



HAL
open science

A quantization tree algorithm: improvements and financial applications for swing options

Anne Laure Bronstein, Gilles Pagès, Benedikt Wilbertz

► **To cite this version:**

Anne Laure Bronstein, Gilles Pagès, Benedikt Wilbertz. A quantization tree algorithm: improvements and financial applications for swing options. 2008. hal-00273790v1

HAL Id: hal-00273790

<https://hal.science/hal-00273790v1>

Preprint submitted on 16 Apr 2008 (v1), last revised 15 Dec 2009 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A quantization tree algorithm: improvements and financial applications for swing options

ANNE LAURE BRONSTEIN ^{*}, GILLES PAGÈS [†] and BENEDIKT WILBERTZ [‡]

April 16, 2008

Abstract

In this paper, we suggest several improvements to the numerical implementation of the quantization method in order to get accurate premium estimations. This technique is applied to derivative pricing in energy markets. Several ways of modeling energy derivatives are described and finally numerical examples are provided to test the procedure accuracy.

Keywords: Quantization, Gaussian process, Lévy process, optimal quantizer.

MSC: 60G15, 60E99.

1 Introduction

In the last decades, optimal quantization arised as an efficient method for pricing complex financial contracts. It has successfully been applied to basket options of american type (see [BPP05]) and energy contracts as swing options (see [BBP07a] and [BBP07b]). In financial institutions, quickness of execution as well as high accuracy are important criteria in the choice of a pricing method. With this observation in mind, we suggest some improvements to the original quantization method. The quantization tree algorithm (or pricing procedure) is divided into three parts: the computation of the quantization grids, the estimation of the transition probabilities and the premium evaluation. As the two first tasks are time consuming, one usually proceeds off-line to these estimations. To reduce drastically the computation time and proceed on-line to these estimations, we suggest the application of a “fast parallel quantization method”, which allows for a parallel implementation of our procedure on a grid. So, we increase enormously the number of processors available to do the computations. We also advice to include a *kd*-tree structure in the fast parallel procedure to diminish complexity. Finally, we appeal to a spatial Richardson-Romberg extrapolation method to improve the convergence rate and the method accuracy.

This improved approach is then applied to swing options pricing. We introduce several models for the underlying asset dynamics. As well as usual Gaussian processes, we consider dynamics driven by Normal Inverse Gaussian (NIG) distribution. The Normal Inverse Gaussian distribution has been introduced in finance by Barndorff-Nielsen in the 90s and has recently been applied to

^{*}Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Paris 6, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5. E-mail: alb@proba.jussieu.fr

[†]Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Paris 6, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5. E-mail: gilles.pages@upmc.fr

[‡]Universität Trier, FB IV-Mathematik, D-54286 Trier, Germany. E-mail: wilbertz@uni-trier.de

energy markets by several authors, e.g. Frestad, Benth and Koekebakker and Benth and Saltyte-Benth (see [BFK07] and [BSB04]). They suggest that the NIG family fits empirical electricity return distribution very well and represents an attractive alternative to the family of normal distribution.

In Section 2, the quantization method is introduced. In Section 3, we recall the quantization tree algorithm associated to the swing option stochastic control problem (see [BBP07a] and [BBP07b]). Some special cases of swing option as Call Strip and Bermuda option are developed. In Section 4, some suggestions are addressed to improve the execution time as well as the accuracy of the algorithm. Section 5 is devoted to financial applications and numerical results.

2 Optimal quantization

Optimal quantization has been developed in the 50s in the field of Signal Processing. Its main purpose consists in approximating a continuous signal by a discrete one on an optimal way. In the 90s, its application has been extended to the field of Numerical Integration to compute some integral estimations by using finite weighted sums. And in the early 2000s, this method has been applied to the field of Numerical Probabilities and Financial Mathematics. This extension has been motivated by the necessity of designing efficient methodologies for pricing and hedging more and more sophisticated financial products. Indeed optimal quantization brought a natural answer to the conditional expectation computations appearing in these financial models.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a given probability space and let X be a random vector defined on this probability space and taking valued in \mathbb{R}^d . Let N be a positive integer, the main idea of quantization consists to discretize X by a $\sigma(X)$ -measurable random vector \hat{X} taking only finitely many values in a grid $\Gamma = \{x_1, \dots, x_N\}$ of \mathbb{R}^d . This grid or codebook is called a N -quantizer of X . Let $x = (x_1, \dots, x_N)$ denotes the N -tuple induced by Γ . One can associate with \hat{X} a Borel function $q_x : \mathbb{R}^d \rightarrow \mathbb{R}^d$ called quantizer such that $\hat{X} = q_x(X)$.

Let $p \in [1, \infty[$ and $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$, optimal quantization consists to study the best L^p -approximation of X . It aims at minimising in q_x the L^p -mean quantization error, that is, it consists in solving

$$\inf\{\|X - q_x(X)\|_p \mid q_x : \mathbb{R}^d \rightarrow \mathbb{R}^d \text{ Borel and } |q_x(X)| \leq N\}. \quad (1)$$

The quantization \hat{X}^x relative to the quantizer function is given by the following definition.

Definition 1. Let $x = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$. A partition $(C_i(x))_{1 \leq i \leq N}$ of \mathbb{R}^d is a Voronoi tessellation of the N -quantizer x , if for every $i \in \{1, \dots, N\}$, $C_i(x)$ is a Borel set satisfying

$$C_i(x) \subset \{u \in \mathbb{R}^d \mid |u - x_i| = \min_{1 \leq j \leq N} |u - x_j|\}, \quad (2)$$

where $|\cdot|$ is the Euclidean norm on \mathbb{R}^d . The nearest neighbor projection on x induced by the Voronoi partition is defined for every $u \in \mathbb{R}^d$ by

$$u \mapsto \sum_{i=1}^N x_i 1_{C_i(x)}(u). \quad (3)$$

The random vector

$$\hat{X}^x = \sum_{i=1}^N x_i 1_{C_i(x)}(X) \quad (4)$$

is called a Voronoi quantization of X .

The minimization problem in x is more subtle and depend on p and some properties of the probabilities distribution. It has at least one solution x^* . And if $\text{supp } \mathbb{P}_X = \infty$ then x^* has pairwise distinct components and $\min_{x \in (\mathbb{R}^d)^N} \|X - \hat{X}^x\|_p$ is decreasing to 0 as N goes to ∞ .

The rate of convergence to zero is given by Zador's Theorem (see [Z82], [GG92], [GL00]).

Theorem 1. (i) Let $X \in L^{p+\eta}(\Omega, \mathcal{A}, \mathbb{P})$, $p \geq 1$, $\eta > 0$, such that $\mathbb{P}_X(du) = \varphi(u)\lambda_d(du) + \nu(du)$, where $\nu \perp \lambda_d$. Then

$$\lim_N N^{\frac{1}{d}} \min_{x \in (\mathbb{R}^d)^N} \|X - \hat{X}^x\|_p = \tilde{J}_{p,d} \left(\int_{\mathbb{R}^d} \varphi^{\frac{d}{p+d}}(u) du \right)^{\frac{1}{p} + \frac{1}{d}}.$$

(ii) Non asymptotic estimate: Let $p \geq 1, \eta > 0$. There exists a real constant $C_{d,p,\eta} > 0$ and an integer $N_{d,p,\eta} \geq 1$ such that for any \mathbb{R}^d -valued random vector X , for $N \geq N_{d,p,\eta}$,

$$\min_{x \in (\mathbb{R}^d)^N} \|X - \hat{X}^x\|_p \leq C_{d,p,\eta} \|X\|_{p+\eta} N^{-\frac{1}{d}}.$$

We refer to the following papers for a study of several algorithms designed to find optimal quantizers: (see [K82, GG92, P97, D98]). The problem dimension and the properties of the law of X might help to determine an efficient algorithm. Note that some optimised quantizers of the normal distribution $\mathcal{N}(0, I_d)$ have been computed and are available at the URL

www.quantize.math-fi.com

for dimensions up to 10 and for several grid sizes.

Since an optimal quantization \hat{X}^x provides the best finite approximation to the distribution of X in the least square sense, it becomes natural to use $\mathbb{E}f(\hat{X}^x)$ as an approximation for $\mathbb{E}f(X)$, where $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is a Borel function.

Note further, that since \hat{X}^x takes only finitely many values, we compute $\mathbb{E}f(\hat{X}^x)$ as the finite sum

$$\sum_{i=1}^N \mathbb{P}(X \in C_i(x)) f(x_i).$$

The weights $p_i := \mathbb{P}(X \in C_i(x))$ of this cubature formula are obtained as a by-product of optimization procedures to generate optimal quantizers like the ones used in [PP03] or [P97].

Assume now that f exhibits some smoothness properties, i.e. f is differentiable with Lipschitz continuous differential Df . If x is a stationary quantizer, i.e.

$$\hat{X}^x = \mathbb{E}(X | \hat{X}^x),$$

which is fulfilled for every optimal L^2 -quantizer, we can conclude from a Taylor expansion of f , that the approximation error of $\mathbb{E}f(\hat{X}^x)$ satisfies

$$|\mathbb{E}f(X) - \mathbb{E}f(\hat{X}^x)| \leq [Df]_{\text{Lip}} \|X - \hat{X}^x\|_2^2.$$

3 The quantization tree algorithm

3.1 Abstract problem formulation

We consider several financial products and we define an abstract model, which will be applied to these derivative products.

Let $(t_k)_{0 \leq k \leq n-1}$ be a given, strictly increasing sequence of exercise dates, such that the first exercise date is set at the origin of the derivative and the last one happens strictly before the option maturity.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a complete probability space and let $p \in [1, \infty[$. On this probability space, one defines an option with maturity T by a sequence of random variables $(V_k)_{0 \leq k \leq n-1}$. For $k \in \{0, \dots, n-1\}$, V_k represents the option payoff obtained by the derivative holder at time t_k . To model all the available market information, one introduces a filtration $\mathcal{F} := (\mathcal{F}_k)_{0 \leq k \leq n-1}$ on $(\Omega, \mathcal{A}, \mathbb{P})$, such that, for $k \in \{0, \dots, n-1\}$, the option payoff V_k is measurable with respect to the σ -field \mathcal{F}_k . One common feature characterizing most of the option is that the product is built on an underlying asset. In our setting, the option owner has the possibility to choose, for each exercise date, the underlying asset amount included in the contract. These decision variables are represented by a sequence $(q_k)_{0 \leq k \leq n-1}$ of \mathcal{F} -adapted random variables and are subjected to constraints given by:

(i) a local constraint:

let q_{\min} and q_{\max} be two given positive constants. For $k = 0, \dots, n-1$,

$$q_{\min} \leq q_k \leq q_{\max} \quad \mathbb{P}\text{-a.s.} \quad (5)$$

(ii) A global constraint:

one defines a cumulative purchased process $(\bar{q}_k)_{0 \leq k \leq n}$ by

$$\bar{q}_0 := 0 \quad \text{and} \quad \bar{q}_k := \sum_{l=0}^{k-1} q_l, \quad k = 1, \dots, n.$$

Then for any couple of \mathcal{F}_0 -measurable, non-negative random variables $Q := (Q_{\min}, Q_{\max})$, the total cumulative volume purchased should satisfy

$$Q_{\min} \leq \bar{q}_n \leq Q_{\max} \quad \mathbb{P}\text{-a.s.} \quad (6)$$

A control process $(q_k)_{0 \leq k \leq n-1}$ which satisfies these local and global conditions is called (\mathcal{F}, Q) -admissible.

Now, one defines the residual global constraints $Q^k = (Q_{\min}^k, Q_{\max}^k)$ at time t_k by

$$Q_{\min}^k = Q_{\min} - \bar{q}_k \quad \text{and} \quad Q_{\max}^k = Q_{\max} - \bar{q}_k.$$

Then, the associated option value at time t_k is given by

$$P_k^n((Q_{\min}^k, Q_{\max}^k)) := \text{esssup} \left\{ \mathbb{E} \left(\sum_{l=k}^{n-1} q_l V_l \middle| \mathcal{F}_k \right), q_l : (\Omega, \mathcal{F}_l) \rightarrow [q_{\min}, q_{\max}], k \leq l \leq n-1, \bar{q}_n - \bar{q}_k \in [Q_{\min}^k, Q_{\max}^k] \right\}. \quad (7)$$

Now, to simplify the numerical computations, one appeals to a decomposition method introduced by Bardou, Bouthemy and Pagès in ([BBP07a]) and ([BBP07b]). At a given time t_k , the option price is split between a swap contract with a closed form solution and a normalized swing one, in which the control variables are $[0, 1]$ -valued. One considers from now on a normalized framework. In this setting, a couple of global constraints Q^k at time t_k is admissible if Q^k is \mathcal{F}_k -measurable and if

$$0 \leq Q_{\min}^k \leq Q_{\max}^k \leq n - k \quad \mathbb{P}\text{-a.s.} \quad (8)$$

For every couple of admissible global constraints Q^k , the derivative price at time t_k satisfies

$$P_k^n((Q_{\min}^k, Q_{\max}^k)) := \text{esssup} \left\{ \mathbb{E} \left(\sum_{l=k}^{n-1} q_l V_l | \mathcal{F}_k \right), q_l : (\Omega, \mathcal{F}_l) \rightarrow [0, 1], k \leq l \leq n-1, \bar{q}_n - \bar{q}_k \in [Q_{\min}^k, Q_{\max}^k] \right\}. \quad (9)$$

To construct the quantization tree algorithm, we appeal to several results summarized below, (a detailed version of these results is available in [BBP07b]):

(i) Bardou, Bouthemy and Pagès show that we loose little generality and gain much simplification by assuming that at time t_k , all couple of admissible global constraints Q^k are deterministic.

