New approach to greedy vector quantization

RANCY EL NMEIR $^{1,2,*}$ HARALD LUSCHGY$^{3,†}$ and GILLES PAGÈS$^{2,‡}$

1 Sorbonne Université, Laboratoire de Probabilités, Statistiques et Modélisations, Campus Pierre et Marie Curie, case 158, 4, pl. Jussieu, F-75252 Paris Cedex 5, France.
E-mail: *rancynmeir@gmail.com; †gilles.pages@upmc.fr
2 Université Saint-Joseph de Beyrouth, Laboratoire de Mathématiques et Applications, B.P. 11-514 Riad El Solh Beyrouth 1107 2050, Liban.
3 Universität Trier, FB IV-Mathematik, D-54286 Trier, Germany.
E-mail: ‡lushg@uni-trier.de

Abstract We extend some rate of convergence results of greedy quantization sequences already inves-
tigated in 2015. We show, for a more general class of distributions satisfying a certain control, that
the quantization error of these sequences has an optimal rate of convergence and that the distortion
mismatch property is satisfied. We will give some non-asymptotic Pierce type estimates. The recursive
character of greedy vector quantization allows some improvements to the algorithm of computation
of these sequences and the implementation of a recursive formula to quantization-based numerical
integration. Furthermore, we establish further properties of sub-optimality of greedy quantization se-
quences.

Keywords: Greedy quantization sequence; rate optimality; Lloyd’s algorithm; distortion mismatch;
quantization-based numerical integration; quasi-Monte Carlo methods

2010 AMS Classification: 60G15, 60G35, 41A25, 65D30

1. Introduction

Let $d \geq 1$, $r \in (0, +\infty)$ and $L^r_{\mathbb{R}^d}(\mathbb{P})$ (or simply $L^r(\mathbb{P})$) the set of $d$-dimensional random variables $X$ defined on the probability space $\Omega, \mathcal{A}, \mathbb{P}$ such that $\mathbb{E}\|X\|^r < +\infty$ where $\|\|$ denotes any norm on $\mathbb{R}^d$. We denote $\mathbb{P} = \mathbb{P}_X$ the probability distribution of $X$. Optimal vector quantization is a technique derived from signal processing, initially devised to optimally discretize a continuous (stationary) signal for its transmission. Originally developed in the 1950s (see Gersho A. and Gray R.M. (1998)), it was introduced as a cubature formula for numerical integration in the early 1990s (see Pagès G. (1998)) and for approximation of conditional expectations in the early 2000s for financial applications (see Bally V. et al. (2001); Bally V. and Pagès G. (2003)). Its goal is to find the best approximation of a continuous probability distribution by a discrete one, or in other words, the best approximation of a multidimensional random vector $X$ by a random variable $Y$ taking at most a finite number $n$ of values.

Let $\Gamma = \{x_1, \ldots, x_n\}$ be a $d$-dimensional grid of size $n$. The idea is to approximate $X$ by $q(X)$, where $q$ is a Borel function defined on $\mathbb{R}^d$ and having values in $\Gamma$. If we consider, for $q$, the nearest neighbor projection $\pi_\Gamma : \mathbb{R}^d \rightarrow \Gamma$ defined by $\pi_\Gamma(\xi) = \sum_{i=1}^n x_i 1_{W_i(\Gamma)}(\xi)$, where $(W_i(\Gamma))_{1 \leq i \leq n}$
is a Borel partition of $\mathbb{R}^d$ satisfying
\begin{equation}
W_i(\Gamma) \subset \{ \xi \in \mathbb{R}^d : \|\xi - x_i\| \leq \min_{j \neq i} \|\xi - x_j\| \}, \quad i = 1, \ldots, n, \tag{1.1}
\end{equation}
and called a Voronoï partition induced by $\Gamma$, then the Voronoï quantization of $X$ is defined by
\[ \hat{X}^\Gamma = \pi_\Gamma(X) := \sum_{i=1}^n x_i 1_{W_i(\Gamma)}(X). \]
We will write, most of the times, $\hat{X}$ instead of $\hat{X}^\Gamma$ when there is no need for specifications.

