Optimal Delaunay and Voronoi quantization schemes for pricing American style options

Gilles Pagès and Benedikt Wilbertz

Abstract We review in this article pure quantization methods for the pricing of multiple exercise options. These quantization methods have the common advantage, that they allow a straightforward implementation of the Backward Dynamic Programming Principle for optimal stopping and stochastic control problems. Moreover we present here for the first time a unified discussion of this topic for Voronoi and Delaunay quantization and illustrate the performances of both methods by several numerical examples.

1 Introduction

This paper is focused on pure quantization method for pricing multi-asset American style options (by contrast with hybrid Monte Carlo-quantization approaches). It continues two goals: it is partly a survey on the pricing of this family of options by optimal Voronoi quantization techniques. It is also an opportunity to present our first attempt to implement in a multi-dimensional setting the new quantization method called dual (or *Delaunay*) quantization recently developed and investigated in [Pagès and Wilbertz 2010a] and [Pagès and Wilbertz 2010b]. This approach relies on the Delaunay triangulation of a grid whereas usual vector quantization relies on its Voronoi diagram, hence its name since the Delaunay triangulation is and Voronoi diagrams are in duality (see [Okabe et al. 2000]). Dual quantization has

Gilles Pagès
LPMA, UPMC, case 188, 4 pl. Jussieu, F-75252 Paris cedex 5; e-mail: gilles.pages@upmc.fr

Benedikt Wilbertz
LPMA, UPMC, case 188, 4 pl. Jussieu, F-75252 Paris cedex 5; e-mail: benedikt.wilbertz@gmx.de

been originally introduced in [Pagès and Wilbertz 2009] to compute the expectation of functionals of nonhomogenous Bernoulli random walks involved in the pricing of CDO's (in a static copula model).

Optimal Voronoi quantization, which is an old story going back to the 1950's has been originally developed for Signal transmission purpose at the Bell Laboratory, has been implemented as a numerical method for the pricing of multi-asset American – strictly speaking Bermuda – options in a series of papers [Bally et al. 2001], [Bally and Pagès 2003a], [Bally and Pagès 2003b], [Bally et al. 2003], [Bally et al. 2005]. Other fields of application have been developed, often in connection with financial problems like numerical integration [Pagès 1993], [Pagès 1998], [Pagès and Printems 2003], non-linear filtering(see [Pagès and Pham 2005], [Pham et al. 2005], [Sellami 2010], [Sellami 2009] with application to stochastic volatility models, stochastic control with application to portfolio management (see [Pagès et al. 2004]) and swing option pricing (see [Bardou et al. 2010a], [Bardou et al. 2010b]), discretization of stochastic PDE's (typically Zakaï and Mc Kean Vlasov equations, see [Gobet et al. 2007], [Gobet et al. 2005]). We also refer to the surveys [Pagès et al. 2003] and [Pagès and Printems 2009] and the references therein, as well as to the website devoted to Optimal quantization (see [Pagès and Printems 2005]).

Quantization methods consist in approximating/discretizing an $\mathbb{R}^d$-valued random vector X by a random vector often denoted $\widehat{X}$ taking values into a grid Γ of size $N \geq 1$ so as to make $\|X - \widehat{X}\|_p$ as small as possible. As concerns Voronoi quantization, $\widehat{X}$ is a projection following the nearest neighbour rule on grid Γ of size N. For dual quantization, $\widehat{X}$ is the result of a random *splitting operator* which projects X on one of the vertices of a "minimal" Γ-valued d-simplex which contains X, with a probability ruled by the barycentric coordinates of X. In a quadratic Euclidean framework optimal Voronoi quantizers satisfy the so-called stationary property $\widehat{X} = \mathbb{E}(X \,|\, \widehat{X})$ whereas *all* dual quantizers satisfy the reverse stationarity property $X = \mathbb{E}(\widehat{X} \,|\, X)$. When X has an unbounded support, one extends the splitting operator by a nearest neighbour projection outside the convex hull of the grid Γ.

In order to solve dynamic optimization problems related to a (discrete time) Markov chain $(X_k)_{0 \leq k \leq n}$, one introduces quantization trees that is quantization grids Γ_k of the marginal X_k and some transition matrices approximating the the Markov transition of the chain. The stationarity of the grids used in the quantization schemes designed on such quantization tree plays a important role to preserve the numerical efficiency/accuracy: the easiest way to get convinced is to check that such grids lead to quantization based cubature formulas of second order (see [Pagès 1993, Pagès and Wilbertz 2010a]). Although not as prominent when dealing with less linear problems (Bermuda option pricing, filtering, stochastic control, etc), stationarity turns out to be crucial when dealing with numerical implementation. Now, only optimal Voronoi quantization grid share this property whereas it is shared by all dual quantization grids. This makes dual quantization more flexi-

ble than the Voronoi one: when switching from a distribution to another like in an iterative calibration procedure, one only has to modify the weights of a dual quantization grid to preserve the stationarity (even if the resulting quantization is no longer optimal). This can be done *on line* by a regular Monte Carlo simulation in a few seconds or even less with the help of high performance massively parallel computation device (GPGPU). When dealing with Voronoi quantization, preserving stationarity requires to re-adjust both the grids and the weights.

In Section 2 we propose in a Markovian framework a unified approach to provide some *a priori* error bounds for Voronoi and Delaunay quantization schemes, relying on a non asymptotic version of Zador's theorem (about the rate of decay of the L^p-quantization error). This improves and simplifies the results in [Bally and Pagès 2003a]. The resulting bound is the (weighted) sum of the quantization errors of the marginals of the Markovian dynamics.

In Section 3, we present with more details both Voronoi and Delaunay quantization. In Section 4, we briefly describe several stochastic optimization methods to optimize grids. Those related to Voronoi quantization are classical (Lloyd's I and *CLVQ*) whereas their counterpart have been recently devised in [Pagès and Wilbertz 2010a] or completely new. In section 6, we propose methods – some of them heuristic – to optimize the structure of the quantization tree. In Section 7, numerical test are carried out on several American payoff functions (swing option, exchange option between geometric indices and call option on minimum of two assets) in a multi-dimensional setting. We determine empirically rates of convergence, discuss several improvement possibilities and finally establish a comparison with the Longstaff-Schwartz algorithm.

In this paper we only consider a (canonical) Euclidean framework although many existence and rate results hold true for general norms. Algorithmic aspects are more Euclidean dependent.

NOTATION: $|.|$ denotes the canonical Euclidean norm on the vector space $\mathbb{R}^d$ of column vectors. $\mathrm{conv}(A)$ denotes the convex hull of $A \subset \mathbb{R}^d$.

2 Quantized Backward Dynamic Programming Principle

Let $(X_k)_{0\le k\le n}$ be an $\mathbb{R}^d$-valued homogeneous Feller Markov chain defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with transition $P(x,dy)$. The homogeneity assumption is essentially made for convenience in order to to alleviate notations but the extension to a non-homogeneous framework is straightforward. We will make the slightly more stringent assumption that the chain is in fact "Lipschitz Feller": this means that the transition is not simply Feller but also preserves uniformly Lipschitz continuous functions: there exists a (finite) real constant $[P]_{\mathrm{Lip}}$ such that

$$\forall f:\mathbb{R}^d\to\mathbb{R}^d,\quad [Pf]_{\mathrm{Lip}}\le [P]_{\mathrm{Lip}}[f]_{\mathrm{Lip}}.$$

where $[f]_{\mathrm{Lip}}=\sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|}$. Without loss of generality we may assume that

$$[P]_{\mathrm{Lip}}=\sup_{[f]_{\mathrm{Lip}}\le 1}[Pf]_{\mathrm{Lip}}.$$

Let $h_k:\mathbb{R}^d\to\mathbb{R}_+$, $0\le k\le n$, be a sequence of Borel functions satisfying

$$\max_{0\le k\le n}\|h_k(X_k)\|_p<+\infty\quad\text{for a } p\in[1,\infty).$$

Let $\mathscr{F}^X=(\mathscr{F}_k^X)_{0\le k\le n}$ denote the natural filtration of the chain X. It is classical background from Optimal Stopping Theory that if one defines by induction a backward sequence of L^p-integrable random variables $(V_k)_{0\le k\le n}$ as follows

$$V_n=h_n(X_n),\quad V_k=\max\Big(h_k(X_k),\mathbb{E}\big(V_{k+1}\,|\,X_k\big)\Big)\tag{1}$$

(called the *Backward Dynamical Programming Principle* (*BDPP*)) then

$$V_0=\sup\Big\{\mathbb{E}\Big(h_\tau(X_\tau)\,|\,\mathscr{F}_0^X\Big),\ \tau:\Omega\to\{0,\dots,n\}\ \mathscr{F}^X\text{-stopping time}\Big\}$$

and more generally

$$V_k=\operatorname{esssup}\Big\{\mathbb{E}\Big(h_\tau(X_\tau)\,|\,\mathscr{F}_k^X\Big),\ \tau:\Omega\to\{k,\dots,n\}\ \mathscr{F}^X\text{-stopping time}\Big\},\ k=0,\dots,n.$$

The sequence $(V_k)_{0\le k\le n}$ is also known as the $(\mathbb{P},\mathscr{F}^X)$-*Snell envelope* of the so-called *obstacle process* $(h(X_k))_{0\le k\le n}$. From a numerical point of view, one is usually interested in $\mathbb{E}V_0$ or $\mathbb{E}V_k$.

The *paradigm* of *Quantized Backward Dynamic Programming Principle* is two folded and can be described as follows:

▷ *Discretization.* As a first step, we consider an abstract approximation process of the Markov Chain $(X_k)_{0\le k\le n}$ by a sequence $(\widehat{X}_k)_{0\le k\le n}$ of the form

$$\widehat{X}_k=\pi_k(X_k,U_k),\quad k=0,\dots,n,$$

where $(U_k)_{0\le k\le n}$ is an i.i.d. sequence of $\mathbb{R}^{d_0}$-valued random vector *independent* of $\mathscr{F}_n^X$ (*i.e.* of $(X_k)_{0\le k\le n}$) and the mappings $\pi_k:\mathbb{R}^d\times\mathbb{R}^{d_0}\to\mathbb{R}^d$ are Borel functions. As concerns numerical implementation we will of course ask the chain $(X_k)_{0\le k\le n}$ and the exogenous simulation noise $(U_k)_{0\le k\le n}$ to be simulatable (at a reasonable cost) and the mapping π_k to take values in finite sets Γ_k (called *grids*).

We will see further on that these random vectors U_k represent an exogenous noise involved in the simulation process of $\widehat{X}_k$ "from" X_k (so will be the case when dealing with *dual* quantization). One can always achieve such a framework by defining the

sequence (U_k) on a probability space $(\Omega_0, \mathscr{A}_0, \mathbb{P}_0)$ and by considering the product probability space $(\widetilde{\Omega}, \widetilde{\mathscr{A}}, \widetilde{\mathbb{P}}) = (\Omega \times \Omega_0, \mathscr{A} \otimes \mathscr{A}_0, \mathbb{P} \otimes \mathbb{P}_0)$.

▷ *Quantized Backward Dynamic Programming Principle.* As a second step, we introduce a dynamic programming formula involving the r.v. $\widehat{X}_k$, obtained by simply mimicking the regular *BDPP* related to the Snell envelope of $(h_k(X_k))_{0\le k\le n}$; in practice this essentially amounts to "forcing" the Markov property although the sequence $(\widehat{X}_k)_{0\le k\le n}$ has no reason to be a Markov chain. To be precise, we assume that $\max_{0\le k\le n} \|h_k(\widehat{X}_k)\|_p < +\infty$ for a $p \in [1,\infty)$ and we define a sequence $(\widehat{V}_k)_{0\le k\le n}$

$$\widehat{V}_n = h_n(\widehat{X}_n), \quad \widehat{V}_k = \max\Big(h_k(\widehat{X}_k), \mathbb{E}\big(\widehat{V}_{k+1} \,|\, \widehat{X}_k\big)\Big). \tag{2}$$

Then the following (new) result holds about the (strong) rate of approximation of the Snell envelope $(V_k)_{0\le k\le n}$ by its quantized counterpart $(\widehat{V}_k)_{0\le k\le n}$, having in mind that $|\mathbb{E}V_k - \mathbb{E}\widehat{V}_k| \le \|V_k - \widehat{V}_k\|_p$ for every $p \ge 1$.

Proposition 2.1 *Let $p \in [1,+\infty)$. Assume that*

$$\max_{0\le k\le n}\Big(\|X_k\|_p + \|\widehat{X}_k\|_p\Big) < +\infty$$

and that all the functions h_k, $k = 0,\dots,n$, are Lipschitz continuous.

(a) If $p = 2$, then, for every $k \in \{0,\dots,n\}$,

$$\|V_k - \widehat{V}_k\|_2 \le \sqrt{2}\left(\sum_{\ell=k}^{n}\Big(C_{n,\ell}([P]_{\mathrm{Lip}},[h_\cdot]_{\mathrm{Lip}})\Big)^2 \|X_\ell - \widehat{X}_\ell\|_2^2\right)^{\frac{1}{2}}$$

(b) If $p \ne 2$, then for every $k \in \{0,\dots,n\}$,

$$\|V_k - \widehat{V}_k\|_p \le 2\sum_{\ell=k}^{n} C_{n,\ell}([P]_{\mathrm{Lip}},[h_\cdot]_{\mathrm{Lip}})\|X_\ell - \widehat{X}_\ell\|_p$$

where

$$C_{n,k}([P]_{\mathrm{Lip}},[h_\cdot]_{\mathrm{Lip}}) = \max_{k\le\ell\le n}\Big([P]_{\mathrm{Lip}}^{\ell-k}[h_\ell]_{\mathrm{Lip}}\Big).$$

Proof. STEP 1. *The functions v_k are Lipschitz.* One first shows by induction using the Markov property that

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

where the functions v_k are Lipschitz continuous satisfying

$$v_n = h_n \quad \text{and} \quad v_k = \max(h_k, Pv_{k+1}),\ k = 0,\dots,n-1.$$

In particular, for every $k = 0,\dots,n$ (with the convention $[v_{n+1}]_{\mathrm{Lip}} = 0$),

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

where we used the elementary inequality $|\sup_{i\in I} a_i - \sup_{i\in I} b_i| \le \sup_{i\in I}|a_i - b_i|$. Then standard computations yield that

$$[v_k]_{\mathrm{Lip}} \le \max_{k\le \ell\le n}\left([P]^{\ell-k}_{\mathrm{Lip}}[h_\ell]_{\mathrm{Lip}}\right).$$

(a) From now on, we focus on the quadratic case $p=2$.

