Optimal quantization for finance: from random vectors to stochastic processes

Gilles Pagès [*]   Jacques Printems [†]

Abstract

In this paper, we present an overview of the recent developments of vector quantization and functional quantization and their applications as a numerical method in Finance, with an emphasis on the quadratic case. Quantization is a way to approximate a random vector or a stochastic process, viewed as a Hilbert-valued random variable, using a nearest neighbour projection on a finite codebook. We make a review of cubature formulas to approximate expectation an conditional expectation, including the introduction of a quantization based Richardson-Romberg extrapolation method. The optimal quadratic quantization of the Brownian motion is presented in full details. A special emphasis is made on the computational aspects and the numerical applications, in particular the pricing of different kinds of options in various fields (swing options on gas, options in a Heston stochastic volatility model).

1   Introduction

Quantization is a way to discretize the path space of a random phenomenon: a random vector in finite dimension, a stochastic process in infinite dimension. Optimal Vector Quantization theory (finite-dimensional) random vectors finds its origin in the early 1950's in order to discretize some emitted signal (see (Gersho and Gray 1992) or (Graf and Luschgy 2000)). It was further developed by specialists in Signal processing and later in Information Theory.

The infinite dimensional case started to be extensively investigated in the early 2000's by several authors (see *e.g.* (Pagès 2000), (Luschgy and Pagès 2002), (Dereich et al. 2003), (Luschgy and Pagès 2004), (Luschgy and Pagès 2006), (Dereich et al. 2006), (Luschgy and Pagès 2004), (Wilbertz 2005), (Graf et al. 2007) etc).

Let us consider a Hilbertian setting. One considers a random vector X defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ taking its values in a separable Hilbert space $(H, (.|.)_H)$ (equipped with its natural Borel σ-algebra) and satisfying $\mathbb{E}|X|^2 < +\infty$. When H is an Euclidean space ($\mathbb{R}^d$), one speaks about *Vector Quantization*. When H is an infinite dimensional space like $L_T^2 := L^2([0,T], dt)$ (endowed with the usual Hilbertian norm $|f|_{L_T^2} := (\int_0^T f^2(t)dt)^{\frac{1}{2}}$) one speaks of *functional quantization* (denoted L_T^2 from now on). A (bi-measurable) stochastic process $(X_t)_{t \in [0,T]}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$ satisfying $|X(\omega)|_{L_T^2} < +\infty$ $\mathbb{P}(d\omega)$-*a.s.* can always be seen, once possibly modified on a $\mathbb{P}$-negligible set, as an L_T^2-valued random variable. Although we will focus on the Hilbertian framework, other choices are possible for H, in particular some more general Banach settings like $L^p([0,T], dt)$ or $\mathcal{C}([0,T], \mathbb{R})$ spaces.

This paper is organized as follows: in Sections 2 we introduce quadratic quantization in a Hilbertian setting. In Section 3, we focus on optimal quantization, including some extensions to

---

[*]Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Paris 6, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5, France. E-mail:gpa@ccr.jussieu.fr

[†]LAMA, Université Paris 12, E-mail:printems@univ-paris12.fr

non quadratic quantization. Section 4 is devoted to some quantized cubature formulae. Section 5.1 provides some classical background on the quantization rate in finite dimension. Section 6 deals with functional quantizations of Gaussian processes, like the Brownian motion, with a special emphasis on the numerical aspects. We present here what is, to our guess, the first large scale numerical optimization of the quadratic quantization of the Brownian motion. We compare it to the optimal product quantization, formerly investigated in (Pagès and Printems 2005). In section, we propose a constructive approach to the functional quantization of scalar or multidimensional diffusions (in the Stratanovich sense). In Section 8, we show how to use functional quantization to price path-dependent options like Asian options (in a Heston stochastic volatility model). We conclude by some recent results showing how to derive universal (often optimal) functional quantization rate from time regularity of a process in Section 9 and by a few clues in Section 10 about the specific methods that produce some lower bounds (this important subject as many others like the connections with small deviation theory is not treated in this numerically oriented overview. As concerns statistical applications of functional quantization we refer to (Tarpey et al. 2003a; Tarpey et al. 2003b).

NOTATIONS. $\bullet$ $a_n \approx b_n$ means $a_n = O(b_n)$ and $b_n = O(a_n)$; $a_n \sim b_n$ means $a_n = b_n + o(a_n)$.

$\bullet$ If $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (H, |\,.\,|_H)$ (Hilbert space), then $\|X\|_2 = (\mathbb{E}|X|_H^2)^{\frac{1}{2}}$.

$\bullet$ $\lfloor x \rfloor$ denotes the integral part of the real x.

2 What is quadratic quantization?

Let $(H, (\,.\,|\,.\,)_H)$ denote a separable Hilbert space. Let $X \in L_H^2(\mathbb{P})$ *i.e.* a random vector $X : (\Omega, \mathcal{A}, \mathbb{P}) \longmapsto H$ (H is endowed with its Borel σ-algebra) such that $\mathbb{E}|X|_H^2 < +\infty$. An N-*quantizer* (or N-*codebook*) is defined as a subset

$$\Gamma := \{x_1, \ldots, x_N\} \subset H$$

with card$\Gamma = N$. In numerical applications, Γ is also called *grid*. Then, one can *quantize* (or simply discretize) X by $q(X)$ where $q : H \mapsto \Gamma$ is a Borel function. It is straightforward that

$$\forall\, \omega \in \Omega, \qquad |X(\omega) - q(X(\omega))|_H \geq d(X(\omega), \Gamma) = \min_{1 \leq i \leq N} |X(\omega) - x_i|_H$$

so that the best pointwise approximation of X is provided by considering for q a nearest neighbour projection on Γ, denoted Proj_Γ. Such a projection is in one-to-one correspondence with the Voronoi partitions (or diagrams) of H induced by Γ *i.e.* the Borel partitions of H satisfying

$$C_i(\Gamma) \subset \left\{ \xi \in H : |\xi - x_i|_H = \min_{1 \leq j \leq N} |\xi - x_j|_H \right\} = \overline{C}_i(\Gamma), \qquad i = 1, \ldots, N,$$

where $\overline{C}_i(\Gamma)$ denotes the closure of $C_i(\Gamma)$ in H (this heavily uses the Hilbert structure). Then

$$\mathrm{Proj}_\Gamma(\xi) := \sum_{i=1}^N x_i \mathbf{1}_{C_i(\Gamma)}(\xi)$$

is a nearest neighbour projection on Γ. These projections only differ on the boundaries of the *Voronoi cells* $C_i(\Gamma)$, $i = 1, \ldots, N$. All Voronoi partitions have the same boundary contained in the union of the median hyperplanes defined by the pairs (x_i, x_j), $i \neq j$. Figure 2 represents the Voronoi diagram defined by a (random) 10-tuple in $\mathbb{R}^2$. Then, one defines a *Voronoi N-quantization* of X by setting for every $\omega \in \Omega$,

$$\widehat{X}^\Gamma(\omega) := \mathrm{Proj}_\Gamma(X(\omega)) = \sum_{i=1}^N x_i \mathbf{1}_{C_i(\Gamma)}(X(\omega)).$$
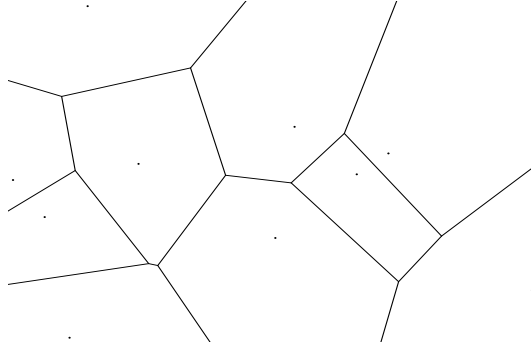
Figure 1: *A 2-dimensional 10-quantizer* $\Gamma = \{x_1, \ldots, x_{10}\}$ *and its Voronoi diagram.*

One clearly has, still for every $\omega \in \Omega$, that

$$|X(\omega) - \widehat{X}^{\Gamma}(\omega)|_H = \operatorname{dist}_H(X(\omega), \Gamma) = \min_{1 \leq i \leq N} |X(\omega) - x_i|_H.$$

The mean (quadratic) quantization error is then defined by

$$e(\Gamma, X, H) = \|X - \widehat{X}^{\Gamma}\|_2 = \sqrt{\mathbb{E}\left(\min_{1 \leq i \leq N} |X - x_i|_H^2\right)}. \tag{2.1}$$

The distribution of $\widehat{X}^{\Gamma}$ as a random vector is given by the N-tuple $(\mathbb{P}(X \in C_i(\Gamma)))_{1 \leq i \leq N}$ of the Voronoi cells. This distribution clearly depends on the choice of the Voronoi partition as emphasized by the following elementary situation: if $H = \mathbb{R}$, the distribution of X is given by $\mathbb{P}_X = \frac{1}{3}(\delta_0 + \delta_{1/2} + \delta_1)$, $N = 2$ and $\Gamma = \{0, 1\}$ since $1/2 \in \partial C_0(\Gamma) \cap \partial C_1(\Gamma)$. However, if $\mathbb{P}_X$ weights no hyperplane, the distribution of $\widehat{X}^{\Gamma}$ depends only on Γ.

As concerns terminology, *Vector Quantization* is concerned with the finite dimensional case – when $\dim H < +\infty$ – and is a rather old story, going back to the early 1950's when it was designed in the field of signal processing and then mainly developed in the community of Information Theory. The term *functional quantization*, probably introduced in (Pagès 2000; Luschgy and Pagès 2002), deals with the infinite dimensional case including the more general Banach-valued setting. The term "functional" comes from the fact that a typical infinite dimensional Hilbert space is the function space $H = L_T^2$. Then, any (bi-measurable) process $X : ([0, T] \times \Omega, Bor([0, T]) \otimes \mathcal{A}) \to (\mathbb{R}, Bor(\mathbb{R}))$ can be seen as a random vector taking values in the set of Borel functions on $[0, T]$. Furthermore, $((t, \omega) \mapsto X_t(\omega)) \in L^2(dt \otimes d\mathbb{P})$ if and only if $(\omega \mapsto X_{\cdot}(\omega)) \in L_H^2(\mathbb{P})$ since

$$\int_{[0,T] \times \Omega} X_t^2(\omega) \, dt \, \mathbb{P}(d\omega) = \int_{\Omega} \mathbb{P}(d\omega) \int_0^T X_t^2(\omega) \, dt = \mathbb{E}\, |X_{\cdot}|_{L_T^2}^2.$$

3 Optimal (quadratic) quantization

At this stage we are lead to wonder whether it is possible to design some optimally fitted grids to a given distribution $\mathbb{P}_X$ *i.e.* which induce the lowest possible mean quantization error among all grids of size at most N. This amounts to the following optimization problem

$$e_N(X, H) := \inf_{\Gamma \subset H, \operatorname{card}(\Gamma) \leq N} e(\Gamma, X, H). \tag{3.1}$$

It is convenient at this stage to make a correspondence between quantizers of size at most N and N-tuples of H^N: to any N-tuple $x := (x_1, \ldots, x_N)$ corresponds a quantizer $\Gamma := \Gamma(x) = \{x_i, i =$
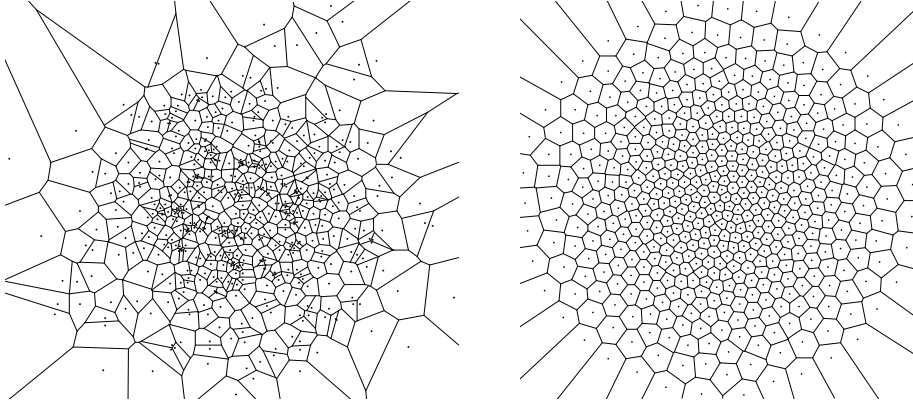
Figure 2: *Two N-quantizers (and their Voronoi diagram) related to bi-variate normal distribution* $\mathcal{N}(0; I_2)$ *(N = 500); which one is the best?*

$1, \ldots, N\}$ (of size at most N). One introduces the quadratic distortion, denoted D_N^X, defined on H^N as a (symmetric) function by

$$D_N^X \quad : \quad H^N \longrightarrow \mathbb{R}_+$$

$$(x_1, \ldots, x_N) \quad \longmapsto \mathbb{E}\left(\min_{1 \leq i \leq N} |X - x_i|_H^2\right).$$

Note that, combining (2.1) and the definition of the distortion, shows that

$$D_N^X(x_1, \ldots, x_N) = \mathbb{E}\left(\min_{1 \leq i \leq N} |X - x_i|_H^2\right) = \mathbb{E}\left(d(X, \Gamma(x))^2\right) = \|X - \widehat{X}^{\Gamma(x)}\|_2^2$$

so that,

$$e_N(X, H) = \inf_{(x_1, \ldots, x_N) \in H^N} \sqrt{D_N^X(x_1, \ldots, x_N)}.$$

The following proposition shows the existence of an optimal N-tuple $x^{(N,*)} \in H^N$ such that $e_N(X, H) = \sqrt{D_N^X(x^{(N,*)})}$. The corresponding optimal quantizer at level N is denoted $\Gamma^{(N,*)} := \Gamma(x^{(N,*)})$. In finite dimensions we refer to (Pollard 1982) and in infinite dimensiononal seetings to (Cuesta-Albertos and Matrán 1988) and (Pärna 1990) ; one may also see (Pagès 1993), (Graf and Luschgy 2000) and (Luschgy and Pagès 2002). For recent developments on existence and pathwise regularity of optimal quantizer see (Graf et al. 2007).

Proposition 3.1 (a) *The function D_N^X is lower semi-continuous for the product weak topology on H^N.*

(b) *The function D_N^X reaches a minimum at a N-tuple $x^{(N,*)}$ (so that $\Gamma^{(N,*)}$ is an* optimal quantizer *at level N).*

    – If $\operatorname{card}(\operatorname{supp}(\mathbb{P}_X)) \geq N$, *the quantizer has full size N (i.e. $\operatorname{card}(\Gamma^{(N,*)}) = N$) and $e_N(X, H) < e_{N-1}(X, H)$.*

    – If $\operatorname{card}(\operatorname{supp}(\mathbb{P}_X)) \leq N$, $e_N(X, H) = 0$.

    Furthermore $\lim_N e_N(X, H) = 0$.

(c) *Any optimal (Voronoi) quantization at level N, $\widehat{X}^{\Gamma^{(N,*)}}$ satisfies*

$$\widehat{X}^{\Gamma^{(N,*)}} = \mathbb{E}(X \,|\, \sigma(\widehat{X}^{\Gamma^{(N,*)}})) \tag{3.2}$$

where $\sigma(\widehat{X}^{\Gamma^{(N,*)}})$ denotes the σ-algebra generated by $\widehat{X}^{\Gamma^{(N,*)}}$.

(d) Any optimal (quadratic) quantization at level N is a best least square (i.e. $L^2(\mathbb{P})$) approximation of X among all H-valued random variables taking at most N values:

$$e_N(X,H) = \|X - \widehat{X}^{\Gamma^{(N,*)}}\|_2 = \min\{\|X - Y\|_2, \ Y : (\Omega, \mathcal{A}) \to H, \ \mathrm{card}(Y(\Omega)) \leq N\}.$$

Proof (sketch of): (a) The claim follows from the l.s.c. of $\xi \mapsto |\xi|_H$ for the weak topology and Fatou's Lemma.

(b) One proceeds by induction on N. If $N = 1$, the optimal 1-quantizer is $x^{(N,*)} = \{\mathbb{E}\,X\}$ and $e_2(X,H) = \|X - \mathbb{E}\,X\|_2$.

Assume now that an optimal quantizer $x^{(N,*)} = (x_1^{(N,*)}, \ldots, x_N^{(N,*)})$ does exist at level N.

– If $\mathrm{card}(\mathrm{supp}(\mathbb{P})) \leq N$, then the $N+1$-tuple $(x^{(N,*)}, x_N^{(N,*)})$ (among other possibilities) is also optimal at level $N+1$ and $e_{N+1}(X,H) = e_N(X,H) = 0$.

– Otherwise, $\mathrm{card}(\mathrm{supp}(\mathbb{P})) \geq N+1$, hence $x^{(N,*)}$ has pairwise distinct components and there exists $\xi_{N+1} \in \mathrm{supp}(\mathbb{P}_X) \setminus \{x_i^{(N,*)}, \ i = 1, \ldots, N\} \neq \emptyset$.

Then, with obvious notations,

$$D_{N+1}^X((x^{(N,*)}, \xi_{N+1})) < D_N^X(x^{(N,*)}).$$

Then, the set $F_{N+1} := \left\{ x \in H^{N+1} \mid D_{N+1}^X(x) \leq D_{N+1}^X((x^{(N,*)}, \xi_{N+1})) \right\}$ is non empty, weakly closed since D_{N+1}^X is l.s.c.. Furthermore, it is bounded in H^{N+1}. Otherwise there would exist a sequence $x_{(m)} \in H^{N+1}$ such that $|x_{(m),i_m}|_H = \max_i |x_{(m),i}|_H \to +\infty$ as $m \to \infty$. Then, by Fatou's Lemma, one checks that

$$\liminf_{m \to \infty} D_{N+1}^X(x_{(m)}) \geq D_N^X(x^{(N,*)}) > D_{N+1}^X((x^{(N,*)}, \xi_{N+1})).$$

Consequently F_{N+1} is weakly compact and the minimum of D_{N+1}^X on F_{N+1} is clearly its minimum over the whole space H^{N+1}. In particular

$$e_{N+1}(X,H) \leq D_{N+1}^X((x^{(N,*)}, \xi_{N+1})) < e_N(X,H).$$

If $\mathrm{card}(\mathrm{supp}(\mathbb{P})) = N+1$, set $x^{(N+1,*)} = \mathrm{supp}(\mathbb{P})$ (as sets) so that t $X = \widehat{X}^{\Gamma^{(N+1,*)}}$ which implies $e_{N+1}(X,H) = 0$.

To establish that $e_N(X,H)$ goes to 0, one considers an everywhere dense sequence $(z_k)_{k \geq 1}$ in the separable space H. Then, $d(\{z_1, \ldots, z_N\}, X(\omega))$ goes to 0 as $N \to \infty$ for every $\omega \in \Omega$. Furthermore, $d(\{z_1, \ldots, z_N\}, X(\omega))^2 \leq |X(\omega) - z_1|_H^2 \in L^1(\mathbb{P})$. One concludes by the Lebesgue dominated convergence Theorem that $D_N^X(z_1, \ldots, z_N)$ goes to 0 as $N \to \infty$.

(c) and (d) Temporarily set $\widehat{X}^* := \widehat{X}^{\Gamma^{(N,*)}}$ for convenience. Let $Y : (\Omega, \mathcal{A}) \to H$ be a random vector taking at most N values. Set $\Gamma := Y(\Omega)$. Since $\widehat{X}^\Gamma$ is a Voronoi quantization of X induced by Γ,

$$|X - \widehat{X}^\Gamma|_H = d(X, \Gamma) \leq |X - Y|_H$$

so that

$$\|X - \widehat{X}^\Gamma\|_2 \leq \|X - Y\|_2.$$

On the other hand, the optimality of $\Gamma^{(N,*)}$ implies

$$\|X - \widehat{X}^*\|_2 \leq \|X - \widehat{X}^\Gamma\|_2.$$

Consequently

$$\|X - \widehat{X}^*\|_2 \leq \min\left\{\|X - Y\|_2, \, Y : (\Omega, \mathcal{A}) \to H, \, \mathrm{card}(Y(\Omega)) \leq N\right\}.$$

The inequality holds as an equality since $\widehat{X}^*$ takes at most N values. Furthermore, considering random vectors of the form $Y = g(\widehat{X})$ (which take at most as many values as the size of $\Gamma^{(N,*)}$) shows, going back to the very definition of conditional expectation, that $\widehat{X}^* = \mathbb{E}(X \mid \widehat{X}^*)$ $\mathbb{P}$-a.s. $\diamond$

Item (c) introduces a very important notion in (quadratic) quantization.