The $L^r$-quantization error associated to the grid $\Gamma$ is defined, for every $r \in (0, +\infty)$, by
\begin{equation}
\varepsilon_r(\Gamma, X) = \|X - \pi_\Gamma(X)\|_r = \min_{1 \leq i \leq n} \|X - x_i\|_r \tag{1.2}
\end{equation}
where $\|\cdot\|_r$ denotes the $L^r(\mathbb{P})$-norm (or quasi-norm if $0 < r < 1$). As $\varepsilon_r(\Gamma, X)$ only depends on the distribution $P$ of the random vector $X$, we will also denote this quantity $\varepsilon_r(\Gamma, P)$. Consequently, the optimal quantization problem comes down to finding the grid $\Gamma$ that minimizes this error among all grids with size at most $n$. This leads to introduce
\[ \varepsilon_{r,n}(X) = \inf_{\Gamma, \text{card}(\Gamma) \leq n} \varepsilon_r(\Gamma, X). \]
It has been shown (see Graf S. and Luschgy H. (2000); Pagès G. (2015, 2018)) that this problem admits a solution, i.e. the infimum holds as a minimum, and that the $L^r$-optimal quantization error $\varepsilon_{r,n}(X)$ converges to 0 when the size $n$ goes to $+\infty$. The rate of convergence is given by two well known results exposed in the following theorem.

**Theorem 1.1.**
(a) Zador’s Theorem (see Zador P.L. (1982)) : Let $X \in L^{r+\eta}_d(\mathbb{P})$, $\eta > 0$, with distribution $P$ having the following decomposition $P = h.\lambda_d + \nu$ where $\lambda_d$ is the Lebesgue measure on $\mathbb{R}^d$ and $\nu \perp \lambda_d$ (singular). Then,
\[ \lim_{n \to +\infty} n^{\frac{1}{2}} \varepsilon_{r,n}(X) = Q_r(P)^{\frac{1}{2}} \]
where $Q_r(P) = J_{r,d}^{\frac{1}{2}}(\|h\|_{L^{r+\eta}(\lambda_d)})$ and $J_{r,d} = \inf_{n \geq 1} n^{\frac{1}{2}} \varepsilon_{r,n}(U([0,1]^d)) \in (0, +\infty)$.
(b) Extended Pierce’s Lemma (see Luschgy H. and Pagès G. (2008)): Let $r, \eta > 0$. There exists a constant $\kappa_{d,r,\eta} \in (0,+\infty)$ such that,
\[ \forall n \geq 1, \quad \varepsilon_{r,n}(X) \leq \kappa_{d,r,\eta} \sigma_{r+\eta}(X)n^{-\frac{1}{2}} \]
where, for every $r \in (0, +\infty)$,
\[ \sigma_r(X) = \inf_{a \in \mathbb{R}^d} \|X - a\|_r \tag{1.3} \]
is the $L^r$-standard deviation of $X$.

However, the numerical implementation of multidimensional optimal quantizers requires the computation of grids of size $N \times d$ which becomes too expensive when $N$ or $d$ increase. Hence,
there is a need to provide a sub-optimal solution to the quantization problem which is easier to handle and whose convergence rate remains similar (or comparable) to that induced by optimal quantizers. A so-called greedy version of optimal vector quantization has been developed in Luschgy H. and Pagèes G. (2015). It consists in building a sequence of points \((a_n)_{n \geq 1}\) in \(\mathbb{R}^d\) which is recursively optimal step by step, in the sense that it minimizes the \(L^r\)-quantization error at each iteration. This means that, having the first \(n\) points \(a^{(n)} = \{a_1, \ldots, a_n\}\) for \(n \geq 1\), we add, at the \((n+1)\)-th step, the point \(a_{n+1}\) solution to