STEP 2. *Induction on* $\|V_k - \widehat{V}_k\|_2^2$. It follows from the quantized *BDPP* that

$$\widehat{V}_k = \widehat{v}_k(\widehat{X}_k) \ \text{ where } \ \widehat{v}_k : \mathbb{R}^d \to \mathbb{R}_+,\ k=0,\dots,n.$$

are Borel functions. Then

$$\begin{aligned}\|V_k - \widehat{V}_k\|_2^2 &\le \|h_k(X_k) - h_k(\widehat{V}_k)\|_2^2 + \|\mathbb{E}(V_{k+1}\,|\,X_k) - \mathbb{E}(\widehat{V}_{k+1}\,|\,\widehat{X}_k)\|_2^2\\ &\le [h_k]^2_{\mathrm{Lip}}\|X_k - \widehat{X}_k\|_2^2 + \|\mathbb{E}(V_{k+1}\,|\,X_k) - \mathbb{E}(\widehat{V}_{k+1}\,|\,\widehat{X}_k)\|_2^2.\end{aligned}$$

where we used that

$$|\max_{i=1,2} a_i - \max_{i=1,2} b_i|^2 \le \max_{i=1,2}|a_i - b_i|^2 \le \sum_{i=1,2}|a_i - b_i|^2.$$

Now, one easily checks that

$$\begin{aligned}\mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\widehat{X}_k\Big) &= \mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,U_k)\Big)\\ &= \int_{\mathbb{R}^{d_0}} \mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\Big)\mathbb{P}_{U_k}(du)\end{aligned}$$

since $\widehat{X}_k = \pi_k(X_k,U_k)$, U_k and $(\widehat{V}_{k+1},X_k)$ are independent (keep in mind that $\widehat{V}_{k+1}$ is $\sigma(\widehat{X}_{k+1})$-measurable and $\sigma(\widehat{X}_{k+1}) \subset \sigma(X_{k+1},U_{k+1})$). It follows

$$\begin{aligned}\Big\|\mathbb{E}(V_{k+1}\,|\,X_k) - \mathbb{E}(\widehat{V}_{k+1}\,|\,\widehat{X}_k)\Big\|_2^2 &= \mathbb{E}\left(\int_{\mathbb{R}^{d_0}}\Big[\mathbb{E}(V_{k+1}\,|\,X_k) - \mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\Big)\Big]\mathbb{P}_{U_k}(du)\right)^2\\ &\le \int_{\mathbb{R}^{d_0}}\mathbb{E}\Big(\mathbb{E}(V_{k+1}\,|\,X_k) - \mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\Big)\Big)^2\mathbb{P}_{U_k}(du)\\ &= \int_{\mathbb{R}^{d_0}}\Big\|\mathbb{E}(V_{k+1}\,|\,X_k) - \mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\Big)\Big\|_2^2\mathbb{P}_{U_k}(du). \quad (3)\end{aligned}$$

Now, for every $u\in\mathbb{R}^{d_0}$, one writes

$$\begin{aligned}&\mathbb{E}\big(V_{k+1}\,|\,X_k\big) - \mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\Big)\\ &\qquad= \mathbb{E}\big(V_{k+1}\,|\,X_k\big) - \mathbb{E}\Big(V_{k+1}\,|\,\pi_k(X_k,u)\Big) + \mathbb{E}\Big(V_{k+1}\,|\,\pi_k(X_k,u)\Big) - \mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\Big)\end{aligned}$$

The random variable

$$\mathbb{E}\Big(V_{k+1}\,|\,X_k\Big)-\mathbb{E}\Big(V_{k+1}\,|\,\pi_k(X_k,u)\Big)=\mathbb{E}\big(V_{k+1}\,|\,X_k\big)-\mathbb{E}\Big(\mathbb{E}\big(V_{k+1}\,|\,X_k\big)\,|\,\pi_k(X_k,u)\Big)$$

and $\mathbb{E}\Big(V_{k+1}\,|\,\pi_k(X_k,u)\Big)-\mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\Big)\in L^2(\sigma(\pi_k(X_k,u))$ are orthogonal owing to the characterization of conditional expectation as an orthogonal projection. Consequently

$$\begin{aligned}
&\Big\|\mathbb{E}\Big(V_{k+1}\,|\,X_k\Big)-\mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\Big)\Big\|_2^2\\
&\qquad\le \big\|\mathbb{E}(V_{k+1}\,|\,X_k)-\mathbb{E}\big(V_{k+1}\,|\,\pi_k(X_k,u)\big)\big\|_2^2+\Big\|\mathbb{E}\big(V_{k+1}-\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\big)\Big\|_2^2\\
&\qquad\le \big\|\mathbb{E}\big(V_{k+1}\,|\,X_k\big)-\mathbb{E}\big(\mathbb{E}\big(V_{k+1}\,|\,X_k\big)\,|\,\pi_k(X_k,u)\big)\big\|_2^2+\|V_{k+1}-\widehat{V}_{k+1}\|_2^2\\
&\qquad= \big\|Pv_{k+1}(X_k)-\mathbb{E}\big(Pv_{k+1}(X_k)\,|\,\pi_k(X_k,u)\big)\big\|_2^2+\|V_{k+1}-\widehat{V}_{k+1}\|_2^2 \qquad (4)
\end{aligned}$$

where we successively used in the last two lines the facts that conditional expectation is an L^p-contraction and that $\mathbb{E}\big(V_{k+1}\,|\,X_k\big)=\mathbb{E}\big(v_{k+1}(X_{k+1})\,|\,X_k\big)=Pv_{k+1}(X_k)$. Now, going back to the very definition of conditional expectation,

$$\big\|Pv_{k+1}(X_k)-\mathbb{E}\big(Pv_{k+1}(X_k)\,|\,\pi_k(X_k,u)\big)\big\|_2\le\big\|Pv_{k+1}(X_k)-Pv_{k+1}(\pi_k(X_k,u))\big\|_2$$

so that finally

$$\begin{aligned}
\Big\|\mathbb{E}(V_{k+1}\,|\,X_k)-\mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\Big)\Big\|_2^2&\le\Big\|\widehat{V}_{k+1}-V_{k+1}\Big\|_2^2\\
&\quad+\big\|Pv_{k+1}(X_k)-Pv_{k+1}(\pi_k(X_k,u))\big\|_2^2\\
&\le\Big\|V_{k+1}-\widehat{V}_{k+1}\Big\|_2^2+[Pv_{k+1}]_{\mathrm{Lip}}^2\,\|X_k-\pi_k(X_k,u)\|_2^2. \qquad (5)
\end{aligned}$$

On the other hand, Fubini's Theorem implies

$$\begin{aligned}
\int_{\mathbb{R}^{d_0}}\|X_k-\pi_k(X_k,u)\|_2^2\,\mathbb{P}_{U_k}(du)&=\int_{\mathbb{R}^{d_0}}\big(\mathbb{E}|X_k-\pi_k(X_k,u)|^2\big)\,\mathbb{P}_{U_k}(du)\\
&\le\mathbb{E}\left(\int_{\mathbb{R}^{d_0}}|X_k-\pi_k(X_k,u)|^2\mathbb{P}_{U_k}(du)\right)\\
&=\Big(\mathbb{E}|X_k-\pi_k(X_k,U_k)|^2\Big)^{\frac12}\\
&=\|X_k-\widehat{X}_k\|_2^2.
\end{aligned}$$

Consequently, plugging this bound in the $\mathbb{P}_U$-integrated form of (5) and the resulting inequality in (3), yields

$$\|V_k-\widehat{V}_k\|_2^2\le\|V_{k+1}-\widehat{V}_{k+1}\|_2^2+\big([h_k]_{\mathrm{Lip}}^2+[Pv_{k+1}]_{\mathrm{Lip}}^2\big)\,\|X_k-\widehat{X}_k\|_2^2.$$

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

$$\begin{aligned}\|V_k-\widehat{V}_k\|_2^2 &\le \sum_{\ell=k}^{n}\left([h_\ell]_{\mathrm{Lip}}^2+[P]_{\mathrm{Lip}}^2[v_{\ell+1}]_{\mathrm{Lip}}^2\right)\|X_\ell-\widehat{X}_\ell\|_2^2\\ &\le 2\sum_{\ell=k}^{n}\Big(C_{n,\ell}([P]_{\mathrm{Lip}},[h_\cdot]_{\mathrm{Lip}})\Big)^2\|X_\ell-\widehat{X}_\ell\|_2^2\end{aligned}$$

owing to the upper bound established in Step 1 for $[v_k]_{\mathrm{Lip}}$.

(b) One mimicks the proof of the above claim (a) but dealing now with $\|X_k-\widehat{X}_k\|_p$ and relying on the generalized Minkowski inequality to establish the counterpart of (3). Then on replaces (4) by

$$\begin{aligned}\Big\|\mathbb{E}\Big(V_{k+1}\,|\,X_k\Big)-\mathbb{E}\Big(\widehat{V}_{k+1}\,|\,\pi_k(X_k,u)\Big)\Big\|_p &\le 2\,\big\|Pv_{k+1}(X_k)-\mathbb{E}\big(Pv_{k+1}(X_k)\,|\,\pi_k(X_k,u)\big)\big\|_p\\ &\quad+\|V_{k+1}-\widehat{V}_{k+1}\|_p.\end{aligned}$$

Finally, one checks that

$$\big\|Pv_{k+1}(X_k)-\mathbb{E}\big(Pv_{k+1}(X_k)\,|\,\pi_k(X_k,u)\big)\big\|_p \le 2\,\|Pv_{k+1}(X_k)-Pv_{k+1}(X_k)\|_p$$

and the conclusion follows. $\diamondsuit$

Example. We consider a jump diffusion solution to

$$dY_t=b(t,Y_t)dt+\sigma(t,Y_t)dW_t+\kappa(t,Y_{t-})dZ_t,$$

where $W=(W_t)_{t\in[0,T]}$ is an l-dimensional standard Brownian motion and $Z=(Z_t)_{t\in[0,T]}$ is an l-dimensional square integrable compensated Lévy process without Brownian component (so that its Lévy measure ν satisfies $\displaystyle\int_{\mathbb{R}^l}|z|^2\nu(dz)<+\infty$).

The processes W and Z are defined on a probability space $(\Omega,\mathscr{A},\mathbb{P})$ and are supposed to be independent. In particular, Z_t is centered, has a second moment and both

$$(Z_t)_{t\in[0,T]}\ \text{ and }\quad \Big(Z_tZ_t^*-t\mathbb{E}(Z_1Z_1^*)\Big)_{t\in[0,T]}$$

are $\mathscr{F}_t^{W,Z}$-martingales (Z_t^* stands for the transpose of Z_t). Assume that $b:[0,T]\times\mathbb{R}^d\to\mathbb{R}$, σ, $\kappa:[0,T]\times\mathbb{R}^d\to\mathscr{M}(d,q)$ are Lipschitz continuous functions in (t,x) (these assumptions are not optimal).

Under these assumptions, the above *SDE* has a strong solution starting from any finite random vector Y_0 independent of (W,Z) defined on $(\Omega,\mathscr{A},\mathbb{P})$.

The "sampled process" $(Y_{t_k^n})_{0\le k\le n}$ at the discretization times $t_k^n=\frac{kT}{n}$, $k=0,\dots,n$, is an homogenous Markov chain with transition $P^{(n)}:=P_{\frac{T}{n}}$ formally reading

$$P_{\frac{T}{n}}(f)(x) = \mathbb{E}_x\Big(f\big(Y_{\frac{T}{n}}\big)\Big).$$

Such a Markov chain is usually not simulatable. However one may always associate to such a diffusion process its Euler scheme with step $\frac{T}{n}$ recursively defined by $\bar{Y}_0 = Y_0$ and, for every $k \in \{0,\dots,n-1\}$,

$$\bar{Y}_{t_{k+1}^n} = \bar{Y}_{t_k^n} + \frac{T}{n} b(t_k^n, Y_{t_k^n}) + \sigma(t_k^n, Y_{t_k^n})(W_{t_{k+1}^n} - W_{t_k^n}) + \kappa(t_k^n, Y_{t_k^n})(Z_{t_{k+1}^n} - Z_{t_k^n}).$$

The sequence $(\bar{Y}_{t_k^n})_{0\le k\le n}$ is a homogeneous Markov chain with transition $\bar{P}^{(n)}$ reading on bounded or non-negative Borel functions f,

$$\bar{P}^{(n)}(f)(x) = \mathbb{E}\left(f\Big(x + b(x)\frac{T}{n} + \sigma(x)\sqrt{\frac{T}{n}}\Xi + \kappa(x)Z_{\frac{T}{n}}\Big)\right) \tag{6}$$

where $\Xi \sim \mathcal{N}(0;I_q)$ is independent of $Z_{\frac{T}{n}}$. For notational convenience we will often note $\bar{P}$ for $\bar{P}^{(n)}$.

Standard computations show that if f is Lipschitz continuous

$$|\bar{P}^{(n)}(f)(x) - \bar{P}^{(n)}(f)(x')|^2 \le [f]_{\mathrm{Lip}}^2\left(1 + [b]_{\mathrm{Lip}}^2\Big(\frac{T}{n}\Big)^2 + C_{\sigma,\kappa,d,Z}\frac{T}{n}\right)|x-x'|^2$$

where $C_{b,\sigma,d,Z} = d[\sigma]_{\mathrm{Lip}}^2 + [\kappa]_{\mathrm{Lip}}^2\mathbb{E}|Z_1|^2$. Similar bounds can be obtained for the jump diffusion at time $\frac{T}{n}$ using Itô's formula with jumps. This leads to the following proposition.

Proposition 2.2 *There exists a real constant $C_{b,\sigma,\kappa,T,d,Z}$ such that,*

$$\forall n \ge 1, \quad [P_{\frac{T}{n}}]_{\mathrm{Lip}} \le 1 + C_{b,\sigma,\kappa,T,d,Z}\frac{T}{n} \quad \textit{and} \quad [\bar{P}^{(n)}]_{\mathrm{Lip}} \le 1 + C_{b,\sigma,\kappa,T,d,Z}\frac{T}{n}.$$

As a consequence, if $P = P_{\frac{T}{n}}$ or $P = \bar{P}^{(n)}$

$$\sup_{n\ge1}\max_{0\le k\le n}[P]_{\mathrm{Lip}}^k \le e^{C_{b,\sigma,\kappa,T,d,Z}} < +\infty.$$

This proposition emphasizes that if one set $X_k = Y_{t_k^n}$ or $X_k = \bar{Y}_{t_k^n}$, $k = 0,\dots,n$, and if, for example, $h_k = e^{-r\frac{kT}{n}}h$, $k = 0,\dots,n$, with $h : \mathbb{R}^d \to \mathbb{R}_+$ a Lipschitz continuous function, then the coefficients $C_{n,k}([P]_{\mathrm{Lip}}, [h_\cdot]_{\mathrm{Lip}})$ introduced in Proposition 2.1 remain uniformly bounded since

$$\sup_{n\ge1}\max_{0\le k\le n} C_{n,k}([P]_{\mathrm{Lip}}, [h_\cdot]_{\mathrm{Lip}}) \le e^{C_{b,\sigma,\kappa,T,d,Z}}[h]_{\mathrm{Lip}} < +\infty.$$

3 Optimal Voronoi and Delaunay quantizations

In this section we deal for a while with a *static problem*: how to optimize the quantization of a fixed $\mathbb{R}^d$-valued random vector X. This is the purpose of optimal quantization which consists in minimizing the L^p-mean approximation error induced by a quantization $\widehat{X}$ of X that takes at most N values. To be more precise, we aim at minimizing $\|X-\widehat{X}\|_p$ over a certain class of discretely valued random vectors $\widehat{X}$.

3.1 Optimal Voronoi quantization

In the case of Voronoi quantization this optimization problem reads

$$e_{p,N}(X) = \inf\Big\{ \|X-\widehat{X}\|_p : \widehat{X} \text{ is a random vector with } \#\widehat{X}(\Omega) \le N \Big\}.$$

It turns out, see *e.g.* [Graf and Luschgy 2000], that this definition is equivalent to the definition of the optimal quantization error as the minimal L^p-distance from X to a finite grid $\Gamma \subset \mathbb{R}^d$ with cardinality $\#\Gamma \le N$, i.e.

$$\begin{aligned} e_{p,N}(X) &= \inf_{\Gamma}\Big\{ \|\mathrm{dist}(X,\Gamma)\|_p : \Gamma \subset \mathbb{R}^d, \#\Gamma \le N \Big\} \\ &= \inf_{\Gamma}\Big\{ \Big(\mathbb{E}\min_{x\in\Gamma}|X-x|^p \Big)^{1/p} : \Gamma \subset \mathbb{R}^d, \#\Gamma \le N \Big\}. \end{aligned}$$

This equivalence is based on the construction of a Voronoi quantization by means of the nearest neighbour projection. Therefore, let $\Gamma = \{x_1,\dots,x_N\} \subset \mathbb{R}^d$ be a grid and denote by $(C_i(\Gamma))_{1\le i\le N}$ a Borel partition of $\mathbb{R}^d$ satisfying

$$C_i(\Gamma) \subset \big\{ \xi \in \mathbb{R}^d : |\xi - x_i| \le \min_{1\le j\le N} |\xi - x_j| \big\}.$$

Such a partition is called a *Voronoi partition* generated by Γ and we may define the corresponding *nearest neighbour projection* as

$$\mathrm{Proj}_\Gamma(\xi) = \sum_{1\le i\le N} x_i \mathbb{1}_{C_i(\Gamma)}(\xi). \tag{7}$$

The discrete random vector

$$\widehat{X}^{\Gamma,\mathrm{Vor}} = \mathrm{Proj}_\Gamma(X) = \sum_{1\le i\le N} x_i \mathbb{1}_{C_i(\Gamma)}(X).$$

is called *Voronoi Quantization* of X induced by Γ and satisfies

$$\mathbb{E}\min_{x\in\Gamma}|X-x|^p = \mathbb{E}|X-\widehat{X}^{\Gamma,\mathrm{Vor}}|^p.$$

At this stage, the purpose of optimal quantization is to prove the existence of optimal grids of size at most N which resulting quantization error attains the minimal L^p-quantization error $e_{p,N}$.

