Optimal quadratic quantization for numerics: the Gaussian case

Gilles PAGÈS* Jacques PRINTEMS†

Abstract

Optimal quantization has been recently revisited in multi-dimensional numerical integration (see [18]), multi-asset American option pricing (see [2]), control theory (see [19]) and nonlinear filtering theory (see [20]). In this paper, we enlighten some numerical procedures in order to get some accurate optimal quadratic quantization of the Gaussian distribution in one and higher dimensions. We study in particular Newton method in the deterministic case (dimension $d = 1$) and stochastic gradient in higher dimensional case ($d \geq 2$). Some heuristics are provided which concern the step in the stochastic gradient method. Finally numerical examples borrowed from mathematical finance are used to test the accuracy of our Gaussian optimal quantizers.

Keywords: Optimal quantization, stochastic gradient methods, numerical integration.

AMS Classification (2000): 94A29 (Secondary: 62L20, 65D30, 65D32, 90C59, 90C52 91B28).

1 Introduction

Although optimal quantization has been extensively investigated for more than fifty years in fields such as Signal Processing and Information Theory (see [11, 13]), it has been recently revisited in the field of Numerical Probability for numerical integration in high dimension (see [18]), multi-asset American option pricing (see [2, 1, 3, 4]) but also in Control Theory (see [19]) and Nonlinear Filtering Theory (see [20])(see also [21] for a survey of applications of optimal quantization methods in finance). In all these fields of application, the access to some very accurate approximation of optimal quantization is crucial. This access has been made possible by the increasing power of modern computers: one can now massively process on a standard personal computer some numerical methods based on massive probabilistic simulation. The most popular one being the regular Monte Carlo method. The aim of this paper is to enlighten the numerical procedures used to get optimal quadratic quantization of random vectors, with a special emphasis on Gaussian vectors.

Let X be a random vector on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking its values in $\mathbb{R}^d$. We denote by $\mathbb{P}_X$ its distribution on $\mathbb{R}^d$. Quantization consists in approximating X by a random vector $q(X)$ taking finitely many values in $\mathbb{R}^d$. Let $q(\mathbb{R}^d) = \{x_1, \ldots, x_N\}$. Among all Borel functions taking their values in the set $\{x_1, \ldots, x_N\}$, one specifies the so-called

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

†INRIA, MathFi project and Centre de Mathématiques, CNRS UMR 8050, Université Paris 12, 61, av. du Général de Gaulle, F-94010 Créteil. E-mail: printems@univ-paris12.fr

Voronoi N-quantizers defined by

$$q_{vor}(\xi) = \sum_{i=1}^{N} x_i \, \mathbf{1}_{C(x_i)}(\xi), \qquad \xi \in \mathbb{R}^d,$$

where $\{C(x_i)\}_{1\le i\le N}$ is a Borel partition of $\mathbb{R}^d$ satisfying

$$C(x_i) \subset \{\xi \in \mathbb{R}^d \mid \ |\xi - x_i| \ \le \ |\xi - x_j|, \ j = 1, \ldots, N\}.$$

Let $p \ge 1$ and $X \in L^p$. One easily checks that these Voronoi N–quantizers minimize the L^p quantization error (to the power p), i.e.

$$\mathbb{E}\,|X - q_{vor}(X)|^p = \min\left\{\mathbb{E}\,|X - q(X)|^p, \ q : \mathbb{R}^d \overset{Borel}{\longrightarrow} \{x_1, \ldots, x_N\}\right\}.$$

From now on, we will only consider Voronoi N–quantizers (and so we will often drop the "Voronoi" term). For these Voronoi N–quantizers, the L^p–error (to the power p) induced by replacing X by its quantizer $q(X)$ reads

$$\mathbb{E}|X - q_{vor}(X)|^p = \int_{\mathbb{R}^d} \min_{1\le i\le N} |x_i - \xi|^p \mathbb{P}_X(\mathrm{d}\xi) \tag{1.1}$$

The right-hand-side of the above equality defines a (symmetric) continuous function $x \mapsto D_N^{X,p}(x)$ on $(\mathbb{R}^d)^N$ of the variable $x := (x_1, \ldots, x_N)$. Such a N-tuple x will often be called N–quantizer as well. The aim of *L^p-optimal quantization* is to find some *optimal quantizer* which minimizes the function $D_N^{X,p}$ over $(\mathbb{R}^d)^N$ (there is always some, see, *e.g.* [13]). Optimal *quadratic* quantization, on which we focus in this paper, stands for $p = 2$.

Let us illustrate by a simple example an application of optimal quantization to numerical integration: one can write, for a regular enough function f and a quantizer $x := (x_1, \ldots, x_N)$:

$$\begin{aligned} \mathbb{E}\, f(X) \ &= \ \sum_{i=1}^{N} \mathbb{P}_X(C(x_i)) f(x_i) + \sum_{i=1}^{N} df(x) \cdot \mathbb{E}\left((x_i - X)\mathbf{1}_{C(x_i)}(X)\right) \\ &\quad + 2^{nd} \text{ order terms.} \end{aligned} \tag{1.2}$$

The first sum in the right-hand-side of the equality can be easily computed provided one knows the x_i's and the $\mathbb{P}_X$-"mass" of their Voronoi cells. Then, one can see that, when for $p = 2$, the first order necessary condition for optimality in (1.1) implies that all the terms $\mathbb{E}((x_i - X)\mathbf{1}_{C(x_i)}(X))$, $i = 1, \ldots, n$, are 0. This improves the numerical accuracy of the approximation of $\mathbb{E}(f(X))$.

In many cases where the random vector X of interest in (1.2) is the d-dimensional Brownian motion B_T at some positive time T (*e.g.* the pricing of an European option in the Black and Scholes model), the crucial step amounts, *modulo* an appropriate dilatation, to optimally quantize the Normal distribution $\mathcal{N}(0; I_d)$. The aim of this paper is to describe in full details some numerical procedures performing optimal quadratic quantization of Gaussian random vectors. We mean by that to give some heuristics concerning efficient choices for the parameters in different gradient-based optimization algorithms proposed to minimize (1.1): Newton's method (in one dimension), a fixed point-like method known as Lloyd's method I (see [14]) and stochastic gradient method (see [8]).

Stochastic gradient methods are based on the integral representation of the gradient of the criterion to be minimized (this is the case of the criterion $D_N^{X,2}$ defined by (1.1)). The

rate of convergence of stochastic gradient methods is ruled by a Central Limit Theorem (CLT). The rate of convergence of stochastic gradient descents is ruled by a Central Limit Theorem (CLT). When the descent step of the procedure is settled to provide the best possible rate, then the variance in the CLT is proportional to the inverse of the lowest eigenvalue of the Hessian $d^2D_N^{X,2}(x^\star)$ at the limiting value $x^\star$.

Hence, we can see that the ill-conditioned nature of $d^2D_N^{X,2}(x^\star)$ is linked to the slowness of the stochastic algorithm. One verifies that this is a crucial problem in practical implementations of such stochastic gradient procedures. This is the reason why we first studied the case of the uniform distribution $U([0,1])$ over the unit interval for which everything can be computed analytically:

$$\min_{x\in(\mathbb{R}^d)^N} D_N^{X,2}(x) = D_N^{X,2}(x^\star) = \frac{1}{2N} \quad \text{with } x^\star = \left(\frac{2k-1}{2N}\right)_{1\le k\le N},$$

and $d^2D_N^{X,2}(x^\star)$ is, up to a normalizing factor, the three points-discretized Laplacian operator which is known to be ill-conditioned. This tells us that the uniform law is some sense the most difficult case for the numerical experiments. When dealing with more general distributions, this is a hint to explain and overcome the numerical difficulties encountered to compute the components of an optimal quantizer close to the modes of the distribution: around these modes, the distribution behaves locally as the uniform distribution. From this study, we will be in position to derive some heuristics concerning the descent step in the stochastic gradient including in higher dimension for the Normal distribution (see Section 3).

The paper is organized as follows. In Section 2, after some definitions, we recall in Theorem 2.1 the asymptotic bound concerning the infimum in (1.1) when N becomes large. Then we recall general facts about the stochastic gradient algorithm and give necessary conditions of convergence in Theorem 2.4 (see [8]). In Section 3, we proceed to the numerical implementation of Newton's Method for the one dimensional case and stochastic gradient in higher dimension (up to 10). In Section 4, we propose some numerical experiments with an example borrowed to mathematical finance. It consists in pricing Put and Put-Spread European options on a geometrical index of Black & Scholes assets using some optimal quadratic quantizers of a d-dimensional Normal distribution for $d \in \{2,\dots,6\}$. This is based upon the above formula (1.2). Its main purposes are to test from a numerical point of view the accuracy of the optimal quantizer obtained in Section 3. Subsequently, it is a way to validate our heuristics concerning the different optimization procedures depicted in Section 3. To this end, we carry out in Section 4 a short comparison with the Monte Carlo method. Several classes of functions are involved depending on their convexity structure and their smoothness. Indeed, as pointed out in Section 2, numerical integration of convex function via optimal quantizer yields a lower bound of the true value. That is why numerical integration of the difference of two convex functions via optimal quantization must yield a better accuracy. Our numerical experiments tend to show that being the difference of two convex functions is more prominent than smoothness. Moreover, in this case, the numerical integration *via* optimal quantization leads to good results both in terms of relative error in percentage and in term of absolute error when we compare it with the standard deviation of the Monte Carlo estimator. In fact, it successfully competes with the Monte Carlo method up to 4-dimension as predicted by theoretical error bounds and seems quite satisfactory even in 5-dimension. Nevertheless, we emphasize that the purpose of this section is essentially to test the accuracy of the optimal quantizer. It is clear that, as far as high dimensional numerical integration is concerned, say $d \ge 6$, Monte Carlo method is especially relevant

when we want to balance accuracy with computational cost. The natural field of application of the quantization method is the computation of a huge number of integrals of regular functions with respect to the same distribution, in medium dimensions (say $1 \leq d \leq 4$ or $d = 5$).

2 Notations and preliminaries

We denote by $|\cdot|$ the Euclidean norm on $\mathbb{R}^d$ and for every Borel set $A \subset \mathbb{R}^d$, we denote by $\mathbf{1}_A$ its indicator function.

2.1 Quantization of random vector

Let X be a square integrable $\mathbb{R}^d$-valued random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $x := (x_i)_{1\leq i\leq N}$ be a N-tuple in $\mathbb{R}^d$ and let $q : \mathbb{R}^d \to \{x_1, \ldots, x_N\}$ be any Borel function. The $\{x_1, \ldots, x_N\}$-valued random vector $q(X)$ is called a *q-quantization* of X. The induced quadratic error $\|X - q(X)\|_2$ is called (quadratic) q-quantization error.

One easily shows that, among all possible $\{x_1, \ldots, x_N\}$-valued functions q, all those defined by

$$q_{vor}(\xi) := \sum_{i=1}^{N} x_i \mathbf{1}_{C(x_i)}(\xi), \qquad \xi \in \mathbb{R}^d,$$

where $\{C(x_i)\}_{1\leq i\leq N}$ is a Borel partition of $\mathbb{R}^d$ satisfying

$$C(x_i) \subset \{\xi \in \mathbb{R}^d \mid \ |\xi - x_i| \leq |\xi - x_j|,\ j = 1, \ldots, N\}$$

minimize the quadratic quantization error. That is

$$\|X - q_{vor}(X)\|_2 = \min\left\{\|X - q(X)\|_2,\ q : \mathbb{R}^d \overset{Borel}{\longrightarrow} \{x_1, \ldots, x_N\}\right\}.$$

Any such partition $\{C(x_i)\}_{1\leq i\leq N}$ of $\mathbb{R}^d$ is called a *Voronoi tessellation* of the N-tuple x and the corresponding function q_{vor} a *Voronoi N-quantizer*. When all the components of the N-tuple x are pairwise distinct, each *cell* $C(x_i)$ contains x_i, its closure is convex and its boundary is included in finite union of hyperplanes. Any q_{vor}-quantization of X where q_{vor} is a Voronoi N-quantizer is called a *Voronoi N-quantization* of X. It is denoted $\widehat{X}^x$ (or simply $\widehat{X}$ when there is no ambiguity). For notational simplicity the N-tuple x itself will often be called (Voronoi) N-quantizer. So, such a Voronoi quantization reads

$$\widehat{X}^x := \sum_{i=1}^{N} x_i \mathbf{1}_{C(x_i)}(X).$$

The resulting quadratic quantization error, to the power 2, that is $\mathbb{E}\,|X - \widehat{X}^x|^2$, is called *quadratic distortion* (this terminology comes from Information Theory and Signal processing and was developed in the early 1950's) and is denoted $D_N^X(x)$. If $\mathbb{P}_X$ denotes the distribution of X, it reads

$$\begin{aligned} D_N^X(x) &:= \mathbb{E}|X - \widehat{X}^x|^2 = \sum_{i=1}^{N} \mathbb{E}(\mathbf{1}_{C(x_i)}(X)|X - x_i|^2) \\ &= \int_{\mathbb{R}^d} \min_{1\leq i\leq N} |x_i - \xi|^2 \, \mathbb{P}_X(\mathrm{d}\xi). \end{aligned}$$

The notation is consistent since the distortion only depends on the N-tuple x and (the distribution of) X. Furthermore, when $\mathbb{P}_X$ is continuous, the Voronoi quantization $\widehat{X}^x$ itself is $\mathbb{P}$-essentially unique.