\[
an_{n+1} \in \arg\min_{\xi \in \mathbb{R}^d} d_r(a^{(n)} \cup \{\xi\}, X),
\]

noting that \(a^{(0)} = \emptyset\), so that \(a_1\) is simply an \(L^r\)-median of the distribution \(P\) of \(X\). The sequence \((a_n)_{n \geq 1}\) is called an \(L^r\)-optimal greedy quantization sequence for \(X\) or its distribution \(P\). The idea to design such an optimal sequence, which will hopefully produce quantizers with a rate-optimal behavior as \(n\) goes to infinity, is very natural and may be compared to sequences with low discrepancy in quasi-Monte Carlo methods when working on the unit cube \([0, 1]^d\).

In fact, such sequences have already been developed and investigated in an \(L^1\)-setting for compactly supported distributions \(P\) as a model of short term experiment planning versus long term experiment planning (see Brancolini A. et al. (2009)). In Luschgy H. and Pagèes G. (2015), the authors investigated independently a greedy version of vector quantization for \(L^r\)-random vectors taking values in \(\mathbb{R}^d\), for numerical integration purposes. They showed that the problem (1.4) admits at least one solution \((a_n)_{n \geq 1}\) when \(X\) is an \(\mathbb{R}^d\)-valued random vector (the existence of such sequences can be proved in Banach spaces but, in this paper, we will only focus on \(\mathbb{R}^d\)). This sequence may not be unique since greedy quantization depends on the symmetry of the distribution (consider for example the \(\mathcal{N}(0,1)\) distribution). However, note that, if the norm \(\|\cdot\|\) is strictly convex and \(r > 1\), then the \(L^r\)-median is unique. They also showed that the \(L^r\)-quantization error converges to 0 when \(n\) goes to infinity and, if \(\text{supp}(P)\) contains at least \(n\) elements, then the sequence \(a^{(n)}\) lies in the convex hull of \(\text{supp}(P)\). The authors also showed that these sequences have an optimal rate of convergence to zero, compared to optimal quantizers, and satisfy the distortion mismatch problem, i.e. the property that the optimal rate of \(L^r\)-quantizers holds for \(L^s\)-quantizers for \(s > r\). The proofs were based on the integrability of the \(b\)-maximal functions associated to an \(L^r\)-optimal greedy quantization sequence \((a_n)_{n \geq 1}\) given, for \(b \in (0, +\infty)\), by

\[
\forall \xi \in \mathbb{R}^d, \quad \Psi_b(\xi) = \sup_{n \in \mathbb{N}} \frac{\lambda_d \left( B(\xi, b \text{dist}(\xi, a^{(n)})) \right)}{P(\xi, b \text{dist}(\xi, a^{(n)}))}.
\]

In this paper, we will extend those rate of convergence and distortion mismatch results to a much larger class of functions. Instead of maximal functions, we will rely on a new micro-macro inequality involving an auxiliary probability distribution \(\nu\) on \(\mathbb{R}^d\). When \(\nu\) satisfies an appropriate control on balls, defined later in section 2, we show that the rate of convergence of the \(L^r\)-quantization error of greedy sequences is \(O(n^{-\frac{1}{2}})\), just like the optimal quantizers. Furthermore, considering appropriate auxiliary distributions \(\nu\) satisfying this control allows us to obtain Pierce type, and hybrid Zador-Pierce type, \(L^r\)-rate optimality results of the error quantization, instead of only Zador type results as given in Luschgy H. and Pagèes G. (2015).