Proposition 1 (Optimal Voronoi quantizer(s)). *(see [Kieffer 1983, Graf and Luschgy 2000, Pagès 1998])* (a) *Let $p \in [1,\infty)$. For every integer $N \geq 1$, there exists at least one optimal grid Γ_N^* of size at most N (or equivalently "at level N") such that*

$$\|X - \widehat{X}^{\Gamma_N^*,Vor}\|_p = e_{p,N}(X)$$

and $N \mapsto e_{p,N}(X)$ is (strictly) decreasing to 0 *(as long as it does not vanish).*

Furthermore $e_{p,N}(X) = 0$ if and only if $\mathrm{supp}(\mathbb{P}_X)$ *has at most N elements and if this support has at least N elements, then any optimal grid Γ_N^* has exactly N pairwise distinct elements.*

(b) *If $p = 2$, any optimal Γ_N^* quantization grid satisfies the* stationary property

$$\mathbb{E}\big(X \,|\, \widehat{X}^{\Gamma_N^*,Vor}\big) = \widehat{X}^{\Gamma_N^*,Vor}. \tag{8}$$

Furthermore, if $d = 1$ and X has an absolutely continuous distribution with a log*-concave probability density, then (see [Abaya and Wise 1982], [Abaya and Wise 1984], [Trushkin 1982], [Kieffer 1983]) there is only one stationary quantizer which is necessarily the unique optimal quantizer of X at level N.*

The stationarity property (8) plays an important role in the numerical aspects of optimal Voronoi quantization although its proof is rather simple for an optimal quantizer: by the very definition of conditional expectation as an $L^2(\mathbb{P})$-orthogonal projection

$$e_{p,N}(X) \leq \|X - \mathbb{E}(X \,|\, \widehat{X}^{\Gamma_N^*,Vor})\|_2 \leq \|X - \widehat{X}^{\Gamma_N^*,Vor}\|_2 = e_{p,N}(X),$$

one derives (by uniqueness) that $\mathbb{E}(X \,|\, \widehat{X}^{\Gamma_N^*,Vor}) = \widehat{X}^{\Gamma_N^*,Vor}$ $a.s.$

For further mathematical insights on optimal vector (or Voronoi) quantization or for more details, we refer to [Graf and Luschgy 2000] and the references therein.

3.2 Optimal Delaunay quantization

By contrast to the above construction of Voronoi quantizations as best possible L^p-mean approximation, optimal Delaunay (or dual) quantization relies on the best approximation which can be achieved by a discrete random vector $\widehat{X}$ that satisfies a certain stationarity assumption on the extended probability space $(\Omega \times \Omega_0, \mathscr{A} \otimes \mathscr{A}_0, \mathbb{P} \otimes \mathbb{P}_0)$. That is we define

$$d_{p,N}(X) = \inf_{\widehat{X}} \Big\{ \|X-\widehat{X}\|_p : \widehat{X} : (\Omega\times\Omega_0, \mathscr{A}\otimes\mathscr{A}_0, \mathbb{P}\otimes\mathbb{P}_0) \to \mathbb{R}^d, \\ \#\widehat{X}(\Omega\times\Omega_0) \le N \text{ and } \mathbb{E}(\widehat{X}|X) = X \Big\}.$$

Then (see [Pagès and Wilbertz 2010a]), one may show that such a definition is equivalent to

$$d_{p,N}(X) = \inf_{\Gamma} \big\{ \|F_p(X;\Gamma)\|_p,\ \Gamma\subset\mathbb{R}^d, \#\Gamma \le N \big\}$$

for the local dual quantization functional

$$F_p(\xi;\Gamma) = \inf_{\lambda} \Big\{ \Big(\sum_{i=1}^N \lambda_i |\xi - x_i|^p \Big)^{1/p},\ (\lambda_i)_{1\le i\le N} \in [0,1]^N \text{ and } \sum_{i=1}^N \lambda_i x_i = \xi, \sum_{i=1}^N \lambda_i = 1 \Big\}.$$

When $p=2$ (quadratic case) and if the grid $\Gamma\subset\mathbb{R}^d$ admits a unique Delaunay triangulation (*e.g.* if Γ contains an affine basis and its points are in *general position*: none of its subset of size $d+1$ lies on the same sphere), then it was proved in [Pagès and Wilbertz 2010a] that we can construct a dual quantization operator which is the counterpart of the nearest neighbour projection for Voronoi quantization. This operator maps the random variable X randomly to the vertices of the Delaunay "triangle" (in fact a d-simplex) in which X falls (see Figure 1 further on), where the probability of mapping X to a given vertex t_i is determined by the i-th barycentric coordinate of X in the (non-degenerated) "hypertriangle" (or d-simplex) $\mathrm{conv}\{t_j : j=1,\dots,d+1\}$. When $p\neq 2$, an extension of the notion of Delaunay can still be defined although slightly more involved (similarly, the Voronoi cells are no longer convex when $p\neq 2$). We refer again to [Pagès and Wilbertz 2010a] for details.

Mathematically speaking, let $(D_k(\Gamma))_{1\le k\le m}$ be a Delaunay partition of the convex hull $\mathrm{conv}(\Gamma)$ of Γ. Let us denote by $\lambda^k(\xi)$ the barycentric coordinates of ξ in the triangle $D_k(\Gamma)$, with the convention $\lambda_i^k(\xi)=0$ if $x_i\notin D_k(\Gamma)$ and set

$$\mathscr{J}_\Gamma^u(\xi) = \sum_{k=1}^m \left[\sum_{i=1}^N x_i \cdot \mathbb{1}_{\left\{ \sum_{j=1}^{i-1} \lambda_j^k(\xi) \le u < \sum_{j=1}^{i} \lambda_j^k(\xi) \right\}} \right] \mathbb{1}_{D_k(\Gamma)}(\xi).$$

Then it holds

$$F_p(\xi;\Gamma) = \Big(\mathbb{E}_{\mathbb{P}_0} |\xi - \mathscr{J}_\Gamma^U(\xi)|^p \Big)^{1/p},$$

where U is defined on $(\Omega_0,\mathscr{A}_0,\mathbb{P}_0)$ with a $\mathscr{U}\big([0,1]\big)$-distributed (so that the operator $\mathscr{J}_\Gamma^u(\xi)$ is defined on this exogenous space). Then we define (on the product probability space $(\widetilde{\Omega},\widetilde{\mathscr{A}},\widetilde{\mathbb{P}})$) the *dual* (or *Delaunay*) *quantization*

$$\widehat{X}^{\Gamma,\mathrm{Del}} = \mathscr{J}_\Gamma^U(X)$$

so that

$$\|F_p(X;\Gamma)\|_p = \|X - \widehat{X}^{\Gamma,\mathrm{Del}}\|_p \quad \text{and} \quad \mathbb{E}(\widehat{X}^{\Gamma,\mathrm{Del}}|X) = X.$$

As a matter of fact, this "strict" *dual* stationarity condition can only be fulfilled if $\mathrm{supp}(\mathbb{P}_X)$ is bounded. To preserve as much intrinsic stationarity for $\widehat{X}^{\Gamma}$ as possible, *i.e.* stationarity on $\mathrm{conv}(\Gamma)$, we introduce the dual quantization for non-compactly supported random vector X as

$$\widehat{\widehat{X}}^{\Gamma,\mathrm{Del}} = \mathscr{J}_{\Gamma}^{U}(X)\mathbf{1}_{\{X\in\mathrm{conv}(\Gamma)\}} + \mathrm{Proj}_{\Gamma}(X)\mathbf{1}_{\{X\notin\mathrm{conv}(\Gamma)\}}.$$

and denote the *optimal dual quantization error* in this case by

$$\bar{d}_{p,N}(X) = \inf_{\Gamma}\big\{\|X - \widehat{\widehat{X}}^{\Gamma,\mathrm{Del}}\|_p,\ \Gamma\subset\mathbb{R}^d, \#\Gamma\le N\big\}.$$

Optimal dual quantizers. In both settings, it is shown in [Pagès and Wilbertz 2010a], under continuity assumption of the distribution of X, that for every $N\ge 1$, there exists at least one *optimal dual quantizer* at level N which has exactly N components for $\bar{d}_{p,N}(X)$. Furthermore $\bar{d}_{p,N}(X)\to 0$ as $N\to\infty$. If the distribution of X is compactly supported the same holds for the modulus $d_{p,N}(X)$ as soon as $N\ge d+1$.

Brief comparison of Delaunay and Voronoi quantization.

To illustrate the difference between Voronoi and Delaunay quantization (in the case $d=p=2$), we compare in Figure 1 below the nearest neighbor projection and the dual quantization operator.

For a given grid $\Gamma\subset\mathbb{R}^d$, the nearest neighbor projection Proj_{Γ} maps $X(\omega)$ entirely to the generator of the Voronoi cell $C_i(\Gamma)$ in which $X(\omega)$ falls. By contrast, the Delaunay random splitting operator $\mathscr{J}_{\Gamma}$ splits up the "weight" 1 of $X(\omega)$ across the vertices of the Delaunay triangle in which $X(\omega)$ falls. Since each vertex receives here a proportion according to the barycentric coordinate of the point $X(\omega)$ in that specific Delaunay triangle, this splitting operator fulfills a backward interpolation property, *i.e.* the "weight" of $X(\omega)$ is given by a convex combination on the vertices of the Delaunay triangle. Finally, this property also implies the intrinsic dual stationarity condition $\mathbb{E}(\widehat{X}^{\Gamma,\mathrm{Del}}|X) = X$

For a comparison in one dimension, we give the example of an optimal quantization for $\mathscr{U}([0,1])$. Following [Pagès and Wilbertz 2010a], Section 5.1, we derive for an optimal dual quantizer of $\mathscr{U}([0,1])$ and size N

$$\Gamma_{\mathrm{Del},N} = \left\{\frac{i-1}{N-1} : i = 1,\dots,N\right\}.$$

On the other hand, it holds in the case of optimal Voronoi quantization

$$\Gamma_{\mathrm{Vor},N} = \left\{ \frac{2i-1}{2N} : i = 1,\dots,N \right\}.$$

so that an optimal Voronoi quantizer of size N is made up by the midpoints of an optimal Delaunay of size $N+1$.

Note, that such a property does not hold for general distributions and in arbitrary dimensions. The asymptotic relationship between the optimal grids for Delaunay and Voronoi quantization is established in the following section 3.3.

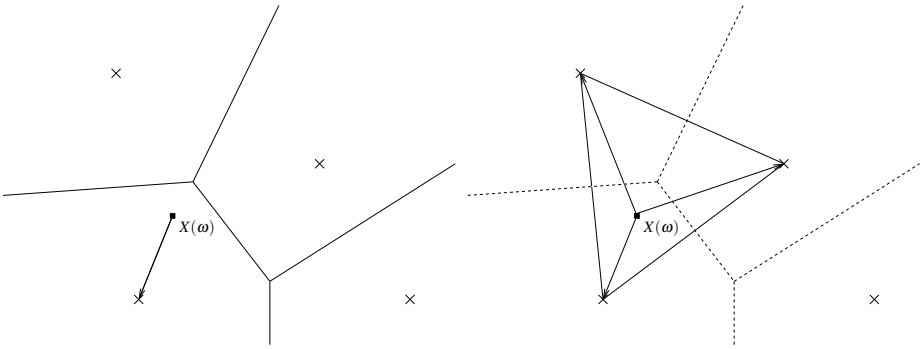

Fig. 1 Voronoi (left) and Delaunay (right) mapping for the realization $X(\omega)$.

3.3 Quantization rates

Both Regular (or Voronoi) and dual (or Delaunay) quantization error moduli satisfy formally the same theorem.

Theorem 3.1 (Optimal Voronoi quantization) *Let $p, p' \in (0,\infty)$, $p < p'$.*

(a) ASYMPTOTIC ERROR BOUND (ZADOR'S THEOREM) *(see e.g. [Zador 1982, Bucklew and Wise 1982, Graf and Luschgy 2000]) Assume $X \in L^{p'}(\Omega, \mathscr{A}, \mathbb{P})$ with a distribution $\mathbb{P}_X(d\xi) = h(\xi)\lambda_d(d\xi) + \nu_X(d\xi)$ where the finite measure ν_X is singular w.r.t. the Lebesgue measure λ_d on $(\mathbb{R}^d, \mathscr{B}or(\mathbb{R}^d))$. Then*

$$\lim_N N^{\frac{1}{d}} e_{p,N}(X) = \widetilde{J}^{vq}_{d,p,\|.\|} \|h\|^{\frac{1}{d}}_{\frac{p}{p+d}}$$

where $\widetilde{J}^{vq}_{d,p,\|.\|} = \inf_{N\geq 1} N^{\frac{1}{d}} e_{p,N}(X) \in (0,\infty)$ corresponds to the uniform distribution over the unit hypercube $[0,1]^d$ when $\mathbb{R}^d$ is equipped with the norm $\|.\|$.

(b) NON-ASYMPTOTIC ERROR BOUND (PIERCE'S LEMMA) *(see e.g. [Luschgy and Pagès 2008]) There exists a real constant $K^{vq}_{d,p,p'} \in (0,\infty)$ such that, for every random vector*

$X:(\Omega,\mathscr{A},\mathbb{P})\to\mathbb{R}^d$,

$$\forall N\geq 1,\qquad e_{p,N}(X)\leq K^{vq}_{d,p,p'}N^{-\frac{1}{d}}\min_{a\in\mathbb{R}^d}\|X-a\|_{p'}.$$

In fact the above non-asymptotic bound is a slight improvement of that established in [Luschgy and Pagès 2008] taking advantage of the obvious invariance of $e_{p,N}(X)$ by translation: $e_{p,N}(X)=e_{p,N}(X+a)$, $a\in\mathbb{R}^d$.

Theorem 3.2 (Optimal dual quantization) *([Pagès and Wilbertz 2010b]) The above theorem for Voronoi quantization also holds true, with appropriate real constants $\widetilde{J}^{dq}_{p,\|.\|}$ ($\geq\widetilde{J}^{vq}_{p,\|.\|}$) and $K^{dq}_{d,p,p'}$ ($\geq K^{vq}_{d,p,p'}$) when replacing $e_{p,N}(X)$ by its counterpart the minimal dual L^p-mean quantization error $\bar{d}_{p,N}(X)$. However, the non-asymptotic claim only holds true for $N\geq N_{d,p,p'}$ (where $N_{d,p,p'}$ only depends on d,p,p').*

When X has a compact support, the theorem holds true for the error modulus $d_{p,N}(X)$ with same constants $\widetilde{J}^{dq}_{p,\|.\|}$ and $K^{dq}_{d,p,p'}$ (with the convention $d_{p,N}(X)=+\infty$ if $N\leq d$). Finally, when $d=1$, $\widetilde{J}^{dq}_{1,p,\|.\|}=\left(\frac{2}{(p+1)(p+2)}\right)^{\frac{1}{p}}=\left(\frac{2^{p+1}}{p+2}\right)^{\frac{1}{p}}\widetilde{J}^{vq}_{1,p,|.|}\geq\widetilde{J}^{vq}_{1,p,|.|}$.

4 How to get optimal Voronoi and Delaunay quantizations

4.1 Optimal quadratic Voronoi Quantization

Throughout this section we focus on the quadratic case, although, at least formally, all proposed algorithms have L^p counterparts for $p\geq 2$.

4.1.1 Original and randomized Lloyd's I algorithm

When the dimension $d=1$ and $p=2$ (quadratic case), one may identify a quantization grid Γ of size N with an N-tuple with increasing components *i.e.* an element of $\mathscr{I}_N:=\{(x_1,\ldots,x_N)\in\mathbb{R}^N\,|\,-\infty<x_1<\cdots<x_N<+\infty\}$. It has been originally shown in [Kieffer 1983] that if the distribution of a random variable X has a log-concave probability density function, then there exists a unique stationary quantizer of size N, denoted $\Gamma^{*,N}$ *i.e.* a quantizer satisfying

$$\mathbb{E}\big(X\,|\,\widehat{X}^{\Gamma^{*,N}}\big)=\widehat{X}^{\Gamma^{*,N}}. \tag{9}$$

Since a quadratic optimal quantizer at level N of an absolutely continuous distribution has exactly N pairwise distinct components and is stationary (see Proposition 1), this stationary quantizer $\Gamma^{*,N}$ is also the unique optimal *quadratic* quantizer.