One crucial feature is that the distortion function $x \mapsto D_N^X(x)$ is continuous, and always reaches (at least) one minimum, at some N-tuple x^* having pairwise distinct components. Let us denote

$$\underline{D}_N^X := \min_{x \in (\mathbb{R}^d)^N} D_N^X(x).$$

Such an optimal quantizer lies in the convex hull of the support of $\mathbb{P}_X$. Furthermore, it is easy to establish that this minimum $\underline{D}_N^X$ decreases to 0 as the size N of the optimal quantizer goes to infinity (see *e.g.* [13, 18] for a proof of these basic properties). The rate of convergence to 0 is a more challenging problem, elucidated in several steps by Zador, Bucklew & Wise and finally Graf & Luschgy (see [13]). It is given by the following theorem.

Theorem 2.1 *Assume $X \in L^{2+\varepsilon}(\Omega, \mathcal{A}, \mathbb{P})$ for some $\varepsilon > 0$. Set $\varphi := \frac{d\mathbb{P}_X}{d\lambda_d}$ the Radon-Nikodym density of the absolutely continuous part of $\mathbb{P}_X$ with respect to the Lebesgue measure λ_d on $\mathbb{R}^d$. Then*

$$\lim_N N^{2/d} \underline{D}_N^X = J_d \|\varphi\|_{\frac{d}{d+2}} \tag{2.1}$$

where $\|\varphi\|_r = (\int_{\mathbb{R}^d} |\varphi|^r d\lambda_d)^{1/r}$ for any $r > 0$. In particular J_d is the limit when $X \sim U([0,1]^d)$ and satisfies $J_d = \min_N N^{2/d} \underline{D}_N^{U([0,1]^d)}$.

The true value of J_d is unknown when $d \geq 3$ but one knows that $J_d \sim \frac{d}{2\pi e}$ ($J_1 = \frac{1}{12}$ and $J_2 = \frac{5}{18\sqrt{3}}$) (see [13]).

It is of high interest to have access to a N-tuple x^* with minimal possible distortion since it provides the best possible quadratic approximation of a random vector X by a random vector taking (at most) N values. This is the purpose of *optimal quantization* which will need in higher dimension to use stochastic procedure of optimization exposed below.

But before getting into these optimization procedures, let us illustrate on a simple example how quantization of random vectors can be used for numerics, namely numerical integration.

2.2 Numerical integration by quantization

The idea is simply to approximate the distribution $\mathbb{P}_X$ on $\mathbb{R}^d$ by that of $\widehat{X}^x$ on Borel functions $f \in L^1(\mathbb{R}^d, \mathbb{P}_X)$ and to use the distortion to evaluate the resulting error. This means comparing

$$\mathbb{E}\, f(X) = \int_{\mathbb{R}^d} f(\xi) \mathbb{P}_X(d\xi) \quad \text{and} \quad \mathbb{E}\, f(\widehat{X}) = \int_{\mathbb{R}^d} f(\xi) \mathbb{P}_{\widehat{X}}(d\xi) = \sum_{i=1}^N f(x_i) \mathbb{P}_X(C(x_i)).$$

From a computational point of view, the numerical computation of the second quantity needs to have access not only to the (hopefully optimal) quantizer x but also to the $\mathbb{P}_X$-mass of the cells of its Voronoi tessellation. One must include this phase in any procedure devised to compute an optimal quantizer (see [18]).

- The basic result is quite simple: if f is *Lipschitz continuous*, then

$$\left| \int_{\mathbb{R}^d} f(\xi) \mathbb{P}_X(d\xi) - \int_{\mathbb{R}^d} f(\xi) \mathbb{P}_{\widehat{X}}(d\xi) \right| = |\mathbb{E} f(X) - \mathbb{E} f(\widehat{X})| \leq \mathbb{E}|f(X) - f(\widehat{X})|$$

$$\begin{aligned} &\leq [f]_{Lip}\mathbb{E}|X-\widehat{X}| \\ &\leq [f]_{Lip}\sqrt{D_N^X(x)}. \end{aligned}$$

This shows that if $x^{(N)}$, $N \geq 1$, denotes a sequence optimal N-quantizers, then $\mathbb{P}_{X^{(N)}}$ weakly converges toward $\mathbb{P}_X$ at optimal rate. (Of course, the weak convergence also holds for any sequence $X^{(N)}$ of N-tuples such that $D_N^X \to 0$ as N goes to infinity).

• When the function f is smoother – *differentiable with a Lipschitz continuous derivative* Df – this error bound can be significantly improved still using the distortion. This relies on a noticeable regularity property of the distortion $D_N^X(x)$ as a function of the N-tuple x: it is *continuously differentiable* on the open set of N-tuples x having pairwise distinct components satisfying

$$\mathbb{P}_X(\cup_{1\leq i\leq N}\partial C(x_i)) = 0 \tag{2.2}$$

(holds for every x if $\mathbb{P}_X$ is continuous), and

$$\frac{\partial D_N^X}{\partial x_i}(x) = 2\int_{C(x_i)}(x_i-\xi)\mathbb{P}_X(d\xi), \quad 1\leq i\leq N. \tag{2.3}$$

Furthermore, one shows (see [13]) that any optimal N-quantizer x^* has pairwise distinct components and satisfies (2.2) provided that $|\mathrm{supp}(\mathbb{P}_X)| \geq N$ (regardless of the continuity of $\mathbb{P}_X$). Consequently x^* is a *stationary quantizer i.e.*

$$\int_{C(x_i)}(x_i^*-\xi)\mathbb{P}_X(d\xi) = 0, \quad 1\leq i\leq N. \tag{2.4}$$

This also holds for any locally optimal quantizer lying inside the support of $\mathbb{P}_X$.

Numerical integration using stationary quantizers has further properties: assume that f is continuously differentiable with a Lipschitz continuous differential $\mathrm{d}f$ ([1]) and that x is a stationary quantizer. Then, the fundamental formula of calculus shows that, for every $i\in\{1,\ldots,N\}$ and every $u\in C(x_i)$

$$|f(\xi)-f(x_i)-\mathrm{d}f(x_i).(\xi-x_i)| \leq [\mathrm{d}f]_{Lip}|\xi-x_i|^2$$

so that, integrating with respect to $\mathbb{P}_X$ on every $C(x_i)$ and summing over i yields

$$\begin{aligned} &\left|\int_{\mathbb{R}^d} f(\xi)\mathbb{P}_X(d\xi) - \sum_{i=1}^N f(x_i)\mathbb{P}_X(C(x_i)) - \sum_{i=1}^N \mathrm{d}f(x_i).\underbrace{\int_{C(x_i)}(x_i-\xi)\mathbb{P}_X(d\xi)}_{=0}\right| \\ &\qquad \leq [\mathrm{d}f]_{Lip}\int_{\mathbb{R}^d}\min_{1\leq i\leq N}|\xi-x_i|^2\mathbb{P}_X(d\xi) \end{aligned}$$

so that

$$\left|\int_{\mathbb{R}^d} f(\xi)\mathbb{P}_X(d\xi) - \sum_{i=1}^N f(x_i)\mathbb{P}_X(C(x_i))\right| \leq [\mathrm{d}f]_{Lip}D_N^X(x). \tag{2.5}$$

When f is twice differentiable with a bounded Hessian d^2f, then the above inequality holds with $\frac{1}{2}\|\mathrm{d}^2f\|_\infty$ instead of $[\mathrm{d}f]_{Lip}$. If x is an optimal N-quantizer, then $D_N^X(x) \ll$

[1] The dual of $\mathbb{R}^d$ is identified with $\mathbb{R}^d$ so that $\mathrm{d}g$ is identified with ∇g from now on.

$\sqrt{D_N^X(x)}$ for N large enough since $\underline{D}_N^X = o(\sqrt{D_N^X})$ as $N \to \infty$. (Also note that $\underline{D}_N^X \leq \underline{D}_1^X = \mathrm{Var}(X)$.)

• A second property of stationary quantizers is of interest for numerical integration: it involves *convex functions*. One starts from the stationary equality (2.4) which also reads, if x denotes a stationary quantizer

$$x_i = \frac{1}{\mathbb{P}_X(C(x_i))} \int_{C(x_i)} \xi \, \mathbb{P}_X(d\xi), \qquad 1 \leq i \leq N. \tag{2.6}$$

Following the definition of $\widehat{X}^x$, this in turn reads

$$\widehat{X}^x = \mathbb{E}(X|\widehat{X}^x).$$

Now the conditional Jensen inequality applied to any convex function f yields

$$\sum_{i=1}^{N} f(x_i)\mathbb{P}_X(C(x_i)) = \mathbb{E}f(\widehat{X}^x) \leq \mathbb{E}f(X). \tag{2.7}$$

since

$$\mathbb{E}f(\widehat{X}^x) = \mathbb{E}f(\mathbb{E}(X|\widehat{X}^x)) \leq \mathbb{E}\left(\mathbb{E}(f(X)|\widehat{X}^x)\right) = \mathbb{E}f(X).$$

Numerical integration by quantization using a stationary quantizer *always yields a lower bound* of the true value $\mathbb{E}f(X)$. For some further error bounds when the function f is simply locally Lipschitz continuous, see [10].

2.3 Stochastic gradient method

Let E be a finite dimensional $\mathbb{R}$-vector space, U a nonempty open subset of E and let μ be a probability measure on $\mathbb{R}^d$. Suppose we are given a continuously differentiable function $g : U \to \mathbb{R}$ with differential $\mathrm{d}g : U \to E$.

Definition 2.2 *We say that* $\mathrm{d}g$ *has an* integral representation on U with respect to μ *if there exists a function* $\mathrm{d}G : U \times \mathbb{R}^d \to E$ *such that* $\mathrm{d}G(x,.) \in L^1(\mu)$ *for every* $x \in U$ *and*

$$\mathrm{d}g(x) = \int_{\mathbb{R}^d} \mathrm{d}G(x,\xi)\, \mu(d\xi).$$

Usually, such a representation formula is obtained by differentiation of a representation formula $g(x) := \displaystyle\int_{\mathbb{R}^d} G(x,\xi)\, \mu(d\xi)$ for g. The principle of stochastic gradient method is to use the function $\mathrm{d}G$ and some independent simulated copies of μ-distributed random vectors to approximate recursively a zero of $\mathrm{d}g$. This procedure can be substituted to the standard gradient descent when the distribution μ can easily be simulated whereas the computation of $\mathrm{d}g(x)$ is out of reach because it requires the computation of integrals with respect to μ in higher dimension. Let us be more specific now. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Following [8] (chapter 2), we have the following definition.

Definition 2.3 *Let* g *be a twice differentiable function from* E *to* $\mathbb{R}$ *such that* $\mathrm{d}g$ *has an integral representation on* E *with respect to* μ. *We call* stochastic gradient method *in* E *for* g, *a triplet of sequences* $((X_n)_{n\geq 0}, (\xi_n)_{n\geq 1}, (\gamma_n)_{n\geq 1})$ *with values respectively in* E, $\mathbb{R}^d$ *and* $[0, +\infty[$ *satisfying for every* $n \geq 1$

$$X_{n+1} = X_n - \gamma_{n+1}\, \mathrm{d}G(X_n, \xi_{n+1}) \tag{2.8}$$

$$(\xi_n)_{n\geq 1} \quad \textit{i.i.d. with} \quad \mathcal{L}(\xi_1) = \mu \tag{2.9}$$

$$\gamma_n > 0 \quad \textit{for every } n \geq 1 \quad \textit{and} \quad \sum_{n\geq 1} \gamma_n = +\infty. \tag{2.10}$$

The sequence $(\gamma_n)_{n\geq 1}$ is called the step *or* gain parameter *sequence.*

This definition is motivated by the following convergence theorem. This result is classical and many variants and generalizations can be found in the literature devoted to Stochastic Approximation Theory ([8], [16], among others).