A very important field of applications is to use these greedy sequences instead of \(n\)-optimal quantizers in quantization-based numerical integration schemes. In fact, the size of the grids
used in these procedures is large in a way that the RAM storing of the quantization tree may exceed the storage capacity of the computing device. Using greedy quantization sequences will dramatically reduce this drawback. One demanding application that we can cite is the approximation of the solutions of Reflected Backward Stochastic Differential Equations (RBSDEs), including the pricing of American options, in El Nmeir R. and Pagès G. (in progress), where greedy quantization proves itself to be quite performing compared to other types of quantization and more generally, to other usual numerical methods. The computation of greedy quantizers is performed by algorithms, detailed in an extended version of Luschgy H. and Pagès G. (2015) on arXiv, allowing also the computation of the weights \( p^n_i \) of the Voronoï cells of the sequence \( a^{(n)} \). These quantities are mandatory for the greedy quantization-based numerical integration to approximate an integral \( I \) of a function \( f \) on \( \mathbb{R}^d \) by the cubature formula

\[
I(f) \approx \sum_{i=1}^{n} p^n_i f(a^{(n)}_i)
\]

where, for every \( i \in \{1, \ldots, n\} \), \( p^n_i = P(X \in W_i(a^{(n)})) \) represents the weight of the \( i^{th} \) Voronoï cell corresponding to the increasing reordering \( \{a^{(n)}_1, \ldots, a^{(n)}_n\} \) of the greedy quantization sequence \( a^{(n)} = \{a_1, \ldots, a_n\} \).

Compared to other methods of numerical integration, such as quasi-Monte Carlo methods (QMC), the quantization-based methods present an advantage in terms of convergence rate, since QMC, for example, is known to induce an \( O\left(\frac{\log(n+1)}{n^2}\right) \) convergence rate when integrating Lipschitz functions (see Proinov P.D. (1988)) while quantization -based numerical integration produces an \( O(n^{-\frac{3}{4}}) \) rate (see Pagès G. (2018)). However, it seems to have a drawback which is the computation of the non-uniform weights \( (p^n_i)_{1 \leq i \leq n} \), unlike the uniform weights in QMC (equal to \( \frac{1}{n} \)). In this paper, we expose how the recursive character of greedy quantization provides several improvements to the algorithm, making it more advantageous. Moreover, this character induces the implementation of a recursive formula for numerical integration, that can replace the usual cubature formula, reducing the time and cost of the computations. This recursive formula will be introduced first in the one-dimensional case, and then extended to the multi-dimensional case for product greedy quantization sequences, computed from one-dimensional sequences, used to reduce the cost of implementations while always preserving the recursive character.

The paper is organized as follows. We first show that greedy quantization sequences are rate optimal in section 2 where we extend the results presented in Luschgy H. and Pagès G. (2015). The distortion mismatch problem will be solved and extended in section 3. In section 4, we present the improvements applied to the algorithm of designing the greedy sequences, as well as the new approach for greedy quantization-based numerical integration. Numerical examples will illustrate the advantages brought by this new approach in section 5. Finally, section 6 is devoted to some numerical conclusions about further properties of greedy quantization sequences such as the sub-optimality, the convergence of empirical measures, the stationarity (or quasi-stationarity) and the discrepancy, to see to what extent greedy sequences can be close to optimality.

**Notations:** We denote by \( \lambda_d \) the Lebesgue measure and by \( B(x, r) \) the ball w.r.t. \( \|\cdot\| \) centered at \( x \) and with radius \( r \) where \( \|\cdot\| \) denotes any norm on \( \mathbb{R}^d \). And we introduce \( V_d = \lambda_d(B(0, 1)) \).
2. Rate optimality: Universal non-asymptotic bounds

We consider a vector space \((\mathbb{R}^d, \| \cdot \|)\). In Luschgy H. and Pagès G. (2015), the authors presented the rate optimality of \(L^r\)-greedy quantizers in the sense of Zador’s Theorem 1.1 based on the integrability of the \(b\)-maximal function \(\Psi_y(\xi)\) defined by (1.5). Here, we present Pierce type non-asymptotic estimates relying on micro-macro inequalities applied to a certain class of auxiliary probability distributions \(\nu\). Different specifications of \(\nu\) lead to various versions of the so-called Pierce’s Lemma.

We recall, first, a micro-macro inequality that will be be used to prove the first result.