(ii) The Backward Dynamic Programming Formula (BDP):

Let $P_n^n \equiv 0$. For every $k \in \{0, \dots, n-1\}$ and every couple of admissible global constraints $Q^k = (Q_{\min}^k, Q_{\max}^k)$ at time t_k , we have

$$P_k^n(Q^k) = \sup \left\{ x V_k + \mathbb{E}(P_{k+1}^n(\chi^{n-k-1}(Q^k, x)) | \mathcal{F}_k), x \in I_{Q^k}^{n-k-1} \right\}$$

with

$$\chi^M(Q^k, x) := \left((Q_{\min}^k - x)^+, (Q_{\max}^k - x) \wedge M \right)$$

and

$$I_{Q^k}^M := [(Q_{\min}^k - M)^+ \wedge 1, Q_{\max}^k \wedge 1].$$

(iii) One defines the set of admissible deterministic global constraints at time 0 by

$$T^+(n) = \{Q^0 \in \mathbb{R}^2 \mid 0 \leq Q_{\min}^0 \leq Q_{\max}^0 \leq n\}. \quad (10)$$

Then, given any admissible integral global constraints at time 0 i.e. $Q^0 \in \mathbb{N}^2 \cap T^+(n)$, there exists an optimal bang-bang control process $q^* = (q_k^*)_{0 \leq k \leq n-1}$ with $q_k^* \in \{0, 1\}$ \mathbb{P} -a.s. for every $k = 0, \dots, n-1$.

The value function is concave, continuous and piecewise affine on the tiling of $T^+(n)$.

So in view of these results, it is sufficient to evaluate the option premium at integral values of $T^+(n)$. Then, it is possible to use a linear interpolation to obtain the derivative price for all admissible constraints, i.e. for all $Q^0 \in T^+(n)$. Hence we may reformulate the BDP Formula for an initial $Q^0 \in \mathbb{N}^2 \cap T^+(n)$ as

$$P_n^n \equiv 0$$

$$P_k^n(Q^k) = \max \left\{ x V_k + \mathbb{E}(P_{k+1}^n(\chi^{n-k-1}(Q^k, x)) | \mathcal{F}_k), x \in \{0, 1\} \cap I_{Q^k}^{n-k-1} \right\}, \quad k = 0, \dots, n-1. \quad (11)$$

In order to simulate the derivative prices obtained in (11), one appeals to the quantization method described in Section 2. At each exercise date t_k , $k = 0, \dots, n-1$, we perform a spatial discretization of the d -dimensional underlying asset and include the resulting discretized process in the BDP Formula. Finally, some convergence theorems are available to obtain some error bounds. However, these theorems require the following assumption:

The underlying d -dimensional process, say $(X_k)_{0 \leq k \leq n-1}$ has a Markovian structure.

Therefore, if one assumes that the payoffs V_k are function of X_k , that is

$$V_k = v_k(X_k), \quad k = 0, \dots, n-1, \quad (12)$$

one obtains

$$\mathbb{E}(v_{k+1}(X_{k+1})|\mathcal{F}_k) = \mathbb{E}(v_{k+1}(X_{k+1})|X_k) = \Theta_k(v_{k+1})(X_k), \quad k = 0, \dots, n-1, \quad (13)$$

where $(\Theta_k)_{0 \leq k \leq n-1}$ is a sequence of Borel transition probabilities on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Therefore (11) becomes

$$\begin{aligned} P_n^n &\equiv 0 \\ P_k^n(Q^k) &= \max \left\{ xv_k(X_k) + \mathbb{E}(P_{k+1}^n(\chi^{n-k-1}(Q^k, x))|X_k), x \in \{0, 1\} \cap I_{Q^k}^{n-k-1} \right\}, \quad k = 0, \dots, n-1. \end{aligned} \quad (14)$$

Now, let \hat{X}_k be a quantized version of X_k of size say N_k . An approximation of the price process is designed by plugging \hat{X}_k in (14) and by forcing the Markov property on \hat{X}_k .

$$\begin{aligned} \hat{P}_n^n &\equiv 0 \\ \hat{P}_k^n(Q^k) &= \max \left\{ xv_k(\hat{X}_k) + \mathbb{E}(\hat{P}_{k+1}^n(\chi^{n-k-1}(Q^k, x))|\hat{X}_k), x \in \{0, 1\} \cap I_{Q^k}^{n-k-1} \right\}, \quad k = 0, \dots, n-1. \end{aligned} \quad (15)$$

The resulting algorithm named a quantization tree is then tractable by computer algorithm. Indeed, it is possible to evaluate the above formula in a recursive way up to $k = 0$ where we arrive at a $\sigma(X_0)$ -measurable solution. Note that for $\mathcal{F}_0 = \{\emptyset, \Omega\}$, the option premium is deterministic. From now on, we assume that \mathcal{F}_0 is trivial.

The following results are concern with the error resulting from the discretization of the underlying price process $(X_k)_{0 \leq k \leq n-1}$.

Theorem 3.1. *Assume that the Markov process $(X_k)_{0 \leq k \leq n-1}$ is Lipschitz Feller in the following sense: for every bounded Lipschitz continuous function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ and every $k \in \{0, \dots, n-1\}$, $\Theta_k(g)$ is a Lipschitz continuous function satisfying $[\Theta_k(g)]_{Lip} \leq [\Theta_k]_{Lip}[g]_{Lip}$. Assume that every function $v_k : \mathbb{R}^d \rightarrow \mathbb{R}$ is Lipschitz continuous with Lipschitz coefficient $[v_k]_{Lip}$. Let $p \in [1, \infty)$ such that $\max_{0 \leq k \leq n-1} \|X_k\| \in L^p(\mathbb{P})$. Then, there exists a real constant $C_p > 0$ such that*

$$\| \sup_{Q \in T^+(n) \cap \mathbb{N}^2} |P_0^n(Q) - \hat{P}_0^n(Q)| \|_p \leq C_p \sum_{k=0}^{n-1} \|X_k - \hat{X}_k\|_p.$$

In view of Zador Theorem (see Theorem 1) and the \mathcal{F}_0 -measurability of P_0^n and \hat{P}_0^n , this also reads

$$\sup_{Q \in T^+(n) \cap \mathbb{N}^2} |P_0^n(Q) - \hat{P}_0^n(Q)| \leq C \sum_{k=0}^{n-1} N_k^{-1/d}.$$

3.2 Complexity and implementationary Notes

For an analysis of the complexity of the quantization tree algorithm in its original form (15), we count the number of multiplications, which occur during the evaluation of $P_0^n(Q^0)$ for a given Q^0 . We implement the quantization tree algorithm in a backward iterative manner, starting at layer k with the computation of $P_k^n(Q^k)$ for every possible residual global constraint $Q^k \in \mathcal{Q}_k^n(Q^0)$ with

$$\mathcal{Q}_k^n(Q^0) = \{((Q_{\min}^0 - l)^+, (Q_{\max}^0 - l)^+ \wedge (n - k)), l = 0, \dots, k\}$$

for given initial global constraints $Q = (Q_{\min}, Q_{\max}) \in T^+(n) \cap \mathbb{N}^2$. This approach results in a complexity proportional to

$$\sum_{k=0}^{n-2} \# \mathcal{Q}_k^n(Q^0) N_k N_{k+1} + N_{n-1}$$

multiplications, which can be estimated from above by

$$\sum_{k=0}^{n-2} (k+1) N_k N_{k+1} + N_{n-1}.$$

If we employ a equal grid size N in each layer for the quantization of the conditional expectations, we finally arrive at an upper bound proportional to $\frac{(n^2-n)}{2} N^2 + N$ multiplications.

3.3 Some special cases

We will study some special cases of swing options. These classes of swing are characterized by particular values of global constraints.

3.3.1 The call strip: case $Q = (0, n)$

In this case the global constraint is always satisfied by the cumulative purchase process, i.e. $\sum_{l=0}^{n-1} q_l \in [Q_{\min}, Q_{\max}]$, as the control process $(q_l)_{0 \leq l \leq n-1}$ is $[0, 1]$ -valued. Hence the optimal control problem is reduced to a maximization problem without constraints:

$$F_k^n(0, n) = \text{esssup} \left\{ \mathbb{E} \left(\sum_{l=k}^{n-1} q_l V_l \middle| \mathcal{F}_k \right), q_l : (\Omega, \mathcal{F}_l) \rightarrow [0, 1], k \leq l \leq n-1 \right\}, \quad k = 0, \dots, n-1, \quad (16)$$

an optimal control for (??) is given by

$$q_l^* := \begin{cases} 1, & \text{on } \{V_l \geq 0\}, \\ 0, & \text{elsewhere,} \end{cases}$$

and the associated price process by

$$F_k^n(0, n) = \sum_{l=k}^{n-1} \mathbb{E} (V_l^+ | \mathcal{F}_k).$$

Appealing to the proceeding of Section (3.1), one obtains the following quantized tree algorithm in a Markovian structure

$$\hat{P}_k^n(0, n) = \sum_{l=k}^{n-1} \mathbb{E} \left((v_l(\hat{X}_l))^+ | \hat{X}_k \right).$$

Note that this case can be considered as a sum of European options. Since explicit formula are available to evaluate European options, the call strip will be used as benchmark for numerical implementations.

3.3.2 The Bermuda option: case $Q = (0, 1)$

For the Bermuda option, one sets $t_{n-1} = T$. In view of result (iii) in Section (3.1), there exists an admissible optimal control $(q_k^*)_{0 \leq k \leq n-1}$, such that q_k^* is $\{0, 1\}$ -valued for all $k \in \{0, \dots, n-1\}$. Since $(q_k^*)_{0 \leq k \leq n-1}$ is admissible, we have

$$0 \leq \sum_{l=0}^{n-1} q_l^* \leq 1,$$

(see (6)). Therefore, there exists at most one $l := l(\omega) \in \{0, \dots, n-1\}$ such that $q_{l(\omega)}^*(\omega) = 1$. So, for $k \in \{0, \dots, n-1\}$

$$\begin{aligned} P_k^n((0, 1)) &= \mathbb{E} \left(\sum_{j=k}^{n-1} q_j^* V_j | \mathcal{F}_k \right) \\ &= \mathbb{E} (q_l^* V_l | \mathcal{F}_k) \\ &= \mathbb{E} (V_l^+ | \mathcal{F}_k). \end{aligned}$$

Hence the optimisation problem can be rewritten in term of an optimal stopping problem

$$P_k^n((0, 1)) = \text{esssup} \{ \mathbb{E} ((V_\tau)^+ | \mathcal{F}_k), \tau \in \theta_k \},$$

where θ_k denotes the family of $\{t_k, \dots, t_{n-1}\}$ -valued \mathcal{F} -stopping times. For the Bermuda case, one observes, for $k = 0, \dots, n-1$, that

$$\begin{aligned} I_{Q^k}^{n-k-1} &= [(Q_{\min}^k - (n-k-1))^+ \wedge 1, Q_{\max}^k \wedge 1] \\ &= [(Q_{\min} - \sum_{l=0}^{k-1} q_l - (n-k-1))^+ \wedge 1, (Q_{\max} - \sum_{l=0}^{k-1} q_l) \wedge 1] \\ &= \begin{cases} [0, 1], & \text{if } q_l = 0, l = 0, \dots, k-1, \\ \{0\}, & \text{if there exists } l \in \{0, \dots, k-1\} \mid q_l = 1. \end{cases} \end{aligned}$$

So, for $k = 0, \dots, n-2$ and for $x \in \{0, 1\}$,

$$\begin{aligned} \chi^{n-k-1}(Q^k, x) &= \left((Q_{\min}^k - x)^+, (Q_{\max}^k - x) \wedge (n-k-1) \right) \\ &= \left((Q_{\min} - \sum_{l=0}^{k-1} q_l - x)^+, (Q_{\max} - \sum_{l=0}^{k-1} q_l - x) \wedge (n-k-1) \right) \\ &= (0, (1 - \sum_{l=0}^{k-1} q_l - x) \wedge (n-k-1)) \\ &= \begin{cases} (0, 1), & \text{if } x = 0 \text{ and if } q_l = 0, l = 0, \dots, k-1, \\ (0, 0), & \text{if } x = 1 \text{ or if there exists } l \in \{0, \dots, k-1\} \mid q_l = 1. \end{cases} \end{aligned}$$

So the BDP Formula (11) can be rewritten as

$$\begin{aligned} P_n^n &\equiv 0 \\ P_k^n((0, 1)) &= \max \{ 1V_k + \mathbb{E}(P_{k+1}^n((0, 0)) | \mathcal{F}_k), 0V_k + \mathbb{E}(P_{k+1}^n((0, 1)) | \mathcal{F}_k) \} \\ &= \max \{ (V_k)^+, \mathbb{E}(P_{k+1}^n((0, 1)) | \mathcal{F}_k) \}, \quad k = 0, \dots, n-1. \end{aligned}$$

Indeed, one observes that if there is at some exercise date t_k , a consumption of quantity 1, then the global constraints are saturated and no further consumption may occur, i.e. $\mathbb{E}(P_{k+1}^n((0,0))|\mathcal{F}_k) = 0$. Note also that the price process defined in a recursive way is a.s. always non-negative. Therefore, one could restrict to positive parts for the option payoff in the (BDP) Formula.