In [Kieffer 1982] is proposed an alternative and more constructive proof of the above facts. It is based on the so-called Lloyd's I procedure which updates recursively a quantization grid $\Gamma_{(m)}$ (of size N) as follows:

$$\widehat{X}^{\Gamma_{(m+1)}} = \mathbb{E}\big(X \,|\, \widehat{X}^{\Gamma_{(m)}}\big),\ m \in \mathbb{N},\ \Gamma_{(0)} \in \mathscr{I}_N \cap \mathscr{H}(\mathbb{P}_X) \tag{10}$$

where $\mathscr{H}(\mathbb{P}_X) = \mathrm{conv}(\mathrm{supp}(\mathbb{P}_X))$. It is proved that the procedure "lives" inside $\mathscr{I}_N \cap \mathscr{H}(\mathbb{P}_X)$ and that, still under the log-concavity assumption, $\Gamma_{(m)}$ *converges exponentially fast* toward the unique stationary N-quantizer $\Gamma^{*,N}$. Written in a more analytical form, (10) reads, if $\Gamma_{(m)} = \{x_{m,1}, \ldots, x_{m,N}\}$,

$$x_{m+1,i} = \mathbb{E}\big(X \,|\, \widehat{X}^{\Gamma_{(m)}} = x_{m,i}\big) = \frac{\displaystyle\int_{C_i(\Gamma_{(m)})} \xi \mathbb{P}_X(d\xi)}{\mathbb{P}_X(C_i(\Gamma_{(m)}))},\ i = 1, \ldots, N,$$

where in this $1D$-setting $C_i(\Gamma_{(m)}) = \big(\frac{x_{m,i-1}+x_{m,i}}{2}, \frac{x_{m,i}+x_{m,i+1}}{2}\big]$, with $x_{m,0} = -\infty$ and $x_{m,N+1} = +\infty$.

It is straightforward that the procedure as defined by (10) can be extended to the d-dimensional setting. One defines recursively the sequence of N-quantizers $\Gamma_{(m)}$, $m \in \mathbb{N}$, by $\Gamma_{(0)} \subset \mathscr{H}(\mathbb{P}_X)$, $\#\Gamma_{(0)} = N$ and

$$x_{m+1,i} = \mathbb{E}\big(X \,|\, \widehat{X}^{\Gamma_{(m)}} = x_{m,i}\big) = \frac{\mathbb{E}\big(X \mathbf{1}_{\{X \in C_i(\Gamma_{(m)})\}}\big)}{\mathbb{P}(X \in C_i(\Gamma_{(m)}))},\ i = 1, \ldots, N,$$

with obvious notations. One easily checks that

$$\begin{aligned} \|X - \widehat{X}^{\Gamma_{(m+1)}}\|_2 &= \big\|X - \mathbb{E}\big(X \,|\, \widehat{X}^{\Gamma_{(m)}}\big)\big\|_2 \\ &= \inf\Big\{ \|X - \varphi(\widehat{X}^{\Gamma_{(m)}})\|_2 :\ \varphi : \mathbb{R}^d \to \Gamma_{(m)}, \varphi \text{ is Borel} \Big\} \le \|X - \widehat{X}^{\Gamma_{(m)}}\|_2 \end{aligned}$$

so that, this multi-dimensional Lloyd's I procedure always makes the quadratic quantization error decrease (except if $\Gamma_{(m)}$ is itself stationary at finite range). Of course, any stationary quantizer is a fixed point for the Lloyd's I procedure and in higher dimension there are always several stationary quantizers. As far as we know, no convincing proof of pointwise convergence to a global minimum has been established so far for the grids $\Gamma_{(m)}$. However, from a practical point of view, one may reasonably hope that this convergence does hold, at least toward a local minimum of the quadratic quantization error functional $\Gamma \mapsto \|X - \widehat{X}^{\Gamma}\|_2$.

As soon as the dimension d of the state of the random vector X is greater than 2 or 3, the Lloyd's I procedure cannot be implemented by analytical means since it becomes impossible to compute integrals like $\displaystyle\int_{C_i(\Gamma)} f(\xi) d\xi$ by any kind of cubature formulas (however see [Wilbertz 2005] for low dimensions). The alternative

solution, when the random vector X is simulatable, is to rely on a Monte Carlo simulation at each step m to compute for every $i \in \{1,\dots,N\}$,

$$\mathbb{E}\big(X \,|\, X \in C_i(\Gamma_{(m)})\big) = a.s.\text{-}\lim_{L\to\infty} \frac{\sum_{\ell=1}^{L} X_\ell \mathbf{1}_{\{X_\ell \in C_i(\Gamma_{(m)})\}}}{\sum_{\ell=1}^{L} \mathbf{1}_{\{X_\ell \in C_i(\Gamma_{(m)})\}}}.$$

Note that $X_\ell \in C_i(\Gamma_{(m)})$ if and only if $x_{m,i}$ is the nearest neighbour of X_ℓ among all components $x_{m,i}$, $i = 1,\dots,N$ of the current grid $\Gamma_{(m)}$ (with appropriate conventions on the boundary). This randomized Lloyd's I procedure has the complexity of L nearest neighbour searches, see Section 4.1.4 for a few comments on (fast) nearest neighbour search. Also note that this phase can be performed *offline* and that each Monte Carlo step can be parallelized.

A huge literature has been devoted to practical aspects of Lloyd's I procedure and its applications in Signal Processing and Data compressing. For further insights in that direction, see *e.g.* [Gersho and Gray 1992]. In Data Analysis (when the underlying distribution of interest is the uniform distribution over the data set (*i.e.* the empirical measure of this data set) the "batch" (for "non-randomized") procedure is known as the *k-means* algorithm. For some applications in Delaunay grid generation see [Du and Gunzburger 2002]. On the other hand little has been done on theoretical aspects, since [Kieffer 1982].

4.1.2 The Competitive Learning Vector Quantization algorithm

The so-called *CLVQ* algorithm is a stochastic gradient algorithm relying on the fact that the squared quadratic quantization error, called *distortion*. We will make the obvious abuse of notation consisting in identifying grids of size at most N and N-tuples with possibly "repeated" components. The distortion is then defined on $(\mathbb{R}^d)^N$ by

$$\Gamma = (x_1,\dots,x_N) \longmapsto \mathrm{Distor}_N(X;\Gamma) := \mathbb{E}\min_{1\le i\le N} |X - x_i|^2.$$

This function is differentiable at every N-tuple $x = (x_1,\dots,x_N) \in (\mathbb{R}^d)^N$ having pairwise distinct components with a gradient $\nabla_x \mathrm{Distor}_N(X;\Gamma)$ given by

$$\nabla_x \mathrm{Distor}_N(X;\Gamma) = 2\left(\mathbb{E}\big((x_i - X)\mathbf{1}_{\{X\in C_i(\Gamma)\}}\big)\right)_{1\le i\le N}.$$

If $\#\mathrm{supp}\mathbb{P}_X \ge N$, the distortion function is differentiable at any minimum since it has pairwise distinct components (see [Graf and Luschgy 2000]). Furthermore as emphasized above its gradient has a representation as an expectation formally reading

$$\nabla_x \mathrm{Distor}_N(X;(x_1,\dots,x_N)) = \mathbb{E}\left(\nabla_x \mathrm{distor}_N\big(X;(x_1,\dots,x_N)\big)\right).$$

The function defined on $\mathbb{R}^d \times (\mathbb{R}^d)^N$ by

$$(\xi, \Gamma) \longmapsto \nabla_x \mathrm{distor}_N\big(X;\Gamma\big)$$

is sometimes called a *local gradient* of the *potential* function Distor_N. Then, the paradigm of stochastic approximation says that under technical assumptions to be specified, the so-called *stochastic gradient descent* defined by

$$\Gamma_{(m+1)} = \Gamma_{(m)} - \gamma_{m+1}\nabla_x \mathrm{distor}_N(X_{m+1};\Gamma_{(m)}),\ m \geq 1,\ \Gamma_{(0)} \subset \mathbb{R}^d,\ \#\Gamma_{(0)} = N,$$

where $(X_m)_{m\geq 1}$ is an i.i.d. sequence of copies of X and $(\gamma_m)_{m\geq 1}$ is a sequence of gain parameter satisfying the *decreasing step assumption"* assumption $\sum_{m\geq 1}\gamma_m = +\infty$ and $\sum_{m\geq 1}\gamma_m^2 < +\infty$ which is standard in Stochastic Approximation Theory, "hopefully" converges toward a local minimum of the distortion function.

From a practical point of view, this abstract formula can be decomposed into two phases: set for convenience $\Gamma_{(m)} = (x_{m,1},\ldots,x_{m,N})$, $m \geq 0$.

(i) *Competitive Phase:* Search of the nearest neighbour $x_{m,i^*(X_{m+1})}$ of X_{m+1} among the components of $x_{m,i}$, $i = 1,\ldots,N$, of $\Gamma_{(m)}$ (using a "winning convention" in case of conflict between two or more components).

(ii) *Learning Phase:* One moves the winning component towards X_{m+1} using a dilatation *i.e.*

$$x_{m+1,i^*(X_{m+1})} = \mathrm{Dilatation}_{[X_{m+1},1-\gamma_{m+1}]}(x_{m,i^*(X_{m+1})})$$

where the dilatation $\mathrm{Dilatation}_{[\xi,\lambda]}$ centered at $\xi \in \mathbb{R}^d$ with ratio $\lambda > 0$ is defined by

$$\forall\, y \in \mathbb{R}^d, \quad \mathrm{Dilatation}_{[\xi,\lambda]}(y) = \xi + \lambda(y-\xi) = (1-\lambda)\xi + \lambda y.$$

All other components stay still.

This procedure is useful for small or medium values of N. For general background on stochastic approximation, we refer to [Benveniste et al. 1990, Duflo 1996, Kushner and Yin 2003]. Unfortunately, the *CLVQ* procedure turns out to be singular in the world of recursive stochastic approximation algorithms: only "conditional *a.s.* convergence" results have been obtained (also known as *a.s.* convergence in the "Kushner-Clark sense") in higher dimension (for compactly supported distributions), see [Pagès 1998]. However, in a 1D framework, regular *a.s.* convergence has been established with a weak rate ruled by a standard Central Limit Theorem, still for distributions with compact support (see [Bouton and Pagès 1993, Benaïm et al. 1998]).

This procedure has also given rise to many empirical investigations and heuristic statements, especially in the artificial neural network community where the *CLVQ* algorithm appears as a degenerate case of the Kohonen self-organizing maps used in

non-linear automatic classification. Its complexity is again closely related to nearest neighbour searches. Parallelized versions based on a stratification of the state space can be used to speed up the procedure

Other optimization procedures have also been implemented like (randomized) evolutionary algorithms (see *e.g.* [Mrad and Ben Hamida 2006]).

4.1.3 Companion parameters

To fully elucidate the distribution of a quantization $\widehat{X}$ of X, not only the grid $\Gamma = \{x_1, \ldots, x_N\}$ is necessary but also the weights $p^i = \mathbb{P}(\widehat{X} = x_i)$. These weights are often called "companion parameters". Other companion parameters may be of interest like the local inertia $\mathbb{E}\big(\mathbf{1}_{X \in \{C_i(\Gamma)\}} |X - x_i|^2\big)$.

$\triangleright$ *Adaptive estimation (CLVQ).* When performing the *CLVQ* algorithm, one may devise a companion procedure to estimate these weights *on-line* by setting

$$p^i_{(m+1)} = p^i_{(m)} - \widetilde{\gamma}_{m+1}\Big(p^i_{(m)} - \mathbf{1}_{\{i^*(X_{m+1})=i\}}\Big),\ i = 1, \ldots, N$$

where $\widetilde{\gamma}_m = \gamma_m$ or $\widetilde{\gamma}_m = 1/m$ (the second choice corresponds to the usual empirical mean but with respect to the "moving grids" $\Gamma_{(m)}$). No significant extra computation is needed since $i^*(X_{m+1})$ is already computed in the core of the *CLVQ* procedure.

$\triangleright$ *Posterior estimation.* From a practical point of view, it seems more efficient to estimate the weights p^i by a standard Monte Carlo simulation posterior to the grid optimization: this amounts to "freezing" $\Gamma_{(m)} = \Gamma$ and setting $\widetilde{\gamma}_m = 1/m$ in the above procedure (still based on repeated nearest neighbour searches).

4.1.4 More on practical aspects

$\triangleright$ **Quasi-Monte Carlo.** For formerly mentioned procedures, one may substitute a sequence of quasi-random numbers – *e.g.* like the Halton or the Sobol' sequences – to the usual sequence pseudo-random numbers. This often speeds up the rate of convergence of the method, although this remains mostly heuristic in Stochastic Approximation (see however [Lapeyre et al. 1990]).

$\triangleright$ **Inductive computation: the splitting method.** The most important step to preserve the accuracy of the quantization as N increases is to use the so-called *splitting method* which finds its origin in the proof of the existence of an optimal N-quantizer: once the optimization of a quantization grid of size N is achieved, one specifies the starting grid for the size $N+1$ or more generally $N+\nu$, $\nu \geq 1$, by merging the optimized grid of size $N$with ν points sampled independently from the distribution having a probability density proportional to $\varphi^{\frac{d}{d+2}}$ where φ denotes

the p.d.f. of the distribution $\mathbb{P}_X$. This rather unexpected choice is motivated by the fact that this distribution provides the lowest *in average* random quantization error (see [Cohort 1998]).

When simulation at a reasonable cost of the distribution $\varphi^{\frac{d}{d+2}}(\xi)\lambda_d(d\xi)$ is impossible, one can still simulate instead $\mathbb{P}_X$-distributed numbers. This is the adopted strategy to compute the grids of the d-dimensional normal distribution available on the website [Pagès and Printems 2005] (see below).

▷ **Nearest neighbour search.** All the above procedures rely on repeated nearest neighbour searches. The complexity of a naive implementation of this procedure grows linearly with $d \times N$ and becomes very demanding as d increases. So reducing its computational cost is strategic.

– The most basic (although quite efficient) method is the *Partial Distance Search*: to check whether a record level L_{rec} is beaten or not by $|x| = ((x^1)^2 + \cdots + (x^d)^2)^{1/2}$ one checks at each step ℓ if $(x^1)^2 + \cdots + (x^\ell)^2 \geq L_{rec}^2$. If so, one rejects x and test a new point.

– A more sophisticate procedure has been originally devised by Bentley and analyzed in a the seminal paper [Friedman et al. 1977] . It is an efficient way to store the data (the N points) based along a search tree called *k-d* tree. It reduces the complexity of the nearest neighbour search down to $O(\log N)$ (after a one shot preprocessing of complexity $O(N \log N)$). An improved version of the *k-d* tree, based on a preliminary *PCA*, has been developed in [McNames 2001] and is known as the *PAT* algorithm (for Principle Axis Tree). Other search trees based on a preliminary "rough" quantization have also been proposed (see [Corlay 2011]). The (relative) efficiency of such methods first increases as the dimension of the state space grows but becomes more limited for large dimension where "brute force" (unfortunately) comes back in the game.

▷ **Still more on practical aspects.** Many practical studies have been carried out, including heuristic considerations about the above described procedures in [Gersho and Gray 1992] with an orientation toward Signal Processing and Data compressing. In [Pagès and Printems 2003] a first numerical study entirely devoted to the multi-variate normal distribution has been developed which finally led to make available optimized grids of multivariate normal distributions on the website [Pagès and Printems 2005] devoted to optimal vector and functional quantization.

These grids have been computed inductively using the splitting method by a combination of *CLVQ* (for medium values of N) and Lloyd's I algorithm, for dimension running from $d = 1$ up to $d = 10$ and sizes N running from 1 up to 10000. For each grid Γ several "companion parameters (see below) are included in the files, especially the weights $w_i = \mathbb{P}(\mathcal{N}(0;I_d) \in C_i(\Gamma))$, $i = 1,\ldots,N$, but also the local L^p-inertia $\left(\mathbb{E}|X - x_i|^p \mathbf{1}_{\{X \in C_i(\Gamma)\}}\right)_{1 \leq i \leq N}$ for $p = 1, 2$.