**Proposition 2.1.** Assume \( \int |x|^r dP(x) < +\infty \). Then, for every probability distribution \( \nu \) on \((\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))\), every \( c \in (0, \frac{1}{2}) \) and every \( n \geq 1 \)

\[
e_r(a^{(n)}, P)^r - e_r(a^{(n+1)}, P)^r \geq \frac{(1 - c)^r - c^r}{(c + 1)^r} \int \nu \left( B \left( x, \frac{c}{c+1} d \left( x, a^{(n)} \right) \right) \right) d(x, a^{(n)})^r dP(x).
\]

**Proof.** Step 1: Micro-macro inequality. Let \( \Gamma \subset \mathbb{R}^d \) be a finite quantizer of a random variable \( X \) with distribution \( P \) and \( \Gamma_1 = \Gamma \cup \{ y \}, y \in \mathbb{R}^d \). For every \( c \in (0, \frac{1}{2}) \), we have \( B(y, cd(y, \Gamma)) \subset W_y(\Gamma_1) \), where \( W_y(\Gamma_1) \) is the Voronoï cell associated to the centroid \( y \) having in mind that \((W_z(\Gamma_1))_{z \in \Gamma_1} \) makes up a partition of \( \mathbb{R}^d \), as defined by (1.1). Hence, for every \( x \in B(y, cd(y, \Gamma)) \), \( d(x, \Gamma) \geq d(y, \Gamma) - \|x - y\| \geq (1 - c)d(y, \Gamma) \). Consequently,

\[
e_r(\Gamma, P)^r - e_r(\Gamma \cup \{y\}, P)^r = \int_{\mathbb{R}^d} (d(x, \Gamma)^r - d(x, \Gamma_1)^r) dP(x)
\]

\[
\geq \int_{W_y(\Gamma_1)} (d(x, \Gamma)^r - \|x - y\|^r) dP(x)
\]

\[
\geq \int_{B(y, cd(y, \Gamma))} ((1 - c)^r - c^r)d(y, \Gamma)^r dP(x).
\]

Finally, we obtain the micro-macro inequality

\[
e_r(\Gamma, P)^r - e_r(\Gamma \cup \{y\}, P)^r \geq (1 - c)^r - c^r \int P B \left( y, cd(y, a^{(n)}) \right) d(x, a^{(n)})^r.
\]

Step 2. We apply the micro-macro inequality (2.1) to the greedy quantization sequence \( a^{(n)} \) and notice that \( e_r(a^{(n+1)}, P) \leq e_r(a^{(n)} \cup \{y\}, P) \) for every \( y \in \mathbb{R}^d \). This yields, for every \( c \in (0, \frac{1}{2}) \) and every \( y \in \mathbb{R}^d \),

\[
e_r(a^{(n)}, P)^r - e_r(a^{(n+1)}, P)^r \geq (1 - c)^r - c^r \int P B \left( y, cd(y, a^{(n)}) \right) d(y, a^{(n)})^r.
\]

We integrate this inequality with respect to \( \nu \) to obtain

\[
e_r(a^{(n)}, P)^r - e_r(a^{(n+1)}, P)^r \geq (1 - c)^r - c^r \int \nu B \left( y, cd(y, a^{(n)}) \right) d(y, a^{(n)})^r d\nu(y).
\]
Lemma 2.2. The proof of the next proposition giving upper bounds of the L⁰ quantization error of the greedy quantization sequence. We begin by a technical lemma which will be used in

In many applications, ε

Proof. We rely on the following Bernoulli inequalities, for every n ≥ 1, d(x, a(n)) ≤ \frac{1}{c+1}d(x, a(n))

and \|x - y\| ≤ \frac{c}{c+1}d(x, a(n)) ≤ cd(y, a(n)). Then,

\int \int P(B(y, cd(y,a(n))))d(y,a(n)) d\nu(y) = \int \int 1_{F_1}(x,y)d(y,a(n))^r d\nu(y)dP(x)

\geq \frac{1}{(c+1)^r} \int \int 1_{F_2}(x,y)d(x,a(n))^r d\nu(y)dP(x)

= \frac{1}{(c+1)^r} \int \nu\left(B\left(x, \frac{c}{c+1}d(x,a(n))\right)\right)d(x,a(n))^r dP(x).

