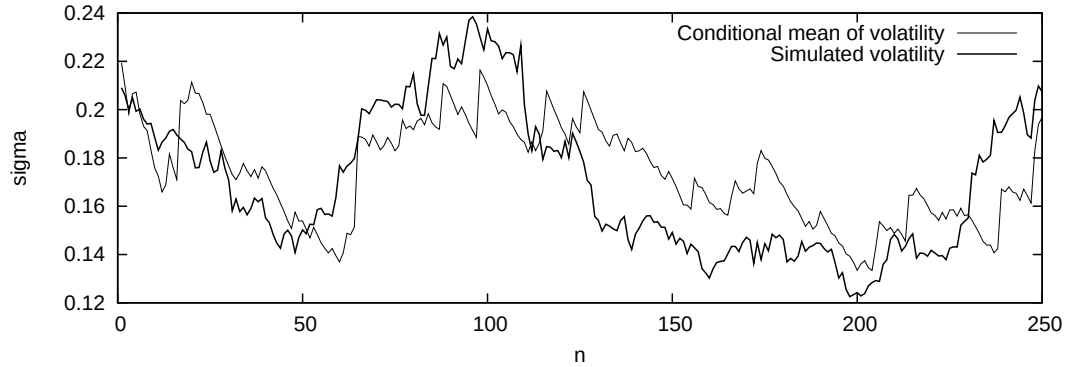
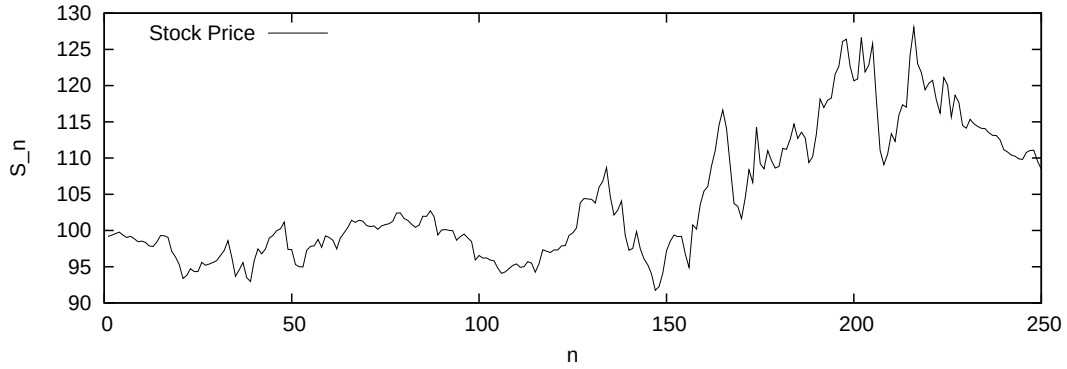


Figure 13: a) Stock price simulated according to (EXP). b) Simulated volatility according to (EXP) (Thick line), Conditional mean of the volatility (Thin line).

a)



b)

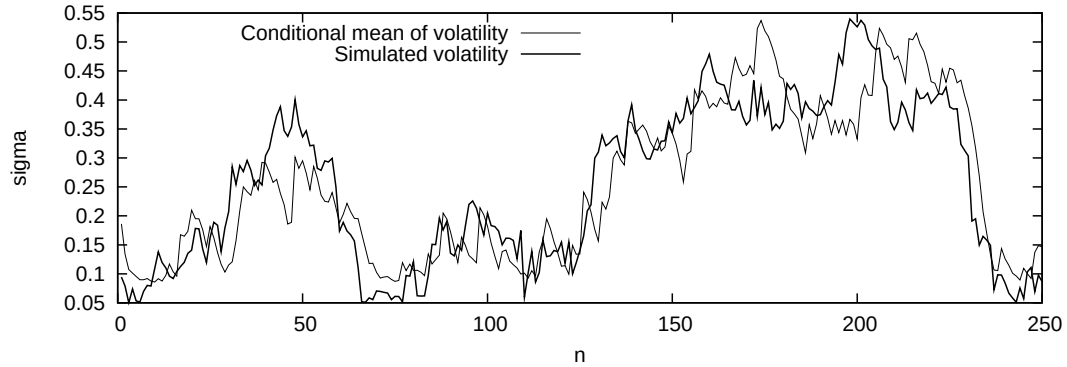
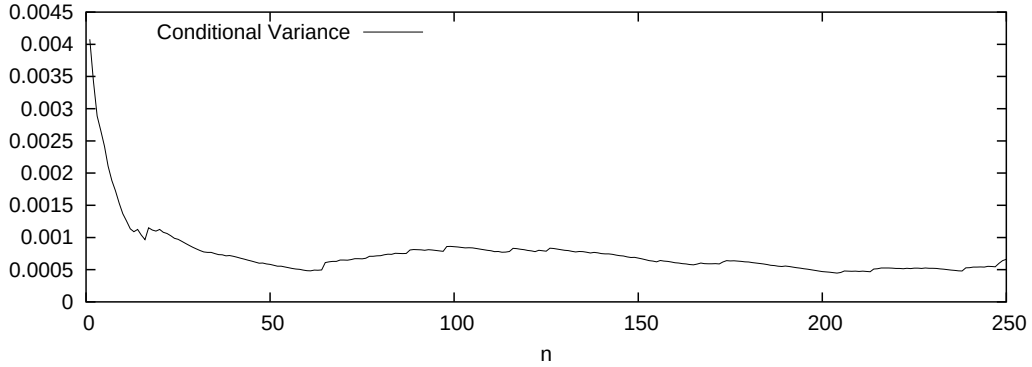