The quantized version in a Markovian setting is then given by

$$\begin{aligned}\hat{P}_n &\equiv 0 \\ \hat{P}_k^n((0,1)) &= \max \left\{ (v_k(\hat{X}_k))^+, \mathbb{E}(\hat{P}_{k+1}^n((0,1))|\hat{X}_k) \right\}, \quad k = 0, \dots, n-1.\end{aligned}$$

3.4 From Bermuda options to American options

Recall that a Bermuda product is in fact a time discretized version of an American option, so we will use this product to approximate the pricing of American options by means of the quantization tree algorithm. Therefore denote by P_t^C the solution of the continuous optimal stopping problem

$$P_t^C := \text{esssup} \left\{ \mathbb{E}((V_\tau^C)^+|\mathcal{F}_t), \tau \text{ is a } [t, T]\text{-valued stopping time} \right\},$$

and suppose that $(V_k)_{0 \leq k \leq n-1}$ is given by an equidistant time discretization of the continuous process $(V_t^C)_{t \in [0, T]}$, i.e. $V_k := V_{k\Delta t}^C$ for $\Delta t = T/(n-1)$ and $k = 0, \dots, n-1$.

Hence, for semiconvex payoff, we have the following time discretization error bound. Let $p \in [1, +\infty)$,

$$\|P_0^C - P_0^n\|_p \leq \frac{C_p}{n},$$

(see [BPP05]). And we may conclude from theorem (3.1) that

$$\|P_0^C - \hat{P}_0^n\|_p \leq \frac{C_p^1}{n} + C_{p,v}^2 \sum_{k=0}^{n-1} \|X_k - \hat{X}_k\|_p,$$

where X_k denotes the underlying Markov process.

Applying Zador's Theorem for the quantization error $\|X_k - \hat{X}_k\|$, this reads

$$\|P_0^C - \hat{P}_0^n\|_p \leq \frac{C_p^1}{n} + C_{p,v}^2 \sum_{k=0}^{n-1} N_k^{-1/d},$$

or in the case of a quantization grid with constant size N

$$\|P_0^C - \hat{P}_0^n\|_p \leq \frac{C_p^1}{n} + C_{p,v}^2 \frac{n}{N^{1/d}}.$$

4 Two main improvements of the quantized tree

4.1 Enhancements for the computations of the transition probabilities

To improve the derivative evaluations, one studies and compares several methods to compute the conditional expectations found in (15). Noting that these expectations have the form of

$$\mathbb{E}(f(\hat{X}_{k+1})|\hat{X}_k)$$

for some Borel functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and take values in finite grids $x^k = (x_1^k, \dots, x_{N_k}^k)$ of \mathbb{R}^d , we have

$$\mathbb{E}(f(\hat{X}_{k+1})|\hat{X}_k = x_i^k) = \sum_{j=1}^{N_{k+1}} f(x_j^{k+1}) \pi_k^{ij}$$

where

$$\pi_k^{ij} = \mathbb{P}(\hat{X}_{k+1} = x_j^{k+1} | \hat{X}_k = x_i^k), \quad k = 0, \dots, n-2,$$

denote the transition probabilities of the quantized process. By definition of the Voronoi quantization \hat{X} in (4), we arrive at

$$\pi_k^{ij} = \mathbb{P}(X_{k+1} \in C_j(x^{k+1}) | X_k \in C_i(x^k)) = \frac{\mathbb{P}(X_{k+1} \in C_j(x^{k+1}), X_k \in C_i(x^k))}{\mathbb{P}(X_k \in C_i(x^k))}. \quad (17)$$

To estimate the transition probabilities, one generates e.g. a Monte Carlo sample of size M of the random vector X_k , that we denote by $(\tilde{X}_k^m)_{1 \leq m \leq M}$, in each layer $k = 1, \dots, n-1$. So π_k^{ij} can be estimated by

$$\tilde{\pi}_k^{ij} := \frac{\sum_{m=1}^M 1_{C_j(x^{k+1})}(\tilde{X}_{k+1}^m) 1_{C_i(x^k)}(\tilde{X}_k^m)}{\sum_{m=1}^M 1_{C_i(x^k)}(\tilde{X}_k^m)}, \quad \text{for } k = 1, \dots, n-2, \quad (18)$$

$$i = 1, \dots, N_k \text{ and } j = 1, \dots, N_{k+1}.$$

Given k , i and j , to determine the value of the indicator functions present in the estimations (18), one checks, for each $m = 1, \dots, M$, if the realizations $\tilde{X}_k^m(\omega)$ and $\tilde{X}_{k+1}^m(\omega)$ belong to the cells $C_i(x^k)$ or $C_j(x^{k+1})$ of the Voronoi tessellation.

This problem consists in finding the nearest neighbor for a query point $q \in \mathbb{R}^d$ in the N -tuple $x = (x_1, \dots, x_N)$ with regard to the Euclidean distance. A commonly known algorithm to optimize this nearest neighbor search, which can in high dimensions become a time consuming procedure, is the kd -tree algorithm.

This procedure consists of two steps: The partitioning of the space, which has an average complexity cost of $\mathcal{O}(n \log n)$ operations. This has to be done only once and is therefore independent of the size M of the Monte Carlo sample. The second part is the search procedure, which has an average complexity cost of $\mathcal{O}(\log n)$ distance computations.

The kd -tree data structure is based on a recursive subdivision of space \mathbb{R}^d into disjoint hyper-rectangular regions called boxes. Each node of the resulting tree is associated with such a region. To ensure a well-defined splitting rule on each node, we equip the root node with a bounding box, which is chosen to contain all data points.

As long as the number of data points within such box is greater than a given threshold, the box is splitted-up into two boxes by an axis-orthogonal hyperplane that intersects the original box. These two boxes are then associated to two children nodes. When the number of points in a box is below the threshold, the associated node is declared a leaf node, and the points are stored within this node.

To find the nearest neighbor for a query point q , we start at the root node and decide at each node due to the cutting hyperplane, to which child node we descend.

When arriving at a leaf node, we compute the minimal distance from all data points of this node to the query point q and ascend back on the tree to the root node as long as it may be possible to find closer points than the previous found minimal distance. During this way back we obviously descend to the so far not visited children and update the minimal distance found so far.

Now, we assume, that $(X_k)_{0 \leq k \leq n-1}$ can be written recursively, i.e.

$$X_{k+1} = A_k X_k + T_k Z_k, \quad k = 0, \dots, n-2, \quad X_0 = x_0,$$

for some positive definite Matrices A_k and T_k , and a deterministic initial value x_0 . Furthermore we suppose, that the distribution of Z_k is either known as density or can be simulated and that it is independent of the distribution of X_k .

These assumptions hold for example for the BS model, where the driving process is a Brownian motion, the Ornstein-Uhlenbeck process, which we will later on formulate as Gaussian first order auto-regressive processes or for certain Lévy processes, e.g. the NIG process. This writing was first initiated by Bardou, Bouthemy and Pagès for Gaussian processes (see [BBP07a]).

4.1.1 Diffusion method

The most straightforward approach consists in generating a Monte Carlo sample of size say M of the random variate Z_k in each layer k , i.e

$$(\tilde{Z}_0^m)_{1 \leq m \leq M}, \dots, (\tilde{Z}_{n-1}^m)_{1 \leq m \leq M},$$

and then in defining the Monte Carlo sample for $(X_k)_{0 \leq k \leq n-1}$ by

$$\begin{aligned} \tilde{X}_0^m &:= x_0, \\ \tilde{X}_{k+1}^m &:= A_k \tilde{X}_k^m + T_k \tilde{Z}_k^m, \quad k = 0, \dots, n-2 \end{aligned} \tag{19}$$

for $m = 1, \dots, M$. So we get an estimate for π_k^{ij} , which writes

$$\tilde{\pi}_k^{ij} := \frac{\sum_{m=1}^M 1_{C_j(x^{k+1})}(\tilde{X}_{k+1}^m) 1_{C_i(x^k)}(\tilde{X}_k^m)}{\sum_{m=1}^M 1_{C_i(x^k)}(\tilde{X}_k^m)}.$$

Note that, in this method, we have to proceed to computations along the discrete trajectories of \tilde{X}^m , that is, step by step from time 0 to the exercise date t_{n-1} , which forbids to do the computations in parallel. But if we may assume, that the X_k 's can be simulated directly, we are able to overcome this obstacle by the so-called *fast parallel quantization* method.

4.1.2 Fast Parallel Quantization method (FPQ)

For this method we generate in each layer k , independently, bivariate Monte Carlo samples by simulating directly from the distributions of X_k and Z_k , i.e.

$$(\tilde{X}_0^m, \tilde{Z}_0^m)_{1 \leq m \leq M}, \dots, (\tilde{X}_{n-2}^m, \tilde{Z}_{n-2}^m)_{1 \leq m \leq M}.$$

The estimated transition probabilities are then given by

$$\tilde{\pi}_k^{ij} := \frac{\sum_{m=1}^M 1_{C_j(x^{k+1})}(\tilde{A}_k \tilde{X}_k^m + T_k \tilde{Z}_k^m) 1_{C_i(x^k)}(\tilde{X}_k^m)}{\sum_{m=1}^M 1_{C_i(x^k)}(\tilde{X}_k^m)}.$$

The structure of this method is ideally suited to a parallelized implementation and therefore it's possible to reduce efficiently the computation time of the transition probabilities procedure.

4.1.3 Completely deterministic method

For a one-dimensional process $(X_k)_{0 \leq k \leq n-1}$, i.e. $X_{k+1} = a_k X_k + t_k Z_k$, for $a_k, t_k \in \mathbb{R} \setminus \{0\}$, it might be even possible to compute the transition probabilities by deterministic integration methods. In

that case we write

$$\begin{aligned}
\pi_k^{ij} &= \frac{\mathbb{P}(X_{k+1} \in C_j(x^{k+1}), X_k \in C_i(x^k))}{\mathbb{P}(X_k \in C_i(x^k))} = \\
&= \frac{\mathbb{E}\left(1_{C_j(x^{k+1})}(a_k X_k + t_k Z_k) 1_{C_i(x^k)}(X_k)\right)}{\mathbb{E}(1_{C_i(x^k)}(X_k))} = \\
&= \frac{\mathbb{E}\left(1_{(C_j(x^{k+1}) - a_k X_k)/t_k}(Z_k) 1_{C_i(x^k)}(X_k)\right)}{\mathbb{E}(1_{C_i(x^k)}(X_k))} = \\
&= \frac{1}{\int_{C_i(x^k)} d\mathbb{P}_{X_k}} \int_{C_i(x^k)} \int_{(C_j(x^{k+1}) - a_k x)/t_k} \mathbb{P}_{Z_k}(dy) \mathbb{P}_{X_k}(dx),
\end{aligned}$$

where we are able, due to Fubini's Theorem, to compute first the expectation with regard to Z_k and afterwards the expectation with regard to X_k .