In order to prove the rate optimality of the greedy quantization sequences and obtain a non-asymptotic Pierce type result, we will consider auxiliary probability distributions \nu satisfying, for every n ≥ 1 by construction of the greedy quantization sequence. We begin by a technical lemma which will be used in the proof of the next proposition giving upper bounds of the L⁰ quantization error c_r(a(n),P) depending on the integral of the inverse of g_r w.r.t. P.

Lemma 2.2. Let C, ρ ∈ (0, +∞) be some real constants and \(x_n\) \(n \geq 1\) be a non-negative sequence satisfying, for every n ≥ 1, \(x_{n+1} \leq x_n - C x_n^{1+\rho}\). Then for every n ≥ 1,

\[(n-1)^{\frac{1}{\rho}} x_n \leq \left(\frac{1}{C\rho}\right)^{\frac{1}{\rho}}.

Proof. We rely on the following Bernoulli inequalities, for every x ≥ -1,

\[ (1 + x)^\rho \geq 1 + \rho x, \quad \text{if } \rho \geq 1, \quad \text{and} \quad (1 + x)^\rho \leq 1 + \rho x, \quad \text{if } 0 < \rho < 1.\]

These inequalities can be obtained by studying the function f defined for every x ∈ (-1, +∞) by f(x) = (1 + x)^\rho -(1 + \rho x). Using that (x_n)_{n \geq 1} is non-increasing and that x_n > 0 for every n ≥ 1, it follows from the assumption made on (x_n)_{n \geq 1} that

\[\frac{1}{x_{n+1}^\rho} \geq \frac{1}{x_n^\rho(1 - C x_n^\rho)} \geq \frac{1}{x_n^\rho(1 + C x_n^\rho)^\rho}.\]
− If $\rho \geq 1$, the Bernoulli inequalities imply $\frac{1}{x_{n+1}^r} \geq \frac{1}{x_n^r} (1 + C \rho x_n^\rho) = \frac{1}{x_n^r} + C \rho$. By induction, one obtains
\[ \frac{1}{x_n^r} \geq \frac{1}{x_1^r} + (n - 1) C \rho \geq (n - 1) C \rho \]
to deduce the result easily.
− If $0 < \rho < 1$, then $-C \rho x_n^\rho \geq -1$ for every $n \geq 1$, and the result is deduced by using the Bernoulli inequality and then reasoning by induction. \hfill \Box

**Proposition 2.3.** Let $P$ be such that $\int_{\mathbb{R}^d} \|x\|^r dP(x) < +\infty$. For any auxiliary distribution $\nu$ and Borel function $g_\epsilon : \mathbb{R}^d \to (0, +\infty)$, $\epsilon \in (0, \frac{1}{3})$, satisfying (2.2),
\[ \forall n \geq 2, \quad e_r(a^{(n)}, P) \leq \varphi_r(\epsilon) \frac{1}{d} V_d \left( \frac{r}{d} \right)^{\frac{1}{r}} \left( \int g_\epsilon x^{-\frac{r}{3}} dP \right)^{\frac{1}{r}} (n - 1)^{-\frac{1}{3}} \] (2.3)
where
\[ \varphi_r(u) = \left( \frac{1}{d} - u^r \right) u^d. \] (2.4)