Figure 14: a) Stock price simulated according to (ABS). b) Simulated volatility according to (ABS) (Thick line), Conditional mean of the volatility (Thin line).

a)



b)

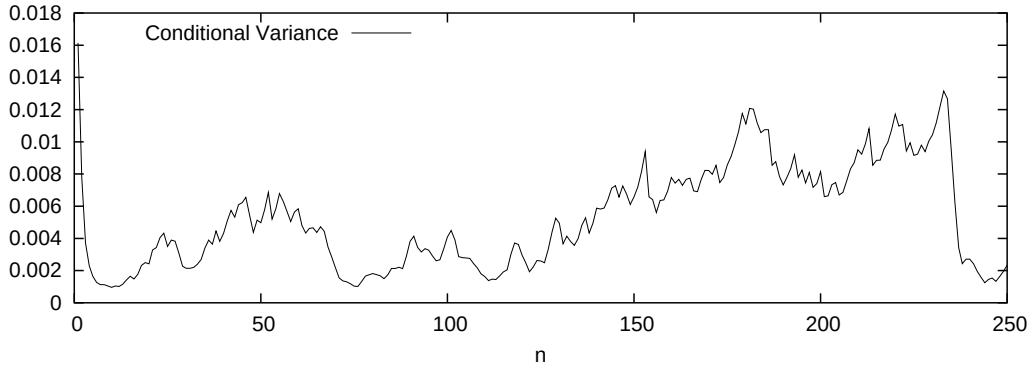


Figure 15: a) Conditional Variance of the Volatility according to (EXP). b) Conditional Variance of the Volatility (ABS).

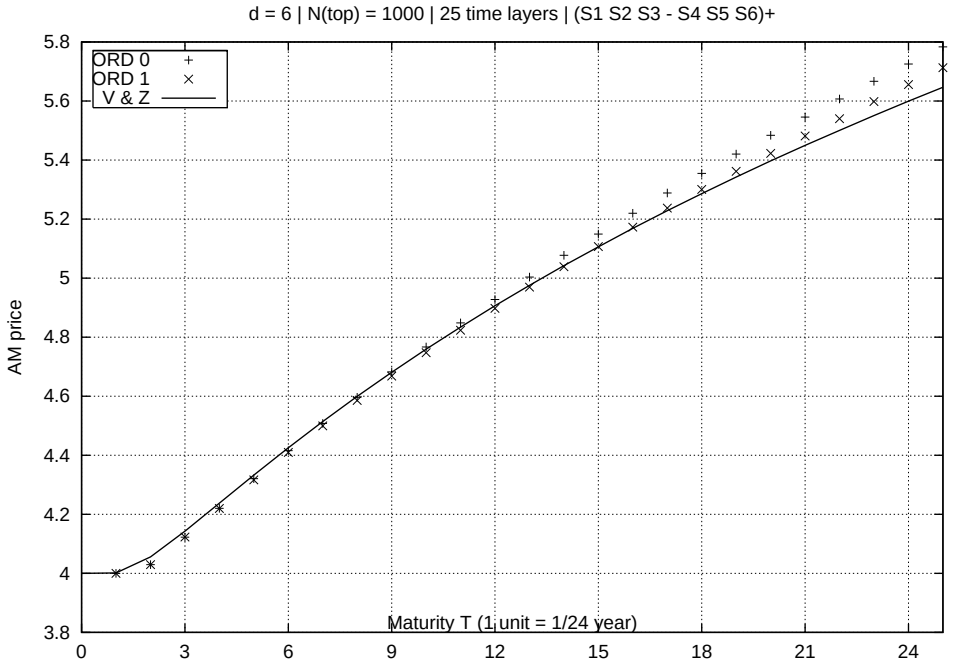


Figure 16: American exchange option in dimension 6 (*in-the-money case*). The reference price is depicted by a line and the quantized prices (order 0 and order 1) by cross lines.