4.1.4 “Methode des gerbes”

Another way to approximate the transition probabilities π_k^{ij} is to replace the random variable X_k by its quantization \hat{X}_k in the identity $X_{k+1} = A_k X_k + T_k Z_k$ as well as in the conditional part $X_k \in C_i(x^k)$ of (17).

Using the equivalence of $\hat{X}_k \in C_i(x^k)$ and $\hat{X}_k = x_i^k$, this leads to the following approximation

$$\begin{aligned}
\pi_k^{ij} &:= \mathbb{P}(A_k \hat{X}_k + T_k Z_k \in C_j(x^{k+1}) | \hat{X}_k \in C_i(x^k)) = \\
&= \mathbb{P}(A_k x_i^k + T_k Z_k \in C_j(x^{k+1})) = \\
&= \mathbb{P}\left(Z_k \in T_k^{-1}(C_j(x^{k+1}) - A_k x_i^k)\right).
\end{aligned}$$

This quantity can in low dimensions and when the density of Z_k is known be computed by deterministic integration methods. Otherwise a Monte Carlo estimate would be given by

$$\tilde{\pi}_k^{ij} := \frac{1}{M} \sum_{m=1}^M 1_{T_k^{-1}(C_j(x^{k+1}) - A_k x_i^k)}(\tilde{Z}_k^m),$$

where

$$(\tilde{Z}_0^m)_{1 \leq m \leq M}, \dots, (\tilde{Z}_{n-2}^m)_{1 \leq m \leq M}$$

denote again the i.i.d. Monte Carlo samples of (Z_k) .

4.2 A faster convergence rate: an extrapolation approach for the quantization tree algorithm

Consider for instance the classical numerical integration by means of optimal quantization, i.e. $\mathbb{E}f(X)$ is approximated by $\mathbb{E}f(\hat{X}^N)$ for sequence of stationary and rate optimal quantizers \hat{X}^N (see Section 2).

Assume that $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is a 3-times differentiable function with bounded derivatives.

If it holds true, that

$$\mathbb{E}(D^2 f(\hat{X}^N)(X - \hat{X}^N)^{\otimes 2}) \sim cN^{-2/d} \quad \text{as } N \rightarrow \infty$$

and

$$\mathbb{E}|X - \hat{X}^N|^3 = \mathcal{O}(N^{-3/d}).$$

Then we can conclude from a Taylor expansion

$$f(X) = f(\hat{X}^N) + Df(\hat{X}^N)(X - \hat{X}^N) + \frac{1}{2}D^2f(\hat{X}^N)(X - \hat{X}^N)^{\otimes 2} + \frac{1}{6}D^3f(\xi)(X - \hat{X}^N)^{\otimes 3}, \quad \xi \in (\hat{X}^N, X),$$

so that

$$\mathbb{E}f(X) = \mathbb{E}f(\hat{X}^N) + cN^{-2/d} + o(N^{-3/d+\varepsilon}) \quad \text{for every } \varepsilon > 0.$$

Motivated by this identity and our numerical tests, we want to assume, that an analogous result holds also (at least approximately) true for the quantization tree algorithm

$$P_0^n = \hat{P}_0^n + cN^{-2/d} + o(N^{-3/d+\varepsilon}) \quad \text{for every } \varepsilon > 0.$$

Hence we perform a Romberg extrapolation on the quantized swing option prices by computing \hat{P}_0^n for two different values N_1, N_2 (e.g. $N_1 \approx 4N_2$) and denote them by $\hat{P}_0^n(N_1)$ and $\hat{P}_0^n(N_2)$. Thus we arrive at

$$P_0^n = \hat{P}_0^n(N_1) + \frac{\hat{P}_0^n(N_1) - \hat{P}_0^n(N_2)}{N_2^{-2/d} - N_1^{-2/d}} N_1^{-2/d} + o(N^{-3/d+\varepsilon}),$$

where

$$P_0^n(N_1, N_2) := \hat{P}_0^n(N_1) + \frac{\hat{P}_0^n(N_1) - \hat{P}_0^n(N_2)}{N_2^{-2/d} - N_1^{-2/d}} N_1^{-2/d}$$

is the improved approximation for P_0^n .

5 Applications

5.1 The American options

One is interested in American exchange option pricing. One considers a Black-Scholes model for the underlying assets dynamics given by

$$dS_t^l = (r - \mu_l)S_t^l dt + \sigma_l S_t^l dW_t^l, \quad t \in [0, T], \quad l = 1, \dots, d.$$

To simplify the computations, w.l.o.g. we will set the interest rate to 0.

The exchange American payoff is defined by

$$v(y) := \max(y^1 \cdot \dots \cdot y^p - y^{p+1} \cdot \dots \cdot y^{2p}, 0), \quad \text{with } d := 2p.$$

To reduce the price estimation variance, one introduces a sequence of control variate variables in the quantized tree algorithm, (see [BPP05], Section 5). In the exchange case, the variable considered is the European exchange option with similar maturity as the American one. The European premium has a closed form solution given by

$$\begin{aligned} \text{Ex}_{\text{BS}}(\Theta, y, y', \tilde{\sigma}, \mu) &:= \text{erf}(d_1) \exp(\mu\Theta)y - \text{erf}(d_1 - \tilde{\sigma}\sqrt{\Theta})y', \\ d_1(y, y', \tilde{\sigma}, \Theta, \mu) &:= \frac{\log(y/y') + (\tilde{\sigma}^2/2 + \mu)\Theta}{\tilde{\sigma}\sqrt{\Theta}} \\ \text{and } \text{erf}(y) &= \int_{-\infty}^y e^{-u^2/2} du / \sqrt{2\pi}, \end{aligned}$$

with

$$\Theta := T - t, \quad \tilde{\sigma} := \left(\sum_{l=1}^d \sigma_l^2 \right)^{1/2}, \quad \mu := \sum_{l=1}^p \mu_l - \sum_{l=p+1}^d \mu_l, \quad y := \prod_{l=1}^p S_t^l, \quad y' := \prod_{l=p+1}^d S_t^l.$$

Following the method described in Sections (3.3.2) and (3.4), one proceeds to a time and space discretization of the underlying diffusion.

Note that in the case of American option, an optimization method is applied to choose the quantizer size at each time step t_k , (see [BPP05], Theorem 2.4). That is, at an exercise date t_k , for $N = \sum_{k=0}^{n-1} N_k$, we have

$$N_k := \left\lceil \frac{t_k^{\frac{d}{2(d+1)}} (N-1)}{t_1^{\frac{d}{2(d+1)}} + \dots + t_{n-1}^{\frac{d}{2(d+1)}}} \right\rceil, \quad (20)$$

where $\lceil x \rceil := \min\{k \in \mathbb{N} : k \geq x\}$.

To compute an American exchange option premium, we consider the following steps: first, we discretize the d -dimensional Brownian motion. Secondly, thanks to a Monte Carlo method, we estimate the transition probabilities, defined for $k \in \{0, \dots, n-2\}$, $i \in \{1, \dots, N_k\}$ and $j \in \{1, \dots, N_{k+1}\}$ by

$$\pi_k^{ij} = \frac{\mathbb{P}(W_{k+1} \in C_j(\bar{x}^{k+1}), W_k \in C_i(\bar{x}^k))}{\mathbb{P}(W_k \in C_i(\bar{x}^k))} \quad (21)$$

where \bar{x}^k is the L^p -optimal N_k -quantizer of the Brownian motion. Finally, we plug these results in the quantization tree algorithm. These three steps could be compute together or separately. Since the first and the last steps are almost instantaneous here to obtain, we concentrate on the second step, which has a significant computation time. So, we test several ways to simulate the transition probabilities and analyse the correlation between quantized premium accuracy and computation time.

We consider two different ways of estimating the transition probabilities, which are both adapted to multidimensional problem (see Section 4.1).

5.1.1 The Diffusion method

We simulate some standard Brownian motion trajectories from $t_0 = 0$ to the maturity of the exchange option $t_{n-1} = T$. This simulation is based on the independence and stationary properties of the Brownian motion increments. Indeed, one will notice that the law of the family $(W_{t_{k+1}} - W_{t_k})_{0 \leq k \leq n-2}$, with $t_k = \frac{kT}{n-1}$, $k = 0, \dots, n-1$, is similar to a family of i.i.d. random vector with $\mathcal{N}(0, \frac{T}{n-1} I_d)$ distribution. Therefore for $k \in \{1, \dots, n-1\}$, W_{t_k} is simulated by

$$\sqrt{\frac{T}{n-1}} \sum_{i=0}^{k-1} \epsilon_{i+1} \quad (22)$$

where ϵ_i , $i \in \{1, \dots, k\}$ are i.i.d random variable, with normal distribution. The Monte Carlo proxies of the theoretical transitions (21) are then simulated for $k \in \{0, \dots, n-2\}$, $i \in \{1, \dots, N_k\}$ and $j \in \{1, \dots, N_{k+1}\}$ by

$$\tilde{\pi}_k^{ij} = \frac{\frac{1}{M} \sum_{m=1}^M 1_{C_j(\bar{x}^{k+1})}(\tilde{W}_{t_{k+1}}^m) 1_{C_i(\bar{x}^k)}(\tilde{W}_{t_k}^m)}{\frac{1}{M} \sum_{m=1}^M 1_{C_i(\bar{x}^k)}(\tilde{W}_{t_k}^m)} \quad (23)$$

and by

$$\tilde{\pi}_0^{1j} = \frac{1}{M} \sum_{m=1}^M 1_{C_j(\bar{x}^1)}(\tilde{W}_{t_1}^m), \quad (24)$$

where for $k \in \{0, \dots, n-1\}$, $(\tilde{W}_{t_k}^m)_{1 \leq m \leq M}$ are M copies of the random vector W_{t_k} given by (22).

Remark 1. Recalling Section 2, there exist several algorithms to compute the L^p -optimal N_k -quantizer \bar{x}^k of the Brownian motion. However, one notices that it also could be obtained by a dilatation of the L^p -optimal N_k -quantizer x^k of the normal distribution. That is for $k \in \{0, \dots, n-1\}$, $\bar{x}^k = \sqrt{t_k} x^k$ where x^k is already known.

5.1.2 The Fast Parallel Quantization method (FPQ)

This method, introduced in the paper of Bardou, Bouthemy and Pagès, (see [BBP07a]) appeals to centered Gaussian first order auto-regressive process. In order to apply this method to the American exchange option, we consider the centered Gaussian first order auto-regressive Brownian motion process in \mathbb{R}^d given by $W_{t_{k+1}} = W_{t_k} + \sqrt{\frac{T}{n-1}} \epsilon_{k+1}$, for $k \in \{0, \dots, n-2\}$ where $(\epsilon_k)_{1 \leq k \leq n-1}$ are i.i.d random vectors with $\mathcal{N}(0, I_d)$ distribution.

Let W be the auto-regressive process described above and let (η_1, η_2) be a couple of independent random vectors normally distributed. Then, the transition probabilities (21) satisfy for $k \in \{0, \dots, n-2\}$, $i \in \{1, \dots, N_k\}$ and $j \in \{1, \dots, N_{k+1}\}$

$$\pi_k^{ij} = \frac{\mathbb{P}(\alpha_{k+1}\eta_1 + \beta_{k+1}\eta_2 \in C_j(x^{k+1}), \eta_1 \in C_i(x^k))}{\mathbb{P}(\eta_1 \in C_i(x^k))} \quad (25)$$

and for $j \in \{1, \dots, N_1\}$ by

$$\pi_0^{1j} = \mathbb{P}(\eta_2 \in C_j(x^1)), \quad (26)$$

where for $k \in \{0, \dots, n-2\}$,

$$\alpha_{k+1} = \sqrt{\frac{k}{k+1}},$$

$$\beta_{k+1} = \frac{1}{\sqrt{k+1}},$$

and x^k is the L^p -optimal N_k -quantizer of the normal distribution.

Remark 2. See [BBP07a] for a proof of this result.

The main interest of this formulation is that the transition probabilities simulations from t_k to t_{k+1} no longer need to be computed recursively. Therefore, they can be computed simultaneously by a parallelized procedure with the same M copies $(\eta_1^m, \eta_2^m)_{1 \leq m \leq M}$ of the couple of independent normally distributed random vectors (η_1, η_2) .

5.1.3 Tests

In this subsection, we present numerical results for the two methods introduced above and observe the consequences resulting on the option premium. We focus on American exchange options in dimension 2 in and out of the money. The model parameters are the following:

- a maturity T of one year,
- a dividend rate μ of 5%
- a volatility σ of 20%

The quantization was proceed with 25 time steps using the optimal dispatching rule (20), with an average number of points per layer of 400. The reference prices were computed by a two dimensional finite difference algorithm devised by Villeneuve et Zanette. The Monte Carlo simulations were computed with 2 millions trials. Several tests are executed for different values of the initial underlying asset $S_0 = (S_0^1, S_0^2)$ and are collected in Table 1.