Definition 3.1 *A quantizer $\Gamma \subset H$ is* stationary *(or* self-consistent*) if (there is a nearest neighbour projection such that $\widehat{X}^\Gamma = \mathrm{Proj}_\Gamma(X)$ satisfying)*

$$\widehat{X}^\Gamma = \mathbb{E}\left(X \mid \widehat{X}^\Gamma\right). \tag{3.3}$$

Note in particular that any stationary quantization satisfies $\mathbb{E}X = \mathbb{E}\widehat{X}^\Gamma$.

As shown by Proposition 3.1(c) any quadratic optimal quantizer at level N is stationary. Usually, at least when $d \geq 2$, there are other stationary quantizers: indeed, the distortion function D_N^X is $|\,.\,|_H$-differentiable at N-quantizers $x \in H^N$ with pairwise distinct components and

$$\nabla D_N^X(x) = 2\left(\int_{C_i(x)} (x_i - \xi)\mathbb{P}_X(d\xi)\right)_{1 \leq i \leq N} = 2\left(\mathbb{E}(\widehat{X}^{\Gamma(x)} - X)\mathbf{1}_{\{\widehat{X}^{\Gamma(x)} = x_i\}}\right)_{1 \leq i \leq N}.$$

hence, any critical points of D_N^X is a stationary quantizer.

Remarks and comments. • In fact (see Theorem 4.2, p. 38, (Graf and Luschgy 2000)), the Voronoi partitions of $\Gamma^{(N,*)}$ always have a $\mathbb{P}_X$-negligible boundary so that (3.3) holds for *any* Voronoi diagram induced by Γ.

• The problem of the uniqueness of optimal quantizer (viewed as a set) is not mentioned in the above proposition. In higher dimension, this essentially never occurs. In one dimension, uniqueness of the optimal N-quantizer was first established in (Fleischer 1964) with strictly log-concave density function. This was successively extended in (Kieffer 1983) and (Trushkin 1982) and lead to the following criterion (for more general "loss" functions than the square function):

If the distribution of X is *absolutely continuous with a* log-*concave density function*, then, for every $N \geq 1$, there exists only one stationary quantizer of size N, which turns out to be the optimal quantizer at level N.

More recently, a more geometric approach to uniqueness based on the Mountain Pass Lemma first developed in (Lamberton and Pagès 1996) and then generalized in (Cohort 1998)) provided a slight extension of the above criterion (in terms of loss functions).

This log-concavity assumption is satisfied by many families of probability distributions like the uniform distribution on compact intervals, the normal distributions, the gamma distributions. There are examples of distributions with a non log-concave density function having a unique optimal quantizer for every $N \geq 1$ (see *e.g.* the Pareto distribution in (Fort and Pagès 2004)). On the other hand simple examples of scalar distributions having multiple optimal quantizers at a given level can be found in (Graf and Luschgy 2000).

• A stationary quantizer can be sub-optimal. This will be emphasized in Section 6 for the Brownian motion (but it is also true for finite dimensional Gaussian random vectors) where some families

of sub-optimal quantizers – the *product quantizers* designed from the Karhunen-Loève basis – are stationary quantizers.

• For the uniform distribution over an interval $[a, b]$, there is a closed form for the optimal quantizer at level N given by $\Gamma^{(N,*)} = \{a + (2k-1)\frac{b-a}{N}, \ k = 1, \dots, N\}$. This N-quantizer is optimal not only in the quadratic case but also for any L^r-quantization (see a definition further on). In general there is no such closed form, either in 1 or higher dimension. However, in (Fort and Pagès 2004) some semi-closed forms are obtained for several families of (scalar) distributions including the exponential and the Pareto distributions: all the optimal quantizers can be expressed using a single underlying sequence $(a_k)_{k \geq 1}$ defined by an induction $a_{k+1} = F(a_k)$.

• In one dimension, as soon as the optimal quantizer at level N is unique (as a set or as an N-tuple with increasing components), it is generally possible to compute it as the solution of the stationarity equation (3.2) either by a zero search (Newton-Raphson gradient descent) or a fixed point (like the specific Lloyd I procedure, see (Kieffer 1982)) procedure.

• In higher dimension, deterministic optimization methods become intractable and one uses stochastic procedures to compute optimal quantizers. We decided to postpone the short overview on these aspects to Section 6, devoted to the optimal functional quantization of the Brownian motion where the case of Gaussian vectors with (diagonal) covariance matrix is considered. All stochastic optimization approaches rely on some repeated nearest neighbour searches: our procedures include some fast (exact) algorithms for that purpose (like K-d-tree, see (Friedman et al. 1977)). So far, the most efficient methods are also based on the so-called *splitting method* which increases progressively the quantization level N (this must be understood when looking for a systematic quantization of a distribution). This method is directly inspired by the induction developed in the proof of claim (b) of Proposition 3.1 since one designs the starting value of the optimization procedure at size $N+1$ by "merging" the optimized N-quantizer obtained at level N with one further point of $\mathbb{R}^d$, usually randomly sampled with respect to an appropriate distribution (see (Pagès and Printems 2003) for a discussion). For normal distributions $\mathcal{N}(0; I_d)$, alternative starting values living on a sphere with an appropriate radius seem to yield the same accuracy for a given size N without splitting (see (Pagès and Sagna 2007)).

• As concerns functional quantization, *e.g.* $H = L_T^2$, there is a close connection between the regularity of optimal (or even stationary) quantizers and that of $t \mapsto X_t$ form $[0, T]$ into $L^2(\mathbb{P})$. Furthermore, as concerns optimal quantizers of Gaussian processes, one shows (see (Luschgy and Pagès 2002)) that they belong to the reproducing space of their covariance operator, *e.g.* to the Cameron-Martin space $H^1 = \{\int_0^{\cdot} \dot{h}_s ds, \ \dot{h} \in L_T^2\}$ when $X = W$. Other properties of optimal quantization of Gaussian processes are established in (Luschgy and Pagès 2002).

Extensions to the $L^r(\mathbb{P})$-quantization of random variables. In this paper, we focus on the purely quadratic framework (L_T^2 and $L^2(\mathbb{P})$-norms), essentially because it is a natural (and somewhat easier) framework for the computation of optimized grids for the Brownian motion and for some first applications (like the pricing of path-dependent options, see section 8). But a more general and natural framework is to consider the functional quantization of random vectors taking values in a separable Banach space $(E, |\,.\,|_E)$. Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, |\,\,|_E)$, such that $\mathbb{E}\,|X|_E^r < +\infty$ for some $r \geq 1$ (the case $0 < r < 1$ can also be taken in consideration).

The N-level $(L^r(\mathbb{P}), |\,.\,|_E)$-quantization problem for $X \in L_E^r(\mathbb{P})$ reads

$$e_{N,r}(X, E) \quad := \quad \inf \left\{ \|X - \widehat{X}^{\Gamma}\|_r, \ \Gamma \subset E, \ \mathrm{card}(\Gamma) \leq N \right\}.$$

The main examples for $(E, |\,.\,|_E)$ are the non-Euclidean norms on $\mathbb{R}^d$, the functional spaces $L_T^p(\mu) := L^p([0, T], \mu(dt)), 1 \leq p \leq \infty$, equipped with its usual norm, $(E, |\,.\,|_E) = (\mathcal{C}([0, T]), \|\,.\,\|_{\sup})$, etc. As concerns, the existence of an optimal quantizer, it holds true for reflexive Banach spaces

(see Pärna (90)) and $E = L_T^1$, but otherwise it may fail even when $N = 1$ (see (Graf et al. 2007)). In finite dimension, the Euclidean feature is not crucial (see (Graf and Luschgy 2000)). In the functional setting, many results originally obtained in a Hilbert setting have been extended to the Banach setting either for existence or regularity results (see (Graf et al. 2007)) or for rates see (Dereich 2005a), (Dereich et al. 2006), (Luschgy and Pagès 2004), (Luschgy and Pagès 2007).

4 Cubature formulae: conditional expectation and numerical integration

Let $F : H \longrightarrow \mathbb{R}$ be a continuous functional (with respect to the norm $|\,.\,|_H$) and let $\Gamma \subset H$ be an N-quantizer. It is natural to approximate $\mathbb{E}(F(X))$ by $\mathbb{E}(F(\widehat{X}^\Gamma))$. This quantity $\mathbb{E}(F(\widehat{X}^\Gamma))$ is simply the finite weighted sum

$$\mathbb{E}\left(F(\widehat{X}^\Gamma)\right) = \sum_{i=1}^{N} F(x_i)\mathbb{P}(\widehat{X}^\Gamma = x_i). \tag{4.1}$$

Numerical computation of $\mathbb{E}\left(F(\widehat{X}^\Gamma)\right)$ is possible as soon as $F(\xi)$ can be computed at any $\xi \in H$ and the distribution $(\mathbb{P}(\widehat{X} = x_i))_{1 \le i \le N}$ of $\widehat{X}^\Gamma$ is known. The induced quantization error $\|X - \widehat{X}^\Gamma\|_2$ is used to control the error (see below). These quantities related to the quantizer Γ are also called *companion parameters*.

Likewise, one can consider *a priori* the $\sigma(\widehat{X}^\Gamma)$-measurable random variable $F(\widehat{X}^\Gamma)$ as a good approximation of the conditional expectation $\mathbb{E}(F(X)\,|\,\widehat{X}^\Gamma)$.

4.1 Lipschitz functionals

Assume that the functional F is Lipschitz continuous on H. Then

$$\left|\mathbb{E}(F(X)\,|\,\widehat{X}^\Gamma) - F(\widehat{X}^\Gamma)\right| \le [F]_{\mathrm{Lip}}\mathbb{E}(|X - \widehat{X}^\Gamma|\,|\,\widehat{X}^\Gamma)$$

so that, for every real exponent $r \ge 1$,

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

(where we applied conditional Jensen inequality to the convex function $u \mapsto u^r$). In particular, using that $\mathbb{E}\,F(X) = \mathbb{E}(\mathbb{E}(F(X)\,|\,\widehat{X}^\Gamma))$, one derives (with $r = 1$) that

$$\begin{aligned}\left|\mathbb{E}\,F(X) - \mathbb{E}\,F(\widehat{X}^\Gamma)\right| &\le \|\mathbb{E}(F(X)\,|\,\widehat{X}^\Gamma) - F(\widehat{X}^\Gamma)\|_1 \\ &\le [F]_{\mathrm{Lip}}\|X - \widehat{X}^\Gamma\|_1.\end{aligned}$$

Finally, using the monotony of the $L^r(\mathbb{P})$-norms as a function of r yields

$$\left|\mathbb{E}\,F(X) - \mathbb{E}\,F(\widehat{X}^\Gamma)\right| \le [F]_{\mathrm{Lip}}\|X - \widehat{X}^\Gamma\|_1 \le [F]_{\mathrm{Lip}}\|X - \widehat{X}^\Gamma\|_2. \tag{4.2}$$

In fact, considering the Lipschitz functional $F(\xi) := d(\xi, \Gamma)$, shows that

$$\|X - \widehat{X}^\Gamma\|_1 = \sup_{[F]_{\mathrm{Lip}} \le 1} \left|\mathbb{E}\,F(X) - \mathbb{E}\,F(\widehat{X}^\Gamma)\right|. \tag{4.3}$$

The Lipschitz functionals making up a characterizing family for the weak convergence of probability measures on H, one derives that, for any sequence of N-quantizers Γ^N satisfying $\|X - \widehat{X}^{\Gamma^N}\|_1 \to 0$ as $N \to \infty$,

$$\sum_{1 \leq i \leq N} \mathbb{P}(\widehat{X}^{\Gamma^N} = x_i^N) \, \delta_{x_i^N} \overset{(H)}{\Longrightarrow} \mathbb{P}_X$$

where $\overset{(H)}{\Longrightarrow}$ denotes the weak convergence of probability measures on $(H, |.|_H)$.

4.2 Convex functionals

If $F : H \to \mathbb{R}$ is a convex functional (assumed to be l.s.c. if H is infinite dimesional) and $\widehat{X}$ a stationary quantization of X, a straightforward application of Jensen inequality yields

$$\mathbb{E}\left(F(X) \,|\, \widehat{X}\right) \geq F(\widehat{X})$$

so that in particular $\mathbb{E}\left(F(\widehat{X})\right) \leq \mathbb{E}\left(F(X)\right)$.

4.3 Differentiable functionals with Lipschitz differentials

Assume now that F is differentiable on H, with a Lipschitz continuous differential DF, and that the quantizer Γ is *stationary* (see Equation (3.3)). A Taylor expansion yields

$$\left|F(X) - F(\widehat{X}^\Gamma) - DF(\widehat{X}^\Gamma).(X - \widehat{X}^\Gamma)\right| \leq [DF]_{\mathrm{Lip}}|X - \widehat{X}^\Gamma|^2.$$

Taking conditional expectation given $\widehat{X}^\Gamma$ yields

$$\left|\mathbb{E}(F(X) \,|\, \widehat{X}^\Gamma) - F(\widehat{X}^\Gamma) - \mathbb{E}\left(DF(\widehat{X}^\Gamma).(X - \widehat{X}^\Gamma) \,|\, \widehat{X}^\Gamma\right)\right| \leq [DF]_{\mathrm{Lip}}\mathbb{E}(|X - \widehat{X}^\Gamma|^2 \,|\, \widehat{X}^\Gamma).$$

Now, using that the random variable $DF(\widehat{X}^\Gamma)$ is $\sigma(\widehat{X}^\Gamma)$-measurable, one has

$$\mathbb{E}\left(DF(\widehat{X}^\Gamma).(X - \widehat{X}^\Gamma)\right) = \mathbb{E}\left(DF(\widehat{X}^\Gamma).\mathbb{E}(X - \widehat{X}^\Gamma \,|\, \widehat{X}^\Gamma)\right) = 0$$

so that

$$\left|\mathbb{E}(F(X) \,|\, \widehat{X}^\Gamma) - F(\widehat{X}^\Gamma)\right| \leq [DF]_{\mathrm{Lip}}\mathbb{E}\left(|X - \widehat{X}^\Gamma|^2 \,|\, \widehat{X}^\Gamma\right).$$

Then, for every real exponent $r \geq 1$,

$$\left\|\mathbb{E}(F(X) \,|\, \widehat{X}^\Gamma) - F(\widehat{X}^\Gamma)\right\|_r \leq [DF]_{\mathrm{Lip}}\|X - \widehat{X}^\Gamma\|_{2r}^2. \tag{4.4}$$

In particular, when $r = 1$, one derives like in the former setting

$$\left|\mathbb{E}F(X) - \mathbb{E}F(\widehat{X}^\Gamma)\right| \leq [DF]_{\mathrm{Lip}}\|X - \widehat{X}^\Gamma\|_2^2. \tag{4.5}$$

In fact, the above inequality holds provided F is $\mathcal{C}^1$ with Lipschitz differential on every Voronoi cell $C_i(\Gamma)$. A similar characterization to (4.3) based on these functionals could be established.

Some variant of these cubature formulae can be found in (Pagès and Printems 2003) or (Graf et al. 2006) for functions or functionals F having only some local Lipschitz regularity.

9

4.4 Quantized approximation of $\mathbb{E}(F(X)\,|\,Y)$

Let X and Y be two H-valued random vector defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $F : H \to \mathbb{R}$ be a Borel functional. The natural idea is to approximate $\mathbb{E}(F(X)\,|\,Y)$ by the quantized conditional expectation $\mathbb{E}(F(\widehat{X})\,|\,\widehat{Y})$ where $\widehat{X}$ and $\widehat{Y}$ are quantizations of X and Y respectively.

Let $\varphi_F : H \to \mathbb{R}$ be a (Borel) version of the conditional expectation *i.e.* satisfying

$$\mathbb{E}(F(X)\,|\,Y) = \varphi_F(Y).$$

Usually, no closed form is available for the function φ_F but some regularity property can be established, especially in a (Feller) Markovian framework. Thus assume that both F and φ_F are Lipschitz continuous with Lipschitz coefficients $[F]_{\mathrm{Lip}}$ and $[\varphi_F]_{\mathrm{Lip}}$. Then

$$\mathbb{E}(F(X)\,|\,Y) - \mathbb{E}(F(\widehat{X})\,|\,\widehat{Y}) = \mathbb{E}(F(X)\,|\,Y) - \mathbb{E}(F(X)\,|\,\widehat{Y}) + \mathbb{E}(F(X) - F(\widehat{X})\,|\,\widehat{Y}).$$

Hence, using that $\widehat{Y}$ is $\sigma(Y)$-measurable and that conditional expectation is an L^2-contraction,

$$
\begin{aligned}
\|\mathbb{E}(F(X)\,|\,Y) - \mathbb{E}(F(X)\,|\,\widehat{Y})\|_2 &= \|\mathbb{E}(F(X)|Y) - \mathbb{E}(\mathbb{E}(F(\widehat{X})|Y)|\widehat{Y})\|_2 \\
&\leq \|\varphi_F(Y) - \mathbb{E}(F(X)|\widehat{Y})\|_2 \\
&= \|\varphi_F(Y) - \mathbb{E}(\varphi_F(Y)|\widehat{Y})\|_2 \\
&\leq \|\varphi_F(Y) - \varphi_F(\widehat{Y})\|_2.
\end{aligned}
$$

The last inequality follows form the definition of conditional expectation given $\widehat{Y}$ as the best quadratic approximation among $\sigma(\widehat{Y})$-measurable random variables. On the other hand, still using that $\mathbb{E}(\,.\,|\sigma(\widehat{Y}))$ is an L^2-contraction and this time that F is Lipschitz continuous yields

$$\|\mathbb{E}(F(X) - F(\widehat{X})\,|\,\widehat{Y})\|_2 \leq \|F(X) - F(\widehat{X})\|_2 \leq [F]_{\mathrm{Lip}}\|X - \widehat{X}\|_2.$$

Finally,

$$\|\mathbb{E}(F(X)\,|\,Y) - \mathbb{E}(F(\widehat{X})\,|\,\widehat{Y})\|_2 \leq [F]_{\mathrm{Lip}}\|X - \widehat{X}\|_2 + [\varphi_F]_{\mathrm{Lip}}\|Y - \widehat{Y}\|_2.$$

In the non-quadratic case the above inequality remains valid provided $[\varphi_F]_{\mathrm{Lip}}$ is replaced by $2[\varphi_F]_{\mathrm{Lip}}$.

5 Vector quantization

5.1 Vector quantization rate ($H = \mathbb{R}^d$)

The fact that $e_N(X, \mathbb{R}^d)$ is a non-increasing sequence that goes to 0 as N goes to ∞ is a rather simple result established in Proposition 3.1. Its (sharp) rate of convergence to 0 is a much more challenging problem. An answer is provided by the so-called Zador Theorem stated below.

This theorem was first stated and established for distributions with compact supports by Zador (see (Zador 1963; Zador 1982)). Then a first extension to general probability distributions on $\mathbb{R}^d$ is developed in (Bucklew and Wise 1982). The first mathematically rigorous proof can be found in (Graf and Luschgy 2000), and relies on a random quantization argument (called upon in a step of the proof sometimes called Pierce Lemma). We also provide a non asymptotic error bound which which can be seen as simple reformulation of this Pierce Lemma. It turns out to be very useful for applications.