4.2 Dual quantization

In general, a grid which has been optimized for Voronoi quantization can also serve as a good grid for Delaunay quantization. As concerns practical applications, the key advantage of dual quantization is its intrinsic (dual) stationarity property

$$\mathbb{E}(\widehat{X}^{\Gamma,\mathrm{Del}}|X) = X \qquad (\text{where } \widehat{X}^{\Gamma,\mathrm{Del}} = \mathscr{J}_\Gamma^U(X))$$

which *holds for any grid* Γ *with* $\mathrm{supp}(\mathbb{P}_X) \subset \mathrm{conv}\{\Gamma\}$ regardless of its optimality with respect to the distribution of X. *Dual stationarity exclusively follows from the way the dual quantization weights are defined* as

$$p^{i,Del} = \mathbb{P}(\widehat{X}^{\Gamma,\mathrm{Del}} = x_i).$$

One way to get (almost) the best from both methods, especially in higher dimension, can to compute for a Voronoi stationary grid both its Voronoi and Delaunay (dual) weights so as to take advantage of both stationarity properties.

Nevertheless, we give here a short sketch of the counterparts of both Lloyd's I procedure and CLVQ algorithm for dual quantization optimization. This is also a way to check that optimal Voronoi and Delaunay quantization grids remain somewhat close, especially as d grows (see Figures 2 and 3).

4.2.1 Lloyd-type algorithm for dual quantization

In order to establish a Lloyd-type algorithm for the optimization of (quadratic) dual quantization grids, we write $\Gamma_{(m)} = \{x_{m,1},\ldots,x_{m,N}\} \subset \mathbb{R}^d$ for $m \in \mathbb{N}$ and denote by $(D_I(\Gamma))_{I\in\mathscr{I}}$ a Delaunay partition of $\mathrm{conv}(\Gamma)$, where the index set $\mathscr{I} = \mathscr{I}(\Gamma) \subset \big\{I \subset \{1,\ldots,N\} : \#I = d+1\big\}$ defines a Delaunay triangulation in Γ. Moreover, if $\xi \in D_I(\Gamma)$, we write $\lambda_{x_i}^I(\xi)$ for the barycentric coordinate of $\xi \in \mathrm{conv}\{x_j : j \in I\}$ with respect to the vertex x_i.

Recall that each Delaunay triangle $D_I(\Gamma)$ is characterized by the center of a sphere spanned by the vertices $\{x_j : j \in I\}$ which contains no point of Γ in its interior. We then denote this center by $z_I = z_I(\Gamma)$ and define a Delaunay center by mapping

$$Z^\Gamma(\xi) = \sum_{I\in\mathscr{I}} z_I \, \mathbb{1}_{D_I(\Gamma)}(\xi). \tag{11}$$

Moreover, note that those Delaunay centers are exactly the vertices of the corresponding Voronoi tessellation since they are at the same distance to the x_j, $j \in J$.

If one considers the optimization problem (still with the same abuse of notation)

$$\Gamma = (x_1,\ldots,x_N) \longmapsto \mathrm{Distor}_N(X;\Gamma) := \mathbb{E}|X - \mathscr{J}_\Gamma^U(X)|^2 \tag{12}$$

then it was shown in [Pagès and Wilbertz 2010a] that the gradient of this function in Γ reads

$$\nabla_\Gamma \mathrm{Distor}_N(X;\Gamma) = 2\Big[\mathbb{E}\big((x_i - Z^\Gamma(X))\mathbb{1}_{\{\mathscr{J}_\Gamma^U(X)=x_i\}}\big)\Big]_{1\le i\le N}.$$

The first order optimality condition therefore writes

$$\mathbb{E}\big(Z^{\Gamma^*}(X)|\,\mathscr{J}_{\Gamma^*}^U(X)\big) = \mathscr{J}_{\Gamma^*}^U(X)$$

and can be regarded as a counterpart to (9). We may therefore define a Lloyd-type method for dual quantization starting at some initial grid $\Gamma_{(0)} \subset \mathbb{R}^d, \#\Gamma_{(0)} = N$ as

$$\widehat{X}^{\Gamma_{(m+1)}} = \mathbb{E}\big(Z^{\Gamma_{(m)}}(X)|\,\mathscr{J}_{\Gamma_{(m)}}^U(X)\big), \quad m \ge 0.$$

Since it holds

$$\mathbb{P}(\mathscr{J}_\Gamma^U(X) = x_i) = \sum_{I\in\mathscr{I}: i\in I} \int_{D_I(\Gamma)} \lambda_{x_i}^I(\xi)\,\mathbb{P}_X(d\xi),$$

we arrive for $m \ge 1$ at

$$x_{m+1,i} = \frac{\sum_{I\in\mathscr{I}: i\in I} z_I \int_{D_I(\Gamma)} \lambda_{x_i}^I(\xi)\,\mathbb{P}_X(d\xi)}{\sum_{I\in\mathscr{I}: i\in I} \int_{D_I(\Gamma)} \lambda_{x_i}^I(\xi)\,\mathbb{P}_X(d\xi)}, \quad i = 1,\dots,N.$$

This means that $x_{m+1,i}$ is chosen as a weighted sum of the Delaunay centers z_I whose triangles share the same vertex $x_{m,i}$ in $\Gamma_{(m)}$. It can be shown that such an algorithm is in fact a Quasi-Newton method and therefore converges to a local minimum of (12) (see *e.g.* [Iri et al. 1984] in the case of the regular Lloyd's I method).

This algorithm, which is new to our knowledge, is the first tool we used to compute optimal dual quantization grids like the one below displayed below for the joint distribution of the Brownian motion and its running supremum at time 1. The second algorithm is the counterpart of the *CLVQ* and is described below.

4.2.2 CLVQ like procedure for dual quantization

Like for the "Voronoi" *CLVQ* algorithm, we consider the *dual distortion function*

$$\Gamma = (x_1,\dots,x_N) \longmapsto \mathrm{Distor}_N(X;\Gamma) := \mathbb{E}|X - \mathscr{J}_\Gamma^U(X)|^2.$$

Referring again to [Pagès and Wilbertz 2010a], it holds for the gradient of the dual distortion function

$$\nabla_\Gamma \mathrm{Distor}_N(X;\Gamma) = 2\Big[\mathbb{E}\big((x_i - Z^\Gamma(X))\mathbb{1}_{\{\mathscr{J}_\Gamma^U(X)=x_i\}}\big)\Big]_{1\le i\le N}.$$

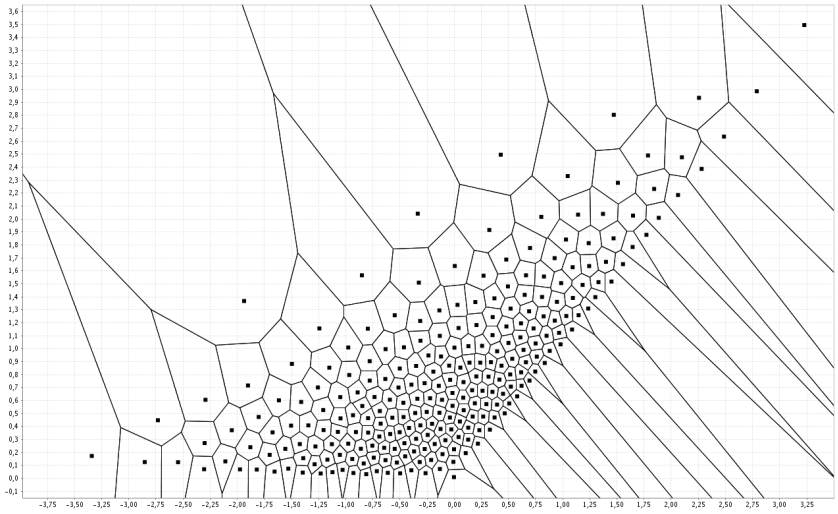

Fig. 2 Voronoi quantization of the joint distribution a standard Brownian motion and its running supremum at time $T = 1$ ($N = 250$).

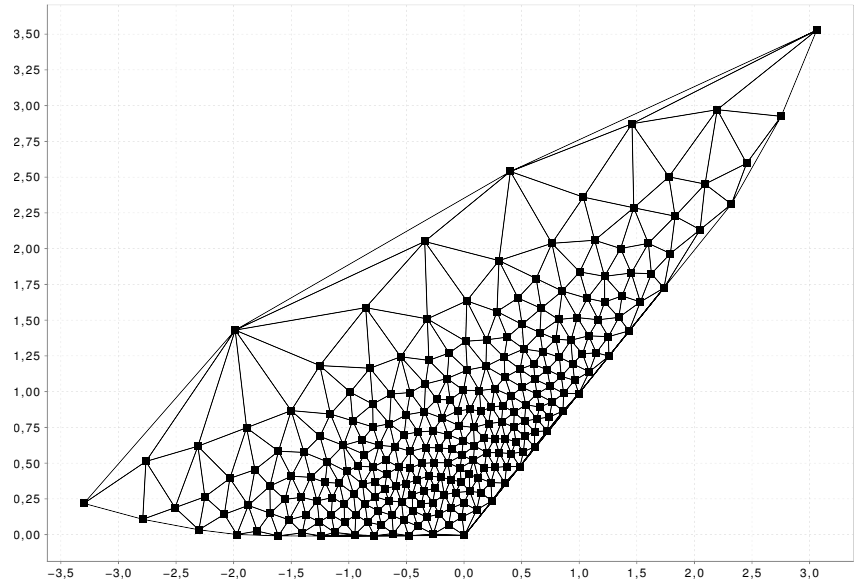

Fig. 3 Delaunay (dual) quantization of the joint distribution a standard Brownian motion and its running supremum at time $T = 1$ ($N = 250$).

As above, the stochastic gradient method is given by

$$\Gamma_{(m+1)} = \Gamma_{(m)} - \gamma_{m+1}\nabla_x \mathrm{distor}_N(X_{m+1};\Gamma_{(m)}),\ m \geq 1,\ \Gamma_{(0)} \subset \mathbb{R}^d,\ \#\Gamma_{(0)} = N$$

where $(X_m)_{m\geq 1}$ is an i.i.d. sequence of copies of X and $(\gamma_m)_{m\geq 1}$ is a sequence of gain parameters satisfying the decreasing step assumption.

In practice that means that we generate a sequence $(X_m)_{m\geq 1}$ of i.i.d copies of X and the two phases of the CLVQ-algorithm read as follows

(i) *Competitive Phase:* Search for the Delaunay triangle $I^*(X_{m+1}) \in \mathscr{I}(\Gamma_{(m)})$ which contains the realization X_{m+1}.

(ii) *Learning Phase:* One moves the winning triangle towards the Delaunay center $Z^{\Gamma_{(m)}}(X_{m+1})$ using a dilatation *i.e.*

$$\forall i \in I^*(X_{m+1}),\ x_{m+1,i} = \text{Dilatation}_{[Z^{\Gamma_{(m)}}(X_{m+1}),1-\gamma_{m+1}]}(x_{m,i}).$$

4.2.3 Search for the matching Delaunay hyper-triangle

A crucial point in both above procedures, as well as in the weight computations later on, is the search for the Delaunay triangle $I^*(\xi) \in \mathscr{I}(\Gamma)$, which contains a point $\xi \in \text{conv}(\Gamma)$. This phase in dual quantization optimization is the exact counterpart of nearest neighbour search for Voronoi quantization. Such a search can be implemented efficiently by a directed search on the Delaunay triangulation of Γ. To be more precise, one starts at a triangle $I_0 \in \mathscr{I}(\Gamma)$ and then moves on to that neighbor triangle of I_0 which lies on the line defined by the Delaunay center z_{I_0} and ξ. It was shown in [Bowyer 1981] that such a procedure reaches the triangle $I^* \in \mathscr{I}(\Gamma)$ which contains ξ in average after $O_d(N^{1/d})$ steps, where N is the number of points in the grid Γ. For more details on such point location procedures in triangulations we refer to [Devroye et al. 2004] and [Muecke et al. 1999].

We did not speak yet about the weight computation in this section although it is a crucial step to fully determine the distribution of $\widehat{X}$ (whatever type of quantization is adopted) which in turn is necessary to produce quantization based cubature formulas. However, since we are interested in American option pricing, we postpone this kind of question to the quantization tree below where we will show how to compute the transition weights of the tree for both types of quantization.

5 Application to cubature formula for numerical integration

Let $\widehat{X}$ be a quantization based approximation of a random vector X taking value in a grid $\Gamma = \{x_1, \ldots, x_N\}$ of size $N \geq 1$ ($\widehat{X} = \text{Proj}_\Gamma(X)$ (Voronoi) or $\mathscr{J}_\Gamma^U(X)$ (Delaunay)) depending on the type of the quantization).

$\triangleright$ *Lipschitz continuous functions*. If $F : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz continuous

$$|\mathbb{E}F(X)-\mathbb{E}F(\widehat{X})| \le [F]_{\mathrm{Lip}}\mathbb{E}|X-\widehat{X}| = \|X-\widehat{X}\|_1.$$

This yields an approximate cubature formula since

$$\mathbb{E}F(\widehat{X}) = \sum_{1\le i\le N} p_i F(x_i) \quad \text{where} \quad p_i = \mathbb{P}(\widehat{X}=x_i),\ i=1,\dots,N.$$

Furthermore, we know that Voronoi quantization is optimal in the following sense

$$\sup\{|\mathbb{E}F(X)-\mathbb{E}F(\widehat{X})|,\ [F]_{\mathrm{Lip}}\le 1\} = e_{1,N}(X).$$

▷ *Functions with Lipschitz continuous differential*. Assume that $\widehat{\Gamma}$ is *stationary* (*i.e.* $\mathbb{E}(X\,|\,\widehat{X})=\widehat{X}$) or "dual stationary" (*i.e.* $\mathbb{E}(\widehat{X}\,|\,X)=X$), then (see [Pagès and Wilbertz 2010a])

$$|\mathbb{E}F(X)-\mathbb{E}F(\widehat{X})| \le [DF]_{\mathrm{Lip}}\mathbb{E}|X-\widehat{X}|^2$$

where DF denotes the (Lipschitz continuous) differential of F. At this stage, one must have in mind that few grids Γ (mainly the optimal quadratic grids) are stationary for Voronoi quantization whereas *all* grids are stationary for dual quantization by construction by construction.

▷ *Convex functions.* If F is convex and Γ is a stationary Voronoi quantizer, then

$$\mathbb{E}F(\widehat{X}^{\Gamma,vor}) \le \mathbb{E}F(X) \quad \text{where} \quad \widehat{X}^{\Gamma,vor} = \mathrm{Proj}_{\Gamma}(X).$$

If X has compact support, for any grid Γ such that $\mathrm{conv}(\Gamma)\supset \mathrm{supp}(\mathbb{P}_X)$,

$$\mathbb{E}F(X) \le \mathbb{E}F(\widehat{X}^{\Gamma,del}) \quad \text{where} \quad \widehat{X}^{\Gamma,del} = \mathscr{J}_{\Gamma}^{U}(X).$$

Combining both quantization approaches yields a *deterministic security interval*.

6 Quantization tree

Let us come back to our Bermuda option pricing problem with the notations introduced in Section 2. At each time $k\in\{0,\dots,n\}$, we consider a grid Γ_k of size N_k supposed to be an optimal (or at least a "good") Voronoi/Delaunay quantization of the Markov chain X_k at time k.

We define the discretization function $\pi_k : \mathbb{R}^d\times[0,1]\to\mathbb{R}^d$ as

- *Voronoi:* A Borel nearest neighbour projection on the grid Γ_k (see (7)) *i.e.*

$$\forall\,\xi\in\mathbb{R}^d,\ \forall\,u\in[0,1],\quad \pi_k(\xi,u) := \mathrm{Proj}_{\Gamma_k}(\xi). \tag{13}$$

- *Delaunay:* A splitting operator on the grid Γ_k

$$\forall \xi \in \mathbb{R}^d,\ \forall u \in [0,1],\ \pi_k(\xi,u) := \mathscr{J}^u_{\Gamma_k}(\xi)\mathbf{1}_{\{\xi \in \mathrm{conv}(\Gamma_k)\}} + \mathrm{Proj}_{\Gamma_k}(\xi)\mathbf{1}_{\{\xi \notin \mathrm{conv}(\Gamma_k)\}}. \tag{14}$$

Definition 6.1 *A quantization tree of the Markov chain $X = (X_k)_{0\le k\le n}$ is a sequence $(\Gamma_k, \mathbf{p}^k)_{0\le k\le n}$ of grids and weight matrices where*

– for $k = 0,\dots,n$, $\Gamma_k \subset \mathbb{R}^d$, $\#\Gamma_k = N_k \ge 1$ whose elements are denoted

$$\Gamma_k = \{x_1^k,\dots,x_N^k\}, \quad k = 0,\dots,n;$$

– for $k = 0,\dots,n-1$, $\mathbf{p}^k = [p_{ij}^k]_{1\le i\le N_k, 1\le j\le N_{k+1}}$, defined by

$$p_{ij}^k = \mathbb{P}\Big(\widehat{X}_{k+1} = x_j^{k+1} \,|\, \widehat{X}_k = x_i^k\Big).$$

with the convention $\mathbf{p}^k = 0$.