Theorem 2.4 (*a*) *A.s.* CONVERGENCE: *Let $g : E \to \mathbb{R}_+$ be a continuously differentiable function whose differential* dg *admits an integral representation on E with respect to μ*

$$\mathrm{d}g(x) = \int_{\mathbb{R}^d} \mathrm{d}G(x,\xi)\mu(d\xi).$$

Assume that dg *and* dG *satisfy*

$$\lim_{|x|\to+\infty} g(x) = +\infty \quad \textit{and} \quad \mathrm{d}g \ \textit{ is Lipschitz continuous} \tag{2.11}$$

$$\int_{\mathbb{R}^d} |\mathrm{d}G(x,\xi)|^2 \mu(d\xi) = O(g(x)) \quad \textit{as} \quad |x| \to \infty. \tag{2.12}$$

Let $((X_n)_{n\geq 0}, (\xi_n)_{n\geq 1}, (\gamma_n)_{n\geq 1})$ be a stochastic gradient method with a positive gain parameter sequence satisfying

$$\sum_{n\geq 1} \gamma_n = +\infty \quad \textit{and} \quad \sum_{n\geq 1} \gamma_n^2 < +\infty. \tag{2.13}$$

Then $g(X_n)$ a.s. converges to some nonnegative random variable $g_\infty \in \mathbb{R}_+$ and X_n a.s. converges toward some random connected component χ^* of $\{\mathrm{d}g = 0\} \cap \{g = g_\infty\}$.

In particular, if $\{\mathrm{d}g = 0\} = \{x^*\}$, then

$$X_n \longrightarrow x^* \quad a.s. \quad \textit{as} \quad n \to +\infty. \tag{2.14}$$

(*b*) RATE OF CONVERGENCE (CLT): *Let x^* be an equilibrium point of $\{\mathrm{d}g = 0\}$. Assume that x^* is attractive, that is g is is twice differentiable at x^* and $\mathrm{d}^2 g(x^*)$ is positive definite. Assume that the "noise" is nondegenerated at x^*, namely that*

$$\Gamma^* := \int_{\mathbb{R}^d} \mathrm{d}G(x^*,\xi)\, {}^t(\mathrm{d}G(x^*,\xi))\mu(d\xi) \ \textit{ is positive definite,} \tag{2.15}$$

where ${}^tA$ is for transpose of A.

Specify the gain parameter sequence as follows

$$\forall\, n \geq 1, \qquad \gamma_n = \frac{a}{b + n^\alpha}, \quad a,\, b > 0, \quad 0 < \alpha < 1.$$

If $\alpha = 1$ assume furthermore that the lowest eigenvalue λ_{min} of $\mathrm{d}^2 g(x^*)$ satisfies

$$a > \frac{1}{2\lambda_{min}}. \tag{2.16}$$

Then, the above a.s. convergence is ruled on the convergence set $\{X_n \to x^*\}$ *by the following Central Limit Theorem*

$$\frac{X_n - x^*}{\sqrt{\gamma_n}} \overset{\mathcal{L}_{stably}}{\longrightarrow} \mathcal{N}(0, \Sigma), \tag{2.17}$$

with $\Sigma := \displaystyle\int_0^{+\infty} e^{-(\mathrm{d}^2 g(x^*) - \rho I_d)u} \Gamma^* e^{-(\mathrm{d}^2 g(x^*) - \rho I_d)u}\, du$ *and* $\rho = \frac{1}{2a}\mathbf{1}_{\{\alpha=1\}}$.

The convergence in (2.17) means that for every bounded continuous function and every $A \in \mathcal{F}$,

$$\mathbb{E}\left(\mathbf{1}_{\{X_n \to x^*\} \cap A}\, f(\frac{X_n - x^*}{\sqrt{\gamma_n}})\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\left(\mathbf{1}_{\{X_n \to x^*\} \cap A}\, f(\sqrt{\Sigma}\,\zeta)\right), \quad \zeta \sim \mathcal{N}(0; I_d).$$

Remark 2.5 • The above formulation is derived from [8]: claim (a) is the combination of Theorem 2.III.4 p.61 and section 3.III.2, p.102. Claim (b) comes from section III., p.160.

• When g is only defined on an (open) domain $U \subset E$, the above convergence still holds when the gain parameter sequence $(\gamma_n)_{n\geq 0}$ takes its values in $(0, \gamma_{max}]$ provided that U is convex, that $x \mapsto x - \gamma_{max}\mathrm{d}G(x,\xi)$ maps U into U for every $\xi \in \mathbb{R}^d$ and that

$$\lim_{d(x,\partial U)\to 0} g(x) = +\infty.$$

This last assumption on g can be relaxed if U *is bounded and if* g *and* $\mathrm{d}g$ *admit a continuous extension on* $\overline{U}$ *and if* $\mathrm{d}G(.,\xi)$, $\xi \in \mathbb{R}^d$ *admit an extension on* $\overline{U}$ *which extends the representation property on* $\overline{U}$.

• The matrix $\mathcal{N}(0;\Sigma)$ is the invariant distribution of the Ornstein-Uhlenbeck diffusion

$$dY_t = -(\mathrm{d}^2 g(x^*) - \rho I_d)Y_t\, dt + \sqrt{\Gamma^*}\, dW_t.$$

• It follows from (2.17) that the fastest possible rate of convergence is $\sqrt{n}$. It is obtained with step sequence $\gamma_n = \frac{a}{b+n}$, $n \geq 1$, a large enough: indeed $\sqrt{n}(X_n - x^*)$ weakly converges toward $\mathcal{N}(0; a\,\Sigma)$. One easily checks that $a\Sigma$ goes to 0 as $a \to \infty$. So the best rate of convergence is obtained for arbitrary large a. Except that the number of iterations needed for this rate of convergence to show up becomes greater and greater. So, an empirical approach is necessary to fit some "reasonable" coefficient a. This could be *e.g.* (in 1-dimension) $a = 1/g''(x^*)$ which then yields a $1/g''(x^*)$ asymptotic variance term. Unfortunately, this quantity is usually out of reach given the fact that we are looking for x^*. Some averaging methods can theoretically provide a solution to that problem but empirical tests were not decisive for the optimal quadratic quantization problem we are dealing with.

Uniform distribution $U([0,1])$: We will illustrate Theorem 2.4 with the quadratic distortion for uniform distribution on $[0,1]$ (see [9]). Set $E = \mathbb{R}^N$, $d = 1$, and

$$g(x_1, \ldots, x_N) := \frac{1}{2}\int_0^1 \min_{1\leq i\leq N} (x_i - \xi)^2\, du.$$

Function g is clearly symmetric, so one may restrict on the open set $U := \{(x_1, \ldots, x_N),\ 0 < x_1 < x_2 < \ldots < x_N < 1\}\}$. On U, g is differentiable and $\mathrm{d}g$ has an integral representation with respect to du given by (2.3). Now $C(x_i) = [x_{i-1/2}, x_{i+1/2}]$, $1 \leq i \leq N$, with $x_{i+1/2} := \frac{x_i + x_{i+1}}{2}$, $1 \leq i \leq N-1$, $x_{1/2} = 0$ and $x_{N+1/2} = 1$. With these conventions, one checks that

$$\mathrm{d}g(x_1, \ldots, x_N) = \left(\int_{x_{i-1/2}}^{x_{i+1/2}} (x_i - \xi)\, d\xi \right)_{1 \leq i \leq N}.$$

These integrals can be computed so that

$$\begin{aligned}
\frac{\partial g}{\partial x_i}(x) &= \frac{1}{8}\left(2x_i - (x_{i+1} + x_{i-1})\right)(x_{i+1} - x_{i-1}), \quad 2 \leq i \leq N-1, \\
\frac{\partial g}{\partial x_1}(x) &= \frac{1}{8}\left(3x_1 - x_2\right)(x_1 + x_2), \\
\frac{\partial g}{\partial x_N}(x) &= \frac{1}{8}\left(3x_N - x_{N-1} - 2\right)(2 - (x_N + x_{N-1})).
\end{aligned}$$

The computation of the Hessian $\mathrm{d}^2 g$ of g is straightforward and we have for a given N-tuple x and for any i such that $2 \leq i \leq N-1$:

$$\mathrm{d}^2 g(x)_{i,i-1} = -\frac{x_i - x_{i-1}}{4}, \quad \mathrm{d}^2 g(x)_{i,i+1} = -\frac{x_{i+1} - x_i}{4},$$

$$\mathrm{d}^2 g(x)_{i,i} = \frac{x_{i+1} - x_{i-1}}{4}.$$

One checks that $\mathrm{d}g(x^*) = 0$ iff $x_i^* = \frac{2i-1}{2N}$ for $i = 1, \ldots, N$. Finally, g satisfies all the assumptions of Theorem 2.4 (with $\gamma_{max} = 1$). Furthermore, the eigenvalues of $\mathrm{d}^2 g(x^*)$ can also be computed and we find

$$\lambda_i = \frac{1}{N} \sin^2\left(\frac{\pi i}{2N}\right), \qquad i \in \{1, \ldots, N\}.$$

so that

$$\lambda_{min} = \frac{1}{N} \sin^2\left(\frac{\pi}{2N}\right) \approx \frac{\pi^2}{4N^3} \qquad \text{when } N \text{ is large.}$$

Thus, Theorem 2.4 shows that a Central Limit Theorem holds for the $a.s.$ convergence $X_n \to x^*$ provided $a > 2N^3/\pi^2$.

Remark 2.6 • One checks that the Hessian $\mathrm{d}^2 g(x^*)$ at x^* is the discrete Laplacian obtained by finite difference on the interval $[0, 1]$ up to a multiplicative factor $N/4$. Here, the ill-conditioned nature of such an operator is directly linked to the (slow) rate of convergence of the algorithm (2.8) through Theorem 2.4. Indeed, the number n of trials necessary to get γ_n close to 0 increases with N.

• This example suggests that when implementing a stochastic gradient to the distortion function of a more general distributions, special attention has to be paid to the points which are close to a mode of the (probability density function of the) distribution μ. There, roughly speaking, the distribution mimics the uniform distribution because of the lack of injectivity and this seems to impose as strong assumptions on the step parameter $(\gamma_n)_{n \geq 1}$ as for the uniform distribution.

PARTIAL EXTENSION TO NON UNIFORM DISTRIBUTIONS ON THE REAL LINE: To conclude this section, let us mention some further results about the quadratic distortion in 1-dimension. One may restrict the distortion function to the open set $U := \{(x_1, \ldots, x_N),\ m < x_1 < x_2 < \ldots < x_N < M\}$ where $m := \inf \mathrm{supp}(\mu)$ and $M := \sup \mathrm{supp}(\mu)$.

– First, when $\mathbb{P}_X$ is absolutely continuous with a log-concave probability density function, then D_N^X has a unique stationary – hence optimal – quantizer x^* *i.e.* $\{\mathrm{d}(D_N^X) = 0\} = \{x^*\}$. This is *e.g.* the case of the Normal distribution $\mu(d\xi) := \exp(-\xi^2/2)/\sqrt{2\pi}$.

– If, furthermore, μ has a compact support, then the above stochastic gradient procedure *a.s.* converges toward x^* (see [14, 17]).

– Some examples of non-uniqueness of the stationary quantizer can be found *e.g.* in [13]. For some examples of uniqueness when the probability density function is not log-concave, see [10].

– No regular *a.s.* convergence result holds for non compactly supported distributions $\mathbb{P}_X$, essentially because the distortion does not go to infinity when $|x|$ goes to infinity.

In higher dimension, uniqueness of stationary quantizers clearly often fails, so Theorem 2.4 must be applied in its general form.

3 Optimal quantization of distributions on $\mathbb{R}^d$. The case of the Normal distribution

Let $d \geq 1$, X be a $\mathbb{R}^d$-valued random vector having an *absolutely continuous* distribution $\mu = P_X$. Let $N \geq 1$ be an integer (in this section D_N^X will always denote the distortion function related to the distribution μ). In this section we deal with the following optimization problem:

$$(\mathcal{P}) \quad \equiv \quad \begin{cases} \text{Find a } N\text{-tuple } x^* = (x_1^*, \ldots, x_N^*) \text{ s.t.} \\ D_N^X(x_1^*, \ldots, x_N^*) \leq D_N^X(x_1, \ldots, x_N), \quad \forall x = (x_1, \ldots, x_N) \in (\mathbb{R}^d)^N, \end{cases}$$

where $D_N^X : (\mathbb{R}^d)^N \to \mathbb{R}^+$ is defined by

$$D_N^X(x) = \sum_{i=1}^N \int_{\mathbb{R}^d} \min_{1 \leq i \leq N} |x_i - \xi|^2 \, \mu(d\xi) = \sum_{i=1}^N \int_{C(x_i)} |x_i - \xi|^2 \, \mu(d\xi), \tag{3.1}$$

where $(C(x_i))_{i=1,\ldots,N}$ denotes the Voronoi tessellation of the N-tuple x in $\mathbb{R}^d$.