Theorem 5.1 (a) SHARP RATE (see (Graf and Luschgy 2000)). Let $r > 0$ and $X \in L^{r+\eta}(\mathbb{P})$ for some $\eta > 0$. Let $\mathbb{P}_X(d\xi) = \varphi(\xi)\, d\xi \overset{\perp}{+} \nu(d\xi)$ be the canonical decomposition of the distribution of X (ν and the Lebesgue measure are singular). Then (if $\varphi \not\equiv 0$),

$$e_{N,r}(X, \mathbb{R}^d) \sim \widetilde{J}_{r,d} \times \left(\int_{\mathbb{R}^d} \varphi^{\frac{d}{d+r}}(u)\, du \right)^{\frac{1}{d}+\frac{1}{r}} \times N^{-\frac{1}{d}} \quad as \quad N \to +\infty. \tag{5.1}$$

where $\widetilde{J}_{r,d} \in (0, \infty)$.

(b) NON ASYMPTOTIC UPPER BOUND (see (Luschgy and Pagès 2007)). Let $d \geq 1$. There exists $C_{d,r,\eta} \in (0, \infty)$ such that, for every $\mathbb{R}^d$-valued random vector X,

$$\forall\, N \geq 1, \qquad e_{N,r}(X, \mathbb{R}^d) \leq C_{d,r,\eta} \|X\|_{r+\eta} N^{-\frac{1}{d}}.$$

Remarks. • The real constant $\widetilde{J}_{r,d}$ clearly corresponds to the case of the uniform distribution over the unit hypercube $[0, 1]^d$ for which the slightly more precise statement holds

$$\lim_N N^{\frac{1}{d}} e_{N,r}(X, \mathbb{R}^d) = \inf_N N^{\frac{1}{d}} e_{N,r}(X, \mathbb{R}^d) = \widetilde{J}_{r,d}.$$

The proof is based on a self-similarity argument. The value of $\widetilde{J}_{r,d}$ depends on the reference norm on $\mathbb{R}^d$. When $d = 1$, elementary computations show that $\widetilde{J}_{r,1} = (r+1)^{-\frac{1}{r}}/2$. When $d = 2$, with the canonical Euclidean norm, one shows (see (Newman 1982) for a proof, see also (Graf and Luschgy 2000)) that $\widetilde{J}_{2,d} = \sqrt{\frac{5}{18\sqrt{3}}}$. Its exact value is unknown for $d \geq 3$ but, still for the canonical Euclidean norm, one has (see (Graf and Luschgy 2000)) using some random quantization arguments,

$$\widetilde{J}_{2,d} \sim \sqrt{\frac{d}{2\pi e}} \approx \sqrt{\frac{d}{17,08}} \quad as \quad d \to +\infty.$$

• When $\varphi \equiv 0$ the distribution of X is purely singular. The rate (5.1) still holds in the sense that $\lim_N N^{\frac{1}{d}} e_{r,N}(X, \mathbb{R}^d) = 0$. Consequently, this is not the right asymptotics. The quantization problem for singular measures (like uniform distribution on fractal compact sets) has been extensively investigated by several authors, leading to the definition of a quantization dimension in connection with the rate of convergence of the quantization error on these sets. For more details we refer to (Graf and Luschgy 2000; Graf and Luschgy 2005) and the references therein.

• A more naive way to quantize the uniform distribution on the unit hypercube is to proceed by *product quantization* i.e. by quantizing the marginals of the uniform distribution. If $N = m^d$, $m \geq 1$, one easily proves that the best quadratic product quantizer (for the canonical Euclidean norm on $\mathbb{R}^d$) is the "midpoint square grid"

$$\Gamma^{sq,N} = \left(\frac{2i_1 - 1}{2m}, \ldots, \frac{2i_d - 1}{2m} \right)_{1 \leq i_1, \ldots, i_d \leq m}$$

which induces a quadratic quantization error equal to

$$\sqrt{\frac{d}{12}} \times N^{-\frac{1}{d}}.$$

Consequently, product quantizers are still *rate optimal* in every dimension d. Moreover, note that the ratio of these two rates remains bounded as $d \uparrow \infty$.

• For a brief discussion and comparison with Quasi-Monte Carlo methods we refer to (Pagès 2007) and the references therein. Let us simply recall that sequences (or sets) with low discrepancy are uniformly distributed sequence over the unit d-dimensional hypercube $[0,1]^d$. When used instead of (pseudo-)random numbers to integrate a function f with *bounded variations* the rate of convergence is theoretically "almost dimension free": it is the product of the variation of f by the discrepancy which behaves like $O\left(\frac{\log(N)^d}{N}\right)$ (for sequences). However such functions become less and less "standard" in higher dimension. When implemented with Lipschitz continuous functions, the QMC method does face the *curse of dimensionality* with theoretical performances which seem to be worse than optimal quantizers of the uniform distribution over $[0,1]^d$ like $O\left(\frac{\log(N)}{N^{\frac{1}{d}}}\right)$ (still for sequences) owing to Proinov's Theorem in (Proinov 1982).

• The non asymptotic Zador Theorem stated and established in (Luschgy and Pagès 2007) is essentially a variant of the so-called Pierce Lemma (see (Graf and Luschgy 2000)). Many developments and heuristics about the rate of convergence of some quantization based algorithms for American option pricing, stochastic control or non linear filtering (see (Pagès et al. 2003)) can be significantly simplified or established rigorously by calling upon this result. This is emphasized by the example below devoted to swing options.

5.2   Examples of application of optimal vector quantization

5.2.1   Numerical integration (II): Richardson-Romberg extrapolation *vs* curse of dimensionality

Combining the above cubature formula (4.1) and the rate of convergence of the (optimal) quantization error, the theoretical critical dimension to use quantization based cubature formulae seems to be $d = 4$ when compared to Monte Carlo simulation (at least for continuously differentiable functions). Several numerical tests have been carried out and reported in (Pagès et al. 2003) and (Pagès and Printems 2003) to evaluate more precisely the effect of the so-called *curse of dimensionality*. The benchmark was made of several European payoffs on a geometric index made of d independent assets in a Black-Scholes model: vanilla put and put spread options and their smoothed versions. No control variate was used. The absence of correlation is not a realistic assumption in finance but is clearly more challenging as a benchmark for numerical integration. Once the dimension d and the quantizer size N have been chosen, we compared the resulting integration error to a symmetric confidence interval with total length equal to two standard deviations of a Monte Carlo (MC) estimator based on N simulated data ($\frac{\sigma_{payoff}}{\sqrt{N}}$). Furthermore, σ_{payoff} has been computed by a Monte Carlo simulation on 10^4 simulated data of the payoff .

The results turned out to be more favourable to quantization than predicted by theoretical bounds, mainly because we carried out our tests with rather small values of N whereas curse of dimensionality is an asymptotic bound. Until the dimension 4, the larger N is, the more quantization outperforms MC simulation. When the dimension $d \geq 5$, quantization always outperforms MC (in the above sense) until a critical size $N_c(d)$ which decreases as d increases.

▷ RICHARDSON-ROMBERG (R-R) EXTRAPOLATION. In this section, we provide a method to push ahead theses critical sizes, at least for smooth enough functionals. Let $F : \mathbb{R}^d \to \mathbb{R}$, be a twice differentiable functional with Lipschitz Hessian D^2F. Let $(\widehat{X}^{(N)})_{N \geq 1}$ be a sequence of optimal quadratic quantizations. Then

$$\mathbb{E}(F(X)) = \mathbb{E}(F(\widehat{X}^{(N)})) + \frac{1}{2}\mathbb{E}\left(D^2F(\widehat{X}^{(N)}).(X - \widehat{X}^{(N)})^{\otimes 2}\right) + O\left(\mathbb{E}|X - \widehat{X}|^3\right) \qquad (5.2)$$

Under some assumptions which are satisfied by most usual distributions (including the normal one), it is proved in (Graf et al. 2006) as a special case of a general theorem about the asymptotic

behaviour of in L^s of sequences of optimal L^r-quantizers for $s \in (r, r + d)$, that

$$\mathbb{E}|X - \widehat{X}|^3 = O(N^{-\frac{3}{d}}) \qquad \text{if } d \geq 2,$$

or $\mathbb{E}|X - \widehat{X}|^3 = O(N^{-\frac{3-\varepsilon}{d}})$, $\varepsilon > 0$, if $d = 2$. If furthermore, we make the conjecture that

$$\mathbb{E}\left(D^2 F(\widehat{X}^{(N)}).(X - \widehat{X}^{(N)})^{\otimes 2}\right) = c_{F,X} N^{-\frac{2}{d}} + O(N^{-\frac{3}{d}}) \tag{5.3}$$

it becomes possible to implement a R-R extrapolation to compute $\mathbb{E}(F(X))$. Namely, one considers two sizes N_1 and N_2 (in practice one often sets $N_1 = N/2$ and $N_2 = N$). Then combining (5.2) with N_1 and N_2,

$$\mathbb{E}(F(X)) = \frac{N_2^{\frac{2}{d}} \mathbb{E}(F(\widehat{X}^{(N_2)})) - N_1^{\frac{2}{d}} \mathbb{E}(F(\widehat{X}^{(N_1)}))}{N_2^{\frac{2}{d}} - N_1^{\frac{2}{d}}} + O\left(\frac{1}{(N_1 \wedge N_2)^{\frac{1}{d}}(N_2^{\frac{2}{d}} - N_1^{\frac{2}{d}})}\right).$$

In Section 8.1, a similar procedure is tested in an infinite dimensional setting: $\mathbb{R}^d$ is replaced by the Hilbert space $H = L^2([0, T], dt)$ viewed as a state of paths for a stochastic process X (namely the Brownian motion).

▷ NUMERICAL ILLUSTRATION: In order to evaluate the effect of the R-R technique described above, numerical computations have been carried out in the case of the regularized versions of some Put Spread options on geometric indices in dimension $d = 4, 6, 8, 10$. By "regularized", we mean that the payoff at maturity $T = 1$ has been replaced by its price function at time $T' < T$. Numerical integration was performed using the Gaussian optimal grids of size $N = 2^k$, $k = 2, \ldots, 12$ (available at the web site www.quantize.maths-fi.com).

We consider again one of the test functions implemented in (Pagès and Printems 2003) (pp. 152 sqq). These test functions were borrowed from classical option pricing in mathematical finance: one considers d independent traded assets $S^1, \ldots, S^d$ following a d-dimensional Black & Scholes dynamics (under its risk neutral probability)

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

where $Z^{i,t} = W_t^i / \sqrt{t}$ and $W = (W^1, \ldots, W^d)$ is a d-dimensional standard Brownian motion. We also assume that $S_0^i = s_0 > 0$, $i = 1, \ldots, d$ and that the d assets share the same volatility $\sigma^i = \sigma > 0$. One considers, the geometric index $I_t = \left(S_t^1 \ldots S_t^d\right)^{\frac{1}{d}}$. One shows that $e^{-\frac{\sigma^2}{2}(\frac{1}{d}-1)}$ has itself a risk neutral Black-Scholes dynamics. We want to test the *regularized Put spread option on this geometric index* with strikes $K_1 < K_2$ (at time $T/2$). Let $\psi(s_0, K_1, K_2, r, \sigma, T)$ the premium at time 0 of a Put spread on any of the assets S^i.

$$
\begin{aligned}
\psi(x, K_1, K_2, r, \sigma, T) &= \pi(x, K_2, r, \sigma, T) - \pi(x, K_1, r, \sigma, T) \\
\pi(x, K, r, \sigma, T) &= Ke^{-rT}\operatorname{erf}(-d_2) - x\operatorname{erf}(-d_1), \\
d_1 = \frac{\log(x/K) + (r + \frac{\sigma^2}{2d})T}{\sigma\sqrt{T/d}}, &\qquad d_2 = d_1 - \sigma\sqrt{T/d}.
\end{aligned}
$$

Using the martingale property of the discounted value of the premium of a European option yields that the premium $e^{-rT}\mathbb{E}((K_1 - I_T)_+ - (K_2 - I_T)_+)$ of the Put spread option on I satisfies on the one hand

$$e^{-rT}\mathbb{E}((K_1 - I_T)_+ - (K_2 - I_T)_+) = \psi(s_0 e^{\frac{\sigma^2}{2}(\frac{1}{d}-1)T}, K_1, K_2, r, \sigma/\sqrt{d}, T)$$

and, one the other hand

$$e^{-rT}\mathbb{E}((K_1 - I_T)_+ - (K_2 - I_T)_+) = \mathbb{E}\,g(Z)$$

where

$$g(Z) = e^{-rT/2}\psi(s_0 e^{\frac{\sigma^2}{2}(\frac{1}{d}-1)\frac{T}{2}}I_{\frac{T}{2}}, K_1, K_2, r, \sigma, T/2)$$

and $Z = (Z^{1,\frac{T}{2}}, \ldots, Z^{d,\frac{T}{2}}) \stackrel{d}{=} \mathcal{N}(0; I_d)$. The numerical specifications of the function g are as follows:

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

The results are displayed below (see Fig. 3) in a log-log-scale for the dimensions $d = 4$, 6, 8, 10.

First, we recover the theoretical rates (namely $-2/d$) of convergence for the error bounds. Indeed, some slopes $\beta(d)$ can be derived (using a regression) for the quantization errors and we found $\beta(4) = -0.48$, $\beta(6) = -0.33$, $\beta(8) = -0.25$ and $\beta(10) = -0.23$ for $d = 10$ (see Fig. 3). Theses rates plead for the implementation of R-R extrapolation. Also note that, as already reported in (Pagès and Printems 2003), when $d \geq 5$, quantization still outperforms MC simulations (in the above sense) up to a critical number $N_c(d)$ of points ($N_c(6) \sim 5000$, $N_c(7) \sim 1000$, $N_c(8) \sim 500$, etc).

As concerns the R-R extrapolation method itself, note first that it always gives better results than "crude" quantization. As regards now the comparison with Monte Carlo simulation, no critical number of points $N_{Romb}(d)$ comes out beyond which MC simulation outperforms R-R extrapolation. This means that $N_{Romb}(d)$ is greater than the range of use of quantization based cubature formulas in our benchmark, namely $5\,000$.

The R-R extrapolation techniques are commonly known to be unstable, and indeed, it has not been always possible to estimate satisfactorily its rate of convergence on our benchmark. But when a significant slope (in a log-log scale) can be estimated from the R-R errors (like for $d = 8$ and $d = 10$ in Fig. 3 (c), (d)), its absolute value is larger than $1/2$, and so, these *extrapolations always outperform the MC method even for large values of N*. As a by-product, our results plead in favour of the conjecture (5.3) and lead to think that R-R extrapolation is a powerful tool to accelerate numerical integration by optimal quantization, even in higher dimension.

5.3 An application to the pricing of swing options

Optimal quantization based algorithms have been already devised to solve several multi-dimensional non-linear problems, from multi-asset American style options ((Bally et al. 2001; Bally and Pagès 2003a; Bally and Pagès 2003b; Bally et al. 2003; Bally et al. 2005) to non-linear filtering and portfolio management (see *e.g.* (Pagès et al. 2003)). Here we present a new application developed by Gaz de France (French gas company) to price swing options contracts. For a detailed version of this section, we refer to the original works (Bardou et al. 2007a; Bardou et al. 2007b). The holder of such a contract daily purchases a quantity of gas, say q_{t_k} at time t_k, $k = 0, \ldots, n-1$, at a price K_{t_k} which can be deterministic or random (*e.g.* an oil based index). These quantities are subject to two kinds of constraints, some *local* daily constraints $q_{\min} \leq q_{t_k} \leq q_{\max}$ and some global constraint about the total amount of purchased gas $Q_{\min} \leq q_0 + \cdots + q_{t_{n-1}} \leq Q_{\max}$ (with $n\,q_{\min} \leq Q_{\min} \leq Q_{\max} \leq n\,q_{\max}$). The spot price S_{t_k} at time t_k of gas is usually not quoted on gas markets but is usually approximated by the day ahead price. The dynamics of the gas price itself is usually multi-factorial which makes it not markovian. However it depends on a multi-dimensional underlying Markov structure process, which can be Gaussian or not. For the sake of simplicity, we assume here that (S_{t_k}) has a Markov dynamics and that the exercise prices K_{t_k} are deterministic (and there is no interest rate). Then, given $q_{\min}, q_{\max}, Q_{\min}, Q_{\max}$ and if $\bar{q}_{t_k} := q_0 + q_{t_1} + \cdots + q_{t_{k-1}}$ denotes the purchased quantity
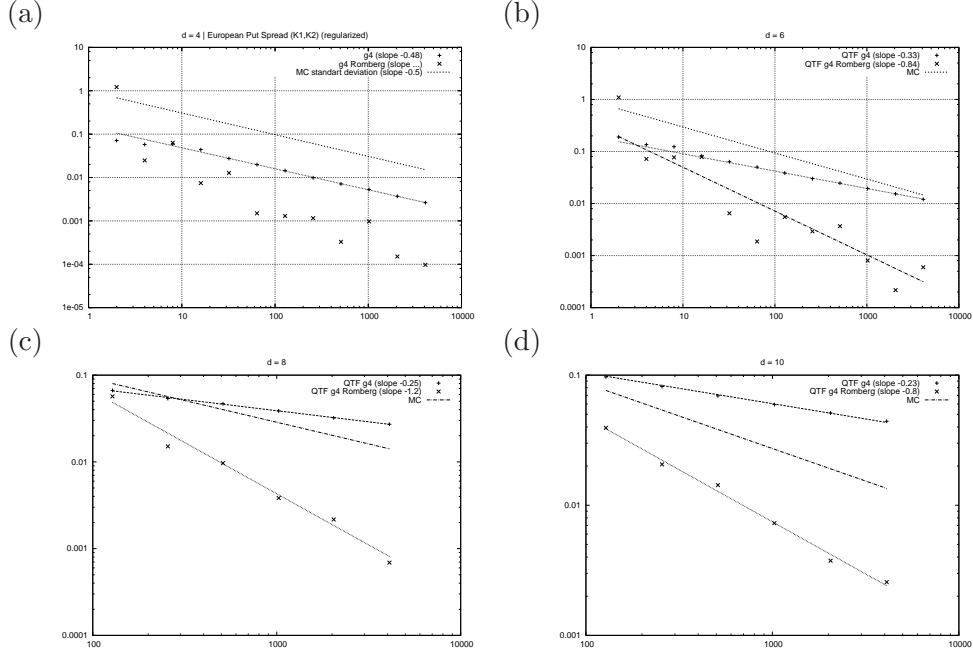
Figure 3: *Errors and standard deviations as functions of the number of points N in a log-log-scale. The quantization error is displayed by the cross + and the R-R extrapolation error by the cross ×. The dashed line without crosses denotes the standard deviation of the Monte Carlo estimator. (a) d = 4, (b) d = 6, (c) d = 8, (d) d = 10.*

prior to time t_k, the price of this contract at time t_k is given by

$$P(t_k, \bar{q}_{t_k}, S_{t_k}) := \inf \left\{ \mathbb{E} \left(\sum_{\ell=k}^{n-1} q_{t_k} (S_{t_\ell} - K_{t_\ell}) \,|\, S_{t_k} \right), q_{t_\ell} \underline{\in} \mathcal{F}_{t_{\ell-1}}^S, q_{\min} \leq q_{t_\ell} \leq q_{\max}, Q_{\min} \leq \bar{q}_{t_\ell} \leq Q_{\max} \right\}$$

where $\mathcal{F}_{t_\ell}^S = \sigma(S_0, S_{t_1}, \dots, S_{t_\ell})$ and $\underline{\in}$ stands for measurability with respect to a σ-field. This formula shows that this pricing problem is a stochastic control problem where the purchased quantity process appears as the control variable. This price satisfies the following dynamic programming principle

$$P(t_k, \bar{q}, S_{t_k}) = \sup_{q \in [q_{\min}, q_{\max}]} \left(q(S_{t_k} - K_{t_k}) + \mathbb{E} \left(P(t_{k+1}, \bar{q} + q, S_{t_{k+1}}) | S_{t_k} \right) \right).$$

It is shown in (Bardou et al. 2007b) that the optimal control does exist under mild integrability assumptions but is not always bang-bang in general (owing to "prediction errors"). However if $\frac{Q_{\max} - n \, q_{\min}}{q_{\max} - q_{\min}}$ and $\frac{Q_{\min} - n \, q_{\min}}{q_{\max} - q_{\min}}$ are integers, then the optimal control $q_{t_k}^*$ is always $\{q_{\min}, q_{\max}\}$-valued. Then one defines the *quantized dynamic programming formula* by setting

$$\widehat{P}(t_k, Q, \widehat{S}_{t_k}) := \max_{q = q_{\min}, q_{\max}} \left(q(S_{t_k} - K_{t_k}) + \mathbb{E} \left(\widehat{P}(t_{k+1}, Q + q, \widehat{S}_{t_{k+1}}) | \widehat{S}_{t_k} \right) \right)$$

where $\widehat{S}_{t_k}$ is an N_k-quantization of S_{t_k} obtained by a nearest neighbour projection on an optimal (quadratic) N_k-quantizer $\Gamma_k = \{x_1^k, \dots, x_{N_k}^k\}$. This quantization approach amounts to approximating the transitions $\mathcal{L}(S_{t_{k+1}} \,|\, S_{t_k})$ by the finitely valued transition $\mathcal{L}(\widehat{S}_{t_{k+1}} \,|\, \widehat{S}_{t_k})$. All these grids and transition weights make up a so-called *quantization tree*. In a Gaussian framework some grids can be obtained from pre-computed normalized optimal grids (available at the web site

`www.quantize.maths-fi.com`). Otherwise, this quantization optimization step can be performed by some stochastic optimization procedures (like randomized Lloyd's I and $CLVQ$ procedures, see Section 6.3 for an example in a Gaussian framework). The next step is the computation of the quantized transitions $\widehat{\pi}_{ij}^k := \mathbb{P}(\widehat{S}_{t_{k+1}} = x_j^{k+1} \,|\, \widehat{S}_{t_k} = x_i^k)$ by a Monte Carlo simulation. Both steps are mainly based on repeated nearest neighbour searches. They can be carried out *off-line* since they *do not depend on the payoff characteristics (the K_{t_k}'s)*. However, using some fast nearest neighbour procedures, typically the K-d-Tree algorithm introduced in (Friedman et al. 1977) or some improved versions like PAT (Principle Axis Tree, see (McNames 2001)), drastically reduces the complexity of this phase (hence its duration) in higher dimension. Now with modern computing devices it becomes possible in many applications to include both phases in the *on-line* computations, which makes it as flexible as Monte Carlo based methods like regression methods.