S_0	Ref VZ	Eur BS	Diff	D Error	FPQ	FPQ Error
(80,40)	40	36.1595	40	$< 10^{-4}\%$	40	$< 10^{-4}\%$
(60,40)	20	17.7432	20	$< 10^{-4}\%$	20	$< 10^{-4}\%$
(44,40)	5.9822	5.5883	5.9827	0,008358%	5.9813	0,01504 %
(40,36)	5.6468	5.2674	5.6468	$< 10^{-4}\%$	5.6455	0.02302%
(36,40)	1.9969	1.9058	1.9969	$< 10^{-4}\%$	1.9965	0,02003%
(40,44)	2.3364	2.2289	2.3375	0.04708%	2.3370	0.02568%
(40,60)	0.31339	0.30429	0.31276	0.2010%	0.31284	0.1755%
(40,80)	0.021208	0.020747	0.021064	0.6790%	0.021076	0.6224%

Table 1: American exchange options in dimension 2.

S_0 : initial value of the underlying asset, Ref VZ: Villeneuve and Zanette reference price, Eur BS: European Black Scholes premium, Diff: quantized premium where the transition probabilities are computed by the diffusion method, D Error: relative error for the diffusion method, FPQ: quantized premium where the transition probabilities are computed by the fast parallel quantization method, FPQ Error: relative error for the fast parallel quantization method.

One interesting criteria of the quantization method is that the same transition probabilities are used for several premium computations. Indeed, in Table 1, all the data estimations of the fourth column, (Diff) are computed with the same transitions given by (23) and (24). Whereas, the premium obtained in column sixth, (FPQ), are computed with transitions given by (25) and (26). We also note that the execution time of the diffusion procedure is more than twice quicker than the execution time of FPQ procedure.

In view of this observation, we notice in the fourth column (Diff) of Table 1 that the couple described by result accuracy and computation time is very efficient in the diffusion method. The relative errors are between less than $10^{-4}\%$ and 0.68% and the execution time is twice faster than in the FPQ method. However, the method FPQ has a significant advantage on the computation time point of view. Indeed, the independence property of the transition probabilities given by (25) and (26) allows for a parallel implementation. Each layer, or even, each transition can be implement and estimate simultaneously, which extremely improve the execution time of the whole procedure. Furthermore, the FPQ method yields also accurate results. The relative errors given by the seventh column, (FPQ Error) of Table 1 show results between less than $10^{-4}\%$ and 0.62%. So, in conclusion, we would suggest the FPQ method for American exchange options pricing for its quickness.

Finally, we have considered the importance sampling method. This method was introduced by Bardou, Bouthemy and Pagès, (see [BBP07a]), in the context of swing options. Following a similar approach as in the FPQ method, they have to compute a 2-dimensional expectation for the transition numerator. They appeal to Fubini Theorem and reduce the problem dimension to a one dimensional expectation computation. Then, they use a Cameron Martin Formula to re-center the simulation in order to improve the accuracy. This idea, very efficient in the swing option pricing, doesn't apply in the exchange option case. We price by quantization several exchange options with this method and conclude that in our case, for a dimension above or equal to 2, this method doesn't work. One reason is that the Fubini change of integration order isn't applicable for a dimension

above 1. In dimension 1, this change of integration order seems to regularize the results obtained for the swing options. Another reason is due to the re-center of the simulations at the origin, indeed in this case, the narrow cells located near the origin are much less visited than the larger one located further, (around 20 on 200 trials reach the cells on the heart of the quantized tree, whereas around 170 on 200 trials reach the cells on the border of the quantized tree). Therefore, the transitions located in the heart of the quantized tree have an inaccurate estimations, which yield an important pricing error.

5.1.4 Spatial Richardson-Romberg extrapolation method.

In this paragraph, we present some premium estimations for American exchange options in several dimensions, ($d = 2$ to 6). We apply the numerical scheme described in Section 5.1. And the transition probabilities are computed by the diffusion method, see (23) and (24). (Indeed we don't focus on execution time on these tests).

The model parameters are similar as in the Section 5.1.3 for maturity, drift and volatility and we consider $S_0 = (40, 36)$. The average number of points per layer, \bar{N} , is fixed and the number of layers, n , is variable (see Table 2). As in Table 1, the Villeneuve and Zanette reference price amounts to 5.6468 and the control variate variable (European Black and Scholes premium) is equal to 5.2674.

	$d = 2$	$d = 4$	$d = 6$
\bar{N}_1	100	750	1000
\bar{N}_2	400	1000	1500
n	5 to 65	4 to 32	4 to 16

Table 2: Quantization parameters.

The results obtained in Figures 1, 2 and 3 emphasize the existence of an optimal couple (\bar{N}, n_{opt}) . I.e., for a fixed average number of points per layer, there exists an optimal number of layer at which the estimation should be proceed.

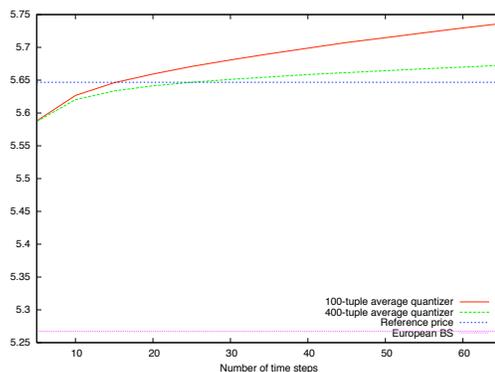


Figure 1: Quantized premium for $d = 2$.

In Figure 1, we observe that in dimension 2, an optimal couple is given by $(\bar{N}, n_{opt}) = (400, 25)$. Indeed the resulting error associated to this couple is negligible and the premium estimation very accurate. For $\bar{N} = 100$ the optimal number of time steps is located around 16. However, for a

given choice of constant c , if one defines an interval error by

$$I(d, \bar{N}) := \{n \mid |p(\bar{N}, n) - p_{VZ}| \leq c\}, \quad (27)$$

where $(\bar{N}, n) \mapsto p(\bar{N}, n)$ models the numerical premium and the constant p_{VZ} the reference one, then one notes that the length of $I(2, 100)$ is shorter than the one of $I(2, 400)$ for a given choice of c reasonably small. So, we suggest a choice of \bar{N} above 300 in dimension 2 to minimize the pricing error. Indeed, a choice of an optimal number of time steps is heuristic (see [BPP05], Section 5) and therefore some possible errors in the choice of n_{opt} could happen. So, a large interval of values of n for which the premium estimation is very closed to the reference price is much preferable.

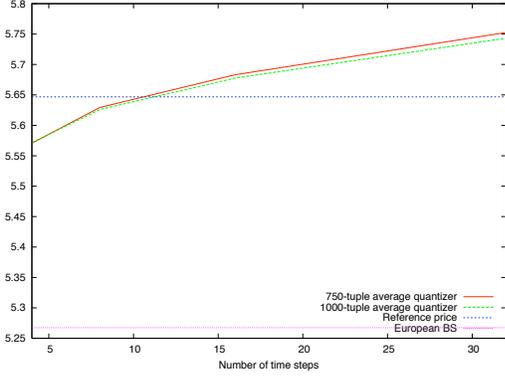


Figure 2: Quantized premium for $d=4$.

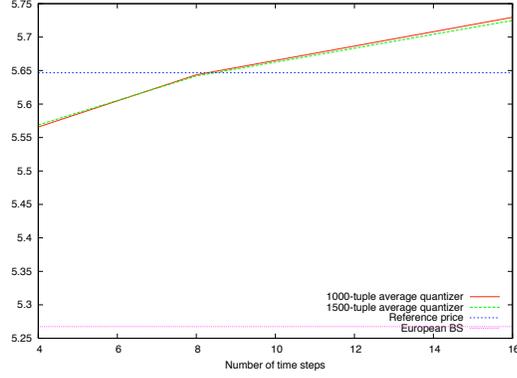


Figure 3: Quantized premium for $d=6$.

In Figures 2 and 3, the optimal number of layers is around 11 in dimension 4 and around 8 in dimension 6. We notice also that an increase of \bar{N} in these figures isn't enough to improve the convergence rate of the trajectories $n \mapsto p(\bar{N}, n)$. Therefore, we suggest the application of a spatial Richardson-Romberg extrapolation method.

First, recalling Section 3.4 and Section 4.2, we fix two average numbers of points per layers \bar{N}_1 and \bar{N}_2 with $\bar{N}_1 < \bar{N}_2$ and we fix a number of time steps n . We compute the quantized premiums $p(\bar{N}_1, n)$ and $p(\bar{N}_2, n)$ and might expect error developments given by

$$|p(\bar{N}_1, n) - p_{VZ}| \sim \frac{c_1}{n} + c_2 \frac{n}{\bar{N}_1^{\frac{2}{d}}}, \quad (28)$$

$$|p(\bar{N}_2, n) - p_{VZ}| \sim \frac{c_1}{n} + c_2 \frac{n}{\bar{N}_2^{\frac{2}{d}}}. \quad (29)$$

Remark 3. *The approximations above have only been proved for an inferior or equal bounds. However several numerical tests tend to suggest that the equivalence holds. Also these tests shows that the convergence rate is closer to $\frac{2}{d}$ instead of $\frac{1}{d}$ given by Theorem 1.*

To improve the convergence rate in dimensions 4 and 6, in view of (28) and (29), we suggest an application of a Richardson-Romberg extrapolation method. Several computations show the ascendancy of the spatial term, $c_2 \frac{n}{\bar{N}^{\frac{2}{d}}}$ over the temporal one, $\frac{c_1}{n}$, in (28) and (29). Therefore we apply a spatial extrapolation to delete the spatial term and obtain

$$p_{VZ} \sim \frac{\bar{N}_2^{\frac{2}{d}} p(\bar{N}_2, n) - \bar{N}_1^{\frac{2}{d}} p(\bar{N}_1, n)}{\bar{N}_2^{2/d} - \bar{N}_1^{2/d}}. \quad (30)$$

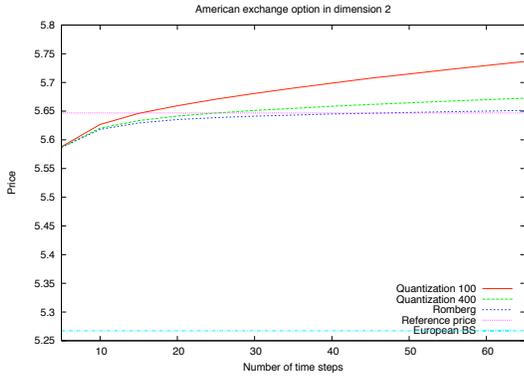


Figure 4: Richardson-Romberg for $d = 2$.

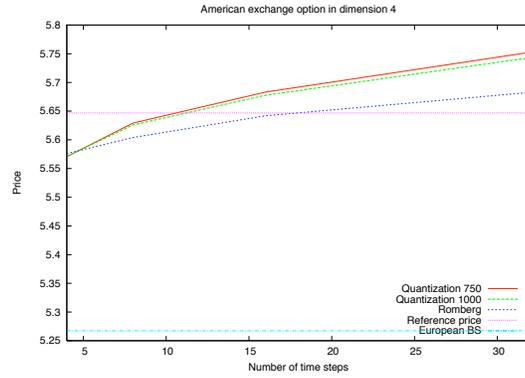


Figure 5: Richardson-Romberg for $d = 4$.

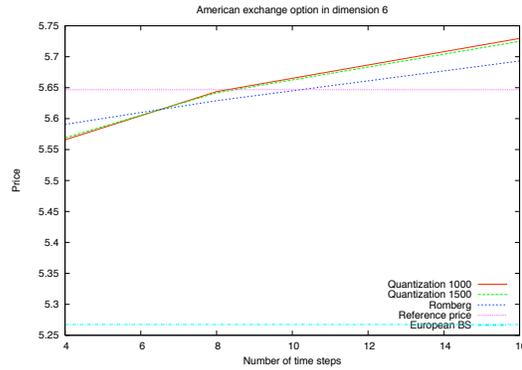


Figure 6: Richardson-Romberg for $d = 6$.

Figures 4, 5 and 6 are obtained from Figures 1, 2 and 3 with application of (30). These graphs show that a spatial Richardson-Romberg method stabilises the results. The price becomes less sensitive to the time discretization step. And therefore the premium estimation is much more accurate.