We have seen that D_N^X is continuously differentiable on the set of N-tuples having pairwise distinct components (see (2.3)) and that every solution x^* of $(\mathcal{P})$ is a stationary quantizer hence satisfying (2.6).

If one looks at problem $(\mathcal{P})$ from a strictly deterministic point of view, several approaches can be processed essentially gradient based methods (including Newton's method) and fixed point methods.

• The gradient descent approach is classical and relies on formula (2.6) for the derivative. One set

$$x^0 := x \quad \text{and} \quad x^{n+1} = x^n - \frac{\gamma}{n} \, \mathrm{d}(D_N^X)(x^n)$$

for a rate parameter $\gamma \in (0, 1)$. One may show, that under assumption (2.12) of Theorem 2.4 it does converge to some zero x^* of $\mathrm{d}(D_N^X)$. It also does with a small enough constant step $\gamma_n = \gamma > 0$ instead of $\frac{\gamma}{n}$ (with a better rate, if convergence does occur).

Newton's method (see paragraph 3.1 below for the scalar Normal distribution) requires to compute the Hessian $\mathrm{d}^2(D_N^X)$: this is done in [9] for quite general 1 and 2-dimensional absolutely continuous distributions.

• The fixed point approach was introduced by Lloyd (in 1-dimension) and consists in writing the following recursive algorithm (so-called Lloyd's method I, see [14]) from the stationarity Equation (2.6): starting from a N-tuple x, one defines recursively a sequence $\{x^n\}_{n\geq 0}$ such that

$$\left\{\begin{array}{ll} x^0 & := x \\ x_i^{n+1} & := \dfrac{1}{\mu(C(x_i^n))}\displaystyle\int_{C(x_i^n)} \xi\,\mu(d\xi), \quad \forall i=1,\ldots,N. \end{array}\right. \tag{3.2}$$

(with $\mu = \mathbb{P}_X$). If one set $\widehat{X}^{n+1} := \widehat{X}^{x^{n+1}}$, one easily checks that Equation (3.2) implies that

$$\widehat{X}^{n+1} = \mathbb{E}(X \,|\, \widehat{X}^n),\ n \geq 0.$$

The very definition of conditional expectation as an orthogonal projection on the space of square integrable $\sigma(\widehat{X}^n)$-measurable random variables shows that

$$\|X - \widehat{X}^{n+1}\|_2 = \|X - \mathbb{E}(X|\widehat{X}^n)\|_2 = \min\left\{\|X - Z\|_2,\ Z \in L^2(\sigma(\widehat{X}^n),\mathbb{P})\right\} < \|X - \widehat{X}^n\|_2$$

(except if $\widehat{X}^n = \mathbb{E}(X|\widehat{X}^n)$) *i.e.* $n \mapsto \|X - \widehat{X}^n\|_2$ is decreasing.

In 1-dimension, when μ is has a strictly log-concave density function, it is established in [14] that $x \mapsto \left(\frac{1}{\mu(C(x_i^n))}\int_{C(x_i^n)} \xi\,\mu(d\xi)\right)_{1\leq i\leq N}$ is a contraction mapping and hence admits a unique fixed point x^* toward which Lloyd's method I converges exponentially fast (this was in fact the first proof for uniqueness of the stationary quantizer in that setting). In higher dimension, the convergence of the procedure is not clearly established in the literature.

As soon as $d \geq 2$, the processing of both methods described above becomes quickly intractable since we have to compute numerically some d-dimensional integrals (on some the elements of the Voronoi tesselation). Furthermore, one checks (see [18]) that the stationary solution of (3.2) is usually not unique in dimension $d \geq 2$. As suggested above, the dimension 1 can be investigated apart since, then, everything can be efficiently computed in both methods. This is the main reason why, in higher dimensions, one needs to look for stochastic procedures instead of deterministic ones.

From now on, we will focus on the Normal distribution $\mu = \mathcal{N}(0; I_d)$, defined for every Borel set A of $\mathbb{R}^d$, by

$$\mu(A) = \int_A \exp\left(-\frac{|\xi|^2}{2}\right)\frac{\lambda_d(d\xi)}{(2\pi)^{d/2}}.$$

We will denote by $\mathrm{erf}(y) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{y} \exp(-u^2/2)\,\mathrm{d}u$ its distribution function in 1D.

3.1 Newton's method for 1-dimensional Normal distribution

3.1.1 Description of the method

In this subsection, we still set $d = 1$, $V := \mathbb{R}^N$ and $U := \{(x_1,\ldots,x_N),\ x_1 < x_2 < \ldots < x_N\}$. Let $N \geq 2$. We can then compute in an pseudo-explicit way the real number $D_N^X(x)$, the vector $\mathrm{d}(D_N^X)(x) \in \mathbb{R}^N$ and the $N \times N$-matrix $\mathrm{d}^2(D_N^X)(x)$ using the tabulation of the

distribution function erf of the scalar Normal distribution in $\mathbb{R}$ (see [9] for more general 1D-distributions). Set $x_{j\pm1/2} := (x_j + x_{j\pm1})/2$, $j = 1, \dots, N-1$, $x_{1/2} = 0$ and $x_{N+1/2} = 0$. Some elementary computations yield, for every $x \in \mathbb{R}^N$, and every i, $j \in \{1, \dots, N\}$,

$$
\begin{aligned}
D_N^X(x) &= \sum_{j=1}^{N} \int_{x_{j-1/2}}^{x_{j+1/2}} (x_j - \xi)^2 \exp(-\xi^2/2) \frac{d\xi}{\sqrt{2\pi}}, \qquad (3.3)\\
&= \sum_{j=1}^{N} \Bigg((1 + x_j^2)(\operatorname{erf}(x_{j+1/2}) - \operatorname{erf}(x_{j-1/2})) \\
&\quad - \frac{1}{\sqrt{2\pi}} (x_{j+1/2} \exp(-x_{j+1/2}^2/2) - x_{j-1/2} \exp(-x_{j-1/2}^2/2)) \\
&\quad + \frac{2}{\sqrt{2\pi}} x_j (\exp(-x_{j+1/2}^2/2) - \exp(-x_{j-1/2}^2/2)) \Bigg),
\end{aligned}
$$

$$
\frac{\partial D_N^X}{\partial x_i}(x) = x_i(\operatorname{erf}(x_{i+1/2}) - \operatorname{erf}(x_{i-1/2})) + (\exp(-x_{i+1/2}^2/2) - \exp(-x_{i-1/2}^2/2))/\sqrt{2\pi}, \qquad (3.4)
$$

$$
\begin{cases}
\dfrac{\partial^2 D_N^X}{\partial x_i \partial x_{i-1}}(x) &= -\dfrac{1}{4\sqrt{2\pi}} (x_i - x_{i-1}) \exp(-x_{i-1/2}^2/2), \\
\dfrac{\partial^2 D_N^X}{\partial x_i^2} &= \operatorname{erf}(x_{i+1/2}) - \operatorname{erf}(x_{i-1/2}) - \dfrac{1}{4\sqrt{2\pi}} (x_{i+1} - x_i) \exp(-x_{i+1/2}^2/2) \\
&\quad - \dfrac{1}{4\sqrt{2\pi}} (x_i - x_{i-1}) \exp(-x_{i-1/2}^2/2) \\
\dfrac{\partial^2 D_N^X}{\partial x_i \partial x_{i+1}}(x) &= -\dfrac{1}{4\sqrt{2\pi}} (x_{i+1} - x_i) \exp(-x_{i+1/2}^2/2),
\end{cases} \qquad (3.5)
$$

We are now able to implement Newton's method in order to find the (single) zero of $\mathrm{d}(D_N^X)$ in $\mathbb{R}^N$. Thus, starting from $x^0 \in \mathbb{R}^N$, we compute recursively

$$
x^{n+1} = x^n - [\mathrm{d}^2(D_N^X)(x^n)]^{-1} \cdot \mathrm{d}(D_N^X)(x^n) \qquad (3.6)
$$

(so we need to invert at every step the matrix $\mathrm{d}^2(D_N^X)(x^n)$).

3.1.2 Numerical results

Computations produced N-optimal quantizers x^* until $D_N^X(x^*)$ is equal to 0.25×10^{-4} (for $N \approx 330$). Then we can say that for such a size

$$
\min_{i \neq j} |x_i^* - x_j^*| \leq 2 \int_{\mathbb{R}^d} \min_{1 \leq i \leq N} |x_i^* - \xi| \, \mu(d\xi) \leq 2\sqrt{D_N^X(x^*))} \approx 10^{-2},
$$

Let us emphasize the importance of the choice of the initial conditions. Thus, we observe that, even for symmetric initial vectors, some components collapse or are rejected far from the others. The following choice gives good results: $x_k^0 := -2 + 2(2k-1)/N$, $1 \leq k \leq N$. Figure 1 displays $n \mapsto -\log_{10}(|\mathrm{d}(D_N^X)(x^n)|)$ for $N = 300$. The Frobenius norm of $\mathrm{d}^2(D_N^X)(x^n)$ is also drawn (thin lines). We can see that even if the problem $(\mathcal{P})$ is not a quadratic optimization problem, it becomes quickly quadratic and then Newton's algorithm converges very quickly (theoretically in one step). In Figure 2 below, we check graphically the quality of the quantizer obtained after convergence of the method by drawing the "weight function" $x_i^* \mapsto \mu(C(x_i^*))$, $i = 1, \dots, N$ (for $N = 50$ and $N = 300$). We rely on

the following result established in [7] which says that optimal N-quantizers of the scalar Normal distribution satisfy

$$\mu(C(x_i^*)) \sim \frac{1}{N}\frac{\exp(-(x_i^*)^2/3)}{\int_{\mathbb{R}}\exp(-\xi^2/3)d\xi} \quad \text{as} \quad N \to \infty.$$

(uniformly on compact sets with respect to x_i^*). This result also holds for more general scalar distributions μ with (positive) density function g and so can be used to test the adequacy of a large size quantizer: it says that $\mu(C(x_i^*)) \sim \frac{1}{N}\frac{g(x_i^*)^{2/3}}{\int_{\mathbb{R}} g(\xi)^{2/3}d\xi}$. It holds as a conjecture in higher dimension in the following form

$$\mu(C(x_i^*)) \sim \frac{1}{N}\frac{g(x_i^*)^{2/(d+2)}}{\int_{\mathbb{R}} g(\xi)^{2/(d+2)}du}.$$

3.2 Stochastic methods in higher dimension

3.2.1 The $CLVQ$ algorithm and its companion procedures

It follows from (2.3) that, if we denote by ξ a $\mathbb{R}^d$-valued Normally distributed random variable,

$$\mathrm{d}(D_N^X)(x) = \mathbb{E}\,(\mathbf{1}_{C(x_i)}(\xi)(x_i - \xi)).$$

Subsequently, the $(\mathbb{R}^d)^N$-valued stochastic gradient procedure for D_N^X used in this subsection can be written as

$$X^{n+1} = X^n - \gamma_{n+1}\mathbf{1}_{C(X_i^n)}(\xi^{n+1})(X_i^n - \xi^{n+1}) \tag{3.7}$$

or, equivalently, if we define $i_0(n+1)$ as the integer such that $\xi^{n+1} \in C(X^n_{i_0(n+1)})$,

$$\begin{cases} X_i^{n+1} = X_i^n - \gamma_{n+1}(X_i^n - \xi^{n+1}) & \text{if } i = i_0(n+1) \\ X_i^{n+1} = X_i^n & \text{if } i \neq i_0(n+1). \end{cases} \tag{3.8}$$

This procedure is known as the *Competitive Learning Vector Quantization algorithm* ($CLVQ$). More recently, it also appeared in the literature as the Kohonen algorithm with 0 neighbour (the initialization of the procedure will be shortly discussed below in subsection 3.2.2). It can be decomposed in two phases:

Competitive phase: Selection of the "winning index" $i_0(n+1)$ using a closest neighbour search.

Learning phase: Updating of the winning component by a *homothety* centered at ξ^{n+1} with ratio $(1-\gamma_{n+1})$.

From a numerical point of view the most time consuming task is to compute the winner index that is the component X_i^n which is the closest to ξ^{n+1}. Some fast (approximate) procedures for the searching of this "nearest neighbour" have been designed (see [12] chapter 10.4, p.332 and chapter 12.16, p.479).