Assume that $S_0 = s_0 \in (0, \infty)$ and that $\widehat{S}_0 = s_0$. It is proved in (Bardou et al. 2007b) that

$$|\widehat{P}(0, 0, s_0) - P(0, 0, s_0)| \leq C \sum_{k=0}^{n-1} \|S_{t_k} - \widehat{S}_{t_k}\|_2.$$

If all the quantizations $\widehat{S}_{t_k}$ of S_{t_k} are optimal (with size $N_k = \bar{N}$) this provides a $O(\frac{n}{\bar{N}^{\frac{1}{d}}})$ rate of convergence In fact, numerical evidences show that the observed rate is usually $O(\frac{n}{\bar{N}^{\frac{2}{d}}})$, *i.e.* somewhat similar to the rates obtained in the cubature formulas for differentiable functions. The choice of the N_k's can be refined like for American options following the lines of (Bally and Pagès 2003a).

A comparison carried out in (Bardou et al. 2007a) suggests that the quantization approach (including the transition computation) is significantly faster than the least squares regression methods à la Longstaff-Schwartz.

NUMERICAL ILLUSTRATION: We consider the one factor toy-model given by

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

where $\sigma = 70\%$, $\alpha = 4$ and $t_k = k/n$. The future prices are real data (day 17/01/2003) corresponding to the first part of the curve in Figure 5.3(b). The contract parameters are $q_{\min} = 0$, $q_{\max} = 6$, $Q_{\min} = 1300$, $Q_{\max} = 1900$, $K_{t_k} = K =$ and $n = 365$ (1 year).

Note the slope in Figure 5.3(d) is $1.96 \approx 2$. The main asset of quantization based is to directly approximate the underlying Markov dynamics (in particular when dealing with multi-factor models): it is a model-driven method which has is its ability to "capture" automatically the correlation structure of the asset which becomes quickly impossible with multinomial trees as the number of factors increases.

6 Optimal quadratic functional quantization of Gaussian processes

Optimal quadratic functional quantization of Gaussian processes is closely related to their so-called Karhunen-Loève expansion which can be seen in some sense as some infinite dimensional Principal Component Analysis (PCA) of a (Gaussian) process. Before stating a general result for Gaussian processes, we start by the standard Brownian motion: it is the most important example in view of (numerical) applications and for this process, everything can be made explicit.
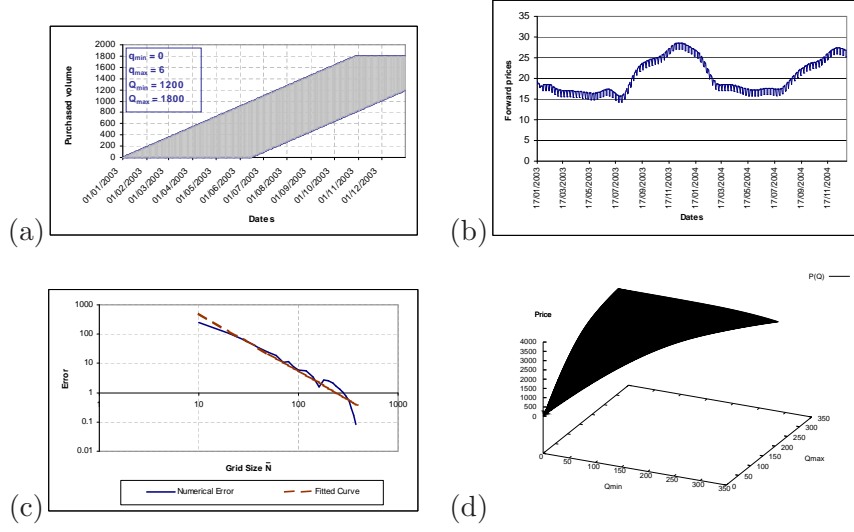
Figure 4: *The parameters are those given in the numerical illustration. (a) Constraint set. (b) Daily forward curve. (c) Numerical convergence as a function of the optimal grid size (log-log-scale) (d) The graph of the price as a function of the global constraints.*

6.1 Brownian motion

One considers the Hilbert space $H = L_T^2 := L^2([0,T], dt)$, $(f|g)_2 = \int_0^T f(t)g(t)dt$, $|f|_{L_T^2} = \sqrt{(f|f)_2}$.

The covariance operator C_W of the Brownian motion $W = (W_t)_{t \in [0,T]}$ is defined on L_T^2 by

$$C_W(f) := \mathbb{E}\left((f,W)_2 W\right) = \left(t \mapsto \int_0^T (s \wedge t) f(s) ds\right).$$

It is a symmetric positive trace class operator which can be diagonalized in the so-called Karhunen-Loève (K-L) orthonormal basis $(e_n^W)_{n \geq 1}$ of L_T^2, with eigenvalues $(\lambda_n)_{n \geq 1}$, given by

$$e_n^W(t) = \sqrt{\frac{2}{T}} \sin\left(\pi(n - \frac{1}{2})\frac{t}{T}\right), \quad \lambda_n = \left(\frac{T}{\pi(n - \frac{1}{2})}\right)^2, \; n \geq 1.$$

This classical result can be established as a simple exercise by solving the functional equation $C_W(f) = \lambda f$. In particular, one can expand W itself on this basis so that

$$W \stackrel{L_T^2}{=} \sum_{n \geq 1} (W|e_n^W)_2 \, e_n^W.$$

Now, the orthonormality of the (K-L) basis implies, using Fubini's Theorem,

$$\mathbb{E}((W|e_k^W)_2 (W|e_\ell^W)_2) = (e_k^W | C_W(e_\ell^W))_2 = \lambda_\ell \delta_{k\ell}$$

where $\delta_{k\ell}$ denotes the Kronecker symbol. Hence the Gaussian sequence $((W|e_n^W)_2)_{n \geq 1}$ is pairwise non-correlated which implies that these random variables are independent. The above identity also implies that $\mathrm{Var}((W|e_n^W)_2) = \lambda_n$. Finally this shows that

$$W \stackrel{L_T^2}{=} \sum_{n \geq 1} \sqrt{\lambda_n} \, \xi_n \, e_n^W \tag{6.1}$$

17

where $\xi_n := (W|e_n^W)_2/\sqrt{\lambda_n}$, $n \geq 1$, is an i.i.d. sequence of $\mathcal{N}(0;1)$-distributed random variables. Furthermore, this K-L expansion converges in a much stronger sense since $\sup_{t\in[0,T]} |W_t - \sum_{k=1}^n \sqrt{\lambda_k}\xi_k e_k^W(t)| \to 0$ $\mathbb{P}$-a.s. and

$$\| \sup_{[0,T]} |W_t - \sum_{1\leq k\leq n} \sqrt{\lambda_k}\xi_k e_k^W(t)|\|_2 = O\left(\sqrt{\log n/n}\right)$$

(see *e.g.* (Luschgy and Pagès 2007)). Similar results (with various rates) hold true for a wide class of Gaussian processes expanded on "admissible" basis (see *e.g.* (Luschgy and Pagès 2007)).

Theorem 6.1 *((Luschgy and Pagès 2002) (2002), (Luschgy and Pagès 2004) (2003) and (Luschgy et al. 2007) (2007)) Let Γ^N, $N \geq 1$, be a sequence of optimal N-quantizers for W.*

(a) For every $N \geq 1$, $\mathrm{span}(\Gamma^N) = \mathrm{span}\{e_1^W, \dots, e_{d(N)}^W\}$ with $d(N) \gtrsim \frac{1}{2}\log N$. Furthermore $\widehat{W}^{\Gamma^N}$ and $W - \widehat{W}^{\Gamma^N}$ are independent.

(b) $e_N(W, L_T^2) = \|W - \widehat{W}^{\Gamma^N}\|_2 \sim \dfrac{T\sqrt{2}}{\pi}\dfrac{1}{\sqrt{\log N}}$ as $N \to \infty$.

Remark. • The fact, confirmed by numerical experiments (see Section 6.3 Figure 6.3.1), that $d(N) \sim \log N$ holds as a conjecture.

• Denoting Π_d the orthogonal projection on $\mathrm{span}\{e_1^W, \dots, e_d^W\}$, one derives from (a) that $\widehat{W}^{\Gamma^N} = \widehat{\Pi_{d(N)}(W)}^{\Gamma^N}$ (optimal quantization at level N) and

$$\begin{aligned}
\|W - \widehat{W}^{\Gamma^N}\|_2^2 &= \|\Pi_{d(N)}(W) - \widehat{\Pi_{d(N)}(W)}^{\Gamma^N}\|_2^2 + \|W - \Pi_{d(N)}(W)\|_2^2 \\
&= e_N\left(Z_{d(N)}, \mathbb{R}^{d(N)}\right)^2 + \sum_{n\geq d(N)+1} \lambda_n
\end{aligned}$$

where $Z_{d(N)} \overset{\mathcal{L}}{\sim} \Pi_{d(N)}(W) \sim \bigotimes_{k=1}^{d(N)} \mathcal{N}(0;\lambda_k)$.

6.2 Centered Gaussian processes

The above Theorem 6.1 devoted to the standard Brownian motion is a particular case of a more general theorem which holds for a wide class of Gaussian processes

Theorem 6.2 *((Luschgy and Pagès 2002), (Luschgy and Pagès 2004) and (Luschgy et al. 2007)) Let $X = (X_t)_{t\in[0,T]}$ be a Gaussian process with K-L eigensystem $(\lambda_n^X, e_n^X)_{n\geq 1}$ (with $\lambda_1 \geq \lambda_2 \geq \dots$ is non-increasing). Let Γ^N, $N \geq 1$, be a sequence of quadratic optimal N-quantizers for X. Assume*

$$\lambda_n^X \sim \frac{\kappa}{n^b} \quad as\ n \to \infty \quad (b > 1).$$

(a) $\mathrm{span}(\Gamma^N) = \mathrm{span}\{e_1^X, \dots, e_{d^X(N)}^X\}$ and $d^X(N) \gtrsim \dfrac{1}{b^{1/(b-1)}}\dfrac{2}{b}\log N$.

(b) $e_N(X, L_T^2) = \|X - \widehat{X}^{\Gamma^N}\|_2 \sim \sqrt{\kappa}\sqrt{b^b(b-1)^{-1}}\,(2\log N)^{-\frac{b-1}{2}}$.

Remarks. • The above result admits an extension to the case $\lambda_n^X \sim \varphi(n)$ as $n \to \infty$ with φ regularly varying, index $-b \leq -1$ (see (Luschgy and Pagès 2004)). In (Luschgy and Pagès 2002), upper or lower bounds are also established when

$$(\lambda_n^X \leq \varphi(n), \quad n \geq 1) \quad \text{or} \quad (\lambda_n^X \geq \varphi(n), \quad n \geq 1).$$

- The sharp asymptotics $d^X(N) \sim \frac{2}{b} \log N$ holds as a conjecture.

Applications to classical (centered) Gaussian processes.

- Brownian bridge: $X_t := W_t - \frac{t}{T} W_T$, $t \in [0,T]$ and $e_n^X(t) = \sqrt{2/T} \sin\left(\pi n \frac{t}{T}\right)$, $\lambda_n = \left(\frac{T}{\pi n}\right)^2$, so that $e_N(X, L_T^2) \sim T \frac{\sqrt{2}}{\pi} (\log N)^{-\frac{1}{2}}$.

- Fractional Brownian motion with Hurst constant $H \in (0,1)$

$$e_N(W^H, L_T^2) \sim T^{H+\frac{1}{2}} c(H) (\log N)^{-H}$$

where $c(H) = \left(\frac{\Gamma(2H) \sin(\pi H)(1+2H)}{\pi}\right)^{\frac{1}{2}} \left(\frac{1+2H}{2\pi}\right)^H$ and $\Gamma(t)$ denotes the Gamma function at $t > 0$.

- Some further explicit sharp rates can be derived from the above theorem for other classes of Gaussian stochastic processes (see (Luschgy and Pagès 2004), 2004) like the fractional Ornstein-Uhlenbeck processes, the Gaussian diffusions, a wide class Gaussian stationary processes (the quantization rate is derived from the high frequency asymptotics of its spectral density, assumed to be square integrable on the real line), for the m-folded integrated Brownian motion, the fractional Brownian sheet, etc.

- Of course some upper bounds can be derived for some even wider classes of processes, based on the above first remark (see *e.g.* (Luschgy and Pagès 2002), 2002).

Extensions to $r, p \neq 2$ When the processes have some self-similarity properties, it is possible to obtain some sharp rates in the non purely quadratic case: this has been done for fractional Brownian motion in (Dereich et al. 2006) using some quite different techniques in which self-similarity properties plays there a crucial role. It leads to the following sharp rates, for $p \in [1, +\infty]$ and $r \in (0, \infty)$

$$e_{N,r}(W^H, L_T^p) \sim T^{H+\frac{1}{2}} c(r,H) (\log N)^{-H}, \quad c(r,H) \in (0, +\infty).$$

6.3 Numerical optimization of quadratic functional quantization

Thanks to the scaling property of Brownian motion, one may focus on the normalized case $T = 1$. The numerical approach to optimal quantization of the Brownian motion is essentially based on Theorem 6.1 and the remark that follows: indeed these results show that quadratic optimal functional quantization of a centered Gaussian process reduces to a finite dimensional optimal quantization problem for a Gaussian distribution with a diagonal covariance structure. Namely the optimization problem at level N reads

$$(\mathcal{O}_N) \equiv \begin{cases} e_N(W, L_T^2)^2 & := e_N(Z_{d(N)}, \mathbb{R}^{d(N)})^2 + \sum_{k \geq d(N)+1} \lambda_k \\ \text{where} \quad Z_{d(N)} \overset{\mathcal{L}}{\sim} \bigotimes_{k=1}^{d(N)} \mathcal{N}(0, \lambda_k). \end{cases}$$

Moreover, if $\beta^N := \{\beta_1^N, \ldots, \beta_N^N\}$ denotes an optimal N-quantizer of $Z_{d(N)}$, then, the optimal N-quantizer Γ^N of W reads $\Gamma^N = \{x_1^N, \ldots, x_N^N\}$ with

$$x_i^N(t) = \sum_{1 \leq \ell \leq d(N)} (\beta_i^N)^\ell e_\ell^W(t), \quad i = 1, \ldots, N. \tag{6.2}$$

The good news is that $(\mathcal{O}_N)$ is in fact a *finite dimensional quantization optimization problem* for each $N \geq 1$. The bad news is that the problem is somewhat ill conditioned since the decrease

of the eigenvalues of W is very steep for small values of n: $\lambda_1 = 0.40528\ldots$, $\lambda_2 = 0.04503\cdots \approx \lambda_1/10$. This is probably one reason for which former attempts to produce good quantization of the Brownian motion first focused on other kinds of quantizers like *scalar product quantizers* (see (Pagès and Printems 2005) and Section 6.4 below) or d-dimensional block product quantizations (see (Wilbertz 2005) and (Luschgy et al. 2007)).

Optimization of the (quadratic) quantization of $\mathbb{R}^d$-valued random vector has been extensively investigated since the early 1950's, first in 1-dimension, then in higher dimension when the cost of numerical Monte Carlo simulation was drastically cut down (see (Gersho and Gray 1992)). Recent application of optimal vector quantization to numerics turned out to be much more demanding in terms of accuracy. In that direction, one may cite (Pagès and Printems 2003), (Mrad and Ben Hamida 2006) (mainly focused on numerical optimization of the quadratic quantization of normal distributions). To apply the methods developed in these papers, it is more convenient to rewrite our optimization problem with respect to the standard d-dimensional distribution $\mathcal{N}(0;I_d)$ by simply considering the Euclidean norm derived from the covariance matrix $\mathrm{Diag}(\lambda_1,\ldots,\lambda_{d(N)})$ *i.e.*

$$(\mathcal{O}_N) \Leftrightarrow \begin{cases} N\text{-optimal quantization of } \displaystyle\bigotimes_{k=1}^{d(N)} \mathcal{N}(0,1) \\ \text{for the covariance norm } |(z_1,\ldots,z_{d(N)})|^2 = \sum_{k=1}^{d(N)} \lambda_k z_k^2. \end{cases}$$

The main point is of course that the dimension $d(N)$ is unknown. However (see Figure 6.3.1), one clearly verifies on small values of N that the conjecture ($d(N) \sim \log N$) is most likely true. Then for higher values of N one relies on it to shift from one dimension to another following the rule $d(N) = d$, $N \in \{e^d,\ldots,e^{d+1}-1\}$.

6.3.1 A toolbox for quantization optimization: a short overview

Here is a short overview of stochastic optimization methods to compute optimal or at least locally optimal quantizers in finite dimension. For more details we refer to (Pagès and Printems 2003) and the references therein. Let $Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0;I_d)$.

Competitive Learning Vector Quantization (CLVQ). This procedure is a recursive stochastic approximation gradient descent based on the integral representation of the gradient $\nabla D_N^Z(x)$, $x \in H^n$ (temporarily coming back to N-tuple notation) of the distortion as the expectation of a *local gradient i.e.*

$$\forall\, x^N \in H^N, \quad \nabla D_N^Z(x^N) = \mathbb{E}(\nabla D_N^Z(x^N,\zeta)), \ \zeta_k \ i.i.d., \ \zeta_1 \overset{\mathcal{L}}{\sim} \mathcal{N}(0,I_d)$$

so that, starting from $x^N(0) \in (\mathbb{R}^d)^N$, one sets

$$\forall\, k \geq 0, \quad x^N(k+1) \;=\; x^N(k) - \frac{c}{k+1}\nabla D_N^Z(x^N(k),\zeta_{k+1})$$

where $c \in (0,1]$ is a real constant to be tuned. As set, this looks quite formal but the operating $CLVQ$ procedure consists of two phases at each iteration:

(*i*) *Competitive Phase:* Search of the nearest neighbor $x^N(k)_{i*(k+1)}$ of ζ_{k+1} among the components of $x^N(k)_i$, $i = 1,\ldots,N$ (using a "winning convention" in case of conflict on the boundary of the Voronoi cells).