The resulting "quantized" dynamical programming principle derived from (2), once written "in distribution", can be written on this tree as follows

$$\begin{aligned}
&\widehat{v}_n(x_i^n) = h_n(x_i^n),\ i = 1,\dots,N_n\\
&\widehat{v}_k(x_i^k) = \max\Big(h_k(x_i^k), \sum_{j=1}^{N_{k+1}} p_{ij}^k \widehat{v}_{k+1}(x_j^{k+1})\Big),\ i = 1,\dots,N_k,\ k = 0,\dots,n-1.
\end{aligned}$$

Remarks. • Once the grids have been settled and the transition weight matrices $\mathbf{p}^k$ have been computed, on can perform the above backward quantization tree descent as many times as necessary for different payoff functions. All the information about the discretization of the Markov dynamics is "stored" in the quantization tree $(\Gamma_k, \mathbf{p}^k)_{0\le k\le n}$.

• The complexity of the backward descent of such a tree is clearly proportional to $\sum_{0\le k\le n-1} N_k N_{k+1}$ for a given global budget of $N = N_0 + \cdots + N_n$ (usually prescribed by the memory limitations of the computing device). Up to edge effects the minimal complexity is attained with constant size trees *i.e.* $N_k = \frac{N}{n+1}$, $k = 0,\dots,n$. If $X_0 = x_0$, then $N_0 = 1$ and $N_k = \frac{N-1}{n}$, $k = 1,\dots,n$. Other considerations (see below) may lead to other specifications for the quantization tree

6.1 Error bounds

By combining the error bounds of Proposition 2.1 and the non asymptotic bounds for optimal quantization(s) we get the following proposition which takes advantage of the non-asymptotic Zador's Theorems (3.1(b) and 3.2(b)). It simplifies the original presentation from [Bally and Pagès 2003a] and extends it to dual quantization.

Proposition 6.3 *Assume the Markov chain satisfies all the assumptions of Proposition 2.1 and that furthermore,* $\max_{0\le k\le n}\|X_k\|_{p'}<+\infty$ *for a* $p'>1$. *Assume that the payoff functions* h_k, $k=0,\dots,n$ *are Lipschitz continuous. Assume the sequence* $(\widehat{X}_k)_{0\le k\le n}$ *is defined either by (13) or by (14) and that, for every* $k=0,\dots,n$, *the quantization size* $N_k\ge N_{d,p,p'}$ *(*$N_{d,p,p'}=1$ *in the Voronoi setting). Then for every* $p\in[1,p')$, *there exists a real constant* $\kappa_{p,p'}>0$ *such that, for every* $k\in\{0,\dots,n\}$,

$$\|v_k(X_k)-\widehat{v}_k(\widehat{X}_k)\|_p\le\kappa_{p,p'}\left(\sum_{\ell=k}^{n}\Big(C_{n,\ell}([P]_{\mathrm{Lip}},[h_\cdot]_{\mathrm{Lip}})\,\sigma_{p'}(X_\ell)\Big)^{\vartheta_p}N_\ell^{-\frac{\vartheta_p}{d}}\right)^{\frac{1}{\vartheta_p}}$$

where $\sigma_p(X_k)=\min_{a\in\mathbb{R}^d}\|X_k-a\|_p$, $k=0,\dots,n$, *and* $\vartheta_p=2$ *if* $p=2$ *and* $\vartheta_p=1$ *otherwise.*

For a second order scheme (based on Voronoi quantization) which takes full advantage of the stationarity, we refer to [Sellami 2010]. For other other applications (cubature formulas, non-linear filtering, stochastic control, etc) we refer to the surveys [Pagès et al. 2003], [Pagès and Printems 2009] and the reference therein (Voronoi quantization) or [Pagès and Wilbertz 2010a] (dual quantization).

6.2 Design of an optimized quantization tree by simulation

6.2.1 Grid sizes

A first step (however not mandatory) is to minimize the error bound (at the origin) obtained in Proposition 6.3 for a given budget of elementary quantizers $N_0+\cdots+N_n\le N$ (where $N\ge n+1$). The choice of N is usually related to the memory devoted to the computation. An elementary optimization under constraint yields for the sizes of the grids

$$N_k=\left\lfloor\frac{a_kN}{a_0+\cdots+a_n}\right\rfloor\vee 1\ \text{ with }\ a_k=\Big(C_{k,n}([P]_{\mathrm{Lip}},[h_\cdot]_{\mathrm{Lip}})\sigma_{p'}(X_k)\Big)^{\frac{\vartheta_p d}{d+1}},\ k=0,\dots,n.$$

with ϑ like in Proposition 6.3. This allocation is payoff-dependent but, if $\max\limits_{0\le k\le n}[h_k]_{\mathrm{Lip}}<+\infty$, one may replace a_k by $\tilde{a}_k=\Big(\max_{0\le\ell\le n-k}[P]^{\ell}_{\mathrm{Lip}}\sigma_{p'}(X_k)\Big)^{\vartheta_p}$ or even $\tilde{a}_k=\sigma^{\vartheta_p}_{p'}(X_k)$ if, one "controls" $\max_{0\le k\le n}[P]^k_{\mathrm{Lip}}$ (like in the example following Proposition 2.1). In the dual setting, this allocation is an heuristic since we have the additional constraint $N_k\ge N_{d,p,p'}$.

Example. Let $X_k=W_{t_k^n}$, W Standard Brownian motion. Then $\sigma_{p'}(X_k)=c_{p'}\sqrt{t_k^n}$, $k=0,\dots,n$ (and $N_0=1$).

6.2.2 Transition weight estimation

▷ *The "diffusion" method.* Like for the grid optimization, a large L-sample $(X^{(\ell)})_{1\le\ell\le L}$ of the chain is generated and sent "through" the grids. Then one estimates each transition weight by

$$p_{ij}^k = a.s.\text{-}\lim_{L\to\infty} \frac{\sum_{\ell=1}^L \mathbb{P}(\pi_k(X_k^{(\ell)},U_k)=x_i^k, \pi_{k+1}(X_{k+1}^{(\ell)},U_{k+1})=x_j^{k+1}\,|\,X_k^{(\ell)},X_{k+1}^{(\ell)})}{\sum_{\ell=1}^L \mathbb{P}(\pi_k(X_k^{(\ell)},U_k)=x_i^k\,|\,X_k^{(\ell)})} \tag{15}$$

where π_k is specified following the quantization type. We may assume that the integration with respect to U_k and U_{k+1} can be performed explicitly by a closed form solution (keeping in mind that (U_k) and (X_k) are independent). This holds trivially true for Voronoi quantization, but also for dual quantization as we will see later on.

The strong consistency follows then from the Strong Law of large Numbers since

$$\begin{aligned}\mathbb{E}\Big(\mathbb{P}\big(\pi_k(X_k^{(\ell)},U_k)=x_i^k, \pi_{k+1}(X_{k+1}^{(\ell)},U_{k+1})=x_j^{k+1}\,|\,X_k^{(\ell)},X_{k+1}^{(\ell)}\big)\Big)\\ = \mathbb{P}\big(\pi_k(X_k^{(\ell)},U_k)=x_i^k, \pi_{k+1}(X_{k+1}^{(\ell)},U_{k+1})=x_j^{k+1}\big)\end{aligned}$$

and

$$\mathbb{E}\Big(\mathbb{P}\big(\pi_k(X_k^{(\ell)},U_k)=x_i^k\,|\,X_k^{(\ell)}\big)\Big) = \mathbb{P}\big(\pi_k(X_k^{(\ell)},U_k)=x_i^k\big).$$

When π_k does not depend on the exogenous noise (like for Voronoi quantization), the above estimator coincide with the naive one, that is

$$p_{ij}^k = a.s.\text{-}\lim_{L\to\infty} \frac{\sum_{\ell=1}^L \mathbf{1}_{\{\pi_k(X_k^{(\ell)},U_k)=x_i^k, \pi_{k+1}(X_{k+1}^{(\ell)},U_{k+1})=x_j^{k+1}\}}}{\sum_{\ell=1}^L \mathbf{1}_{\{\pi_k(X_k^{(\ell)},U_k)=x_i^k\}}}.$$

∘ To be precise, in the case of Voronoi quantization, it holds

$$\pi_k(X_k^{(\ell)},U_k)=x_i^k \quad\Longleftrightarrow\quad X_k^{(\ell)}\in C_i(\Gamma_k),$$

where $C_i(\Gamma_k)$, $i=1,\dots,N_k$, denotes a Voronoi partition of $\mathbb{R}^d$, so that (15) finally reads

$$p_{ij}^k = a.s.\text{-}\lim_{L\to\infty} \frac{\sum_{\ell=1}^L \mathbf{1}_{\{X_k^{(\ell)}\in C_i(\Gamma_k)\cap X_{k+1}^{(\ell)}\in C_j(\Gamma_{k+1})\}}}{\sum_{\ell=1}^L \mathbf{1}_{\{X_k^{(\ell)}\in C_i(\Gamma_k)\}}}.$$

Note here, as far as implementation is concerned, we do not need to construct the whole Voronoi diagram of the grids Γ_k. It is sufficient to perform a Nearest Neighbor search to estimate the transition probabilities as it can be seen in Algorithm 1.

∘ As for dual quantization, it holds for $X_k\in\mathrm{conv}(\Gamma_k)$, with the notation from Section 4.2,

Algorithm 1 Transition probability estimation for Voronoi quantization

for $\ell = 1,\dots,L$ **do**
  $x \leftarrow x_0$, $i \leftarrow 0$, $p_1^i \leftarrow 1$
  for $k = 1,\dots,n$ **do**
    Simulate X_k^ℓ given X_{k-1}^ℓ
    Find Nearest Neighbor-Index j of X_k^ℓ in Γ_k
    Set
      $p_{ij}^k += 1$
      $p_j^{k+1} += 1$
    $i \leftarrow j$
  end for
end for
Set $\mathbf{p_{ij}^k} \leftarrow \frac{p_{ij}^k}{p_i^k}, \quad 1 \le i,j \le N_k, 1 \le k \le n$

$$\mathbb{P}(\pi_k(X_k^{(\ell)}, U_k) = x_i^k) = \sum_{I \in \mathscr{I}(\Gamma_k): i \in I} \int_{D_I(\Gamma_k)} \lambda_{x_i^k}^I(\xi)\, \mathbb{P}_X(d\xi),$$

where $D_I(\Gamma_k)$, $I \in \mathscr{I}(\Gamma_k)$ denotes a Delaunay partition of $\mathrm{conv}(\Gamma_k)$ and $\lambda_{x_i^k}^I(\xi)$, $i \in I$, denotes the barycentric coordinates of ξ with respect to "its" Delaunay d-simplex.

The estimation of the transition probabilities p_{ij}^ks then can be implemented as shown in Algorithm 2.

Algorithm 2 Transition probability estimation for dual quantization

for $\ell = 1,\dots,L$ **do**
  $x \leftarrow x_0$, $i \leftarrow 0$, $p_1^i \leftarrow 1$
  for $k = 1,\dots,n$ **do**
    Simulate X_k^ℓ given X_{k-1}^ℓ
    Find Delaunay hyper-triangle τ_k of X_k^ℓ in Γ_k
    Update $p_{\cdot,\cdot}^k$ w.r.t. barycentric coordinates of (X_{k-1}^ℓ, X_k^ℓ) (τ_{k-1}, τ_k)
    Update $p_{\cdot}^{k+1}$ w.r.t. barycentric coordinates of X_k^ℓ in τ_k
  end for
end for
Set $\mathbf{p_{ij}^k} \leftarrow \frac{p_{ij}^k}{p_i^k}, \quad 1 \le i,j \le N_k, 1 \le k \le n$

Although this transition probability estimation by Monte-Carlo simulation is usually the most time consuming part of the quantization tree algorithm in practice, one has to emphasize here, that both above algorithms can be parallelized very efficiently. This is indeed of special importance since the availability of massive parallel computing device at very low price like as *GPGPU*s. It was shown in [Pagès and Wilbertz 2011], that the computational time for transition probability estimation can be reduced by a factor 200 when implemented on a *GPGPU* device.

▷ *The spray method.* One can decouple the computation of the transitions at each time step by noting that

$$\mathscr{L}\Big(\pi_{k+1}(X_{k+1},U_{k+1})=x_j^{k+1}\,|\,\pi_k(X_k,U_k)=x_i^k\Big)\approx\mathscr{L}\Big(\pi_{k+1}(X_{k+1},U_{k+1})=x_j^{k+1}\,|\,X_k=x_i^k\Big).$$

The distribution on the right hand side is easy to simulate (since the chain is supposed to be simulatable). Consequently one can perform a Monte Carlo simulation based on this distribution to estimate (approximately) the p_{ij}^ks. As concerns Voronoi quantization, it has been shown in [Pagès et al. 2003] that the error induced by such an approximation is of second order if the grids Γ_k are stationary.

Decoupling the estimation of the successive transition matrices makes possible to perform a new parallelization of the estimation procedure (see [Bronstein et al. 2010]) with again a significant reduction of the computation time down to a few seconds on a *GPGPU* device.

6.3 Martingale correction: an efficient heuristics

When the structure process $(X_k)_{0\le k\le n}$is a martingale (*e.g.* a discounted set of d risky assets under a risk neutral martingale probability, or a Brownian motion at times $t_k^n=\frac{kT}{n}$, etc) and $X_0=x_0$, the quantization based approaches do not preserve naturally this property (or any dynamical property). One way to proceed is to slightly modify the grids Γ_k as follows:

– Define by a backward induction $\widetilde{\Gamma}_n=\Gamma_n$ and for every $k=0,\ldots,n-1$,

$$\widetilde{\Gamma}_k=\Big\{x_1^k,\ldots,x_{N_k}^k\Big\}\quad\text{where}\quad\widetilde{x}_i^k=\sum_{j=1}^{N_{k+1}}p_{ij}^k\widetilde{x}_j^{k+1},\ i=1,\ldots,N_k.$$

– Re-center the grids by setting

$$\Gamma_k^{mart}=\widetilde{\Gamma}_k+x_0-\widetilde{x}_0.$$

The resulting quantization tree $(\Gamma_k^{mart},\mathbf{p}^k)_{0\le k\le n}$ has the distribution of a martingale starting at x_0 at time 0. Although it often significantly improves numerical results, theoretical error bounds no longer hold. It is observed in practice that the translation $x_0-\widetilde{x}_0$ is negligible.

7 Numerical experiments

7.1 Swing Options

We begin the numerical illustrations by the example of the pricing of swing options in a 2-factor Gaussian model. Such a problem consists in solving the normalized stochastic control problem (interest rate is neglected)

$$\operatorname{esssup}\left\{\mathbb{E}\left(\sum_{k=0}^{n-1} q_k\big(v_k(X_k)-K\big)|\mathscr{F}_0\right), q_k:(\Omega,\mathscr{F}_k)\to[0,1], \bar{q}_n\in[Q_{\min},Q_{\max}]\right\} \tag{16}$$

for global consumption couple $(Q_{\min},Q_{\max})\in\mathbb{N}^2$ and a cumulated consumption before time k given by $\bar{q}_k := \sum_{l=0}^{k-1} q_l$. The sequence $(X_k)_{0\le k\le n}$ is two-dimensional Gaussian Markov process specified below and$S_{t_k} = v_k(X_k)$ stands for the price of the underlying risky asset at time $t_k = \frac{kT}{n}$ (interest rates are assumed to be 0). As shown in [Bardou et al. 2010b] there exists an optimal bang-bang control for this problem, which leads, in combination with the *BDPP*, to

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

with admissible set $I_{Q^k}^M := [(Q_{\min}^k - M)_+ \wedge 1, Q_{\max}^k \wedge 1]$ and
$\chi^M(Q^k,x) := \big((Q_{\min}^k - x)_+, (Q_{\max}^k - x)\wedge M\big)$ so that $P_0^n(Q_{\min},Q_{\max})$ is a solution to (16).