An attractive feature of this procedure is that, as a by-product, one can compute the μ-masses $\mu(C(x_i))$, $i = 1, \ldots, N$ of the Voronoi cells and the distortion. To evaluate them, one simply increments a counter k_i^n as follows:

$$k_i^{n+1} = k_i^n + \mathbf{1}_{\{i=i_0(n+1)\}}, \quad i = 1, \ldots, N.$$

Then $k_i^n/n \to \mu(C(x_i^*))$ on the event $\{X^n \to x^*\}$ as n goes to infinity. Other "on line" approximation procedure for these weights involve the gain parameter like γ_n

$$\alpha_i^{n+1} = \alpha_i^n + \gamma_{n+1}(\mathbf{1}_{\{i=i_0(n+1)\}} - \alpha_i^n), \quad \alpha_i^0 = 1/N, \qquad i = 1, \ldots, N.$$

which converges toward $\mu(C(x_i^*))$ on $\{X^n \to x^*\}$ as well.

Concerning the distortion, one proceeds similarly by setting

$$D_N^{X,n+1} = D_N^{X,n} + |X_{i_0(n+1)}^n - \xi^{n+1}|^2, \qquad D_N^{X,0} = 0.$$

so that $D_N^{X,n}/n \to D_N^X(x^*)$ on the event $\{X^n \to x^*\}$ as n goes to infinity. One can also update using the step sequence $(\gamma_n)_{n\geq 0}$ like for the weights.

A slowlier and less sophisticated procedure consists in freezing the CLVQ procedure for n large enough and to process afterwards a standard Monte Carlo simulation.

After the processing of the $CLVQ$ procedure, one may refine the produced N-quantizer by processing M_{Lloyd} *randomized Lloyd's method I.* By randomized Lloyd's method I, we mean that all expectations w.r.t. to the Normal random vectors in Equation (3.2) are computed by a (short) Monte Carlo simulation. Usually $M_{Lloyd} \approx 10$.

3.2.2 Heuristic specifications for the $CLVQ$ procedure and illustrations

We will now turn the discussion about three kinds of problems which arise in practise. The first one concerns the quantization of a distribution near its modes (when some). The second one concerns the quantization of non compactly supported distributions. The third problem is the initialization of both the quantizer and the step.

Concerning the first point, it has been pointed out in Remark 2.6 that not any parameter sequence $(\gamma_n)_{n\geq 0}$ can be chosen here. In fact to take into account the mode of the Normal d-dimensional distribution, one essentially specifies the step as if we wish to quantize the uniform distribution on $[0,1]^d$. We adopt the following heuristic: we infer from the uniform quantization of $[0,1]$ with $N^{1/d}$ points our choice of step γ_n for the uniform quantization of $[0,1]^d$ with N points. Consequently the parameter sequence $(\gamma_n)_{n\geq 0}$ will be set equal to

$$\gamma_n = \gamma_0 \frac{a}{a + \gamma_0 b\, n}, \tag{3.9}$$

where a and b are equal to

$$a = 4N^{1/d}, \quad b = \pi^2 N^{-2/d}. \tag{3.10}$$

Thus $\gamma_n \sim \frac{a}{bn} \sim \frac{4N^{3/d}}{\pi^2\, n}$ so that $\gamma_n > \frac{2N^{3/d}}{\pi^2\, n}$ which is the critical step for the uniform distribution to get a Central Limit Theorem for large enough n. This explains our choice for the ratio a/b. The balance between a and b (in particular $a \gg b$) implies that the procedure first behaves like a constant step algorithm. Now, the constant step version of the procedure is known to be positively (even geometrically) recurrent (see [5]) so that it visits every open set of the state space, especially the attracting basin of the optimal quantizer. Hopefully it may remain in it when γ_n finally goes to 0. Some simulated annealing version of the procedure can be implemented instead of this (almost) constant step phase. However it seems not to give significant results. Let us illustrate the choice of a and b in 1-dimension. In Figure 3, we have represented two different results for two different choices of the parameter γ_n when $N = 100$. In both cases, we have computed 10^7

trials in order to be sure that we get convergence. The value of the distortion obtained are very close in the two cases. In Figure 3 a), we have taken $\gamma_0 = 1$, $a = 400$ and $b = 0.1$. The counters $\{k_i\}$ are plotted as function of the quantizer $\{x_i\}$. We can see that the distribution obtained is far from the Normal distribution. In Figure 3 b), γ_0, a are the same as above but now $b = 10^{-3}$ which is close to $\pi^2/10^4$.

Concerning the second problem, the simulation of points with too large norms may cause dramatic effects on the $CLVQ$ procedure when the step is not yet small enough (cf. Eq. (3.7)). In order to avoid this, we will (first) simulate some spherically truncated Normal variables (calibrating the threshold radius so as to keep at least 99 % of the mass). This truncation has a stabilizing effect on the procedure. Then, to get a quantization of the original Normal distribution, one can complete the optimization by processing once the randomized Lloyd's method I with *nontruncated Normally distributed random numbers*. One verifies that, when the number of points is large, this only affects the location of the peripheral points. On the other hand, as expected, it slightly increases the distortion (but it produces more accurate results for numerical integration of course). In Figure 6 are displayed 2D quantizers with $N = 500$. In Figure 6 (a) the depicted quantizer has been obtained using an extended splitting initialization method described below and truncated simulated Normal random variables. Its distortion is $D_N^X((a)) = 7.08(-3)$. The quantizer depicted in Figure 6 (b) has been obtained from that in (a) by simply processing one randomized Lloyd's method I with a nontruncated Normal distribution as described in (3.2). Its distortion is $D_N^X((b)) = 8.55(-3)$.

Let us come now to the initialization of the N-quantizer in the $CLVQ$ procedure. When N is small ($N \leq 10$) we adopted a *random initialization* so that $X^0 \sim (\mathcal{N}(0; I_d))^{\otimes N}$ ([2]). When N gets larger we passed to the so-called *splitting initializing method*, consisting in adding one further point (usually the optimal 1-quantizer *i.e.* the origin $0_{\mathbb{R}^d}$) in order to obtain the starting quantizer of the $CLVQ$ procedure with $N+1$ components. This $N+1$-quantizer is not optimal. So, we then processed a $CLVQ$ algorithm (3.7). In Figure 5, we compare the N-quantizer ($N = 14$) obtained from a splitting method (in (a)) based on the 13-quantizer depicted in the former Figure 4 on one hand and from a random (Normal) initialization (in (b)) on the other hand. Two "pseudo"-locally optimal quantizers seem to exist simultaneously. The added component at $0_{\mathbb{R}^d}$ has moved the pentagon into a hexagon whereas in (b) the fourteenth point has moved to the outside circle. In fact both 14-quantizers have not the same distortion: $D_N^X((a)) = 2.38(-1)$ and $D_N^X((b)) = 2.35(-1)$. So the 14-quantizer in (a) is only a local minimum. This emphasizes that, in higher dimension, the distortion function has a more intricate shape than in 1-dimension. This also shows that the splitting method may provide only sub-optimal stationary quantizers. Overall, it turns out to be a good compromise between stability and efficiency.

The splitting initializing method can be extended to the initialization of a $N+N'$ $CLVQ$ procedure by simply "aggregate" an optimal N'-quantizer to an optimal N-quantizer, $N' \ll N$. This has been done successfully up to $d = 10$ to cut down computation time when dealing with quantizers having many components (we set $N' = 10$ if $100 \leq N \leq 1000$ and $N' = 100$ if $N \geq 1\,000$).

Finally, as far as splitting methods are concerned, the step parameter γ_0 is chosen equal either to the square root of the quadratic distortion computed at the last step or to 1 if the

[2] Other choices are possible taking into account some results about random quantization (see [6]) which could suggest to sample $(X_i^0)_{1 \leq i \leq N}$ following the (Gaussian) probability distribution whose density is proportional to $(f_d^{\otimes N})^{\frac{d}{d+2}}$ distribution where f_d is the density of the Normal distribution on $\mathbb{R}^d$.

distortion is greater than 1. This choice is suggested (or motivated) by the inequality

$$\min_{i\neq j}|x_i - x_j| \leq 2\mathbb{E}|X - x_i| \leq 2(\mathbb{E}|X - x_i|^2)^{1/2}.$$

As a matter of fact, since we start from an optimal N-quantizer, this choice seems quite appropriate to preserve the past computations in the splitting method.

In Figure 8 is depicted a 1000-quantizer of $\mathcal{N}(0, I_3)$ in 3-dimension. The distortion is $D_N^X = 5.45(-2)$.

3.2.3 Numerical and geometrical features of optimal quantizers in dimension greater than 4.

To evaluate the quality of a computed N-quantizer in dimension $d \geq 4$ we can no longer use the graphic approach either directly or using the μ-masses of the Voronoi cells like for 1-dimensional distributions.

Concerning the purely numerical aspects, we rely on Inequality (2.7) for convex functions which says that

$$\sum_{i=1}^{N} f(x_i)\mu(C(x_i)) \leq \int f d\mu$$

for any stationary N-quantizer $x = (x_1, \ldots, x_N)$. Thus, as far as Normal distribution is concerned, *i.e.* $\mu = \mathcal{N}(0; I_d)$, one may choose the convex function $f(x) := |x|^2$ and reject any N-quantizer x such that $\sum_{1\leq i\leq N} |x_i|^2 \mu(C(x_i)) > d$. One can refine this test by considering other convex functions like $f(x) := (w|e^j)^2$, $j = 1, \ldots, d$ where $(e^1, \ldots, e^d)$ denotes the canonical basis of $\mathbb{R}^d$, $f(x) = |x|^{1+\rho}$, etc.

Concerning the geometrical aspects, we computed the norms of each component in $\mathbb{R}^d$ and sorted them in increasing order. These curves are displayed in Figure 9. In (a), we can distinguish four regions of slow growth for $N = 1220$, the first one around 100, the second one between 200 and 400, the third one between 500 and 800 and the last beyond 800. It suggests that the mass seems to be located on a finite number of spheres (4). In (b), this number decreases to 3. In (c), it is 2 and in (d) there is only one flat line beyond 100. The conclusion is that the mass of the Gaussian measure tends to be more and more localized as dimensions increases. This is related with the fact that, by the strong Law of the Large Numbers, if $X_d \sim \mathcal{N}(0, I_d)$ then $|X_d|^2 \sim d$ as $d \to \infty$: a χ^2 distribution with d degrees of freedom tends to be concentrated (with a suitable normalization) on a sphere when d increases.

4 Evaluation of a Put Spread European option

The aim of this section is to test the optimal quantizers that we obtained by the numerical methods described in subsection 3.2.2 in dimensions $2 \leq d \leq 6$. Simultaneously, we aim to illustrate the performances of vector quantization for numerical integration. That is why we carry out a short comparison between quantization method and Monte Carlo method on a simple numerical integration problem.

The strong Law of Large Number implies that, given a Normally distributed random vector X and a sequence of i.i.d. random vectors $(\xi_k)_{k\geq 1}$ with common Normal distribution $\mathcal{N}(0; I_d)$,

$$\mathbb{P}(d\omega)\text{-}a.s. \quad \frac{f(\xi_1(\omega)) + \cdots + f(\xi_N(\omega))}{N} \overset{N\to+\infty}{\longrightarrow} \mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(\xi)\exp\left(-|\xi|^2/2\right)\frac{d\xi}{(2\pi)^{d/2}}.$$

for every $f \in \mathcal{L}^1(\mathbb{R}^d, \mathbb{P}_X)$. The Monte Carlo method consists in generating on a computer a path $(\xi_k(\omega))_{k\geq 1}$ to compute the above Gaussian integral. Roughly speaking, the Law of the Iterated Logarithm says that if f is square integrable, the above convergence $a.s.$ holds at a

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

rate where $\sigma(f(X))$ is the standard deviation of $f(X)$. When f is twice differentiable, this is to be compared to the error bound provided by (2.5) when using a quadratic optimal N-quantizer $x^* := (x_1^*, \ldots, x_N^*)$, namely

$$[\mathrm{d}f]_{Lip} D_N^{\mathcal{N}(0;I_d)} \approx \left(J_{2,d}(1+2/d)^{1+d/2}[\mathrm{d}f]_{Lip}\right) N^{-2/d}.$$

Consequently the dimension $d = 4$ appears as the critical dimension for the numerical integration of such functions by quantization for a given computational complexity (quantization formulae involving higher order differentials yield better rates): we assume that the optimal quantizers have been formerly computed and that the computation time of a (Gaussian) random number or a weight is negligible with respect to the computation of a value of f.

The test is processed in each dimension d with four random variables $g_i(X)$, $X \sim \mathcal{N}(0; I_d)$, $i = 0, 1, 2, 3, 4$ where the g_i's are five functions with compact support satisfying respectively

– g_0 is a (bounded) interval indicator (hence discontinuous);

– g_1 is Lipschitz continuous and the composition of two convex functions;

– g_2 is twice differentiable and the composition of two convex functions;

– g_3 is difference of two convex functions (via Call-Put parity) and is Lipschitz continuous;

– g_4 is difference of two convex functions (via Call-Put parity) and is twice differentiable.