(*ii*) *Cooperative Phase:* One moves the winning component toward ζ_{k+1} using a dilatation *i.e.* $x^N(k+1)_{i*(k+1)} = \mathrm{Dilatation}_{\zeta_{k+1},1-\frac{c}{k+1}}(x^N(k)_{i*(k+1)})$.

This procedure is useful for small or medium values of N. For an extensive study of this procedure, which turns out to be singular in the world of recursive stochastic approximation algorithms, we refer to (Pagès 1998). For general background on stochastic approximation, we refer to (Kushner and Yin 2003; Benveniste et al. 1990).

The randomized "Lloyd I procedure". This is the randomization of the stationarity based fixed point procedure since any optimal quantizer satisfies (3.3):

$$\widehat{Z}^{x^N(k+1)} = \mathbb{E}(Z \,|\, \widehat{Z}^{x^N(k)}), \qquad x^N(0) \subset \mathbb{R}^d.$$

At every iteration the conditional expectation $\mathbb{E}(Z \,|\, \widehat{Z}^{x^N(k)})$ is computed using a Monte Carlo simulation. For more details about practical aspects of Lloyd I procedure we refer to (Pagès and Printems 2003). In (Mrad and Ben Hamida 2006), an approach based on genetic evolutionary algorithms is developed.

For both procedures, one may substitute a sequence of quasi-random numbers to the usual pseudo-random sequence. This often speeds up the rate of convergence of the method, although this can only be proved (see (Lapeyre et al. 1990)) for a very specific class of stochastic algorithms (to which $CLVQ$ does not belong).

The most important step to preserve the accuracy of the quantization as N (and $d(N)$) increase 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 resulting from the former procedure with ν points sampled independently from the normal distribution with probability density proportional to $\varphi^{\frac{d}{d+2}}$ where φ denotes the p.d.f. of $\mathcal{N}(0; I_d)$. This rather unexpected choice is motivated by the fact that this distribution provides the lowest *in average* random quantization error (see (Cohort 1998)).

As a result, to be downloaded on the web site (Pagès and Printems 2005):

www.quantize.maths-fi.com

$\triangleright$ *Optimized stationary codebooks for* W: in practice, the N-quantizers β^N of the distribution $\otimes_{k=1}^{d(N)} \mathcal{N}(0; \lambda_k)$, $N = 1$ up to $10\,000$ ($d(N)$ runs from 1 up to 9).

$\triangleright$ *Companion parameters*:

– distribution of $\widehat{W}^{\Gamma^N}$: $\mathbb{P}(\widehat{W}^{\Gamma^N} = x_i^N) = \mathbb{P}(\widehat{Z}_{d(N)}^{\beta^N} = \beta_i^N)$ ($\leftarrow$ in $\mathbb{R}^{d(N)}$).

– The quadratic quantization error: $\|W - \widehat{W}^{\Gamma^N}\|_2$.

Remarks. • Both stochastic optimization procedures that we described above can of course be implemented to produce optimal (or optimized) grids of any multi-dimensional probability distribution on $\mathbb{R}^d$, having in mind that as the dimension d increases the second one becomes the most efficient.

• These procedure are based on a nearest neighbour search among N points. A naive implementation of such a procedure has a linear complexity in N and becomes very demanding in higher dimension. So to drastically reduce this optimization phase as well as that devoted to the weight estimation of the resulting optimal quantizer, one can call upon some fast nearest neighbour procedure like that originally developed by Bentley and analyzed a the seminal paper (Friedman et al. 1977) by Friedman, Bentley and Finkel which is based on the notion of k-d tree introduced for that purpose by the authors. It reduces the complexity of the search down to $O(\log N)$. The (relative) efficiency of the method increases as the dimension of the state space increases.

6.4 An alternative: product functional quantization

Scalar Product functional quantization is a quantization method which produces rate optimal sub-optimal quantizers. They were used *e.g.* in (Luschgy and Pagès 2002) to provide exact rate
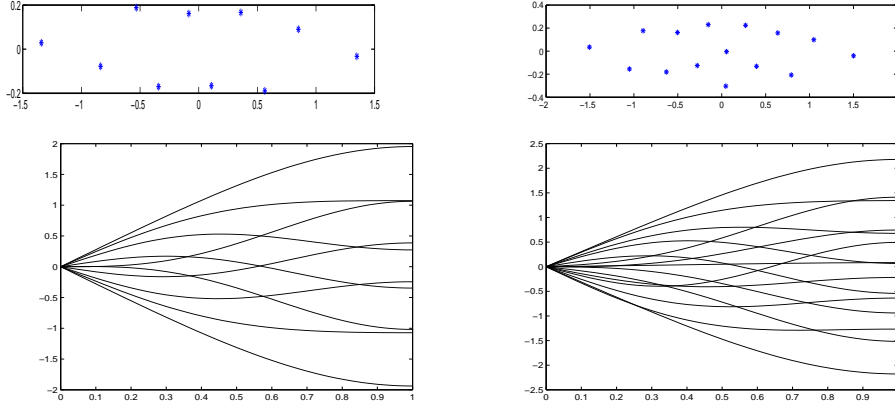
Figure 5: *Optimized functional quantization of the Brownian motion W for $N = 10$, 15 ($d(N) = 2$). Top: β^N depicted in $\mathbb{R}^2$. Bottom: the optimized N-quantizer Γ^N.*

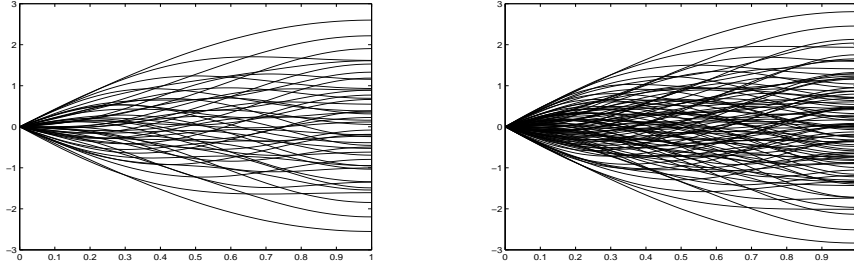
Figure 6: *Optimized functional quantization of the Brownian motion W. The N-quantizers Γ^N. Left: $N = 48$ ($d(N) = 3$). Right: $N = 96$, $d(96) = 4$.*
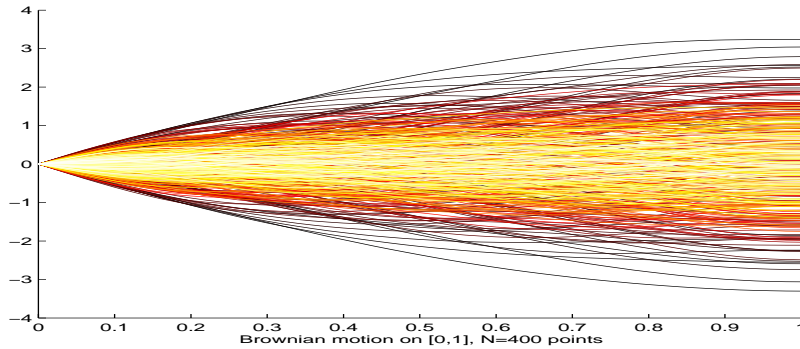

Figure 7: *Optimized N-quantizer Γ^N of the Brownian motion W with $N = 400$. The grey level of the paths codes their weights.*

22
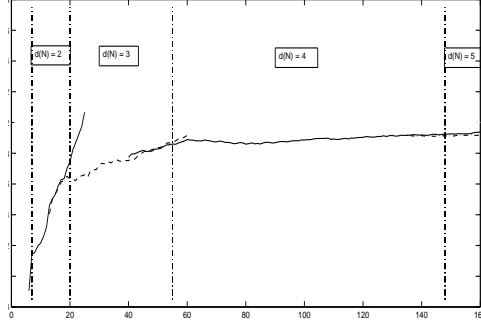
Figure 8: *Optimal functional quantization of the Brownian motion.* $N \mapsto \log N \, (e_N(W, L_T^2))^2$, $N \in \{6, \dots, 160\}$. *Vertical dashed lines: critical dimensions for* $d(N)$, $e^2 \approx 7$, $e^3 \approx 20$, $e^4 \approx 55$, $e^5 \approx 148$.

(although not sharp) for a very large class of processes. The first attempts to use functional quantization for numerical computation with the Brownian motion was achieved with these quantizers (see (Pagès and Printems 2005)). We will see further on their assets. What follows is presented for the Brownian motion but would work for a large class of centered Gaussian processes.

Let us consider again the expansion of W in its *K-L* basis :

$$ W \overset{L_T^2}{=} \sum_{n \geq 1} \sqrt{\lambda_n} \, \xi_n \, e_n^W $$

where $(\xi_n)_{n \geq 1}$ is an i.i.d. sequence $\mathcal{N}(0; 1)$-distributed random variables (keep in mind this convergence also holds *a.s.* uniformly in $t \in [0, T]$). The idea is simply to quantize these (normalized) random coordinates ξ_n: for every $n \geq 1$, one considers an optimal N_n-quantization of ξ_n, denoted $\widehat{\xi}_n^{(N_n)}$ ($N_n \geq 1$). For $n > m$, set $N_n = 1$ and $\widehat{\xi}_n^{(N_n)} = 0$ (which is the optimal 1-quantization). The integer m is called the *length* of the product quantization. Then, one sets

$$ \widehat{W}_t^{(N_1, \dots, N_m, prod)} \quad := \quad \sum_{n \geq 1} \sqrt{\lambda_n} \, \widehat{\xi}_n^{(N_n)} \, e_n^W(t) = \sum_{n=1}^m \sqrt{\lambda_n} \, \widehat{\xi}_n^{(N_n)} \, e_n^W(t). $$

Such a quantizer takes $\prod_{n=1}^m N_n \leq N$ values.

If one denotes by $\alpha^M = \{\alpha_1^M, \dots, \alpha_M^M\}$ the (unique) optimal quadratic M-quantizer of the $\mathcal{N}(0; 1)$-distribution, the underlying quantizer of the above quantization $\widehat{W}^{(N_1, \dots, N_m, prod)}$ can be expressed as follows (if one introduces the appropriate multi-indexation): for every multi-index $\underline{i} := (i_1, \dots, i_m) \in \prod_{n=1}^m \{1, \dots, N_n\}$, set

$$ x_{\underline{i}}^{(N)}(t) := \sum_{n=1}^m \sqrt{\lambda_n} \, \alpha_{i_n}^{(N_n)} e_n^W(t) \text{ and } \Gamma^{N_1, \dots, N_m, prod} := \left\{ x_{\underline{i}}^{(N)}, \, \underline{i} \in \prod_{n=1}^m \{1, \dots, N_n\} \right\}. $$

Then the product quantization $\widehat{W}^{(N_1, \dots, N_m, prod)}$ can be rewritten as

$$ \widehat{W}_t^{(N_1, \dots, N_m, prod)} = \sum_{\underline{i}} \mathbf{1}_{\{W \in C_{\underline{i}}(\Gamma^{N_1, \dots, N_m, prod})\}} x_{\underline{i}}^{(N)}(t). $$

where the Voronoi cell of $x_{\underline{i}}^{(N)}$ is given by

$$ C_{\underline{i}}(\Gamma^{N_1, \dots, N_m, prod}) = \prod_{n=1}^m (\alpha_{i_n - \frac{1}{2}}^{(N_n)}, \alpha_{i_n + \frac{1}{2}}^{(N_n)}), \qquad \alpha_{i \pm \frac{1}{2}}^{(M)} := \frac{\alpha_i^{(M)} + \alpha_{i \pm 1}^{(M)}}{2}, \, \alpha_0 = -\infty, \, \alpha_{M+1} = +\infty. $$

6.4.1 Quantization rate by product quantizers

It is clear that the optimal product quantizer is the solution to the optimal integral bit allocation

$$\min\left\{\|W - \widehat{W}^{(N_1,\ldots,N_m,prod)}\|_2, N_1,\ldots,N_m \geq 1, N_1 \times \cdots \times N_m \leq N, m \geq 1\right\}. \qquad (6.3)$$

Expanding $\|W - \widehat{W}^{(N_1,\ldots,N_m,prod)}\|_2^2 = \||W - \widehat{W}^{(N_1,\ldots,N_m,prod)}|_{L_T^2}\|_2^2$ yields

$$\|W - \widehat{W}^{(N_1,\ldots,N_m,prod)}\|_2^2 = \sum_{n\geq 1} \lambda_n \|\widehat{\xi}_n^{(N_n)} - \xi_n\|_2^2 \qquad (6.4)$$

$$= \sum_{n=1}^m \lambda_n(e_{N_n}^2(\mathcal{N}(0;1),\mathbb{R}) - 1) + \frac{T^2}{2} \qquad (6.5)$$

since

$$\sum_{n\geq 1} \lambda_n = \mathbb{E}\sum_{n\geq 1}(W \mid e_n^W)_2^2 = \mathbb{E}\int_0^T W_t^2 dt = \int_0^T t\, dt = \frac{T^2}{2}.$$

Theorem 6.3 *(see (Luschgy and Pagès 2002)) For every $N \geq 1$, there exists an optimal scalar product quantizer of size at most N (or at level N), denoted $\widehat{W}^{(N,prod)}$, of the Brownian motion defined as the solution to the minimization problem (6.3). Furthermore these optimal product quantizers make up a rate optimal sequence: there exists a real constant $c_W > 0$ such that*

$$\|W - \widehat{W}^{(N,prod)}\|_2 \leq \frac{c_W T}{(\log N)^{\frac{1}{2}}}.$$

Proof (sketch of). By scaling one may assume without loss of generality that $T = 1$. Combining (6.4) and Zador's Theorem shows

$$\|W - \widehat{W}^{(N_1,\ldots,N_m,prod)}\|_2^2 \leq C\left(\sum_{n=1}^m \frac{1}{n^2 N_n^2}\right) + \sum_{n\geq m+1} \lambda_n$$

$$\leq C'\left(\sum_{n=1}^m \frac{1}{n^2 N_n^2} + \frac{1}{m}\right)$$

with $\prod_n N_n \leq N$. Setting $m := m(N) = [\log N]$ and $N_k = \left[\frac{(m! N)^{\frac{1}{m}}}{k}\right] \geq 1$, $k = 1,\ldots,m$, yields the announced upper-bound. $\diamond$

Remarks. • One can show that the length $m(N)$ of the optimal quadratic product quantizer satisfies

$$m(N) \sim \log N \qquad \text{as} \qquad N \to +\infty.$$

• The most striking fact is that very few ingredients are necessary to make the proof work as far as the quantization rate is concerned. We only need the basis of L_T^2 on which W is expanded to be orthonormal *or* the random coordinates to be orthogonal in $L^2(\mathbb{P})$. This robustness of the proof has been used to obtain some upper bounds for very wide classes of Gaussian processes by considering alternative orthonormal basis of L_T^2 like the Haar basis for processes having self-similarity properties (see (Luschgy and Pagès 2002)), or trigonometric basis for stationary processes (see (Luschgy and Pagès 2002)). More recently, combined with the non asymptotic Zador's Theorem, it was used to provide some connections between mean regularity of stochastic processes and quantization rate (see Section 9 and (Luschgy and Pagès 2007)).

• Block quantizers combined with large deviations estimates were used to provide the sharp rate obtained in Theorem 6.1 in (Luschgy and Pagès 2004).

• d-dimensional block quantization is also possible, possibly with varying block size, providing a constructive approach to sharp rate, see (Wilbertz 2005) and (Luschgy et al. 2007).

• A similar approach can also provide some $L^r(\mathbb{P})$-rates for product quantization with respect to the sup-norm over $[0, T]$, see (Luschgy and Pagès 2007).

6.4.2 How to use product quantizers for numerical computations ?

For numerics one can assume by a scaling argument that $T = 1$. To use product quantizers for numerics we need to have access to the quantizers (or grid) at a given level N, their weights (and the quantization error). All these quantities are available with product quantizers. In fact the first attempts to use functional quantization for numerics (path dependent option pricing) were carried out with product quantizers (see (Pagès and Printems 2005)).

• The optimal product quantizers (denoted $\Gamma^{(N,prod)}$) at level N are explicit, given the optimal quantizers of the scalar normal distribution $\mathcal{N}(0; 1)$. In fact the optimal allocation of the size N_i of each marginal has been already achieved up to very high values of N. Some typical optimal allocation (and the resulting quadratic quantization error) are reported in the table below.

N	$N_{\rm rec}$	Quant. Error	Opti. Alloc.
1	1	0.7071	1
10	10	0.3138	5-2
100	96	0.2264	12-4-2
1 000	966	0.1881	23-7-3-2
10 000	9 984	0.1626	26-8-4-3-2-2
100 000	97 920	0.1461	$34 - 10 - 6 - 4 - 3 - 2 - 2$

• The weights $\mathbb{P}(\widehat{W}^{(N,prod)} = x_{\underline{i}})$ are explicit too: the normalized coordinates ξ_n of W in its K-L basis are independent, consequently

$$
\begin{aligned}
\mathbb{P}(\widehat{W}^{(N,prod)} = x_{\underline{i}}) &= \mathbb{P}(\widehat{\xi}_n^{(N_n)} = \alpha_{i_n}^{(N_n)}, \, n = 1, \ldots, m(N)) \\
&= \prod_{n=1}^{m(N)} \underbrace{\mathbb{P}(\widehat{\xi}_n^{(N_n)} = \alpha_{i_n}^{(N_n)})}_{1D \ (tabulated) \ weights} \, .
\end{aligned}
$$

• Equation (6.5) shows that the (squared) quantization error of a product quantizer can be straightforwardly computed as soon as one knows the eigenvalues and the (squared) quantization error of the normal distributions for the N_i's.

The optimal allocations up to $N = 12\,000$ can be downloaded on the website (Pagès and Printems 2005) as well as the necessary 1-dimensional optimal quantizers (including the weights and the quantization error) of the scalar normal distribution (up to a size of 500 which quite enough for this purpose).

For numerical purpose we are also interested in the stationarity property since such quantizers produce lower (weak) errors in cubature formulas.

Proposition 6.1 *(see (Pagès and Printems 2005)) The product quantizers obtained from the K-L basis are stationary quantizers (although sub-optimal).*

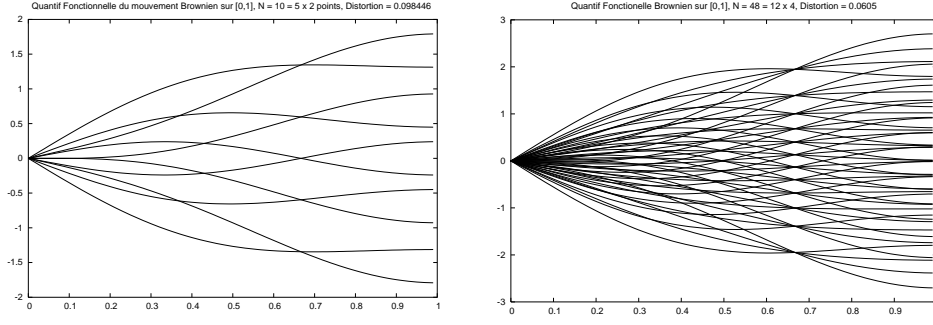Figure 9: *Product quantization of the Brownian motion: the N_{rec}-quantizer $\Gamma^{(N,prod)}$. $N = 10$: $N_{\text{rec}} = 10$ and $N = 50$: $N_{\text{rec}} = 12 \times 4 = 48$.*
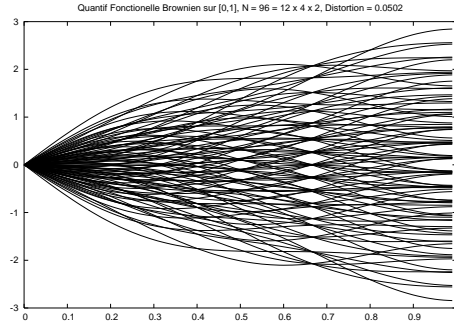

Figure 10: *Product quantization of the Brownian motion: the N_{rec}-quantizer $\Gamma^{(N,prod)}$. $N = 100$: $N_{\text{rec}} = 12 \times 4 \times 2 = 96$.*

Proof. Firstly, note that

$$\widehat{W}^{N,prod} \;=\; \sum_{n \geq 1} \sqrt{\lambda_n}\, \widehat{\xi}_n^{(N_n)} e_n(t)$$

so that $\sigma(\widehat{W}^{N,prod}) = \sigma(\widehat{\xi}_k^{(N_k)},\ k \geq 1)$. Consequently

$$
\begin{aligned}
\mathbb{E}(W \,|\, \widehat{W}^{N,prod}) &= \mathbb{E}(W \,|\, \sigma(\widehat{\xi}_k^{(N_k)},\ k \geq 1)) \\
\mathbb{E}(W \,|\, \widehat{W}^{N,prod}) &= \sum_{n \geq 1} \sqrt{\lambda_n}\, \mathbb{E}\left(\xi_n \,|\, \sigma(\widehat{\xi}_k^{(N_k)},\ k \geq 1)\right) e_n^W \\
&\overset{i.i.d.}{=} \sum_{n \geq 1} \sqrt{\lambda_n}\, \mathbb{E}\left(\xi_n \,|\, \widehat{\xi}_n^{(N_n)}\right) e_n^W \\
&= \sum_{n \geq 1} \sqrt{\lambda_n}\, \widehat{\xi}_n^{(N_n)} e_n^W \;=\; \widehat{W}. \qquad \diamond
\end{aligned}
$$

Remarks. • This result is no longer true for product quantizers based on other orthonormal basis.
• This shows the existence of non optimal stationary quantizers.