A straightforward quantization of this problem then reads

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

and error bounds have been established in [Bardou et al. 2010b]. Note here that the computation of the conditional expectations $\mathbb{E}\big[\hat{P}_{k+1}^n(\chi^{n-k-1}(Q^k,x))|\hat{X}_k = x_i^k\big]$ becomes straightforward owing to Section 6 since it holds $\mathbb{E}(f(\hat{X}_{k+1})|\hat{X}_k = x_i^k) = \sum_{j=1}^{N_{k+1}} p_{ij}^k f(x_j^{k+1})$.

Furthermore we will focus here on the case $Q_{\min}=0$, $Q_{\max}=n$ so that the solution P_0^n has the representation

$$P_0^n = \sum_{k=1}^{n}(v_k(X_k)-K)_+.$$

We therefore may hope that due to this simple structure as a strip of calls and in view of section 5 that stationarity may play an important role for the numerical results.

The structure Markov process $(X_k)_{0\le k\le n}$ is specified as in [Bronstein et al. 2010] by

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

so that the 2-factor underlying risky asset is given at time t_k by $v_k(X_k)$ with $v_k(x_1,x_2) = s_0 \exp\left(\sigma_1 x_1 + \sigma_2 x_2 - \frac{1}{2}\mu_{t_k}\right)$ where μ_{t_k} is chosen so that $\mathbb{E}(S_{t_k} = s_0$, $0 \le k \le n$. The numerical parameters here read in detail as

$$s_0 = 20,\ \alpha_1 = 1.11,\ \alpha_2 = 5.4,\ \sigma_1 = 0.36,\ \sigma_2 = 0.21,\ \rho = -0.11,\ n = 30$$

i.e. we have a Gaussian process (X_k) with a true correlation. Note that in such a setting the transformation of an optimal and stationary Voronoi quantization grid for the bivariate standard normal distribution into one with correlation ρ destroys already the stationarity property in the transformed grid. In the case of dual quantization, stationarity for the transformed grid is at least preserved on $\mathrm{conv}(\Gamma)$.

As shown in Figures 4 and 5 the dual methods outperforms clearly the Voronoi approach, which is mainly caused by the intrinsic stationarity of the Delaunay quantization mapping.

Moreover, we already observe that Dual quantization tends to lead to an upper bound whereas Voronoi quantization is approaching from below. (Both those observations hold true in general for convex functions F and stationary quantizers $\widehat{X}$.)

7.2 Bermuda options

First we recall the following basic fact: in classical non-arbitrage theory of contingent claims, it is well-known that, in a complete market, the discounted fair price of a Bermuda option with payoff process $(h_k(S_{t_k}))_{0\le k\le n}$, $0 = t_0 < t_1 < \ldots < t_k \ldots < t_n = T$, is the Snell envelope of the discounted payoff process so that

$$\frac{\text{Premium}_{t_k}}{S^0_{t_k}} = \text{Snell}_{\mathbb{P}^*}\left(\frac{h_k(S_{t_k})}{S^0_{t_k}}\right)_{0\le k\le n}$$

where $(S^0_t)_{t\in[0,T]}$ is the (positive) *numéraire* (also called "riskless asset" with $S^0_0 = 1$) and $S_t = (S^1_t,\ldots,S^d_t)_{t\in[0,T]}$ is the risky asset price $(0,\infty)^d$-valued process and $\mathbb{P}^*$ is a/the risk-neutral probability. Strictly speaking, we assume this "numéraire" to be deterministic to fit the scope of this paper. In what follows Bermuda options

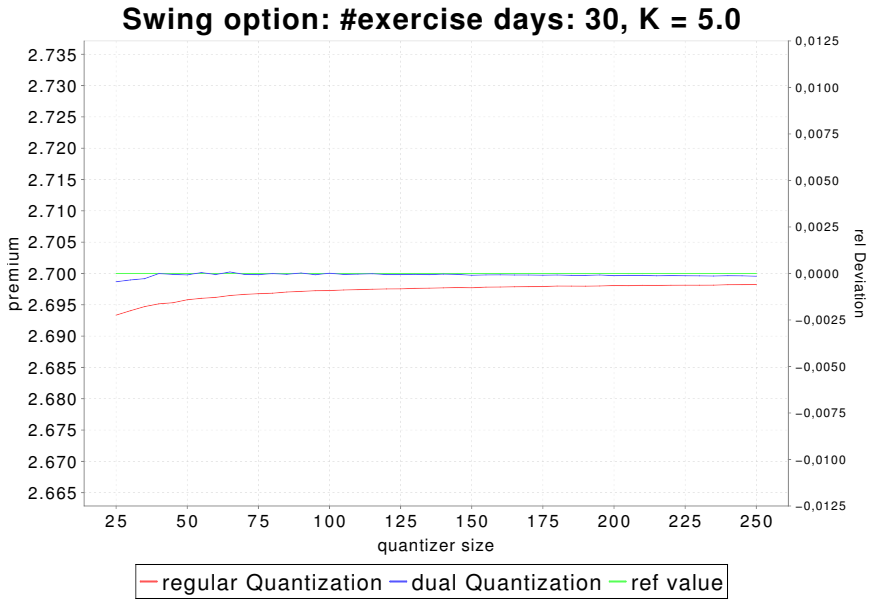

Fig. 4 Convergence of the quantization methods as function of the average grid size N.

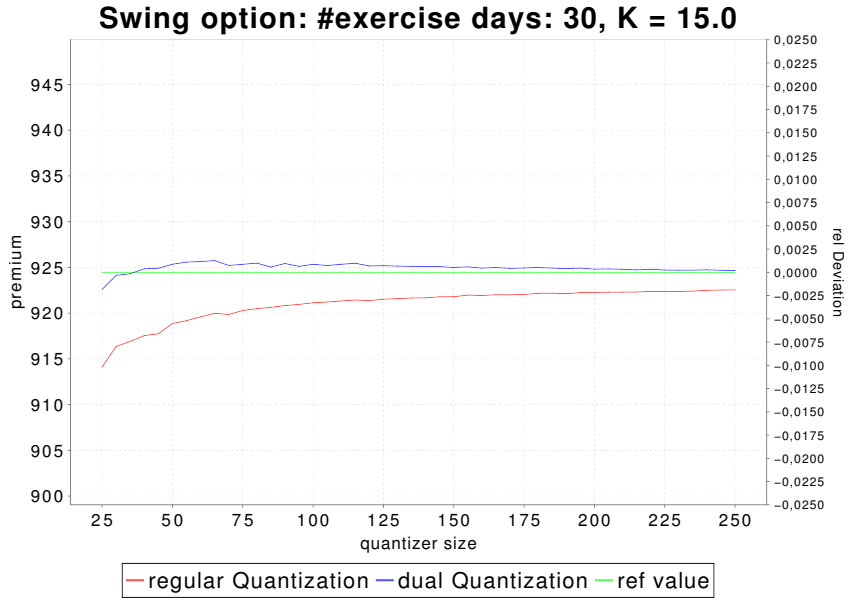

Fig. 5 Convergence of the quantization methods as function of the average grid size N.

appear as time approximation of American options (see [Bally and Pagès 2003b] for various time discretization error bounds).

7.2.1 Geometric Exchange Option

We now consider the case of a geometric exchange put option in a multi-dimensional Black Scholes model with maturity T and 11 exercise dates $k\frac{T}{10}$, $k = 0,\dots,10$. That means that $S_t^0 = e^{-rt}$ and that the the underlyings $(S_t^i)_{t\in[0,T]}$, $i = 1,\dots,d$, are given by the (uncorrelated) Black-Scholes dynamics:

$$S_t^i = s_0^i \exp\Big((r-\delta_i-\frac{\sigma_i^2}{2})t+\sigma_i W_t^i\Big),\ s_0^i>0,$$

$W = (W^1,\dots,W^d)$ standard Brownian motion, and the payoff of this option reads for $d = 2k$

$$\varphi(S_t^1,\dots,S_t^d) = \Big(\prod_{i=1}^{k} S_t^i - \prod_{i=k+1}^{d} S_t^i\Big)_+.$$

Example 1. As parameters we have chosen a Bermudan option with maturity $T = 1$, 11 exercise dates: $k/10$, $k = 0,\dots,10$, and

$$s_0^i = 40^{\frac{2}{d}},\ i=1,\dots,k, \quad s_0^i = 40^{\frac{2}{d}},\ i=k+1,\dots,d, \quad r = 0.05,$$
$$\sigma_i = 0.2,\ i=1,\dots,d, \quad \delta_i = 0.05, i=1,\dots,k, \quad \delta_i = 0.0,\ i=k+1,\dots,d.$$

These settings can be reduced for any d to a 2-dimensional exchange option for which we computed reference values using a Boyle-Evnine-Gibbs tree with 10.000 time steps.

The resulting log-log plots of the convergence for Voronoi and Dual quantization can be found in Figures 6 and 7.

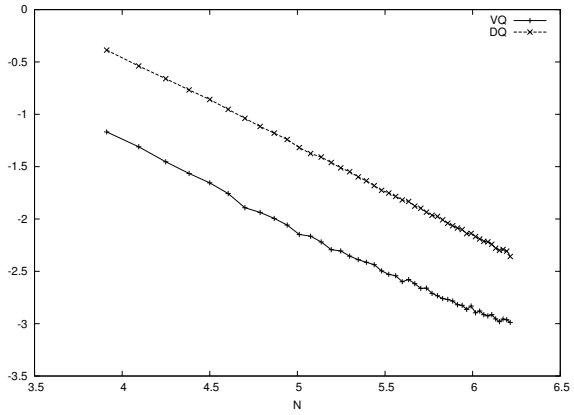

Fig. 6 Log-Log plot of quantization methods for the geometric exchange option in dimension 2.

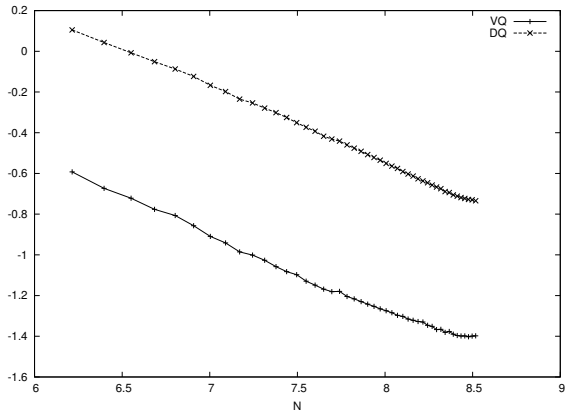

Fig. 7 Log-Log plot of quantization methods for the geometric exchange option in dimension 4.

One observes here again that dual quantization approach yields a slightly better rate (cf. Table 1) than the Voronoi quantization approximation.

	2d	4d
Voronoi Quantization	0.73	0.36
Dual Quantization	0.86	0.38

Table 1 Rates of convergence for the exchange option.

Note moreover that the upper bound in Proposition 6.3 promises only an optimal rate of 0.5 in dimension 2 and 0.25 in dimension 4. Therefore it seems that also in this example there is some more smoothness to capture which leads in practice to better rates than those for the worst case error within class of Lipschitz functionals.

Due to the very smooth convergence seen in Figures 6 and 7, we furthermore apply a Richardson-Romberg extrapolation on the error expansion

$$\mathbb{E}F(X) \approx \mathbb{E}F(\widehat{X}) + \kappa N^{-\alpha},$$

which is a pure heuristic but has a theoretical justification for stationary quantizer (see, *e.g.*, [Pagès and Printems 2009]). We therefore use the rates α from Table 1 and extrapolate the unknown κ using two different grids sizes N_1 and N_2. As a result, we obtain in the above setting for

$$\hat{P}_0^{\mathrm{Rom}} = \hat{P}_0^{N_1} + \frac{\hat{P}_0^{N_1} - \hat{P}_0^{N_2}}{N_2^{-\alpha} - N_1^{-\alpha}} N_1^{-\alpha}$$

a stable and fast convergence as shown in Figures 8 and 9 for dimensions 2 and 4. These experiments suggest to adopt the *mid-price* $0.5 \times (Price_{VQ} + Price_{DQ})$.

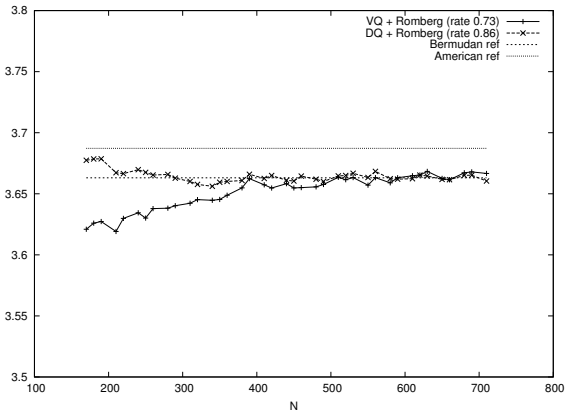

Fig. 8 Convergence of the extrapolated quantization methods for the geometric exchange option in dimension 2.

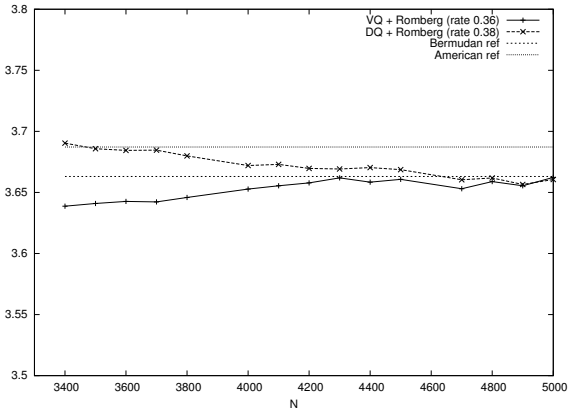

Fig. 9 Convergence of the extrapolated quantization methods for the geometric exchange option in dimension 4.

Alternatively, following the commonly shared idea of (temporarily) including the payoff in the regression basis of Longstaff-Schwartz's algorithm, one may use the European price of the exchange option as a control variate. This means that the BDPP reads

$$\tilde{V}_n = \varphi_{t_n}(X_n) - C^{\mathrm{Eur}}_{T-t_n}(X_n)$$
$$\tilde{V}_k = \max\Big\{ \varphi_{t_k}(X_k) - C^{\mathrm{Eur}}_{T-t_k}(X_k), \mathbb{E}\big(\tilde{V}_{k+1}\big|X_k\big) \Big\},\ 0 \le k \le n-1,$$

where $C^{\mathrm{Eur}}_t(x)$ is the European price for maturity t and initial Stock price x.

Consequently, the true price V_0 is given by

$$V_0 = \tilde{V}_0 + C_T^{\mathrm{Eur}}(X_0).$$

Numerical results for the above setting are given in Figures 10 and 11.

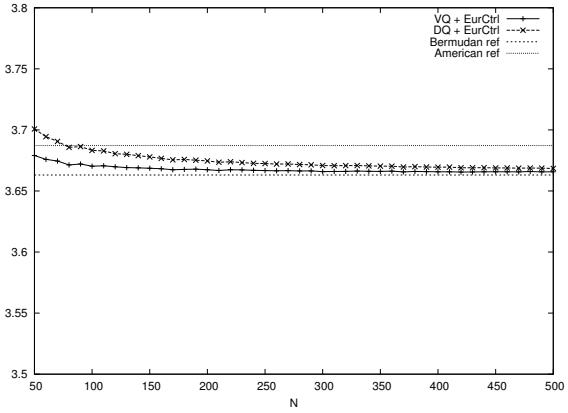

Fig. 10 Convergence of quantization methods with European control variate for the geometric exchange option in dimension 2.

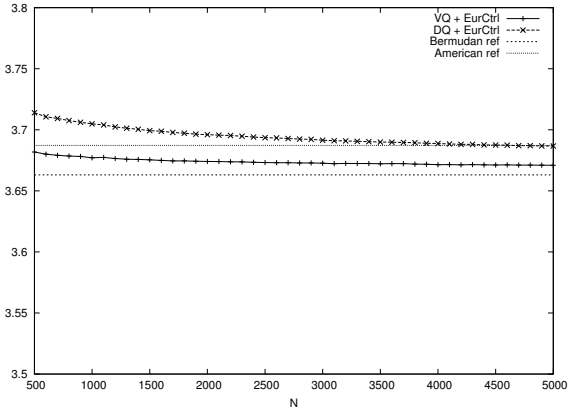

Fig. 11 Convergence of quantization methods with European control variate for the geometric exchange option in dimension 4.