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

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

then one considers, still at time T, the geometric index

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

Then, one specifies the random variables $g_i(\xi)$ as follows

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

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

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

$$\mathbb{E}\, g_1(X) = \pi(I_0, K_1, r, \sigma, T) \quad \text{and} \quad \mathbb{E}\, g_3(X) = \psi(I_0, K_1, K_2, r, \sigma, T)$$

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

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

and
$$\psi(x, K_1, K_2, r, \sigma, T) = \pi(x, K_2, r, \sigma, T) - \pi(x, K_1, r, \sigma, T).$$

Then, one sets

$$\begin{aligned} g_2(X) &= e^{-rT/2}\pi(I_{\frac{T}{2}}, K_1, r, \sigma, T/2), \\ g_4(X) &= e^{-rT/2}\psi(I_{\frac{T}{2}}, K_1, K_2, r, \sigma, T/2). \end{aligned}$$

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

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

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

$$g_0(X) = -e^{-rT}\frac{I_T}{I_0}\mathbf{1}_{\{I_T \in [K_1, K_2]\}}. \tag{4.1}$$

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

$$s_0 = 100, \qquad K_1 = 98, \qquad K_2 = 102, \qquad r = 5\%, \qquad \sigma = 20\%, \qquad T = 2. \tag{4.2}$$

Finally, let $x^N = (x_j^N)$ be the N–optimal quantizer of X. We will compute the quantized versions of $\mathbb{E}g_i(X)$, $i = 0, \ldots, 4$:

$$\mathbb{E}\, g_i(\widehat{X}) = \sum_{j=1}^{N} \mathbb{P}_X(C_j(x^N))\, g_i(x_j^N), \tag{4.3}$$

where $\widehat{X}$ denotes the Voronoï quantization of ξ. The comparison with the Monte Carlo estimator

$$\widehat{\mathbb{E}\, g_i(X)}_N = \frac{1}{N}\sum_{k=1}^{N} g_i(\xi_k), \qquad \xi_k \ i.i.d., \ \xi_1 \sim \mathcal{N}(0; I_d), \tag{4.4}$$

is carried out as follows: we computed (a proxy of the) the standard deviation $\widehat{\sigma(g_i)(X)}_N$ of the above estimator (4.4) using a $M = 10\,000$ trial Monte Carlo simulation and we compared it with the quantization error.

• Graphical tests (dimensionality effect): one sets

$$\text{Absolute error}(N) = |B\&S \text{ Reference value} - \text{ Quantized value(N)}|.$$

In the figures 10 and 11 below is drawn the graph of the $N \mapsto$ Absolute error(N) in a log-log scale for functions g_2 and g_4 in dimensions $d = 2, 3, 4, 5, 6$, its least square regression line (dotted line) and the $\log(N) \mapsto -\frac{1}{2}\log(N) + \log \sigma(\widehat{g_i(X)})_N$ (continuous line). The theoretical slope of the regression line should be $1/d$ or $2/d$ according to the regularity of the function g_i. In the smooth case, this theoretical $2/d$ slope appears in the convex case (g_2, see Figure 10) but is significantly improved in the case of the difference of two convex functions (g_4, see Figure 11). In the Lipschitz continuous setting (corresponding to functions g_1 and g_3 not depicted here), one observes that the slopes are closer to $2/d$ than to $1/d$: this is probably due to the fact that functions g_1 and g_3 are "essentially" smooth except for one single point. This is in fact a very common situation in applications. Furthermore, one verifies in Figure 11 (e) that, in the case of the difference of two convex functions, numerical quantization behaves better than the Monte Carlo method – for the accuracy threshold set at one standard deviation – in dimension $d = 6$ as long as N is lower than a critical number $N_{6,c}$. This is a very common feature of the method which may justify in some special cases the use of optimal quantization for numerical integration in dimensions higher than $d = 4$ (when many integrals have to be computed with respect to the same distribution measure).

• Numerical tests: In Table 1 below we extract some of the above results to provide numerical values for the errors. In the second column are displayed the B&S price using the numerical values specified in (4.2). In the third and fourth columns are displayed the quantized values computed owing to (4.3) and the relative errors with respect to the B&S price. Finally, in the two last columns, we have written down a proxy of the standard deviation of estimator (4.4) and the *ratio*

$$\frac{|B\&S \text{ Reference value} - \text{Quantized value(N)}|}{\sigma(\widehat{g_i(X)})_N}$$

to measure the error induced by the quantization in the scale of the MC estimator standard deviation. The lines of Table 1 represent the different functions g_i labelled with respect to their structures and their smoothness.

Table 1 illustrates a phenomenon widely observed when integrating functions by quantization: differences of convex (*DiffConv*) functions behave better than (composition of) convex (*Conv*) functions, $\mathcal{C}^1_{Lip}$ (in fact $\mathcal{C}^\infty$) functions behave better than Lipschitz continuous (*Lip*) functions, as predicted by (2.5). These numerical tests suggest that being the difference of two convex functions is more prominent than smoothness. The behaviour of quantized integration along discontinuous functions (like the indicator function g_0, *Disc*) seems to highly depend on the integrated function itself and it seems difficult to draw general rules at this stage.

References

[1] V. Bally, G. Pagès (2000). A quantization algorithm for solving multi-dimensional discrete time optimal stopping problems, pre-print n^0 628, Laboratoire de Probabilités & Modèles aléatoires, Université Paris 6 (France), to appear in *Bernoulli*.

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

[3] V. Bally, G. Pagès, J. Printems (2002). A quantization tree method for pricing and hedging multi-dimensional American options, pre-print n^{0}753, Laboratoire de Probabilités & Modèles aléatoires, Université Paris 6 (France), submitted to *Mathematical Finance*.

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

[5] C. Bouton, G. Pagès (1997), About the multidimensional Competitive Learning Vector Quantization algorithm with constant gain, *The Annals of Applied Probability*, **7**, n^{0}3, pp.679-710.

[6] P. Cohort (2003), Limit Theorems for the Random Normalized Distortion, forthcoming in *The Annals of Applied Probability*.

[7] S. Delattre, J.C. Fort, G. Pagès (2002), Local distortion and μ-mass of the cells of one dimensional asymptotically optimal quantizers, pre-print n^{0}710, Laboratoire de Probabilités & Modèles aléatoires, Université Paris 6 (France), submitted to *Communication in Statistics*.

[8] M. Duflo (1998), *Algorithmes stochastiques*, coll. Mathématiques & Applications, **23**, Springer-Verlag, Berlin, 1996.

[9] J.C. Fort, G. Pagès (1995), On the *a.s.* convergence of the Kohonen algorithm with a general neighborhood function, *The Ann. of Applied Proba.*, **5**, $n^0$4, pp.1177-1216.

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

[11] A. Gersho, R. Gray (eds.) (1982), *IEEE Transactions on Information Theory, Special Issue on Quantization*, **28**.

[12] A. Gersho, R. Gray (1992, 6^{th} edition, 1999), *Vector Quantization and Signal Compression*, Kluwer, Boston, 732p.

[13] S. Graf, H. Luschgy (2000), *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics n^{0}1730, Springer, Berlin, 230p.

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

[15] H.J. Kushner, D.S. Clark (1978), *Stochastic Approximation for Constrained and Unconstrained Systems*, Applied Math. Science Series, **26**, Springer.

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

[17] D. Lamberton, G. Pagès (1996), On the critical points of the 1-dimensional Competitive Learning Vector Quantization Algorithm, *Proceedings of the ESANN'96*, Bruges, D Facto, Brussels, (Belgium).

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

[19] G. Pagès, H. Pham (2001), A quantization algorithm for multidimensional stochastic control problems, pre-print n^{0}697, Laboratoire de Probabilités & Modèles aléatoires, Universités Paris 6/7 (France), submitted to *Stochastics & Dynamics*.

[20] G. Pagès, H. Pham (2002), Optimal quantization methods for nonlinear filtering with discrete-time observations, pre-print n^{0}778, Laboratoire de Probabilités & Modèles aléatoires, Universités Paris 6/7 (France), submitted to *Bernoulli*.

[21] G. Pagès, H. Pham, J. Printems (2003), Optimal quantization methods and applications to numerical problems in finance, pre-print n^{0}813, Laboratoire de Probabilités & Modèles aléatoires, Universités Paris 6/7 (France), to appear in *Handbook of Numerical Methods in Finance*, Birkhauser.

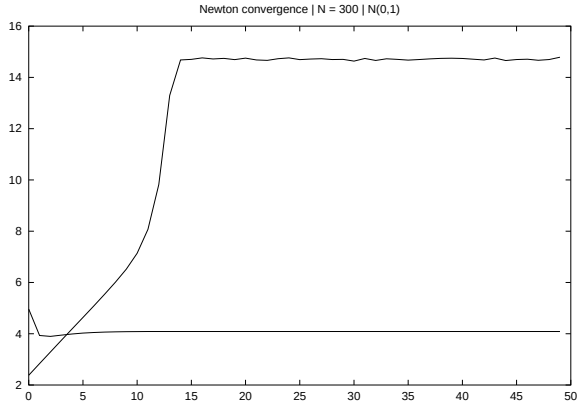


Figure 1: $n \mapsto \log_{10} |\mathrm{d}D_N^X(x^n)|$ *for* $N = 300$.

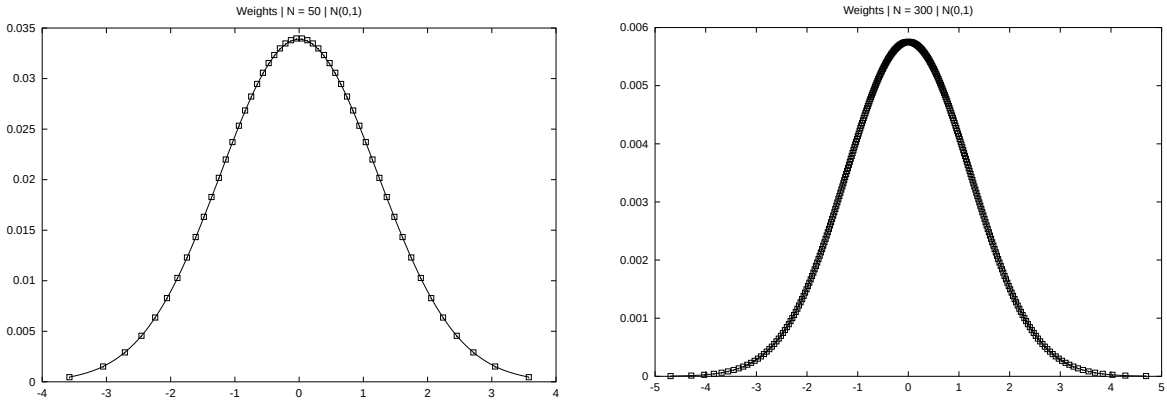


Figure 2: $\mathbb{P}_X$*-mass of the Voronoi cells* $C(x_i^*)$ *as a function of the quantizer components* x_i^* ($\square$), $i = 1, \ldots, N$, $N = 50$ *and* 300. *Functions* $x \mapsto \exp(-x^2/3)/29.5$ ($N = 50$, —) *and* $x \mapsto \exp(-x^2/3)/173$ ($N = 300$, —).

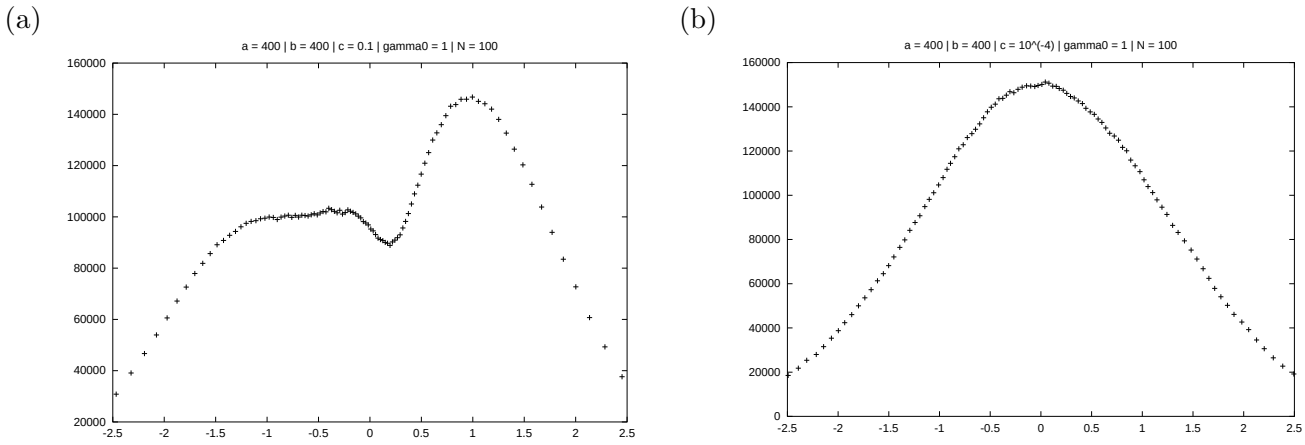


Figure 3: (a) Counter k_i plotted as a function of x_i obtained in dimension 1 after 10^7 trials with $\gamma_0 = 1$, $a = 400$ and $b = 10^{-1}$. The value of $D_N(x^*)$ is 1.60(-2). (b) Quantizer obtained in 1-dimension after 10^7 trials with $\gamma_0 = 1$, $a = 400$ and $b = 10^{-3}$. The value of $D_N(x^*)$ is 1.57(-2).