6.5 Optimal *vs* product quadratic functional quantization ($T = 1$)

▷ (Numerical) Optimized Quantization: By scaling, we can assume without loss of generality that $T = 1$. We carried out a huge optimization task in order to produce some *optimized* quanti-

zation grids for the Brownian motion by solving numerically $(\mathcal{O}_N)$ for $N = 1$ up to $N = 10\,000$.

$$e_N(W, L_T^2)^2 \approx \frac{0.2195}{\log N}, \qquad N = 1, \ldots, 10\,000.$$

This value (see Figure 11(left)) is significantly greater than the theoretical (asymptotic) bound given by Theorem 6.1 which is

$$\lim_N \log N e_N(W, L_T^2)^2 = \frac{2}{\pi^2} = 0.2026...$$

Our guess, supported by our numerical experiments, is that in fact $N \mapsto \log N e_N(W, L_T^2)^2$ is possibly not monotone but unimodal.

$\triangleright$ OPTIMAL PRODUCT QUANTIZATION: as displayed on Figure 11(right), one has approximately

$$\min\left\{\| \, |W - \widehat{W}|_{L_T^2} \|_2^2, \ 1 \leq N_1 \cdots N_m \leq N, \ m \geq 1\right\} = \|W - \widehat{W}^{(N, prod)}\|_2^2 \approx \frac{0.245}{\log N}.$$

$\triangleright$ OPTIMAL d-DIMENSIONAL BLOCK PRODUCT QUANTIZATION: let us briefly mention this approach developed in (Wilbertz 2005) in which product quantization is achieved by quantizing some marginal blocks of size 1, 2 or 3. By this approach, the corresponding constant is approximately 0.23, *i.e.* roughly in between scalar product quantization and optimized numeric quantization.

The conclusion, confirmed by our numerical experiments on option pricing (see Section 8), is that
– Optimal quantization is significantly more accurate on numerical experiments but is much more demanding since it needs to keep off line or at least to handle large files (say 1 GB for $N = 10\,000$).
– Both approaches are included in the option pricer PREMIA (MATHFI Project, Inria). An online benchmark is available on the website (Pagès and Printems 2005).

7 Constructive functional quantization of diffusions

7.1 Rate optimality for Scalar Brownian diffusions

One considers on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ an homogenous Brownian diffusion process:

$$dX_t = b(X_t)dt + \vartheta(X_t)\,dW_t, \quad X_0 = x_0 \in \mathbb{R},$$

where b and ϑ are continuous on $\mathbb{R}$ with at most linear growth (*i.e.* $|b(x)| + |\sigma(x)| \leq C(1 + |x|)$) so that at least a weak solution to the equation exists.

To devise a constructive way to quantize the diffusion X, it seems natural to start from a rate optimal quantization of the Brownian motion and to obtain some "good" (but how good?) quantizers for the diffusion by solving an appropriate ODE. So let $\Gamma^N = (w_1^N, \cdots, w_N^N)$, $N \geq 1$, be a sequence of stationary rate optimal N-quantizers of W. One considers the following (non-coupled) Integral Equations:

$$dx_i^{(N)}(t) = \left(b(x_i^{(N)}(t)) - \frac{1}{2}\vartheta\theta'(x_i^{(N)}(t))\right)dt + \vartheta(t, x_i^{(N)}(t))\,dw_i^N(t), \quad i = 1, \ldots, N. \qquad (7.1)$$

Set

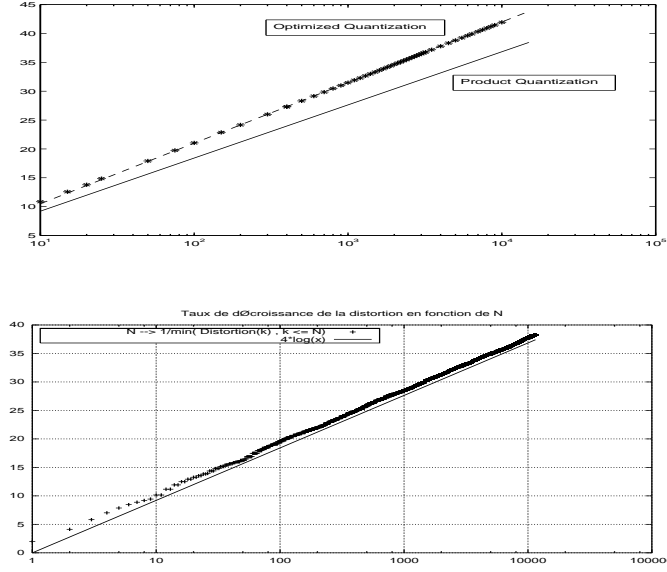$$\widetilde{X}_t^{x^{(N)}} = \sum_{i=1}^N x_i^{(N)}(t)\mathbf{1}_{\{\widehat{W}^{\Gamma^N} = w_i^N\}}.$$

Figure 11: *Numerical quantization rates. Top (Optimal quantization). Line +++: $\log N \mapsto (\|W - \widehat{W}^N\|_2)^{-2}$. Dashed line: $\log N \mapsto \log N/0.2194$. Solid line: $\log N \mapsto \log N/0.25$. Bottom (Product quantization). Line +++: $\log N \mapsto (\min_{1 \le k \le N} \|W - \widehat{W}^{k,prod}\|_2^2)^{-1}$. Solid line: $\log N \mapsto \log N/0.25$.*

The process $\widetilde{X}^{x^{(N)}}$ is a *non-Voronoi* quantizer (since it is defined using the Voronoi diagram of W). What is interesting is that it is a *computable quantizer* (once the above integral equations have been solved) since the weights $\mathbb{P}(\widehat{W}^{\Gamma^N} = w_i^N)$ are known. The Voronoi quantization defined by $x^{(N)}$ induces a lower quantization error but we have no access to its weights for numerics. The good news is that $\widetilde{X}^{x^{(N)}}$ is already rate optimal.

Theorem 7.1 *((Luschgy and Pagès 2006)) Assume that b is differentiable, ϑ is positive twice differentiable and that $b' - b\frac{\vartheta'}{\vartheta} - \frac{1}{2}\vartheta\vartheta''$ is bounded. Then*

$$e_N(X, L_T^2) \le \| X - \widetilde{X}^{x^{(N)}} \|_2 = O((\log N)^{-\frac{1}{2}}).$$

If furthermore, $\vartheta \ge \varepsilon_0 > 0$, then $e_N(X, L_T^2) \approx (\log N)^{-\frac{1}{2}}$.

Remarks. • For some results in the non homogenous case, we refer to (Luschgy and Pagès 2006). Furthermore, the above estimates still hold true for the $(L^r(\mathbb{P}), L_T^p)$-quantization, $1 < r, p < +\infty$ provided $\|\,\|W - \widehat{W}^{\Gamma^N}|_{L_T^p}\|_r = O((\log N)^{-\frac{1}{2}})$.

• This result is closely connected to the Doss-Sussman approach (see *e.g.* (Doss 1977)) and in fact the results can be extended to some classes multi-dimensional diffusions (whose diffusion coefficient is the inverse of the gradient of a diffeomorphism) which include several standard multi-dimensional financial models (including the Black-Scholes model).

• A sharp quantization rate $e_{N,r}(X, L_T^p) \sim c(\log N)^{-\frac{1}{2}}$ for scalar elliptic diffusions is established in (Dereich 2005a; Dereich 2005b) using a non constructive approach, $1 \le p \le \infty$.

EXAMPLE: Rate optimal product quantization of the Ornstein-Uhlenbeck process.

$$dX_t = -kX_t dt + \vartheta dW_t, \qquad X_0 = x_0.$$

One solves the non-coupled integral (linear) system

$$x_i(t) = x_0 - k \int_0^t x_i(s)\,ds + \vartheta w_i^N(t),$$

28

where $\Gamma^N := \{w_1^N, \ldots, w_N^N\}, N \geq 1$ is a *rate optimal* sequence of quantizers

$$w_i^N(t) = \sqrt{\frac{2}{T}} \sum_{\ell \geq 1} \varpi_{i,\ell} \frac{T}{\pi(\ell - 1/2)} \sin\left(\pi(\ell - 1/2)\frac{t}{T}\right), \quad i \in I_N.$$

If Γ^N is optimal for W then $\varpi_{i,\ell} := (\beta_i^N)^\ell$, $i = 1, \ldots, N$, $1 \leq \ell \leq d(N)$ with the notations introduced in (6.2). If Γ^N is an optimal product quantizer (and $N_1, \ldots, N_\ell, \ldots$ denote the optimal size allocation), then $\varpi_{i,\ell} = \alpha_{i_\ell}^{(N_\ell)}$, where $i := (i_1, \ldots, i_\ell, \ldots) \in \prod_{\ell \geq 1}\{1, \ldots, N_\ell\}$. Elementary computations show that

$$x_i^N(t) = e^{-kt}x_0 + \vartheta \sum_{\ell \geq 1} \chi_{i_\ell}^{(N_\ell)} \widetilde{c}_\ell \, \varphi_\ell(t)$$

with
$$\widetilde{c}_\ell = \frac{T^2}{(\pi(\ell - 1/2))^2 + (kT)^2}$$

and
$$\varphi_\ell(t) := \sqrt{\frac{2}{T}}\left(\frac{\pi}{T}(\ell - 1/2)\sin\left(\pi(\ell - 1/2)\frac{t}{T}\right) + k\left(\cos\left(\pi(\ell - 1/2)\frac{t}{T}\right) - e^{-kt}\right)\right).$$

7.2 Multi-dimensional diffusions for Stratanovich SDE's

The correcting term $-\frac{1}{2}\vartheta\vartheta'$ coming up in the integral equations suggest to consider directly some diffusion in the Stratanovich sense

$$dX_t = b(t, X_t)\, dt + \vartheta(t, X_t) \circ dW_t \qquad X_0 = x_0 \in \mathbb{R}^d, \qquad t \in [0, T].$$

(see *e.g.* (Revuz and Yor 1999) for an introduction) where $W = (W^1, \ldots, W^d)$ is a d-dimensional standard Brownian Motion.

In that framework, we need to introduce the notion of p-variation: a continuous function $x : [0, T] \to \mathbb{R}^d$ has finite p-variations if

$$Var_{p,[0,T]}(x) := \sup\left\{\left(\sum_{i=0}^{k-1}|x(t_i) - x(t_{i+1})|^p\right)^{\frac{1}{p}}, 0 \leq t_0 \leq t_1 \leq \cdots \leq t_k \leq T, \ k \geq 1\right\} < +\infty.$$

Then $d_p(x, x') = |x(0) - x'(0)| + Var_{p,[0,T]}(x - x')$ defines a distance on the set of functions with finite p-variations. It is classical background that $Var_{p,[0,T]}(W(\omega)) < +\infty \ \mathbb{P}(d\omega)$-a.s. for every $p > 2$.

One way to quantize W at level (at most) N is to quantize each component W^i at level $\lfloor \sqrt[d]{N} \rfloor$. One shows (see (Luschgy and Pagès 2004)) that $\|W - (\widehat{W}^{1,\lfloor \sqrt[d]{N} \rfloor}, \ldots, \widehat{W}^{d,\lfloor \sqrt[d]{N} \rfloor})\|_2 = O((\log N)^{-\frac{1}{2}})$.

Let $\mathcal{C}_b^r([0, T] \times \mathbb{R}^d)$, $r > 0$, denote the set of $\lfloor r \rfloor$-times differentiable bounded functions $f : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ with bounded partial derivatives up to order $\lfloor r \rfloor$ and whose partial derivatives of order $\lfloor r \rfloor$ are $(r - \lfloor r \rfloor)$-Hölder.

Theorem 7.2 *(see (Pagès and Sellami 2007)) Let $b, \vartheta \in \mathcal{C}_b^{2+\alpha}([0, T] \times \mathbb{R}^d)$ $(\alpha > 0)$ and let $\Gamma^N = \{w_1^N, \ldots, w_N^N\}$, $N \geq 1$, be a sequence of N-quantizers of the standard d-dimensional Brownian motion W such that $\|W - \widehat{W}^{\Gamma^N}\|_2 \to 0$ as $N \to \infty$. Let*

$$\widetilde{X}_t^{x^{(N)}} := \sum_{i=1}^N x_i^{(N)}(t)\mathbf{1}_{\{\widehat{W} = w_i^N\}}$$

where, for every $i \in \{1, \dots, N\}$, $x_i^{(N)}$ is solution to

$$ODE_i \quad \equiv \quad dx_i^{(N)}(t) = b(t, x_i^{(N)}(t))dt + \vartheta(t, x_i^{(N)}(t))dw_i^N(t), \quad x_i^{(N)}(0) = x.$$

Then, for every $p \in (2, \infty)$,

$$Var_{p,[0,T]}(\widetilde{X}^{x^{(N)}} - X) \xrightarrow{\mathbb{P}} 0 \quad as \quad N \to \infty.$$

Remarks. • The keys of this results are the Kolmogorov criterion, stationarity (in a slightly extended sense) and the connection with rough paths theory (see (Lejay 2003) for an introduction to rough paths theory, convergence in p-variation, etc).

• In that general setting we have no convergence rate although we conjecture that $\widetilde{X}^{x^{(N)}}$ remains rate optimal if $\widehat{W}^{\Gamma^N}$ is.

• There are also some results about the convergence of stochastic integrals of the form $\int_0^t g(\widehat{W}_s^N)\, d\widehat{B}_s^N \to \int_0^t g(W_s) \circ dB_s$, with some rates of convergence when $W = B$ or W and B independent (depending on the regularity of the function g, see (Pagès and Sellami 2007)).

7.3 About the quantization of multi-dimensional Brownian motion

We assume $\mathbb{R}^d$ is equipped with the canonical Euclidean norm $|(x^1, \dots, x^d)|^2 = (x^1)^2 + \cdots + (x^d)^2$. Let $W = (W^1, \dots, W^d)$ be a d-dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

$\triangleright$ The most elementary way to quantize W is to quantize each marginal component W^i at a N^i-level so that $N^1 \cdots N^d \leq N$. This appears as a *spatial product quantization*. This is a simple, somewhat flexible approach for applications since N^i can be chosen different of $\lfloor N^{\frac{1}{d}} \rfloor$. However it is clearly not optimal . Furthermore, it suffers as any product like quantization of the difficulty to "control" the product $\prod_{1 \leq i \leq d} N^i$. One easily shows that if $N^i = \lfloor N^{\frac{1}{d}} \rfloor$, $i = 1, \dots, d$, then $\| \, |W - (\widehat{W}^{(N^i)})_{1 \leq i \leq d}| \, \|_2 = O((\log N)^{\frac{1}{2}})$ (see (Luschgy and Pagès 2006)).

$\triangleright$ An alternative way is to quantize the K-L expansion of W given by

$$W^i = \sum_{n \geq 1} \sqrt{\lambda_n^W} \xi_n^i e_n^W, \quad i = 1, \dots, d, \quad \xi_n = (\xi^1 - n, \dots, \xi_n^d) \sim \mathcal{N}(0; I_d),$$

by setting

$$(\widehat{W})^i = \sum_{n \geq 1} \sqrt{\lambda_n^W} \pi^i(\widehat{\xi}_n^{(N_n)}) e_n^W,$$

where $\pi^i(x^1, \dots, x^d) = x^i$ and $\xi_n^{(N_n)}$ is an optimal N_n-quantization of ξ_n.

8 Applications to path-dependent option pricing

The typical functionals F defined on $(L_T^2, |.|_{L_T^2})$ for which $\mathbb{E}(F(W))$ can be approximated by the cubature formulae (4.2), (4.5) are of the form $F(\omega) := \varphi \left(\int_0^T f(t, \omega(t))dt \right) \mathbf{1}_{\{\omega \in \mathcal{C}([0,T], \mathbb{R})\}}$ where $f : [0, T] \times \mathbb{R} \to \mathbb{R}$ is *locally Lipschitz continuous* in the second variable, namely

$$\forall t \in [0, T], \, \forall u, v \in \mathbb{R}, \, |f(t, u) - f(t, v)| \leq C_f |u - v|(1 + g(|u|) + g(|v|))$$

(with $g : \mathbb{R}_+ \to \mathbb{R}_+$ is increasing, convex and $g(\sup_{t \in [0,T]} |W_t|) \in L^2(\mathbb{P})$) and $\varphi : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous. One could consider for ω some càdlàg functions as well. A classical example is the Asian payoff in a Black-Scholes model

$$F(\omega) = \exp(-rT) \left(\frac{1}{T} \int_0^T s_0 \exp(\sigma \omega(t) + (r - \sigma^2/2)t)dt - K \right)_+ .$$

8.1 Numerical integration (III): Richardson-Romberg log-extrapolation

Let $F : L_T^2 \longrightarrow \mathbb{R}$ be a 3 times $|\,.\,|_{L_T^2}$-differentiable functional with bounded differentials. Assume $\widehat{W}^{(N)}$, $N \geq 1$, is a sequence of a rate-optimal stationary quantizations of the standard Brownian motion W. Assume furthermore that

$$\mathbb{E} \left(D^2 F(\widehat{W}^{(N)}).(W - \widehat{W}^{(N)})^{\otimes 2} \right) \sim \frac{c}{\log N} \quad \text{as} \quad N \to \infty \tag{8.1}$$

and

$$\mathbb{E} \, |W - \widehat{W}^{(N)}|_{L_T^2}^3 = O\left((\log N)^{-\frac{3}{2}} \right). \tag{8.2}$$

Then, a higher order Taylor expansion yields

$$
\begin{aligned}
F(W) &= F(\widehat{W}^{(N)}) + DF(\widehat{W}^{(N)}).(W - \widehat{W}^{(N)}) + \frac{1}{2} D^2 F(\widehat{W}^{(N)}).(W - \widehat{W}^{(N)})^{\otimes 2} \\
&\quad + \frac{1}{6} D^2(\zeta).(W - \widehat{W}^{(N)})^{\otimes 3}, \qquad \zeta \in (\widehat{W}^{(N)}, W), \\
\mathbb{E} \, F(W) &= \mathbb{E} F(\widehat{W}^{(N)}) + \frac{c}{2 \log N} + o\left((\log N)^{-\frac{3}{2}+\varepsilon} \right).
\end{aligned}
$$

Then, one can design a log-R-R extrapolation by considering N, N', $N < N'$ (e.g. $N' \approx 4 N$), so that

$$\mathbb{E}(F(W)) = \frac{\log N' \times \mathbb{E}(F(\widehat{W}^{(N')})) - \log N' \times \mathbb{E}(F(\widehat{W}^{(N)}))}{\log N' - \log N} + o\left((\log N)^{-\frac{3}{2}+\varepsilon} \right).$$

For practical implementation, it is suggested in (Wilbertz 2005) to replace $\log N$ by the more consistent "estimator" $\|W - \widehat{W}^{(N)}\|_2^{-2}$.