**Proof.** We may assume that $\int g_\epsilon x^{-\frac{r}{3}} dP < +\infty$. Assume $c \in (0, \frac{\epsilon}{1 - \frac{\rho}{3}}) \cap (0, \frac{1}{2})$ so that $\frac{c}{c + 1} \leq \epsilon$. Moreover $d(x, a^{(n)}) \leq d(x, a_1)$ since $a_1 \in a^{(n)}$. Consequently, for any such $c$, $\frac{c}{c + 1} d(x, a^{(n)}) \leq \epsilon \|x - a_1\|$ so that, by (2.2), there exists a function $g_\epsilon$ such that
\[ \nu \left( B \left( x, \frac{c}{c + 1} d(x, a^{(n)}) \right) \right) \geq V_d \left( \frac{c}{c + 1} \right)^d d(x, a^{(n)}) g_\epsilon(x). \]
Then, noting that $\left( \frac{c}{c + 1} \right)^d - c^d \geq \frac{1}{d} \left( \frac{c}{c + 1} \right)^r > 0$, since $c \in (0, \frac{1}{2})$, Proposition 2.1 yields
\[ e_r(a^{(n)}, P)^r - e_r(a^{(n+1)}, P)^r \geq V_d \varphi_r \left( \frac{c}{c + 1} \right) \int g_\epsilon(x) d(x, a^{(n)})^d + r dP(x) \] (2.5)
where $\varphi_r(u) = \left( \frac{1}{d} - u^r \right) u^d$, $u \in (0, \frac{1}{2})$. Applying the reverse Hölder inequality with the conjugate Hölder exponents $p = -\frac{d}{r}$ and $q = \frac{r}{1 - r}$ yields
\[ e_r(a^{(n)}, P)^r - e_r(a^{(n+1)}, P)^r \geq V_d \varphi_r \left( \frac{c}{c + 1} \right) \left( \int g_\epsilon(x)^{-\frac{d}{r}} dP(x) \right)^{-\frac{d}{r}} (e_r(a^{(n)}, P)^r)^{1 + \frac{d}{r}}. \]
Then, applying Lemma 2.2 to the sequence $x_n = e_r(a^{(n)}, P)^r$ with $\rho = \frac{d}{r}$ and $C = V_d \varphi_r \left( \frac{c}{c + 1} \right)$
\[ \times \left( \int g_\epsilon(x)^{-\frac{d}{r}} dP(x) \right)^{-\frac{d}{r}}, \]
and then raising to the power $r$, one obtains, for every $c \in (0, \frac{1}{2})$,
\[ e_r(a^{(n)}, P) \leq V_d \frac{1}{d} \left( \frac{r}{d} \right)^{\frac{1}{r}} \varphi_r \left( \frac{c}{c + 1} \right) \left( \int g_\epsilon x^{-\frac{r}{3}} dP \right)^{\frac{1}{r}} (n - 1)^{-\frac{1}{3}}. \]
To deduce the result, we specify $c$ as $c = \frac{\epsilon}{1 - \frac{\rho}{3}}$, so that $\frac{c}{c + 1} = \epsilon$. This choice is explained by the fact that, since in most applications $\epsilon \mapsto \left( \int g_\epsilon x^{-\frac{r}{3}} dP \right)^{\frac{1}{r}}$ is increasing on $(0, 1/3)$, we are led to
study \( \varphi_r\left(\frac{c}{c+1}\right)^{-\frac{1}{3}} \) subject to the constraint \( c \in (0, \frac{1}{c+1}) \cap (0, \frac{1}{2}) \). And, since \( \varphi_r \) is increasing in the neighborhood of 0 and \( \varphi_r(0) = 0 \), one has, for every \( \varepsilon \in (0, \frac{1}{3}) \) small enough, \( \varphi_r\left(\frac{c}{c+1}\right) \leq \varphi_r(\varepsilon) \), for \( c \in (0, \frac{1}{c+1}) \).

By specifying the measure \( \nu \) and the function \( g_\varepsilon \), we obtain two first natural versions of the Pierce Lemma.