7.2.2 Put-On-The-Min option

A final comparison is taken out on the example of an put-on-the-min option in a two dimensional Black Scholes model. The payoff of this option reads

$$\varphi(S_t^1, S_t^2) = \Big(K - \min(S_t^1, S_t^2)\Big)_+.$$

Here again the reference values were computed using a Boyle-Evnine-Gibbs tree with 10 000 time steps.

We compare the dual quantization approach including the martingale correction of Section 6.3 to the Longstaff-Schwartz (*L-S*) approach from the Premia software package, see [Premia (Inria)]. For the *L-S* procedure, we have chosen a family of 22 independent functions (21 monomial functions + the payoff function) and plotted in Figure 12 a Monte Carlo simulation with an increasing number of sample paths ranging from 10.000 to 100.000 and its 95% confidence interval.

This setting was chosen to arrive at approximately equal computational times for the *L-S* approach and the dual quantization method.

One clearly sees in Figure 12 that the quantization approach with martingale correction provides already for small N a very good approximation to the true value of the Bermuda option. In addition, the *L-S* approach suffers from a higher volatility, since it is more depending on the Monte Carlo error than the quantization tree approach, which contains the critical MC-Simulation only in the weight estimation.

Furthermore we have also plotted in Figure 12 the Monte Carlo estimation by an *L-S* approach from the *Premia* software package in order to compare results.

Example 2. 2-asset (correlated) Black-Scholes model with maturity $T = 1$ and 11 exercise times, $k\frac{T}{10}$, $k = 0, \ldots, 10$,

$$s_0^1 = s_0^2 = 40,\ r = 0.05,\ \sigma_1 = 0.2,\ \sigma_2 = 0.3,\ \rho = 0.5,\ K = 40,$$

for a put on the min, *i.e.* payoff

$$\varphi(S_t^1, S_t^2) = \Big(K - \min(S_t^1, S_t^2)\Big)_+.$$

As underlying Markov process X_k we have chosen a 2-dimensional Brownian Motion $W = (W^1, W^2)$ with correlation ρ.

As a global conclusion, optimal quantization methods show their efficiency in various fields of Applied Probability (American pricing, stochastic control, nonlinear filtering, etc) in medium dimensions, say $1 \le d \le 5$, and sometimes higher ones when using and, if necessary, combining in an appropriate way speeding up methods like Romberg extrapolation, martingale correction, control variate like procedures, etc. We refer to survey papers devoted to other applications like [Pagès et al. 2003] for more numerical experiments. *Quantization trees appear as space discretizations of the global underlying Markov dynamics*. Such methods can take advantage either of the opportunity of an offline pre-processing or of recent massive parallelization techniques (*GPGPU*). The second (on-line) phase, consisting of a tree descent, is in any case instantaneous at a human scale.

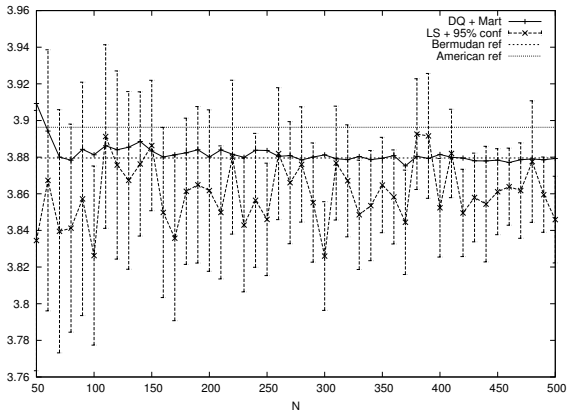

Fig. 12 Convergence of quantization methods for a put-on-the-min option in dimension 2.

In higher dimensions, recent works on quantization based stratified sampling (see [Corlay and Pagès 2010]) suggest that quantization could also be used to optimally stratify a forward Monte Carlo simulation.

COMPUTATION DEVICE. All numerical illustrations were computed on GNU Linux 2.6.27.56 and `SUN Java SE 6` JVM. For numerical experiments involving *GPGPU* (only for Voronoi quantization) we refer to [Bronstein et al. 2010] and [Pagès and Wilbertz 2011].

ACKNOWLEDGEMENT. Parts of this work has benefited from helpful discussions with S. Bouthemy and N. Casini (GDF-SUEZ).

References

[Abaya and Wise 1982] ABAYA, E.F. AND WISE, G.L. [1982]: On the existence of optimal quantizers. *IEEE Trans. Inform. Theory*, **28**, 937-940.

[Abaya and Wise 1984] ABAYA, E.F. AND WISE, G.L. [1984]: Some remarks on the existence of optimal quantizers. *Statistics and Probab. Letters*, **2**: 349-351.

[Bally et al. 2001] BALLY, V., PAGÈS, G. AND PRINTEMS, J. [2001]: A Stochastic quantization method for nonlinear problems, *Monte Carlo Methods and Appl.*, **7**(1):21-34.

[Bally and Pagès 2003a] BALLY, V., PAGÈS, G. [2003]: A quantization algorithm for solving discrete time multidimensional optimal stopping problems, *Bernoulli*, **9**(6):1003-1049.

[Bally and Pagès 2003b] V. BALLY, G. PAGÈS [2003]: Error analysis of the quantization algorithm for obstacle problems, *Stochastic Processes & Their Applications*, **106**(1):1-40.

[Bally et al. 2003] BALLY, V., PAGÈS, G. AND PRINTEMS, J. [2003]: First order schemes in the numerical quantization method, *Mathematical Finance* **13**(1):1-16.

[Bally et al. 2005] BALLY, V., PAGÈS, G. AND PRINTEMS, J. [2005]: A quantization tree method for pricing and hedging multidimensional American options, *Mathematical Finance*, **15**(1):119-168.

[Bardou et al. 2010a] BARDOU, O., BOUTHEMY, S. AND PAGÈS, G. [2009]: Optimal quantization for the pricing of swing options, *Applied Mathematical Finance*, **16**(2):183-217.

[Bardou et al. 2010b] BARDOU, O., BOUTHEMY, S. AND PAGÈS, G. [2010]: When are swing option bang-bang?, *International Journal for Theoretical and Applied Finance*, **13**(6):867-899.

[Benaïm et al. 1998] BENAÏM, M., FORT, J.C. AND PAGÈS, G. [1998]: About the convergence of the one dimensional Kohonen algorithm, *Advances in Applied Probability*, **30**(3):850-869.

[Benveniste et al. 1990] BENVENISTE, A., MÉTIVIER, M. AND PRIOURET, P. [1990]: *Adaptive algorithms and stochastic approximations*, Translated from the French by Stephen S. Wilson. Applications of Mathematics **22**, Springer-Verlag, Berlin, 365 p.

[Bouton and Pagès 1993] BOUTON, C. AND PAGÈS, G. [1993]: Self-organization and *a.s.* convergence of the 1-dimensional Kohonen algorithm with non uniformly distributed stimuli, *Stochastic Processes and their Applications*, **47**:249-274 .

[Bowyer 1981] BOWYER, A. [1981]: Computing Dirichlet tessellations. *The Computer Journal*, **24**(2):162-166.

[Bronstein et al. 2010] BRONSTEIN A.L., PAGÈS, G., WILBERTZ, B.[2010]: A quantization tree algorithm: improvements and financial applications for swing options, *Quantitative Finance*, **10**(9):995-1007.

[Bucklew and Wise 1982] BUCKLEW, J.A. AND WISE, G.L. [1982]: Multidimensional asymptotic quantization theory with r^{th} power distortion. *IEEE Trans. Inform. Theory*, **28**(2):239-247.

[Cohort 1998] COHORT, P. [1998]: Limit theorems for random normalized distortion, *Annals of Applied Probability*, **14**(1):118-143.

[Corlay 2011] CORLAY, S. [2011]: A fast nearest neighbour search algorithm based on vector quantization, PhD Thesis, in progress.

[Corlay and Pagès 2010] CORLAY, S. PAGÈS, G. [2010] : Functional quantization based stratified sampling methods. Pre-pub PMA-1341.

[Devroye et al. 2004] DEVROYE, L. LEMAIRE, C.AND MOREAU, J.-M. [2004]: Expected time analysis for Delaunay point location, *Computational Geometry*, **29**(2):61-89

[Du and Gunzburger 2002] DU, Q. AND GUNZBURGER, M. [2002]: Grid generation and optimization based on centroidal Voronoi tessellations, *Appl. Math. and Comput.*, **133**(4):591-607.

[Duflo 1996] DUFLO, M. [1996]: *Algorithms stochastiques*, coll. SMAI Mathématiques & Applications, **23**, Springer, 319p.

[Friedman et al. 1977] FRIEDMAN, J. H., BENTLEY, J.L. AND FINKEL R.A. [1977]: An Algorithm for Finding Best Matches in Logarithmic Expected Time, *ACM Transactions on Mathematical Software*, **3**(3):209-226.

[Gobet et al. 2007] GOBET, E., PAGÈS, G. PHAM, H. AND PRINTEMS, J. [2007]: Discretization and simulation of the Zakai Equation, *SIAM J. on Numerical Analysis*, **44**(6):2505-2538.

[Gobet et al. 2005] GOBET, E., PAGÈS, G. PHAM, H. AND PRINTEMS, J. [2005]: Discretization and simulation for a class of SPDEs with applications to Zakai and McKean-Vlasov equation, pre-pub. PMA-958.

[Gersho and Gray 1992] GERSHO, A. AND GRAY, R.M. [1992]: *Vector Quantization and Signal Compression*. Kluwer, Boston.

[Graf and Luschgy 2000] GRAF, S. AND LUSCHGY, H. [2000]: *Foundations of Quantization for Probability Distributions*. Lect. Notes in Math. 1730, Springer, Berlin, 230p.

[Iri et al. 1984] IRI, M., MUROTA, K., AND OHYA, T.[1984]: A fast Voronoi-diagram algorithm with applications to geographical optimization problems. In P. Throft-Christensen, editor, *Proceedings of the 11th IFIP Conference Copenhagen*, *Lecture Notes in Control and Information Science*, **59**, 273–288.

[Kieffer 1982] KIEFFER, J.C. [1982]: Exponential rate of convergence for Lloyd's Method I, *IEEE Trans. Inform. Theory*, **28**(2), 205-210.

[Kieffer 1983] KIEFFER, J.C. [1983]: Uniqueness of locally optimal quantizer for log-concave density and convex error weighting functions, *IEEE Trans. Inform. Theory*, **29**, 42-47.

[Kushner and Yin 2003] KUSHNER, H. J., YIN, G. G. [2003]: *Stochastic approximation and recursive algorithms and applications*. Second edition. Applications of Mathematics **35**. Stochastic Modelling and Applied Probability. Springer-Verlag, New York, 474p.

[Lapeyre et al. 1990] LAPEYRE, B., SAB, K. AND PAGÈS, G. [1990]: Sequences with low discrepancy. Generalization and application to Robbins-Monro algorithm, *Statistics*, **21**(2): 251-272.

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

[Luschgy and Pagès 2008] LUSCHGY, H., PAGÈS, G. [2008]: Functional Quantization Rate and mean regularity of processes with an application to Lévy Processes, *Annals of Applied Probability*, **18**(2):427-469.

[McNames 2001] MCNAMES, J. [2001]: A Fast Nearest-Neighbor Algorithm Based on a Principal Axis Search Tree, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(9), 964-976.

[Mrad and Ben Hamida 2006] MRAD, M., BEN HAMIDA, S. [2006]: Optimal Quantization: Evolutionary Algorithm *vs* Stochastic Gradient, *Proceedings of the 9th Joint Conference on Information Sciences*.

[Muecke et al. 1999] MÜCKE, E.P., SAIAS, I. AND ZHU, B. [1999]: Fast randomized point location without preprocessing in two- and three-dimensional Delaunay triangulations. *Computational Geometry*, **12**(1-2), 63-83.

[Newman 1982] NEWMAN, D.J. [1982]: The Hexagon Theorem. *IEEE Trans. Inform. Theory*, **28**, 137-138.

[Okabe et al. 2000] OKABE, A. BOOTS, B. SUGIHARA K. AND CHIU S.N. [2000]: *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd Edition, Wiley, New York, 696p.

[Pagès 1993] PAGÈS, G. [1993]: Voronoi tessellation, space quantization algorithm and numerical integration. *Proceedings of the ESANN'93*, M. Verleysen Ed., Editions D Facto, Bruxelles, 221-228.

[Pagès 1998] PAGÈS, G. [1998]: A space vector quantization method for numerical integration, *J. Computational and Applied Mathematics*, **89**:1-38.

[Pagès et al. 2003] PAGÈS, G., PHAM, H. AND PRINTEMS, J. [2003]: Optimal quantization methods and applications to numerical methods in finance. *Handbook of Computational and Numerical Methods in Finance*, S.T. Rachev ed., Birkhäuser, Boston, 429p.

[Pagès and Printems 2003] PAGÈS, G. AND PRINTEMS, J. [2003]: Optimal quadratic quantization for numerics: the Gaussian case, *Monte Carlo Methods and Appl.*, **9**(2):135-165.

[Pagès et al. 2004] PAGÈS, G., PHAM, H. AND PRINTEMS, J. [2004]: An Optimal Markovian Quantization Algorithm for Multidimensional Stochastic Control Problems, *Stochastics and Dynamics*, **4**(4):501-545.

[Pagès and Printems 2005] PAGÈS, G. AND PRINTEMS, J. [2005]: `www.quantize.maths-fi.com`, website devoted to optimal vector and functional quantization.

[Pagès and Pham 2005] PAGÈS, G., AND PHAM, H. [2005]: Optimal quantization methods for nonlinear filtering with discrete-time observations, *Bernoulli*, **11**(5):893-932.

[Pagès and Printems 2009] PAGÈS, G., PRINTEMS, J. [2009]: Optimal quantization for finance: from random vectors to stochastic processes, chapter in *Mathematical Modeling and Numerical Methods in Finance* (special volume) (A. Bensoussan, Q. Zhang guest eds.), coll. Handbook of Numerical Analysis (P.G. Ciarlet Editor), North Holland, 595-649.

[Pagès and Wilbertz 2009] PAGÈS, G. AND WILBERTZ W. [2009]: Dual Quantization for random walks with application to credit derivatives, pre-pub PMA-1322, to appear in *Journal of Computational Finance*.

[Pagès and Wilbertz 2010a] PAGÈS, G. AND WILBERTZ W. [2010]: Intrinsic stationarity for vector quantization: Foundation of dual quantization, pre-pub PMA-1393.

[Pagès and Wilbertz 2010b] PAGÈS, G. AND WILBERTZ W. [2010]: Sharp rate for the dual quantization problem, pre-pub PMA-1402.

[Pagès and Wilbertz 2011] PAGÈS, G. AND WILBERTZ W.[2011]: GPGPUs in computational finance: Massive parallel computing for American style options, pre-pub PMA 1385, to appear in Concurrency and Computable: Practice and Experience.

[Pärna 1990] PÄRNA, K. [1990]: On the existence and weak convergence of k-centers in Banach spaces, *Tartu Ülikooli Toimetised*, **893**:17-287.

[Pham et al. 2005] PHAM, H. SELLAMI, A. AND RUNGGALDIER W. [2005] :Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation, *Monte Carlo Methods and Applications*, **11**(1):57-81.

[Pollard 1982] POLLARD, D. [1982]: Quantization and the method of k-means. *IEEE Trans. Inform. Theory*, **28**(2):199-205.

[Premia (Inria)] Premia software by MATHFI team (Inria), `www-rocq.inria.fr/mathfi/Premia/index.html`.

[Sellami 2010] SELLAMI A. [2010]: Quantization Based Filtering Method Using First Order Approximation, *SIAM J. on Num. Anal.*, **47**(6):4711-4734.

[Sellami 2009] SELLAMI, A. [2010]: Comparative survey on nonlinear filtering methods: the quantization and the particle filtering approaches, *Journal of Statistical Computation and Simulation*, **78**(2):93-113.

[Trushkin 1982] TRUSHKIN, A.V. [1982]: Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions, *IEEE Trans. Inform. Theory*, **28**(2):187-198.

[Wilbertz 2005] WILBERTZ, B. [2005]: Computational aspects of functional quantization for Gaussian measures and applications, diploma thesis, Univ. Trier (Germany).

[Zador 1963] ZADOR, P.L. [1963]: Development and evaluation of procedures for quantizing multivariate distributions. Ph.D. dissertation, Stanford Univ. (USA).

[Zador 1982] ZADOR, P.L. [1982]: Asymptotic quantization error of continuous signals and the quantization dimension, *IEEE Trans. Inform. Theory*, **28**(2), 139-149.