In fact Assumption (8.1) holds true for optimal product quantization when F is polynomial function F, $d^0 F = 2$. Assumption (8.2) holds true in that case as well (see (Graf et al. 2006)). As concerns optimal quantization, these statements are still conjectures. However, given that $\widehat{W}$ and $W - \widehat{W}$ are independent (see (Luschgy and Pagès 2002)), (8.1) is equivalent to the simple case where $D^2 F(\widehat{W}^{(N)})$ is constant.

Note that the above extrapolation or some variants can be implemented with other stochastic processes in accordance with the rate of convergence of the quantization error.

8.2 Asian option pricing in a Heston stochastic volatility model

In this section, we will price an Asian call option in a Heston stochastic volatility model using some optimal (or at least optimized) functional quantizations of the two Brownian motions that drive the diffusion. This model has already been considered in (Pagès and Printems 2005) in which numerical aspects of functional quantization were investigated for the first time. The implementation was made with some optimal product quantizations of the Brownian motions. The Heston stochastic volatility model was introduced in (Heston 1993) to model stock price dynamics. Its popularity partly comes from the existence of semi-closed forms for vanilla European options, based on inverse

Fourier transform and from its ability to reproduce some skewness shape of the implied volatility surface. We consider it under its risk-neutral probability measure, namely

$$
\begin{aligned}
dS_t &= S_t(r\,dt + \sqrt{v_t}dW_t^1), \qquad S_0 = s_0 > 0, \quad \text{(risky asset)} \\
dv_t &= k(a - v_t)dt + \vartheta\sqrt{v_t}\,dW_t^2, \; v_0 > 0 \text{ with } d\langle W^1, W^2\rangle_t = \rho\,dt, \; \rho \in [-1, 1].
\end{aligned}
$$

where ϑ, k, a such that $\vartheta^2/(4ak) < 1$. We consider the Asian Call payoff with maturity T and strike K. No closed form is available for its premium

$$
\mathrm{AsCall}^{Hest} = e^{-rT}\mathbb{E}\left(\frac{1}{T}\int_0^T S_s ds - K\right)_+.
$$

We briefly recall how to proceed (see (Pagès and Printems 2005) for details): first, one projects W^1 on W^2 so that $W^1 = \rho W^2 + \sqrt{1 - \rho^2}\,\widetilde{W}^1$ and decompose S_t as

$$
\begin{aligned}
S_t &= s_0 \exp\left((r - \frac{1}{2}\bar{v}_t)t + \rho\int_0^t \sqrt{v_s}dW_s^2\right)\exp\left(\sqrt{1 - \rho^2}\int_0^t \sqrt{v_s}d\widetilde{W}_s^1\right) \\
&= s_0 \exp\left(t\left((r - \frac{\rho ak}{\vartheta}) + \bar{v}_t(\frac{\rho k}{\vartheta} - \frac{1}{2})\right) + \frac{\rho}{\vartheta}(v_t - v_0)\right)\exp\left(\sqrt{1 - \rho^2}\int_0^t \sqrt{v_s}d\widetilde{W}_s^1\right),
\end{aligned}
$$

where we have used the notation $\bar{v}_t = \dfrac{1}{t}\int_0^t v_s ds$ and where we have used the dynamics of (v_t) in the second line. The chaining rule for conditional expectations yields

$$
\mathrm{AsCall}^{Hest}(s_0, K) = e^{-rT}\mathbb{E}\left(\mathbb{E}\left(\left(\frac{1}{T}\int_0^T S_s ds - K\right)_+ |\sigma(W_t^2, 0 \le t \le T)\right)\right).
$$

Combining these two expressions and using that $\widetilde{W}^1$ and W^2 are independent imply that $\mathrm{AsCall}^{Hest}(s_0, K)$ is a functional of $(\widetilde{W}_t^1, v_t)$ (as concerns the squared volatility process v, only v_T and $\bar{v}_T$ are involved).

Let $\Gamma^{N_k} = \{w_1^{N_k}, \dots, w_{N_k}^{N_k}\}$ be an N_k-quantizer of the Brownian motion for $k = 1, 2$ and set the (non-Voronoi) (N_1, N_2)-quantization of (v_t, S_t) by

$$
\begin{aligned}
\widehat{v}_t^{N_1} &= \sum_{i=1}^{N_1} y_i^{N_1}(t)\mathbf{1}_{C_i(\Gamma^{N_1})}(W^2) \\
\widehat{S}_t^{N_1,N_2} &= \sum_{1 \le i \le N_1}\sum_{1 \le j \le N_2} s_{i,j}^{N_1,N_2}(t)\,\mathbf{1}_{C_i(\Gamma^{N_1})}(W^2)\mathbf{1}_{C_j(\Gamma^{N_2})}(\widetilde{W}^1),
\end{aligned}
$$

where $y_i = y_i^{N_1}$ and $s_{i,j} = s_{i,j}^{N_1,N_2}$, for $i = 1, \dots, N_1$ and $j = 1, \dots, N_2$, are the solutions of the following ordinary differential equations:

$$
\frac{dy_i}{dt}(t) = k\left(a - y_i(t) - \frac{\vartheta^2}{4k}\right) + \vartheta\sqrt{y_i(t)}\frac{dw_i^{N_1}}{dt}(t), \quad y_i(0) = v_0, \tag{8.3}
$$

$$
\frac{ds_{i,j}}{dt}(t) = s_{i,j}(t)\left(r - \frac{1}{2}y_i(t) - \frac{\rho\vartheta}{4} + \sqrt{y_i(t)}\left(\rho\frac{dw_i^{N_1}}{dt}(t) + \sqrt{1 - \rho^2}\frac{dw_j^{N_2}}{dt}(t)\right)\right), \quad s_{i,j}(0) = s_0.
$$

These ODE's are solved using e.g. a Runge-Kutta scheme. Note that these formulæ require the computation of the $N_1 \times N_2$ quantized stochastic integrals $\int_0^t \sqrt{y_i^{N_1}(s)}dw_j^{N_2}(s)$ (which corresponds to the independent case). Let us pointed out that it is well known that (8.3) can be solved

more or less in an explicit way for special sets of parameter of the model (when $\theta^2 = 4ak$) as emphasized in (Rogers 1995), in this case, there is no more time integration error and the computations can be made significantly faster (see (Pagès and Printems 2005)). It is not the case of the selected parameters in the present paper (see further on). The weights of the product cells $\{\widetilde{W}^1 \in C_j(\Gamma^{N_2}), W^2 \in C_i(\Gamma^{N_1})\}$ is given by

$$\mathbb{P}(\widetilde{W}^1 \in C_j(w^{N_2}), W^2 \in C_i(w^{N_1})) = \mathbb{P}(\widetilde{W}^1 \in C_j(\Gamma^{N_2}))\mathbb{P}(W^2 \in C_i(\Gamma^{N_1}))$$

owing to the independence. Eventually the premium is approximated by

$$
\begin{aligned}
\widehat{\text{AsCall}}^{Hest}_{N_1,N_2}(s_0, K) &= e^{-rT}\,\mathbb{E}\left(\frac{1}{T}\int_0^T \widehat{S}_t^{N_1,N_2}\,dt - K\right)_+ \\
&= e^{-rT}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{1}{T}\int_0^T s_{i,j}^{N_1,N_2}(t)dt - K\right)_+ \mathbb{P}(\widetilde{W}^1 \in C_j(\Gamma^{N_2}))\mathbb{P}(W^2 \in C_i(\Gamma^{N_1})).
\end{aligned}
$$

We follow the guidelines of the methodology introduced in (Pagès and Printems 2005): we compute the *crude* quantized premium for two sizes (N_1, N_2) and (N_1', N_2'), then proceed a space R-R log-extrapolation where $\log(N)$ (resp. $\log(N')$) is replaced by $\log(N_1 N_2)$ (resp. by $\log(N_1' N_2')$). Finally, we make a K-linear interpolation based on the (Asian) forward moneyness $s_0 e^{rT}(\frac{1-e^{-rT}}{rT}) \approx s_0 e^{rT}$ and the Asian Call-Put parity formula

$$\text{AsianCall}^{Hest}(s_0, K) = \text{AsianPut}^{Hest}(s_0, K) + s_0\left(\frac{1 - e^{-rT}}{rT}\right) - Ke^{-rT}.$$

To be precise, we set with obvious notations

(i) $\qquad \widehat{\text{AsCallParity}}^{Hest}_{N_1,N_2}(s_0, K) = \widehat{\text{AsPut}}^{Hest}_{N_1,N_2}(s_0, K) + s_0\left(\frac{1 - e^{-rT}}{rT}\right) - Ke^{-rT},$

(ii) $\qquad \widehat{\text{AsCallInterp}}^{Hest}_{N_1,N_2} = \dfrac{(K_{max} - K)\widehat{\text{AsCallParity}}^{Hest}_{N_1,N_2} + (K - K_{min})\widehat{\text{AsCall}}^{Hest}_{N_1,N_2}}{K_{max} - K_{min}}.$

The *anchor strikes* $K_{\min}$ and $K_{\max}$ of the interpolation are chosen symmetric with respect to the forward moneyness. At $K_{\max}$, the Call is deep out-of-the-money: we use the R-R extrapolated FQ computation; at $K_{\min}$ the Call is deep in-the-money: on computes the Call by parity. Between $K_{\min}$ and $K_{\max}$, we proceed a linear interpolation in K. This interpolation yields the best results within the range of variance of our model, compared to other extrapolations like the quadratic regression approach. Further comments on this step are made at the end off the section.

- $\circ$ *Parameters of the Heston model*: $s_0 = 100$, $k = 2$, $a = 0.01$, $\rho = 0.5$, $v_0 = 10\%$, $\vartheta = 20\%$.

- $\circ$ *Parameters of the option portfolio*: $T = 1$, $K = 99, \cdots, 111$ (13 strikes).

- $\circ$ *The reference price* has been computed by a Monte Carlo simulation of size $M_{MC} = 10^8$ (including a time R-R extrapolation of the Euler scheme with $2n = 256$).

- $\circ$ *The differential equations (8.3)* are solved with the parameters of the quantization cubature formulae $\Delta t = 1/32$, with couples of quantization levels $(N_1, N_2, N_1', N_2') = (1200, 1200, 100, 100)$, $(3200, 3200, 1100, 1100)$.

Functional Quantization can compute a whole vector (more than 10) option premia for the Asian option in the Heston model with 1 **cent accuracy in less than** 1 **second** (implementation in C on a 2.5 GHz processor).
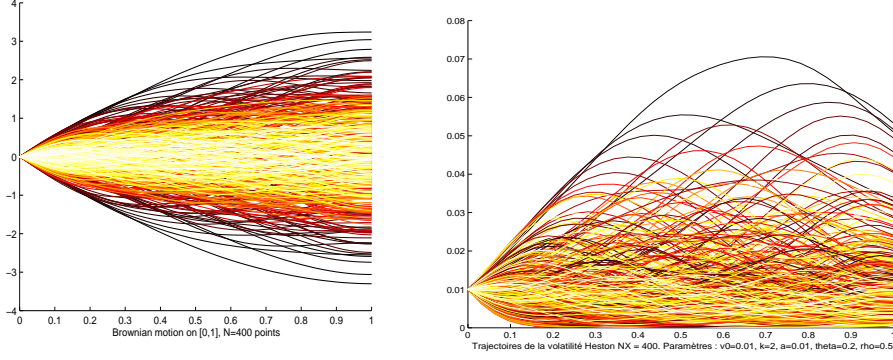
Figure 12: *N-quantizer of the Heston squared volatility process* (v_t) *(N = 400) resulting from an (optimized) N-quantizer of W.*
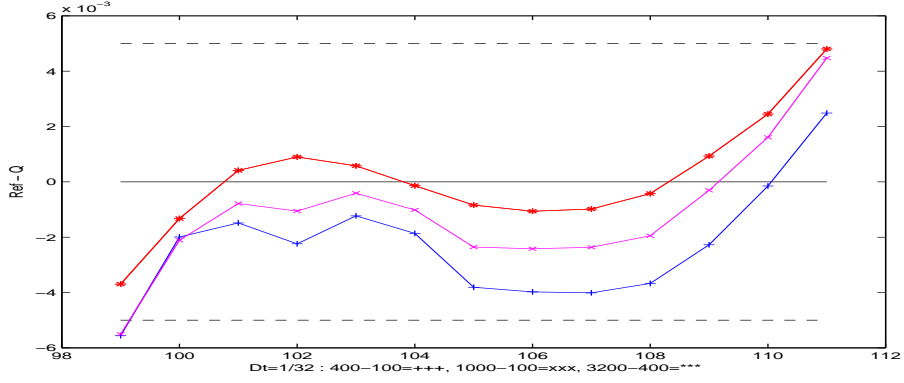

Figure 13: *Quantized diffusions based on optimal functional quantization: Pricing by K-Interpolated-*log*-R-R extrapolated-FQ prices as a function of K: absolute error with* $(N_1, N_2, , N_1', N_2') = (1200, 1200, 100, 100)$ *and* $(3200, 3200, 1100, 1100)$. $T = 1$, $s_0 = 100$, $K \in \{99, \ldots, 111\}$; $k = 2$, $a = 0.01$, $\rho = 0.5$, $\vartheta = 0.2$.
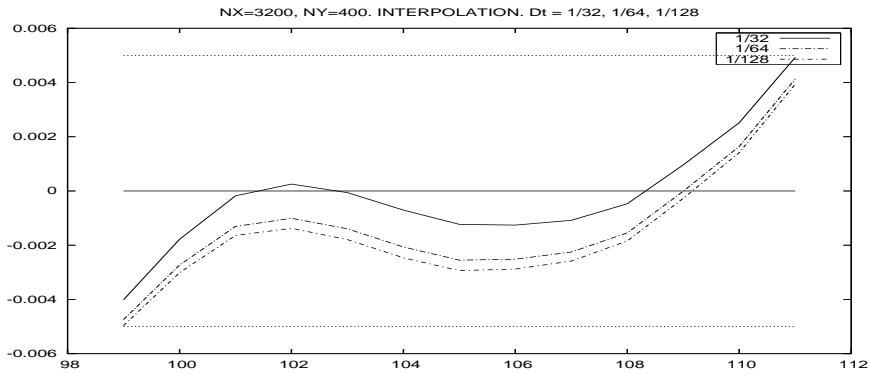

Figure 14: *Quantized diffusions based on optimal functional quantization: Pricing by K-Interpolated-*log*-R-R-extrapolated-FQ price as a function of K: convergence as* $\Delta t \in \{1/8, 1/16, 1/32\}$ *with* $(N_1, N_2, N_1', N_2') = (3200, 3200, 1100, 1100)$ *(absolute error).* $T = 1$, $s_0 = 100$, $K \in \{99, \ldots, 111\}$; $k = 2$, $a = 0.01$, $\rho = 0.5$, $\vartheta = 0.2$.

Further numerical tests carried out or in progress with the B-S or $SABR$ models (vanilla, Asian European options) confirmed the efficiency of the model is not model dependent.

Let us now give some insights about the choice of the couple (N_1, N_2) that has to be done in (8.3). Other numerical experiments suggest that it depends on both the standard deviation $\mathrm{StDev}(S_T)$ of S_T and the sign of ρ. For low values of $\mathrm{StDev}(S_T)$ (let us say 10% of s_0), square quantizers (i.e. $N_1 = N_2$) are relevant. It means that, for a given size $N_1 \times N_2$, a good precision can be obtained using a square quantizer. When $\mathrm{StDev}(S_T)$ increases, further numerical simulations not reproduced here show that this choice of a square quantizer can be optimized. To be precise, we have done numerical simulations with a various 4-tuples (N_1, N_2, N_1', N_2') for two sets of model parameters such that the variance is about of $\bar{S}_T$ is about 200 in the two cases ($s_0 = 100$, $r = 0.1$, $k = 2$, $(a, \rho) \in \{(0.06; -0.6), (0.047; 0.6)\}$). The primes denote the second couple used in the R-R extrapolation.

Then, we have selected the 4-tuples (N_1, N_2, N_1', N_2') giving both "good" Call prices (the reference prices are still computed by a Monte Carlo simulation of size $M_{MC} = 10^8$ and "good" prices means that they stand within one standard deviation interval) around the forward moneyness ($K \in \{95, 100, 105, 110\}$) and have the smallest quantization errors as regards AsCall$-$AsCallParity, $i.e.$

$$\mathbb{E} \int_0^T (S_t - \widehat{S}_t^{N_1, N_2}) \, dt.$$

This second criterion has been chosen since it is more "parameter free" than the option prices themselves.

In order to discuss the results, we set $\alpha := (N_1 - N_2)/(N_1 + N_2)$ and $\alpha' := (N_1' - N_2')/(N_1' + N_2')$ and we displayed in a plane (α, α') the 4-tuples selected above. Two conclusions can then be drawn. The first is that the "good" prices lies in the (α, α') along the diagonal $\alpha' = \alpha$. Not surprisingly, it means that during the R-R extrapolation, the couples (N_1, N_2, N_1', N_2') have to be chosen so that the relative parts of the quantizations of the volatility and the asset remain the same. The second information is that, in the case $\rho < 0$, most 4-tuples lie along the positive part of the diagonal $\alpha \sim \alpha' > 0$ whereas, when $\rho > 0$, most of them lies in the negative part $\alpha \sim \alpha' < 0$. It means that a good choice is $N_1 > N_2$ when $\rho < 0$ and $N_1 < N_2$ when $\rho > 0$.

Let us also remark that the choice of the K-interpolation in (ii) should be tempered when the variance of $\bar{S}_T$ increases. For high variances, a better choice is to use a least square regression based (LSR) variance reduction technique inherited from Monte Carlo method. Namely, we interpolate AsCall and AsCallParity (as given by (ii)) by computing (using quantization)

$$\mathrm{LSR}\widehat{\mathrm{AsCAll}} := \widehat{\mathrm{AsCall}} - \frac{\widehat{\mathrm{Cov}}(\widehat{\mathrm{AsCall}}, \widehat{\mathrm{AsCall}} - \widehat{\mathrm{AsCallParity}})}{\mathrm{Var}(\widehat{\mathrm{AsCall}} - \widehat{\mathrm{AsCallParity}})}(\widehat{\mathrm{AsCall}} - \widehat{\mathrm{AsCallParity}}) \qquad (8.4)$$

given that $\widehat{\mathrm{AsCall}} - \widehat{\mathrm{AsCallParity}} = e^{-rT}\frac{1}{T}\int_0^T S_t \, dt - s_0\frac{1-e^{-rT}}{rT}$ (^{Hest} and the size of the grid are dropped for notational convenience). This approach needs further computations (covariances and variances). When the variance of S_T is not too high, the extremal values of λ at anchor strikes are nearly 0 or 1 respectively and the LSR interpolation coincides with the (slightly faster) K-linear interpolation.