Remark 4. An additional attribute of Richardson-Romberg method is the following: we observe in Figures 4, 5 and 6, for \bar{N}_1 and \bar{N}_2 fixed, a translation of n_{opt} to the right. This feature is interesting for high dimensions as a minimal number of time steps is necessary to approximate the diffusion.

So, we recommend a spatial Richardson-Romberg application for dimension above or equal to 4 in the pricing of American exchange option.

5.2 Swing options

We will consider in the section three different ways of modeling for the underlying dynamics, which all fulfil the assumptions of the quantization tree algorithm. These are namely the 1- and 2-factor Gaussian model and an exponential Lévy model.

From now on, we assume an equidistant time discretization $0 = t_0 < t_1 < \dots < t_n = T$, i.e. $t_k = k\Delta t, k = 0, \dots, n$ with $\Delta t := T/n$.

5.2.1 Gaussian 1-Factor Model

In this case, we consider a dynamic of the underlying forward curve $(F_{t,T})_{t \in [0,T]}$ given by the SDE

$$dF_{t,T} = \sigma e^{-\alpha(T-t)} F_{t,T} dW_t,$$

which yields

$$S_t = F_{0,t} \exp\left(\sigma \int_0^t e^{-\alpha(t-s)} dW_s - \frac{1}{2} \Delta_t^2\right)$$

for

$$\Delta_t^2 = \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha t}).$$

If we set $X_k := \int_0^{k\Delta t} e^{-\alpha(k\Delta t-s)} dW_s, k = 0, \dots, n-1, (X_k)_{0 \leq k \leq n-1}$ is a discrete time Markov process and we get $(S_{k\Delta t} - K)^+ = v_k(X_k)$ with Lipschitz-continuous $v_k(x) = (F_{0,k\Delta t} \exp(\sigma x - \frac{1}{2} \Delta_{k\Delta t}^2) - K)$. So the formal requirements for the application of the quantization tree algorithm are fulfilled. For the quantization and the computation of the transition probabilities of X_k we need the following results:

Proposition 5.1. *The discrete time Ornstein-Uhlenbeck process $X_k = \int_0^{k\Delta t} e^{-\alpha(k\Delta t-s)} dW_s$ can be written as first order auto-regressive process*

$$X_{k+1} = e^{-\alpha\Delta t} X_k + \sqrt{\frac{1 - e^{-2\alpha\Delta t}}{2\alpha}} \epsilon_k, \quad k = 0, \dots, n-2, \quad X_0 := 0,$$

for an i.i.d sequence $(\epsilon_k), \epsilon_1 \sim \mathcal{N}(0, 1)$.

Especially we have

$$X_k \sim \mathcal{N}\left(0, \frac{1 - e^{-2\alpha t_k}}{2\alpha}\right).$$

Since affine transformations of a one-dimensional random variate can be transformed one-to-one on its optimal quantizers, the quantizers for X_k can be constructed as a dilatation of optimal quantizers for $\mathcal{N}(0, 1)$ by the factor $\sqrt{\frac{1 - e^{-2\alpha t_k}}{2\alpha}}$.

For the numerical results later on we have chosen the parameters

$$\sigma = 0.7, \alpha = 4, F_{0,t} = 20.$$

5.2.2 Gaussian 2-Factor Model

Furthermore we have also considered a Gaussian 2-factor model, where the dynamic of the forward curve $(F_{t,T})_{t \in [0,T]}$ is given by the SDE

$$dF_{t,T} = F_{t,T} \left(\sigma_1 e^{-\alpha_1(T-t)} dW_t^1 + \sigma_2 e^{-\alpha_2(T-t)} dW_t^2 \right)$$

for two Brownian motions W^1 and W^2 with correlation coefficient ρ .

This yields

$$S_t = F_{0,t} \exp\left(\sigma_1 \int_0^t e^{-\alpha_1(t-s)} dW_s^1 + \sigma_2 \int_0^t e^{-\alpha_2(t-s)} dW_s^2 - \frac{1}{2} \Delta_t^2\right)$$

for

$$\Delta_t^2 = \frac{\sigma_1^2}{2\alpha_1} (1 - e^{-2\alpha_1 t}) + \frac{\sigma_2^2}{2\alpha_2} (1 - e^{-2\alpha_2 t}) + 2\rho \frac{\sigma_1 \sigma_2}{\alpha_1 + \alpha_2} (1 - e^{-(\alpha_1 + \alpha_2)t}).$$

In this case we have to choose

$$X_k := \left(\int_0^{k\Delta t} e^{-\alpha_1(k\Delta t-s)} dW_s^1, \int_0^{k\Delta t} e^{-\alpha_2(k\Delta t-s)} dW_s^2 \right), k = 0, \dots, n-1,$$

as underlying Markov structure process with

$$v_k(x_1, x_2) = (F_{0,k\Delta t} \exp(\sigma_1 x_1 + \sigma_2 x_2 - \frac{1}{2} \Delta_k^2) - K).$$

Applying Proposition 5.1 on the two components of $(X_k)_{0 \leq k \leq n-1}$ allows us to write it as a first order auto-regressive process

$$X_{k+1} = A_k X_k + T_k \epsilon_k$$

with

$$A = \begin{pmatrix} e^{-\alpha_1 \Delta t} & 0 \\ 0 & e^{-\alpha_2 \Delta t} \end{pmatrix}$$

and

$$T = \begin{pmatrix} \sqrt{\frac{1}{2\alpha_1} (1 - e^{-2\alpha_1 \Delta t})} & 0 \\ \sqrt{\frac{1}{2\alpha_2} (1 - e^{-2\alpha_2 \Delta t})} r & \sqrt{\frac{1}{2\alpha_2} (1 - e^{-2\alpha_2 \Delta t})} \sqrt{1 - r^2} \end{pmatrix}$$

where

$$r = \rho \frac{\frac{1}{\alpha_1 + \alpha_2} (1 - e^{-(\alpha_1 + \alpha_2) \Delta t})}{\sqrt{\frac{1}{4\alpha_1 \alpha_2} (1 - e^{-2\alpha_1 \Delta t}) (1 - e^{-2\alpha_2 \Delta t})}}.$$

For the numerical results we have chosen the parameters

$$\sigma_1 = 0.36, \alpha_1 = 0.21, \sigma_2 = 1.11, \alpha_2 = 5.4, \rho = -0.11, F_{0,t} = 20.$$

5.2.3 NIG-Model

In this case we assume, that the dynamics of the underlying is driven by the exponential of a special Lévy process, the so-called *Normal Inverse Gaussian* process (NIG). This means, that we assume, that the dynamic of the underlying is given by

$$S_t = S_0 \exp(L_t)$$

for a NIG process $(L_t)_{t \in [0, T]}$. This special model for financial data has been first proposed by Barndorff-Nielsen in [BN98] and has been later also applied to financial contracts on energy markets (see e.g. [BFK07] and [BSB04]), on which we also aim in this paper.

The NIG process has beneath its Lévy property, which makes it a Markov process, the convenient property, that it is completely determined by its distribution at time $t=1$, the so-called NIG distribution, $\text{NIG}(\alpha, \beta, \delta, \mu)$, with parameters $\alpha > 0, |\beta| < \alpha, \delta > 0$ and $\mu \in \mathbb{R}$, i.e.

$$L_1 \sim \text{NIG}(\alpha, \beta, \delta, \mu) \quad \text{and} \quad L_t \sim \text{NIG}(\alpha, \beta, t\delta, t\mu)$$

Thus, as soon as we have knowledge on some properties of the NIG distribution like the density or the characteristic function, we know it already for $(L_t)_{t \in [0, T]}$ at any timescale.

So, the density of $(L_t)_{t \in [0, T]}$ is for example given by

$$f_t^{\text{NIG}(\alpha, \beta, \delta, \mu)}(x) = \alpha \delta t e^{t\delta \sqrt{\alpha^2 - \beta^2} + \beta(x - t\mu)} \frac{K_1(\alpha \sqrt{t^2 \delta^2 + (x - t\mu)^2})}{\pi \sqrt{t^2 \delta^2 + (x - t\mu)^2}},$$

where K_1 denotes the Bessel function of third kind with index 1.

These facts will help to the computation of optimal quantizers for L_t at given time-points $\{k\Delta t, k = 0, \dots, n-1\}$.

Another useful property of the NIG process is the fact, that if we model our underlying as

$$S_t = S_0 \exp(L_t)$$

with respect to the real-world measure \mathbb{P} , we can use the Esscher transform to construct an equivalent martingal measure \mathbb{Q} (see [GS94] and [GS95]), for a first occurrence of this method, which preserve the NIG structure of the driving Lévy process.

Hence the change from \mathbb{P} to \mathbb{Q} can be completely performed by adjusting the parameters α, β, δ and μ of the NIG process, so that from a numerical point of view it makes no difference, if we are modeling under the real-world measure or the risk-neutral one.

5.2.4 Computation of optimal quantizers for the NIG distribution

Recall that we may rewrite the L^2 -quantization problem of the NIG distribution as

$$\min_{x \in \mathbb{R}^N} \sum_{i=1}^N \int_{C_i(x)} (\xi - x_i)^2 f^{\text{NIG}}(\xi) d\xi \quad (31)$$

where $\{C_i(x), i = 1, \dots, N\}$ denotes a Voronoi partition induced by $x = (x_1, \dots, x_N)$.

The fact, that in the one-dimensional setting the Voronoi cells $C_i(x)$ are just intervals in \mathbb{R} and the uniqueness of the optimal quantizer due to the unimodality of f^{NIG} , makes the quantization problem in this case very straightforward to solve.

Now assume $x = (x_1, \dots, x_N) \in \mathbb{R}^N$ to be ordered increasingly and denote by

$$x_{1/2} := -\infty, \quad x_{i\pm 1/2} := \frac{x_i + x_{i\pm 1}}{2} \quad \text{for } 2 \leq i \leq N-1, \quad x_{N+1/2} := +\infty$$

the midpoints between the quantizer elements respectively $\pm\infty$. A Voronoi partition of Γ is therefore given by

$$\begin{aligned} C_1(x) &= (-\infty, x_{1+1/2}), \\ C_i(x) &= (x_{i-1/2}, x_{i+1/2}], \quad 2 \leq i \leq N-1, \\ C_N(x) &= (x_{N-1/2}, +\infty). \end{aligned}$$

so that (31) now reads

$$\min_{x \in \mathbb{R}^N} \sum_{i=1}^N \int_{x_{i-1/2}}^{x_{i+1/2}} (\xi - x_i)^2 f^{\text{NIG}}(\xi) d\xi. \quad (32)$$

This is a N -dimensional optimization problem, which can be easily solved by Newton's method as soon as we have access to the first and second order derivatives of D_N .

In fact, we can calculate the first order derivative of D_N (see e.g. [GL00], Lemma 4.10 or [P82], Lemma C) as

$$\frac{\partial D_N}{\partial x_i}(x) = 2 \int_{x_{i-1/2}}^{x_{i+1/2}} (x_i - \xi) f^{\text{NIG}}(\xi) d\xi, \quad 1 \leq i \leq N. \quad (33)$$

Moreover the Hessian matrix turns out to be a symmetric tridiagonal matrix with diagonal entries

$$\begin{aligned}\frac{\partial^2 D_N}{\partial x_1^2}(x) &= 2 \int_{x_{1/2}}^{x_{1+1/2}} f^{\text{NIG}}(\xi) d\xi - \frac{x_2 - x_1}{2} f^{\text{NIG}}(x_{1+1/2}) \\ \frac{\partial^2 D_N}{\partial x_i^2}(x) &= 2 \int_{x_{i-1/2}}^{x_{i+1/2}} f^{\text{NIG}}(\xi) d\xi - \frac{x_i - x_{i-1}}{2} f^{\text{NIG}}(x_{i-1/2}) - \frac{x_{i+1} - x_i}{2} f^{\text{NIG}}(x_{i+1/2}), \quad 2 \leq i \leq N \\ \frac{\partial^2 D_N}{\partial x_N^2}(x) &= 2 \int_{x_{N-1/2}}^{x_{N+1/2}} f^{\text{NIG}}(\xi) d\xi - \frac{x_N - x_{N-1}}{2} f^{\text{NIG}}(x_{N-1/2})\end{aligned}\tag{34}$$

and the super- respectively sub-diagonals

$$\frac{\partial^2 D_N}{\partial x_i \partial x_{i-1}}(x) = \frac{\partial^2 D_N}{\partial x_{i-1} \partial x_i}(x) = -\frac{x_i - x_{i-1}}{2} f^{\text{NIG}}(x_{i-1/2}), \quad 2 \leq i \leq N.$$

As a consequence of (33), the remaining entries vanish

$$\frac{\partial^2 D_N}{\partial x_i \partial x_j}(x) = \frac{\partial^2 D_N}{\partial x_j \partial x_i}(x) = 0, \quad 1 \leq i < j - 1 \leq N - 1.$$

For the evaluation of the integrals occurring in the expressions (33) and (34) we employed high-order numerical integration methods which gave satisfying results.