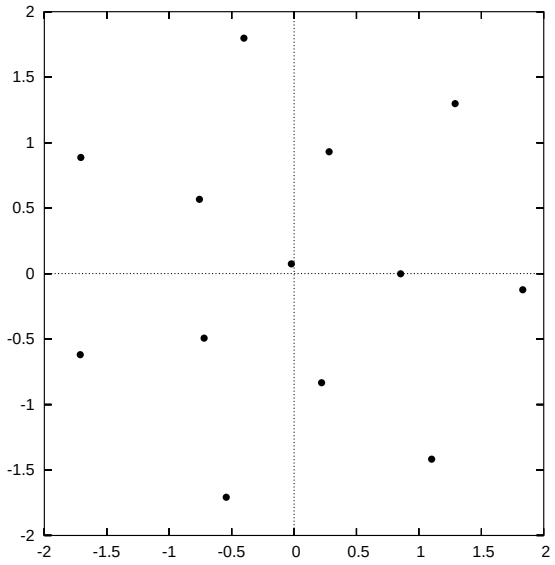


Figure 4: Quantizer with $N = 13$ obtained after 10^6 trials of the randomly initialized $CLVQ$ algorithm (3.7) followed by $M_{Lloyd} = 10$ Lloyd's method I. Except for the origin, its components make up a regular centered pentagon and a regular centered heptagon.

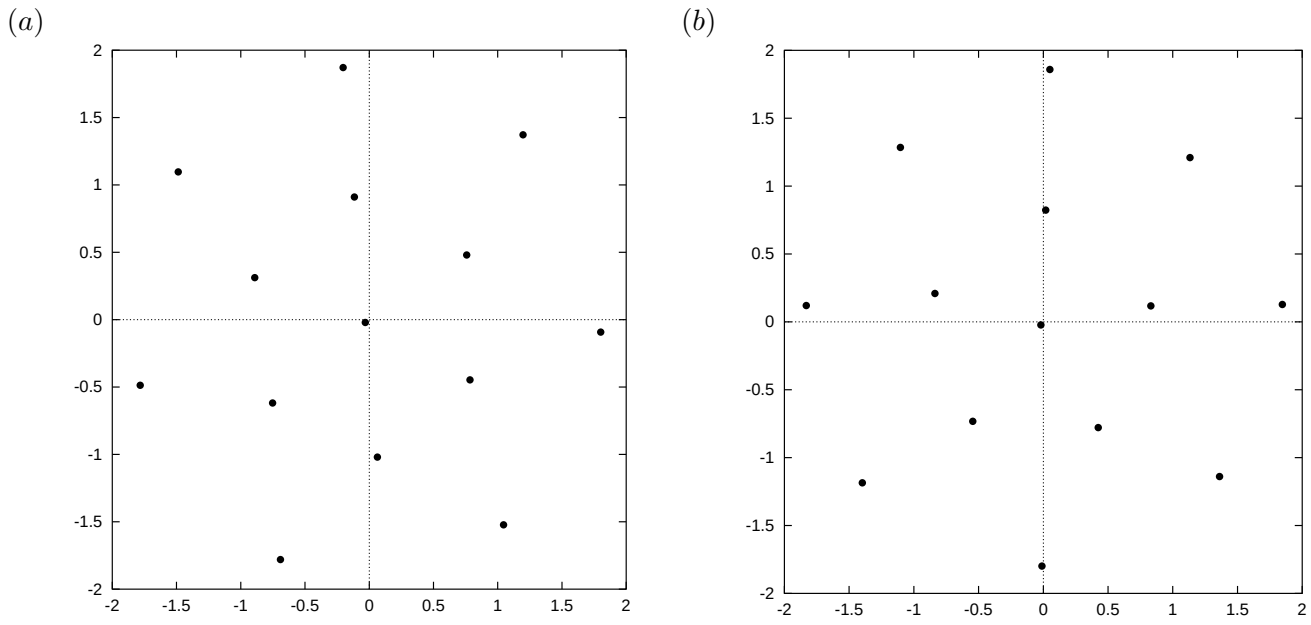


Figure 5: (a) Quantizer with $N = 14$ obtained from the one with $N = 13$ and the point 0 after 10^6 trials of $CLVQ$ algorithm (3.7) with $\gamma_0 \approx \min_{i \neq j} |x_i - x_j|/2$ followed by 10 Lloyd's method I. (b) Quantizer with $N = 14$ obtained after 10^6 trials of the $CLVQ$ algorithm (3.7) followed by $M_{Lloyd} = 10$ Lloyd's methods I.

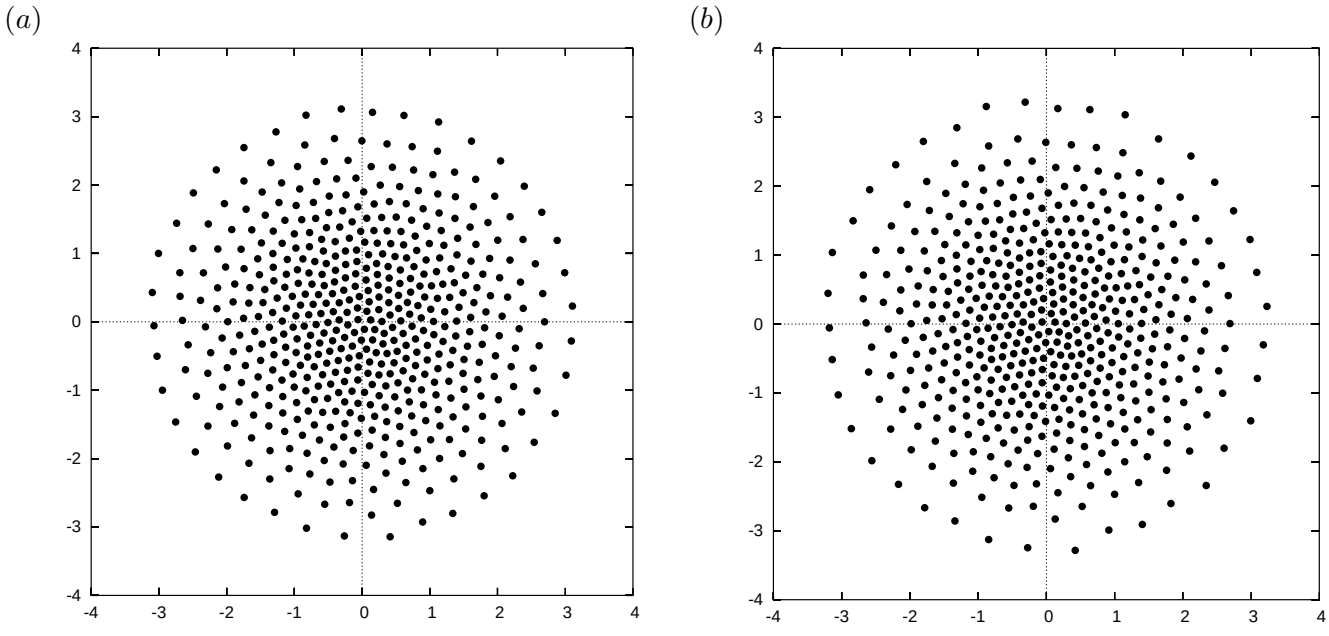


Figure 6: (a) Quantizer with $N = 500$. $D_N^X = 7.08(-3)$. Truncated case. (b) Quantizer with $N = 500$. $D_N^X = 8.56(-3)$. Non-truncated case.

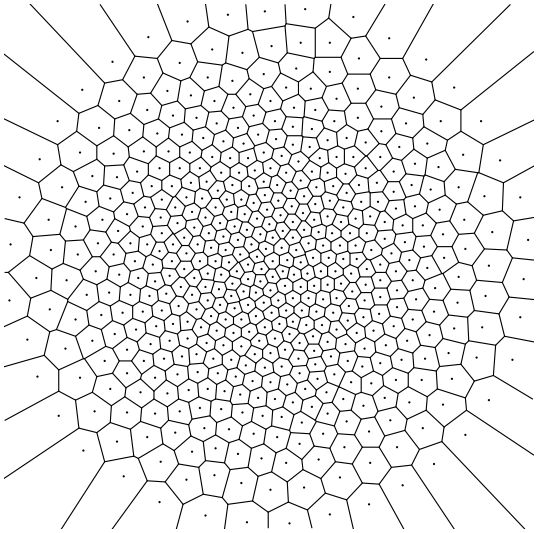


Figure 7: Optimal quantizer of Figure 6(a) with its Voronoi tessellation. Truncated case.

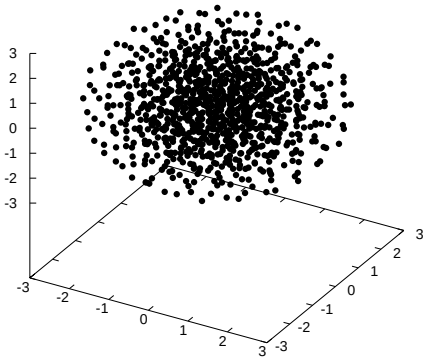


Figure 8: 1000–quantizer of the Normal law in $\mathbb{R}^3$. The value of the distortion obtained is $D_N^X = 5.45(-2)$. Non truncated case.