8.3 Comparison: optimized quantization vs (optimal) product quantization

The comparison is balanced and probably needs some further $in\ situ$ experiments since it may depend on the modes of the computation. However, it seems that product quantizers (as those implemented in (Pagès and Printems 2005)) are from 2 up to 4 times less efficient than optimal

K	95	100	105	110
Reference price	11.200	7.948	5.244	3.170
$\widehat{LSRAsCall}^{Hest}$	11.204 (0.036%)	7.947 (0.012%)	5.246 (0.038%)	3.172 (0.063%)

Table 1: *Asian Heston prices with the parameters : $s_0 = 100$, $r = 0.1$, $k = 2$, $v_0 = a = 0.06$, $\theta = 0.5$ and $\rho = -0.6$ and where K here denotes the strikes prices. Here $(N_1, N_2, N_1', N_2') = (1700, 800, 2000, 900)$, i.e. $\alpha \sim \alpha' \sim 0.36$. The final prices have been computing using formula (8.4).*
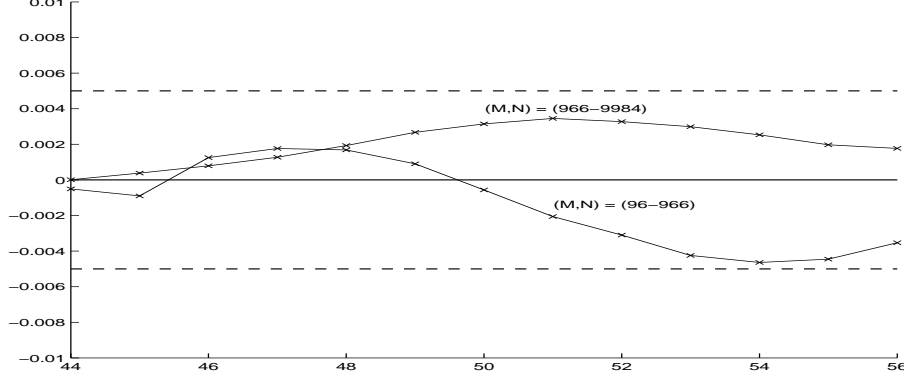

Figure 15: *Quantized diffusions based on optimal product quantization: Pricing by K-linear interpolation of R-R log-extrapolations as un function of K (absolute error) with $(M, N) = (96, 966)$, $(966, 9984)$. $T = 1$, $s_0 = 50$, $K \in \{44, \ldots, 56\}$; $k = 2$, $a = 0.01$, $\rho = 0.5$, $\vartheta = 0.1$. .*

quantizers within our range of application (small values of N). On the other hand, the design of product quantizer from 1-dim scalar quantizers is easy and can be made from some light elementary "bricks" (the scalar quantizer up to $N = 35$ and the optimal allocation rules). Thus, the whole set of data needed to design all optimal product quantizers up to $N = 10\,000$ is approximately 500 KB whereas one optimal quantizer with size $10\,000 \approx 1\ MB \ldots$

9 Universal quantization rate and mean regularity

The following theorem points out the connection between functional quantization rate and mean regularity of $t \mapsto X_t$ from $[0, T]$ to $L^r(\mathbb{P})$.

Theorem 9.1 *(Luschgy and Pagès 2007) Let $X = (X_t)_{t \in [0,T]}$ be a stochastic process. If there is $r^* \in (0, \infty)$ and $a \in (0, 1]$ such that*

$$X_0 \in L^{r^*}(\mathbb{P}), \quad \|X_t - X_s\|_{L^{r^*}(\mathbb{P})} \leq C_X |t - s|^a,$$

for some positive real constant $C_X > 0$, then

$$\forall\, p, r \in (0, r^*), \quad e_{N,r}(X, L_T^p) = O((\log N)^{-a}).$$

The proof is based on a constructive approach which involves the Haar basis (instead of K-L basis), the non asymptotic version Zador Theorem and product functional quantization. Roughly speaking, we use the unconditionality of the Haar basis in every L_T^p (when $1 < p < \infty$) and its wavelet feature *i.e.* its ability to "code" the path regularity of a function on the decay rate of its coordinates.

EXAMPLES (see (Luschgy and Pagès 2007)): • d-dimensional Itô processes (includes d-dim diffusions with sublinear coefficients) with $a = 1/2$.
• General Lévy process X with Lévy measure ν with square integrable big jumps. If X has a Brownian component, then $a = 1/2$, otherwise if $\beta(X) > 0$ where $\beta(X) := \inf\{\theta : \int |y|^\theta \nu(dy) < +\infty\} \in (0, 2)$ (Blumenthal-Getoor index of X), then $a = 1/\beta(X)$. This rate is the *exact rate i.e.*

$$e_{N,r}(X, L_T^p) \approx (\log N)^{-a}$$

for many classes of Lévy processes like symmetric stable processes, Lévy processes having a Brownian component, etc (see (Luschgy and Pagès 2007) for further examples).
• When X is a compound Poisson processes, then $\beta(X) = 0$ and one shows, still with constructive methods, that

$$e_N(X) = O(e^{-(\log N)^\vartheta}), \qquad \vartheta \in (0, 1),$$

which is in-between the finite and infinite dimensional settings.

10 About lower bounds for functional quantization

In this overview, we gave no clue toward lower bounds for FQ although most of the rates we mentioned are either exact ($\approx$) or sharp ($\sim$) (we tried to emphasize the numerical aspects). Several approaches can be developed to get some lower bounds. Historically, the first one was to rely on subadditivity property of the quantization error derived from self-similarity of the distribution: this works with the uniform distribution over $[0, 1]^d$ but also in an infinite dimensional framework (see *e.g.* (Dereich et al. 2006) for the fractional Brownian motion).

A second approach consists in pointing out the connection with the Shannon-Kolmogorov entropy (see *e.g.* (Luschgy and Pagès 2002)) using that the entropy of a random variable taking at most N values is at most $\log N$.

A third connection can be made with small deviation theory (see (Dereich et al. 2003), (Graf et al. 2003) and (Luschgy and Pagès 2007)). Thus, in (Graf et al. 2003), a connection is established between (functional) quantization and small ball deviation for Gaussian processes. In particular this approach provides a method to derive a lower bound for the quantization rate from some upper bound for the small deviation problem. A careful reading of the proof of Theorem 1.2 in (Graf et al. 2003) shows that this small deviation lower bound holds for any *unimodal* (w.r.t. 0) non zero process. To be precise: assume that $\mathbb{P}_X$ is L_T^p-unimodal *i.e.* there exists a real $\varepsilon_0 > 0$ such that

$$\forall x \in L_T^p, \ \forall \varepsilon \in (0, \varepsilon_0], \qquad \mathbb{P}(|X - x|_{L_T^p} \le \varepsilon) \le \mathbb{P}(|X|_{L_T^p} \le \varepsilon).$$

For centered Gaussian processes (or processes "subordinated" to Gaussian processes) this follows from the Anderson Inequality (when $p \ge 1$). If

$$G(-\log(\mathbb{P}(|X|_{L_T^p} \le \varepsilon))) = \Omega(1/\varepsilon) \quad \text{as} \quad \varepsilon \to 0$$

for some increasing unbounded function $G : (0, \infty) \to (0, \infty)$, then

$$\forall c > 1, \quad \liminf_N G(\log(cN)) e_{N,r}(X, L_T^p) > 0, \qquad r \in (0, \infty). \tag{10.1}$$

This approach is efficient in the non quadratic case as emphasized in (Luschgy and Pagès 2007) where several universal bounds are shown to be optimal using this approach.

11 Toward new applications: a guided Monte Carlo method.

This section provides some preliminary (and theoretical) elements about a quantization based stratification method to reduce the variance of a Monte Carlo simulation. It can be seen as a guided Monte Carlo method or a hybrid Quantization/Monte Carlo method. This method has been introduced in (Pagès and Printems 2005) for Lipschitz functionals of the Brownian motion. Here we will mainly focus on the finite dimensional case and consider some more regular functions. Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ be square integrable random vector. For more details and some simulation results, we refer to (Pagès and Printems 2007).

$\triangleright$ *Lipschitz functions.* Let $F : \mathbb{R}^d \to \mathbb{R}^d$ be a Lipschitz function. In order to compute $\mathbb{E}(F(X))$, one writes

$$
\begin{aligned}
\mathbb{E}(F(X)) &= \mathbb{E}(F(\widehat{X}^N)) + \mathbb{E}\left(F(X) - F(\widehat{X}^N)\right) \\
&= \underbrace{\mathbb{E}(F(\widehat{X}^N))}_{(a)} + \underbrace{\frac{1}{M}\sum_{m=1}^{M} F(X^{(m)}) - F(\widehat{X^{(m)}}^N)}_{(b)} + R_{N,M}
\end{aligned}
\tag{11.1}
$$

where $X^{(m)}$, $m = 1, \ldots, M$ are M independent copies of the standard Brownian motion, $\widehat{}^N$ denotes the nearest neighbour projection on a fixed N-quantizer and $R_{N,M}$ is a remainder term defined by (11.1). Term (a) can be computed by quantization and Term (b) can computed by a Monte Carlo simulation. Then,

$$
\|R_{N,M}\|_2 = \frac{\sigma(F(X) - F(\widehat{X}^N))}{\sqrt{M}} \leq \frac{\|F(X) - F(\widehat{X}^N)\|_2}{\sqrt{M}} \quad \text{and} \quad \sqrt{M}\, R_{N,M} \xrightarrow{\mathcal{L}} \mathcal{N}(0; \mathrm{Var}(F(X) - F(\widehat{X}^N)))
$$

as $M \to +\infty$ where $\sigma(Y)$ denotes the standard deviation of a random variable Y. Consequently, if F is simply a Lipschitz function and if $(\widehat{X}^N)_{N \geq 1}$ is a rate optimal sequence quantization of X, then

$$
\|F(X) - F(\widehat{X}^N)\|_2 \leq \frac{[F]_{\mathrm{Lip}} C_X}{N^{\frac{1}{d}}} \quad \text{and} \quad \|R_{N,M}|_{L^2_T}\|_2 \leq \frac{[F]_{\mathrm{Lip}} C_X}{M^{\frac{1}{2}} N^{\frac{1}{d}}}.
$$

The main gap to implement this variance reduction method is the nearest neighbour search needed at each step to simulate $\widehat{X^{(m)}}^N$ from $X^{(m)}$ (assuming that the quantizer is known as well as its weights). If one uses a product quantization, this search reduces to the marginals and the complexity of a nearest number search on the real line based on dichotomy is approximately $\frac{\log N}{\log 2}$ (once the N points of interest are sorted). As concerns non-product quantizations fast nearest neighbour search procedures can also be implemented (if N is not too small, see (Friedman et al. 1977)).

$\triangleright$ *Quantization based stratification.* This paragraph sums up an that will be fully idea developed and tested in (Pagès and Printems 2007). In many natural situations (*e.g.* $d = 1$ or $d \geq 2$ with a product quantizer X^N), one has an explicit expression for the conditional distribution of X given $\widehat{X}^N = \mathrm{Proj}_{\Gamma_N}(X)$ that is, we can write

$$
X = \varphi(\widehat{X}^N, U)
$$

where U is independent of $\widehat{X}^N$ with a distribution $\mu := \mathbb{P}_U$ on $\mathbb{R}^q$ and $\varphi : \Gamma_N \times \mathbb{R}^q \to \mathbb{R}^d$ a Borel function. The probability distribution μ is supposed to be easy to simulate and φ easy to compute. In fact, from a theoretical viewpoint, one may always assume that U is uniformly distributed on a unit hypercube $[0, 1]^q$ (or even the unit interval $[0, 1]$). Then, one derives that for every Borel

function $F : \mathbb{R}^d \to \mathbb{R}$ in $L^1(\mathbb{P}_X)$,

$$
\begin{aligned}
\mathbb{E}\, F(X) &= \mathbb{E}\left(\mathbb{E}\left(F(\varphi(\widehat{X}^N, U)) \,|\, U\right)\right) \\
&= \mathbb{E}(\Phi_F(U))
\end{aligned}
\tag{11.2}
$$

with

$$
\Phi_F(u) := \mathbb{E}(F(X)\,|\, U = u) = \sum_{i=1}^{N} \mathbb{P}(\widehat{X}^N = x_i) F(\varphi(x_i, u)), \quad u \in [0,1]^q
$$

since $\widehat{X}^N$ and U are independent. Then, one derives that

$$
\begin{aligned}
\mathrm{Var}(\Phi_F(U)) &= \mathrm{Var}(\Phi_F(U) - \mathbb{E}\, F(\widehat{X}^N)) \\
&= \mathrm{Var}(\mathbb{E}\left(F(\varphi(\widehat{X}^N, U)) - F(\widehat{X}^N)\,|\, U\right)) \\
&\leq \mathrm{Var}(F(X) - F(\widehat{X}^N))
\end{aligned}
\tag{11.3}
$$

where we used that conditional expectation is an $L^2(\mathbb{P})$-contraction, preserves expectation and again that $\widehat{X}^N$ and U are independent. Furthermore the *above last inequality is strict* except if $F(X) = F(\widehat{X}^N) + G(U)$. As expected this stratification produces an additional variance reduction with respect to the above regular hybrid Monte Carlo-quantization method.

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. 2007a] BARDOU, O., BOUTHEMY, S. AND PAGÈS, G. [2007]: Pricing swing options using optimal quantization, pre-print LPMA-1146, submitted.

[Bardou et al. 2007b] BARDOU, O., BOUTHEMY, S. AND PAGÈS, G. [2007]: When are swing option bang-bang and how to use it, pre-print LPMA-1141, submitted.

[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.

[Bouleau and Lépingle 1994] BOULEAU, N., LÉPINGLE, D. [1994]: *Numerical methods for stochastic processes*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 359 pp. ISBN: 0-471-54641-0.

[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): A geometric method for uniqueness of locally optimal quantizer. Preprint LPMA-464 and Ph.D. Thesis, *Sur quelques problèmes de quantification*, 2000, Univ. Paris 6.

[Cuesta-Albertos and Matrán 1988] CUESTA-ALBERTOS, J.A., MATRÁN, C. [1988]: The strong law of large numbers for k-means and best possible nets of Banach valued random variables, *Probab. Theory Rel. Fields*, **78**, 523-534.

[Delattre et al. 2004] DELATTRE, S., FORT, J.-C. AND PAGÈS, G. [2004]: Local distortion and μ-mass of the cells of one dimensional asymptotically optimal quantizers, *Communications in Statistics*, **33**(5), 1087-1118.

[Dereich et al. 2003] DEREICH, S., FEHRINGER, F., MATOUSSI, A. AND SCHEUTZOW, M. [2003]: On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces, *J. Theoretical Probab.*, **16**, pp.249-265.

[Dereich 2005a] DEREICH, S. [2005]: The coding complexity of diffusion processes under $L^p[0,1]$-norm distortion, pre-print, to appear in *Stoch. proc. and their Appl.*.

[Dereich 2005b] DEREICH, S. [2005]: The coding complexity of diffusion processes under supremum norm distortion, pre-print, to appear in *Stoch. proc. and their Appl.*.

[Dereich et al. 2006] DEREICH, S., SCHEUTZOW, M. [2006]: High resolution quantization and entropy coding for fractional Brownian motions, *Electron. J. Probab.*, **11**, 700-722.

[Doss 1977] DOSS, H. [1977]: Liens entre équations différentielles stochastiques et ordinaires, *Ann. I.H.P.*, section B, **13**(2), 99-125.

[Fleischer 1964] FLEISCHER, P.E. [1964]: Sufficient conditions for achieving minimum distortion in a quantizer. *IEEE Int. Conv. Rec.*, part I, 104-111.

[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.

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

[Fort and Pagès 2004] FORT, J.-C. AND PAGÈS, G. [2004]: Asymptotics of optimal quantizers for some scalar distributions, *Journal of Computational and Applied Mathematics*, **146**, 253-275, 2002.

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

[Graf and Luschgy 2005] GRAF, S. AND LUSCHGY, H. [2005]: The point density measure in the quantization of self-similar probabilities. *Math. Proc. Cambridge Phil. Soc.*. **138**, 513-531.

[Graf et al. 2003] GRAF, S., LUSCHGY H. AND PAGÈS, G. [2003]: Functional quantization and small ball probabilities for Gaussian processes, *J. Theoret. Probab.*, **16**[(4), 1047-1062.

[Graf et al. 2006] GRAF, S., LUSCHGY, H. AND PAGÈS, G. [2006]: Distortion mismatch in the quantization of probability measures, pre-pub PMA. to appear in *ESAIM P&S*.

[Graf et al. 2007] GRAF, S., LUSCHGY, H., PAGÈS, G. [2007]: Optimal quantizers for Radon random vectors in a Banach space, *J. of Approximation Theory*, **144**, 27-53.

[Heston 1993] HESTON, S.L. [1993]: A closed-form solution for options with stochastic volatility with applications to bond and currency options, *The review of Financial Studies*, **6**(2), 327-343.

[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.

[Lamberton and Pagès 1996] LAMBERTON, D. AND PAGÈS, G. [1996]: On the critical points of the 1-dimensional Competitive Learning Vector Quantization Algorithm. *Proceedings of the ESANN'96*, (ed. M. Verleysen), Editions D Facto, Bruxelles, 97-106.

[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.

[Lejay 2003] LEJAY, A. [2003]: An introduction to rough paths, *Séminaire de Probabilités XXXVII*, Lecture Notes in Mathematics 1832, Springer, Berlin, 1-59.

[Luschgy and Pagès 2002] LUSCHGY, H., PAGÈS, G. [2002]: Functional quantization of Gaussian processes, *Journal of Functional Analysis*, **196**(2), 486-531.

[Luschgy and Pagès 2004] LUSCHGY, H., PAGÈS, G. [2004]: Sharp asymptotics of the functional quantization problem for Gaussian processes, *The Annals of Probability*, **32**(2), 1574-1599.

[Luschgy and Pagès 2006] LUSCHGY, H., PAGÈS, G. [2006]: Functional quantization of a class of Brownian diffusions: A constructive approach, *Stochastic Processes and Applications*, **116**, 310-336.

[Luschgy and Pagès 2007] LUSCHGY, H., PAGÈS, G. [2007]: High-resolution product quantization for Gaussian processes under sup-norm distortion, *Bernoulli*, **13**(3), 653-671.

[Luschgy and Pagès 2007] LUSCHGY, H., PAGÈS, G. [2006]: Functional Quantization Rate and mean regularity of processes with an application to Lévy Processes, pre-print LPMA-1048, to appear in *Annals of Applied Probability*.

[Luschgy and Pagès 2007] LUSCHGY, H., PAGÈS, G. [2007]: Expansion of Gaussian processes and Hilbert frames, in progress.

[Luschgy et al. 2007] LUSCHGY, H., PAGÈS, G. AND WILBERTZ, B. [2007]: Asymptotically optimal quantization schemes for Gaussian processes, technical report.

[McNames 2001] MCNAMES, J. [2001]: J. McNames. 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*.

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

[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 2000] PAGÈS, G. [2000]: Functional quantization: a first approach, pre-print CMP12-04-00, Univ. Paris 12 (France).

[Pagès 2007] Pagès, G. [2007]: Quadratic optimal functional quantization methods and numerical applications, *Proceedings of MCQMC, Ulm'06*, Springer, Berlin, 101-142.

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

[Pagès and Printems 2005] Pagès, G., Printems, J. [2005]: Functional quantization for numerics with an application to option pricing, *Monte Carlo Methods and Appl.*, **11**(4), 407-446.

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

[Pagès and Printems 2007] Pagès, G., Printems, J. [2007]: A hybrid Quantization - Monte Carlo method, in progress.

[Pagès and Sagna 2007] Pagès, G. and Sagna, A. [2007]: Asymptotics of the radius of an L^r-optimal sequence of quantizers, prep-pub. LPMA, Univ. Paris 6 (France), submitted.

[Pagès and Sellami 2007] Pagès, G., Sellami, A. [2007]: Convergence of multi-dimensional quantized SDE's. Pré-print LPMA-1196.

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

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

[Proinov 1982] Proinov, P.D. [1988]. Discrepancy and integration of continuous functions, *J. of Approx. Theory*, **52**, 121-131.

[Revuz and Yor 1999] Revuz, D., Yor, M. [1999]: *Continuous martingales and Brownian motion*, Third edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], **293**, Springer-Verlag, Berlin, 1999, 602 p.

[Rogers 1995] Rogers L.C.G. (1995), Which model for the term structure of interest rates should one use?, *Mathematical Finance, IMA* **65**, 93-116.

[Sagna 2007] Sagna, A. [2007]: Universal L^s-rate-optimality of L^r-optimal quantizers by dilatation-contraction, prep-pub. LPMA, Univ. Paris 6 (France), to appear in *ESAIM P&S*.

[Tarpey et al. 2003a] Tarpey, T., Kinateder, K.K.J. [2003]: Clustering functional data, *J. Classification*, **20**, 93-114.

[Tarpey et al. 2003b] Tarpey, T., Petkova, E., Ogden, R.T. [2003]: Profiling Placebo responders by self-consistent partitioning of functional data, *J. Amer. Statist. Association*, **98**, 850-858.

[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.