But some attention should be paid to the initialization of the Newton's method, since it converges only locally. We achieved a fast convergence using a equidistant placed N -tuple in the interval $[\mathbb{E}L_{t_k} - 2 \text{Var} L_{t_k}, \mathbb{E}L_{t_k} + 2 \text{Var} L_{t_k}]$ as starting vector, i.e.

$$\left[t_k \left(\mu + \frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}} - 2 \frac{\delta\alpha^2}{(\alpha^2 - \beta^2)^{3/2}} \right), \quad t_k \left(\mu + \frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}} + 2 \frac{\delta\alpha^2}{(\alpha^2 - \beta^2)^{3/2}} \right) \right]$$

for the case of the NIG, so that we reached the stopping criterion of $\|\nabla D_N(x^*)\| \leq 10^{-8}$ within 20 – 30 iterations.

5.2.5 Simulation of ΔL_t

For the simulation of a NIG($\alpha, \beta, \delta, \mu$) process it is useful to regard L_t as a subordinated Brownian Motion, i.e.

$$L_t = \beta\delta^2 I_t + \delta W_{I_t} + \mu t,$$

for an *Inverse Gaussian* process I_t , that is again a Lévy process with Inverse Gaussian distribution, IG(a, b), and parameters $a = t$ and $b = \sqrt{\alpha^2 - \beta^2}$.

A way to simulate the IG(a, b) distribution was proposed in [CT03] page 182.

To arrive now at a simulation of a whole sample path of $(L_t)_{t \in [0, T]}$ at given time-points $\{k\Delta t, k = 0, \dots, n - 1\}$, recall that by definition of a Lévy process we have

$$L_{k\Delta t} = L_0 + \sum_{j=1}^k L_{j\Delta t} - L_{(j-1)\Delta t}$$

with

$$L_{k\Delta t} - L_{(k-1)\Delta t} \sim L_{\Delta t}.$$

Hence by simulating independent increments $L_{\Delta t}$ we get a simulation of the whole sample path (see [CT03] page 184).

Thus $(L_{k\Delta t})_{0 \leq k \leq n}$ gives a simulated trajectory of a NIG process with parameters $\alpha, \beta, \delta, \mu$.

5.2.6 Parameters

For the numerical results we have chosen the parameters

$$\alpha = 50, \quad \beta = -2.0, \quad \delta = 0.02, \quad \mu = 0.001, \quad s_0 = 20.$$

5.2.7 Numerical results

In the above explained models we performed some test on the pricing of Swing options with parameters

$$n = 30, \quad Q = (100, 150) \quad \text{and} \quad n = 365, \quad Q = (1300, 1900)$$

for local constraints

$$q_{\min} = 0, \quad q_{\max} = 6$$

and strikes

$$K = 5, 10, 15, 20.$$

In the tables below, we present numerical data for Gaussian 1-factor and NIG Lévy models. For the Gaussian 2-factors model, the experimental results are a bit less decisive. As benchmark, we used the case without global constraints, i.e. $Q=(0, n)$, since as shown in Section 3.3.1, a reference price can be calculated as sum of plain call options. For the computation of the transition probabilities, we implemented as far as possible all the methods proposed in Section 4.1.

Note especially that due to the high accuracy of the deterministic integration methods in dimension one for estimation of the transition probabilities by the completely Deterministic method (cD), we may assume that the for the Gaussian 1-factor model, the error curve of this method represents solely the error of the quantization tree algorithm and we can therefore distinguish between that error and the one, which is caused by the estimation of the transition probabilities by the remaining methods proposed in Section 4.1 (see Figures 7 - 10).

Furthermore we observe, that the Diffusion method (dif) error curve is in all cases very close to the completely deterministic one. So it offers in our results the smallest quantization error.

Unfortunately it does not allow a layer-wise parallelization of the transition probabilities computations as the Fast Parallel Quantization method (fpQ). The price of this feature to pay is a loss in the consistency of the Monte Carlo estimates, which results in a slightly larger quantization error compared to the diffusion method.

Solely in the “in-the-money”-case ($K = 20$) of the NIG Lévy model (see Figure 10), we see a substantial larger loss in accuracy.

The error behavior of the “Methode des Gerbes” (MG) reveals a different convergence speed, which is induced by the systematic error replacing X_k by \hat{X}_k in (17).

Nevertheless, the error of the deterministic “Methode des Gerbes” (dMG) behaves much more stable than the Monte Carlo based methods, which makes it more important for an extrapolation attempt, (see Section 4.2).

The sample size for the Monte Carlo based estimates was chosen as $M = 10^7$.

5.2.8 Numerical data

N	$K = 5$			
	dMG	cD	fpQ	dif
50	2338.14	2338.41	2338.28	2338.26
75	2338.19	2338.32	2338.18	2338.16
100	2338.22	2338.29	2338.16	2338.13
125	2338.23	2338.27	2338.13	2338.11
150	2338.23	2338.27	2338.13	2338.11
175	2338.24	2338.26	2338.12	2338.11
200	2338.24	2338.26	2338.12	2338.10
250	2338.24	2338.26	2338.12	2338.10
300	2338.25	2338.25	2338.12	2338.10

Table 3: Gaussian 1-Factor-model for $n = 30$ and $Q = (100, 150)$

N	$K = 5$							
	dMG	rel. Err.	cD	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50	2699.67	0.0121%	2699.90	0.0039%	2699.72	0.0103%	2699.73	0.0100%
75	2699.85	0.0057%	2699.95	0.0017%	2699.77	0.0085%	2699.79	0.0079%
100	2699.91	0.0033%	2699.97	0.0010%	2699.81	0.0072%	2699.81	0.0071%
125	2699.95	0.0019%	2699.99	0.0005%	2699.81	0.0071%	2699.82	0.0066%
150	2699.96	0.0015%	2699.99	0.0004%	2699.81	0.0072%	2699.82	0.0065%
175	2699.97	0.0011%	2699.99	0.0003%	2699.81	0.0072%	2699.83	0.0064%
200	2699.98	0.0009%	2699.99	0.0002%	2699.82	0.0068%	2699.83	0.0063%
250	2699.98	0.0006%	2700.00	0.0002%	2699.82	0.0066%	2699.83	0.0062%
300	2699.99	0.0004%	2700.00	0.0001%	2699.82	0.0067%	2699.83	0.0062%

Table 4: Gaussian 1-Factor-model for $n = 30$ and $Q = (0, 180)$

N	$K = 5$			
	dMG	cD	fpQ	dif
50	29357.00	29389.62	29386.93	29388.69
75	29355.40	29372.52	29369.92	29371.59
100	29355.75	29366.25	29364.07	29365.31
125	29356.34	29362.52	29360.49	29361.58
150	29356.53	29361.64	29359.28	29360.71
175	29356.79	29360.65	29358.34	29359.71
200	29356.99	29360.00	29357.66	29359.06
250	29357.25	29359.24	29356.82	29358.30
300	29357.41	29358.82	29356.59	29357.88

Table 5: Gaussian 1-Factor-model for $n = 365$ and $Q = (1300, 1900)$

N	$K = 5$							
	dMG	rel. Err.	cD	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50	32827.47	0.0686%	32848.71	0.0039%	32845.65	0.0132%	32847.59	0.0073%
75	32836.90	0.0399%	32849.42	0.0018%	32846.51	0.0106%	32848.30	0.0052%
100	32841.59	0.0256%	32849.67	0.0010%	32847.23	0.0084%	32848.55	0.0044%
125	32844.88	0.0156%	32849.82	0.0006%	32847.52	0.0075%	32848.70	0.0040%
150	32845.72	0.0130%	32849.85	0.0004%	32847.17	0.0086%	32848.73	0.0039%
175	32846.73	0.0099%	32849.89	0.0003%	32847.28	0.0083%	32848.77	0.0037%
200	32847.42	0.0078%	32849.92	0.0003%	32847.24	0.0084%	32848.80	0.0037%
250	32848.28	0.0052%	32849.95	0.0002%	32847.22	0.0085%	32848.83	0.0036%
300	32848.77	0.0037%	32849.96	0.0001%	32847.45	0.0078%	32848.85	0.0035%

Table 6: Gaussian 1-Factor-model for $n = 365$ and $Q = (0, 2190)$

N	$K = 5$		
	dMG	fpQ	dif
50	2720.46	2721.75	2720.91
75	2720.80	2721.90	2720.92
100	2720.88	2721.27	2720.92
125	2720.92	2721.90	2720.92
150	2720.93	2722.01	2720.92
175	2720.93	2721.37	2720.92
200	2720.94	2721.14	2720.92
250	2720.94	2721.66	2720.93
300	2720.94	2721.06	2720.93

Table 7: exponential Lévy-NIG model for $n = 30$ and $Q = (100, 150)$

N	$K = 5$					
	dMG	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50	2720.46	0.0179%	2721.75	0.0295%	2720.91	0.0012%
75	2720.80	0.0053%	2721.90	0.0349%	2720.92	0.0010%
100	2720.88	0.0023%	2721.27	0.0118%	2720.92	0.0009%
125	2720.92	0.0010%	2721.90	0.0352%	2720.92	0.0008%
150	2720.93	0.0008%	2722.01	0.0391%	2720.92	0.0008%
175	2720.93	0.0005%	2721.37	0.0155%	2720.92	0.0008%
200	2720.94	0.0004%	2721.14	0.0069%	2720.92	0.0008%
250	2720.94	0.0002%	2721.66	0.0263%	2720.93	0.0008%
300	2720.94	0.0001%	2721.06	0.0043%	2720.93	0.0008%

Table 8: exponential Lévy-NIG model for $n = 30$ and $Q = (0, 180)$

N	K = 5		
	dMG	fpQ	dif
50	1.59348E23	37923.88	36184.02
75	3.04967E15	37037.65	36185.10
100	3060198091	36665.25	36185.50
125	61922.96	36513.57	36185.73
150	40574.91	36381.55	36185.80
175	39492.80	36301.25	36185.86
200	39298.66	36263.60	36185.89
250	39022.80	36315.16	36185.94
300	38814.24	36251.46	.

Table 9: exponential Lévy-NIG model for $n = 365$ and $Q = (1300, 1900)$

N	dMG	rel. Err.	K = 5			
			fpQ	rel. Err.	dif	rel. Err.
50	1.59348E23	4.40228E20%	37923.88	4.7716%	36184.02	0.0351%
75	3.073E15	8.48971E12%	37037.77	2.3235%	36185.10	0.0321%
100	3159246147	8727886.57%	36665.94	1.2963%	36185.51	0.0310%
125	62940.69	73.8850%	36514.43	0.8777%	36185.74	0.0304%
150	40615.87	12.2087%	36382.38	0.5129%	36185.80	0.0302%
175	39493.80	9.1087%	36302.01	0.2908%	36185.86	0.0300%
200	39298.87	8.5702%	36264.23	0.1865%	36185.90	0.0299%
250	39023.04	7.8082%	36315.63	0.3285%	36185.95	0.0298%
300	38814.60	7.2323%	36251.81	0.1521%	.	. %

Table 10: exponential Lévy-NIG model for $n = 365$ and $Q = (0, 2190)$

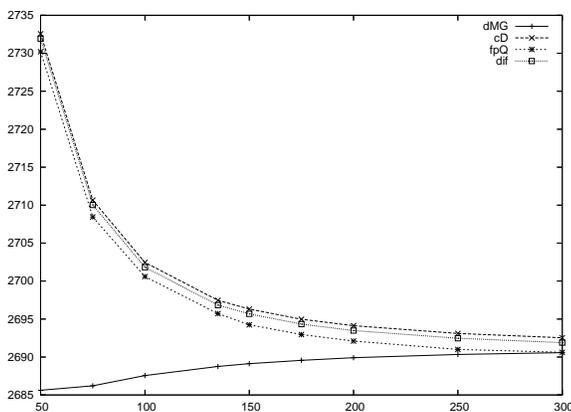
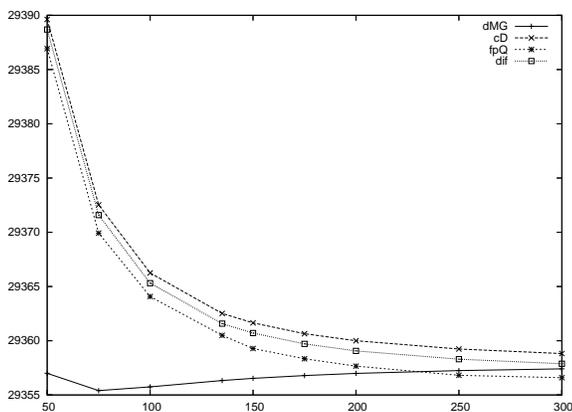