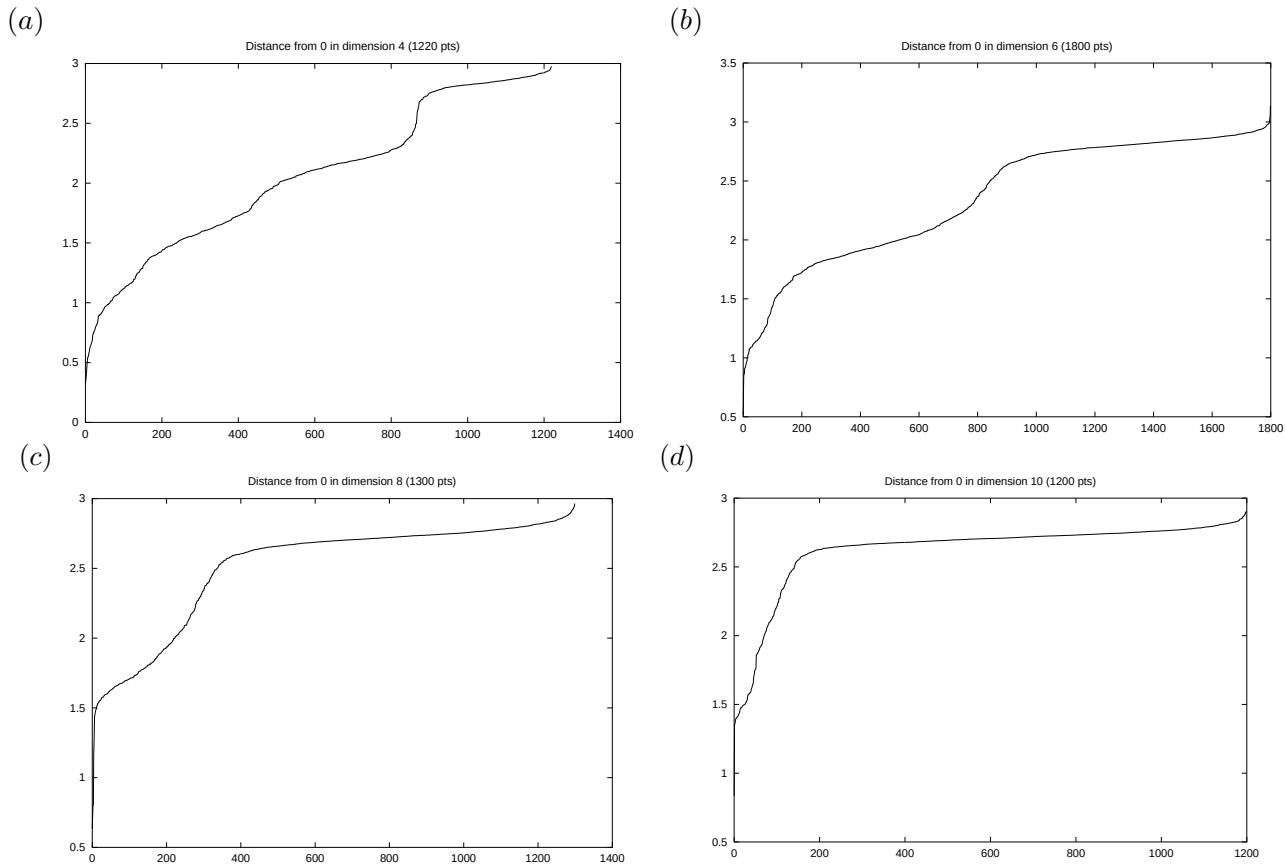
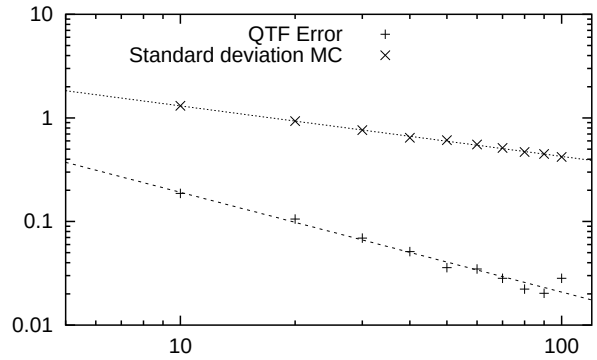
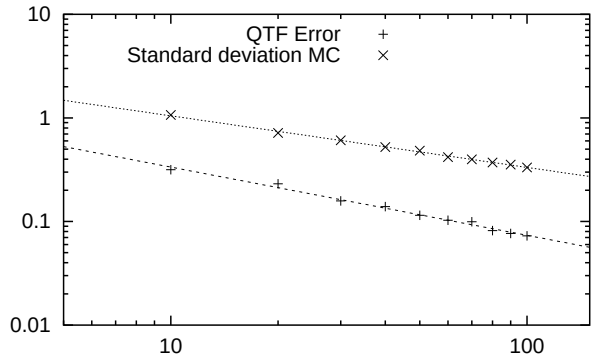


Figure 9: Radii of optimal quantizers in dimension $d = 4, 6, 8, 10$ with N from 1 200 to 1 800. Drawing of $i \mapsto |x_i|$, we can guess 4 layers of points in dimension 4, 3 in dimension 6, 2 in dimension 8 and 1 in dimension 10.

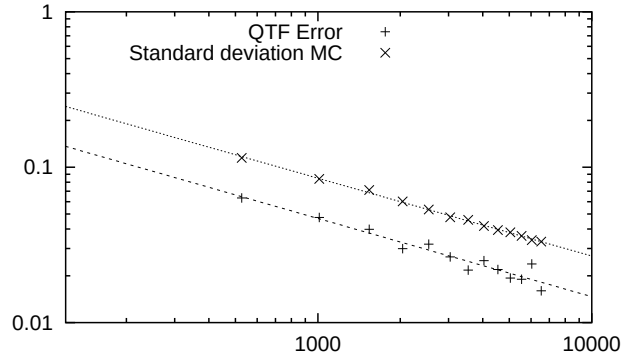
(a) $d = 2$. Slope of + plot : $0.963 \approx 1.926/d$



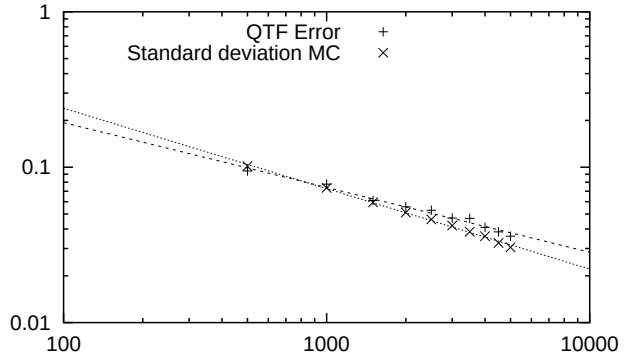
(b) $d = 3$. Slope of + plot : $0.658 \approx 1.974/d$



(c) $d = 4$. Slope of + plot : $0.504 \approx 2.016/d$



(d) $d = 5$. Slope of + plot : $0.417 \approx 2.085/d$



(e) $d = 6$. Slope of + plot : $0.337 \approx 2.022/d$

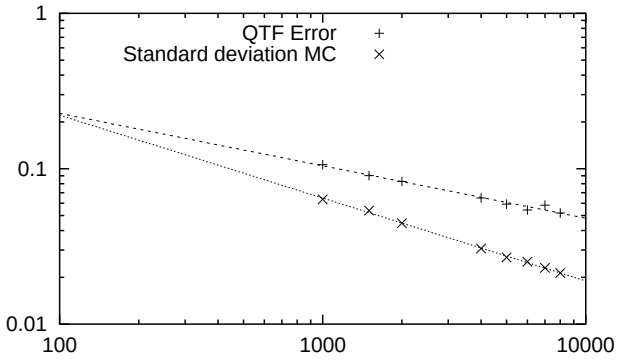
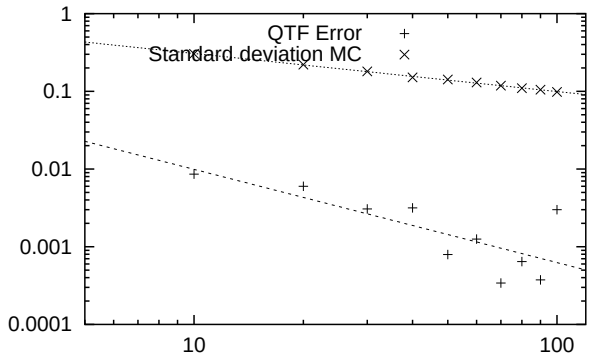
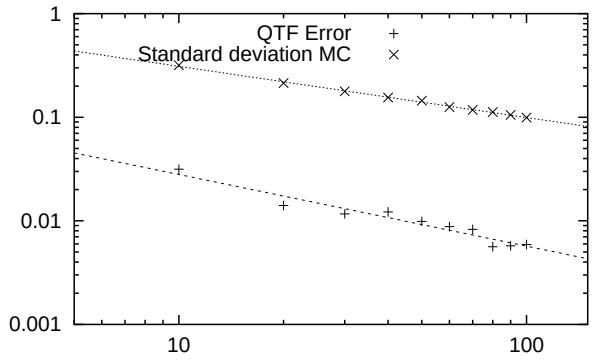


Figure 10: Linear regression in log-log scale of $N \mapsto |\mathbb{E}g_2(\widehat{Z}) - \widehat{\mathbb{E}g_2(Z)}_N|$. In a) $d = 2$; b) $d = 3$; c) $d = 4$; d) $d = 5$; e) $d = 6$.

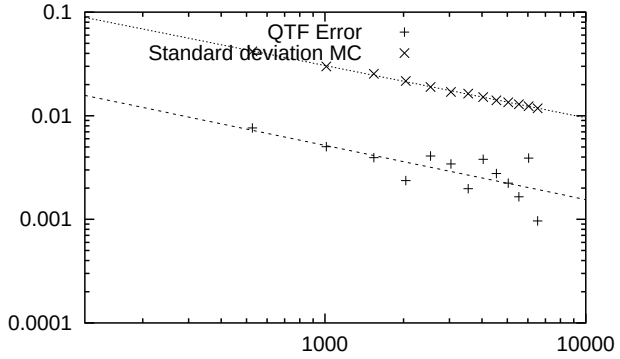
(a) $d = 2$. Slope of + plot : $1.20 \approx 2.40/d$



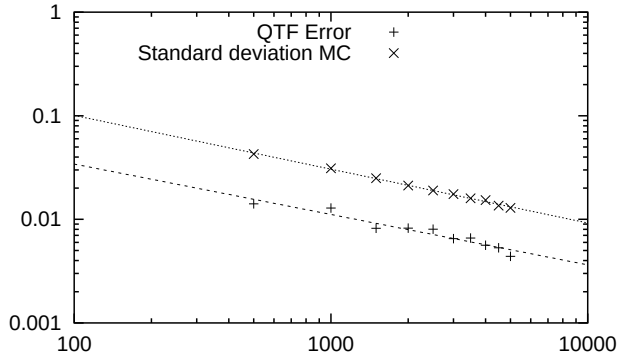
(b) $d = 3$. Slope of + plot : $0.692 \approx 2.076/d$



(c) $d = 4$. Slope of + plot : $0.523 \approx 2.092/d$



(d) $d = 5$. Slope of + plot : $0.487 \approx 2.435/d$



(e) $d = 6$. Slope of + plot : $0.379 \approx 2.274/d$.

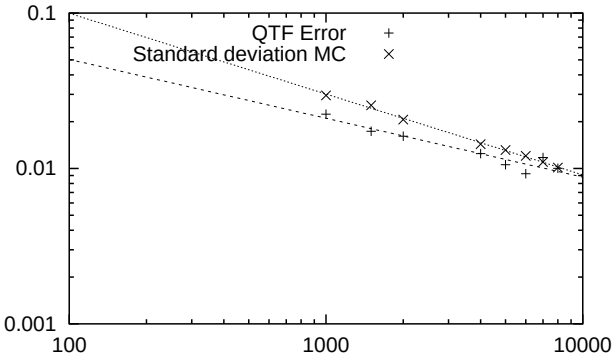


Figure 11: Linear regression in log-log scale of $N \mapsto |\mathbb{E}g_4(\widehat{Z}) - \widehat{\mathbb{E}g_4(Z)}_N|$. In ($a$) $d = 2$; (b) $d = 3$; (c) $d = 4$; (d) $d = 5$; (e) $d = 6$.

Table 1: Value of absolute error with respect to the MC standard deviation for maximal values of N in dimension 2, 4 and 6.

$d = 2$ & $N = 600$ $\mathbb{E}\, g_i(Z)$	$B\&S$ Reference value	Quantized value	Relative error	MC St Dev	Absolute Error/StD
Lip & *Conv* $(i = 1)$	3.672905	3.66233	0.29 %	2.76(-1)	0.03827
$\mathcal{C}^\infty$ & *Conv* $(i = 2)$	3.672905	3.66776	0.14 %	1.77(-1)	0.02904
Lip & *DiffConv* $(i = 3)$	1.383143	1.38388	0.05 %	6.93(-2)	0.01063
$\mathcal{C}^\infty$ & *DiffConv* $(i = 4)$	1.383143	1.38310	0.003 %	4.21(-2)	0.00102
Disc $(i = 0)$	-0.068907	-0.0689169	0.01 %	9.73(-3)	0.00102
$d = 4$ & $N = 6540$ $\mathbb{E}\, g_i(Z)$	$B\&S$ Reference	Quantized value	Relative error	MC St Dev	Absolute Error/StD
Lip & *Conv* $(i = 1)$	2.076954	2.04709	1.44 %	5.46(-2)	0.54762
$\mathcal{C}^\infty$ & *Conv* $(i = 2)$	2.076954	2.06092	0.77 %	3.32(-2)	0.48193
Lip & *DiffConv* $(i = 3)$	1.216210	1.21303	0.26 %	2.09(-2)	0.15215
$\mathcal{C}^\infty$ & *DiffConv* $(i = 4)$	1.216210	1.21524	0.08 %	1.18(-2)	0.08186
Disc $(i = 0)$	$-0,093039$	$-0,0908095$	2.40 %	3.46(-3)	-0.6446
$d = 6$ & $N = 8000$ $\mathbb{E}\, g_i(Z)$	$B\&S$ Reference	Quantized value	Relative error	MC St Dev	Absolute Error/StD
Lip & *Conv* $(i = 1)$	1.395727	1.29660	7.10 %	3.80(-2)	2.60789
$\mathcal{C}^\infty$ & *Conv* $(i = 2)$	1.395727	1.34381	3.72 %	2.14(-2)	2.42523
Lip & *DiffConv* $(i = 3)$	1.094376	1.08037	1.28 %	1.83(-2)	0.76503
$\mathcal{C}^\infty$ & *DiffConv* $(i = 4)$	1.094376	1.08436	0.91 %	1.01(-2)	0.99010
Disc $(i = 0)$	-0.108825	-0.109751	0.85 %	3.19(-3)	0.29028