Figure 7: Gaussian 1-Factor-model for $n = 365$, $Q = (1300, 1900)$ and $K = 5$ Figure 8: Gaussian 1-Factor-model for $n = 365$, $Q = (1300, 1900)$ and $K = 20$

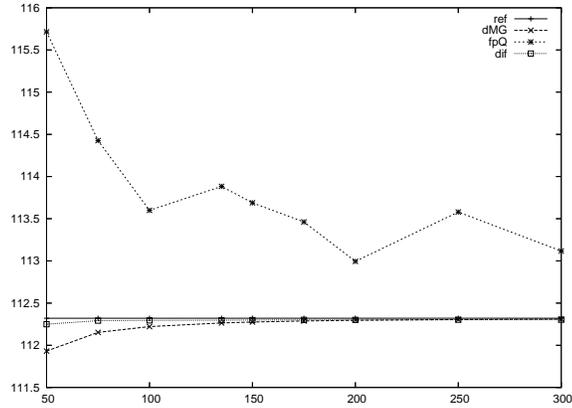
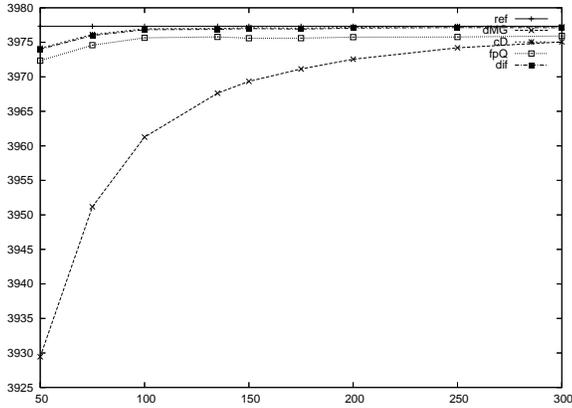


Figure 9: Gaussian 1-Factor-model for $n = 365$, Figure 10: exponential Lévy NIG model for $n = 30$, $Q = (0, 2190)$ and $K = 20$

5.2.9 Numerical data

N	$K = 5$				$K = 10$			
	dMG	cD	fpQ	dif	dMG	cD	fpQ	dif
50 - 100	2338.24	2338.25	2338.12	2338.09	1588.54	1588.55	1588.42	1588.40
100 - 200	2338.25	2338.25	2338.11	2338.09	1588.55	1588.55	1588.41	1588.39

N	$K = 15$				$K = 20$			
	dMG	cD	fpQ	dif	dMG	cD	fpQ	dif
50 - 100	862.19	862.21	862.09	862.07	224.96	224.97	224.87	224.87
100 - 200	862.19	862.19	862.07	862.06	224.96	224.96	224.86	224.86

Table 11: Gaussian 1-Factor-model for $n = 30$ and $Q = (100, 150)$

N	$K = 5$							
	dMG	rel. Err.	cD	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	2699.99	0.0004%	2700.00	0.0000%	2699.83	0.0062%	2699.84	0.0061%
100 - 200	2700.00	0.0001%	2700.00	0.0000%	2699.82	0.0067%	2699.84	0.0061%

N	$K = 10$							
	dMG	rel. Err.	cD	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	1800.32	0.0006%	1800.33	0.0000%	1800.16	0.0093%	1800.16	0.0091%
100 - 200	1800.32	0.0001%	1800.33	0.0000%	1800.14	0.0101%	1800.16	0.0091%

N	$K = 15$							
	dMG	rel. Err.	cD	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	937.34	0.0011%	937.35	0.0026%	937.21	0.0124%	937.23	0.0109%
100 - 200	937.32	0.0007%	937.33	0.0004%	937.17	0.0166%	937.20	0.0140%

N	$K = 20$							
	dMG	rel. Err.	cD	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	320.28	0.0093%	320.30	0.0140%	320.26	0.0030%	320.26	0.0030%
100 - 200	320.24	0.0024%	320.25	0.0017%	320.20	0.0143%	320.21	0.0130%

Table 12: Gaussian 1-Factor-model for $n = 30$ and $Q = (0, 180)$

N	$K = 5$				$K = 10$			
	dMG	cD	fpQ	dif	dMG	cD	fpQ	dif
50 - 100	29355.33	29358.46	29356.46	29357.51	19878.93	19882.36	19880.35	19881.41
100 - 200	29357.40	29357.92	29355.52	29356.98	19881.21	19881.81	19879.41	19880.86
N	$K = 15$				$K = 20$			
	dMG	cD	fpQ	dif	dMG	cD	fpQ	dif
50 - 100	10709.54	10713.79	10711.88	10712.97	2688.20	2692.42	2690.72	2691.78
100 - 200	10712.12	10712.77	10710.51	10711.96	2690.69	2691.35	2689.27	2690.71

Table 13: Gaussian 1-Factor-model for $n = 365$ and $Q = (1300, 1900)$

N	$K = 5$							
	dMG	rel. Err.	cD	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	32846.29	0.0113%	32849.99	0.0000%	32847.76	0.0068%	32848.88	0.0034%
100 - 200	32849.37	0.0019%	32850.00	0.0000%	32847.25	0.0084%	32848.88	0.0034%
N	$K = 10$							
	dMG	rel. Err.	cD	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	21900.05	0.0183%	21904.06	0.0000%	21901.82	0.0102%	21902.94	0.0051%
100 - 200	21903.36	0.0032%	21904.06	0.0000%	21901.31	0.0125%	21902.94	0.0051%
N	$K = 15$							
	dMG	rel. Err.	cD	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	11407.65	0.0453%	11413.12	0.0026%	11411.04	0.0156%	11412.24	0.0051%
100 - 200	11411.88	0.0083%	11412.77	0.0005%	11410.25	0.0225%	11411.88	0.0082%
N	$K = 20$							
	dMG	rel. Err.	cD	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	3971.88	0.1371%	3977.90	0.0142%	3976.76	0.0144%	3977.76	0.0108%
100 - 200	3976.31	0.0257%	3977.26	0.0018%	3975.78	0.0389%	3977.13	0.0052%

Table 14: Gaussian 1-Factor-model for $n = 365$ and $Q = (0, 2190)$

N	$K = 5$			$K = 10$		
	dMG	fpQ	dif	dMG	fpQ	dif
50 - 100	2721.02	2721.11	2720.93	1821.00	1821.11	1820.93
100 - 200	2720.95	2721.09	2720.93	1820.95	1821.09	1820.93
N	$K = 5$			$K = 10$		
	dMG	fpQ	dif	dMG	fpQ	dif
50 - 100	920.97	921.11	920.93	20.94	21.11	20.93
100 - 200	920.95	921.09	920.93	20.95	21.09	20.93

Table 15: exponential Lévy-NIG model for $n = 30$ and $Q = (100, 150)$

		$K = 5$				
N	dMG	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	2721.02	0.0029%	2721.11	0.0059%	2720.93	0.0007%
100 - 200	2720.95	0.0003%	2721.09	0.0053%	2720.93	0.0008%
		$K = 10$				
N	dMG	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	1821.00	0.0028%	1821.11	0.0088%	1820.93	0.0011%
100 - 200	1820.95	0.0003%	1821.09	0.0079%	1820.93	0.0011%
		$K = 15$				
N	dMG	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	921.04	0.0024%	921.18	0.0172%	921.00	0.0022%
100 - 200	921.02	0.0002%	921.16	0.0146%	921.00	0.0022%
		$K = 20$				
N	dMG	rel. Err.	fpQ	rel. Err.	dif	rel. Err.
50 - 100	112.32	0.0024%	112.89	0.5106%	112.31	0.0100%
100 - 200	112.32	0.0003%	112.79	0.4202%	112.30	0.0155%

Table 16: exponential Lévy-NIG model for $n = 30$ and $Q = (0, 180)$

		$K = 5$		$K = 10$	
N		fpQ	dif	fpQ	dif
50 - 100		36245.70	36185.99	25295.70	25235.99
100 - 200		36129.71	36186.02	25179.71	25236.02
		$K = 15$		$K = 20$	
N		fpQ	dif	fpQ	dif
50 - 100		14345.70	14285.99	3395.70	3335.99
100 - 200		14229.71	14286.02	3279.71	3336.02

Table 17: exponential Lévy-NIG model for $n = 365$ and $Q = (1300, 1900)$

$K = 5$				
N	fpQ	rel. Err.	dif	rel. Err.
50 - 100	36246.63	0.1378%	36186.00	0.0297%
100 - 200	36130.33	0.1834%	36186.03	0.0296%
$K = 10$				
N	fpQ	rel. Err.	dif	rel. Err.
50 - 100	25315.39	0.2097%	25251.64	0.0426%
100 - 200	25195.48	0.2649%	25251.66	0.0425%
$K = 15$				
N	fpQ	rel. Err.	dif	rel. Err.
50 - 100	14795.79	0.3341%	14736.50	0.0679%
100 - 200	14674.84	0.4860%	14736.50	0.0679%
$K = 20$				
N	fpQ	rel. Err.	dif	rel. Err.
50 - 100	6563.16	0.9242%	6495.77	0.1121%
100 - 200	6454.92	0.7402%	6495.83	0.1112%

Table 18: exponential Lévy-NIG model for $n = 365$ and $Q = (0, 2190)$

References

- [BPP05] BALLY V., PAGÈS G. AND PRINTEMS J. (2005), A quantization tree method for pricing and hedging multi-dimensional American options, *Mathematical Finance*, vol **15**, n^0 1, pp. 119-168.
- [BBP07a] BARDOU O., BOUTHEMY S. AND PAGÈS G., (2007), Optimal quantization for the pricing of swing options, LPMA preprint.
- [BBP07b] BARDOU O., BOUTHEMY S. AND PAGÈS G., (2007), When are swing options bang-bang and how to use it, LPMA preprint.
- [BN98] BARNDORFF-NIELSEN O.E. (1998), Processes of Normal Inverse Gaussian Type, *Finance & Stochastics*, vol **2**, pp. 41-68.
- [BSB04] BENTH F.E. AND SALTYTE-BENTH J. (2004), The Normal Inverse Gaussian distribution and spot price modelling in energy markets, *International Journal of Theoretical and Applied Finance*, vol **7**, n^0 2, pp. 177-192..
- [BFK07] BENTH F.E., FRESTAD D. AND KOEKEBAKKER S. (2007), Modeling the term structure dynamics in the Nordic Electricity swap market, Preprint.
- [D98] DUFLO M. (1998), *Algorithmes stochastiques*, coll. *Mathématiques & Applications*, Springer-Verlag, Boston.
- [CT03] CONT R. AND TANKOV P. (2003), *Financial modelling with jump processes*, Chapman & Hall / CRC Press.

- [**GS94**] GERBER H.U. AND SHIU E.S.W. (1994), Option pricing by Esscher transforms, Transactions of society of actuaries, vol **46**.
- [**GS95**] GERBER H.U. AND SHIU E.S.W. (1995), Actuarial approach to option pricing, Insurance: Mathematics and Economics, vol **16**, n^0 3, pp. 288-288(1).
- [**GG92**] GERSHO A. AND GRAY R.M. (1992), Vector quantization and signal compression, Kluwer, Boston.
- [**GL00**] GRAF S. AND LUSCHGY H. (2000), Foundations of quantization for probability distributions, Lecture Notes in Mathematics n^0 1730, Springer, Berlin.
- [**K82**] KIEFER J. (1982), Exponential rate of convergence for the Lloyd's method I, IEEE Trans. Info. Theory, Special issue on Quantization, vol **28**, n^0 2 pp. 205-210.
- [**P97**] PAGÈS G., (1997) A space vector quantization method for numerical integration, Journal of Applied and Computational Mathematics, vol **89**, pp. 1-38.
- [**PP03**] PAGÈS G. AND PRINTEMS J., (2003) Optimal quadratic quantization for numerics: the Gaussian case, Monte Carlo Methods and Applications, vol **9**, n^0 2 pp. 135-166.
- [**P82**] POLLARD D. (1982) A central limit theorem method for k -means clustering, Ann. Proba., vol **10**, n^0 4, pp. 919-926.
- [**Z82**] ZADOR P.L. (1982), Asymptotic quantization error of continuous signals and the quantization dimension, IEEE Trans. Info. Theory, Special issue on Quantization, vol **28**, n^0 2, pp. 139-149.