**Theorem 2.1** (Pierce’s Lemma). (a) Assume \( \int_{\mathbb{R}^d} \|x\|^r dP(x) < +\infty \). Let \( \delta > 0 \). Then \( e_r(a_1, P) \) is equal to the \( L^{r+\delta} \)-standard deviation \( \sigma_r(P) \) of \( P \) defined by (1.3) and

\[
\forall n \geq 2, \quad e_r(a(n), P) \leq \kappa_{d,\delta,r}^{G,P} \sigma_r(P)(n-1)^{-\frac{1}{3}}
\]

where \( \kappa_{d,\delta,r}^{G,P} \leq V_d^{-\frac{1}{3}} \left( \frac{r}{d} \right)^{\frac{1}{d}} \left( 1 + \frac{r}{\delta} \right)^{1+\frac{2}{d}} \left( \int_{\mathbb{R}^d} (\|x\| \vee 1)^{-d-\frac{dr}{\delta}} dx \right)^{\frac{1}{d}} \min_{\varepsilon \in (0,\frac{1}{3})} \left( (1 + \varepsilon) \varphi_r(\varepsilon)^{-\frac{1}{6}} \right).

(b) Assume \( \int_{\mathbb{R}^d} \|x\|^r dP(x) < +\infty \). Let \( \delta > 0 \). Then

\[
\forall n \geq 2, \quad e_r(a(n), P) \leq \kappa_{d,r,\delta}^{G} \left( \int (\|x - a_1\| \vee 1)^{r} (\log(\|x - a_1\| \vee \varepsilon))^\frac{r}{r+\delta} dP(x) \right)^\frac{1}{r} (n-1)^{-\frac{1}{3}}
\]

where \( \kappa_{d,r,\delta}^{G} \leq 2 \left( \frac{r}{d} \right)^{\frac{1}{d}} \left( \log(\|x\| \vee \varepsilon) \right)^{-\frac{1}{r+\delta}} (\int (\|x\| \vee 1)^{-d} (\log(\|x\| \vee \varepsilon) \left[ (1 + \varepsilon) \varphi_r(\varepsilon)^{-\frac{1}{6}} \right] (\int (1 \vee \|x\|)^{-d} (\log(\|x\| \vee \varepsilon) \left[ (1 + \varepsilon) \varphi_r(\varepsilon)^{-\frac{1}{6}} \right]) \right)^{\frac{1}{r}}

In particular, if \( \int_{\mathbb{R}^d} \|x\|^r dP(x) < +\infty \), then

\[
\limsup_n n^{\frac{1}{3}} \text{sup}\{e_r(a(n), P) : (a_n) \text{ L}^r \text{-optimal greedy sequence for P}\} < +\infty.
\]

**Proof.** (a) Let \( \delta > 0 \) be fixed. We set \( \nu(dx) = \gamma_{r,\delta}(x) \lambda_d(dx) \) where

\[
\gamma_{r,\delta}(x) = \frac{K_{\delta,r}}{(1 \vee \|x - a_1\|)^{d(1 + \frac{1}{r})}} \quad \text{with} \quad K_{\delta,r} = \left( \int \frac{dx}{(1 \vee \|x\|)^{d(1 + \frac{1}{r})}} \right)^{-1} < +\infty
\]

is a probability density with respect to the Lebesgue measure on \( \mathbb{R}^d \).

Let \( \varepsilon \in (0,1) \) and \( t > 0 \). For every \( x \in \mathbb{R}^d \) such that \( \varepsilon \|x - a_1\| \geq t \) and every \( y \in B(x,t) \), \( \|y - a_1\| \leq \|y - x\| + \|x - a_1\| \leq (1 + \varepsilon)\|x - a_1\| \) so that

\[
\nu(B(x,t)) \geq \frac{K_{\delta,r} V_d t^d}{(1 \vee [(1 + \varepsilon)\|x - a_1\|])^{d(1 + \frac{1}{r})}}.
\]

Hence, (2.2) is verified with \( g_\varepsilon(x) = \frac{K_{\delta,r}}{(1 \vee [(1 + \varepsilon)\|x - a_1\|])^{d(1 + \frac{1}{r})}} \), so we can apply Proposition 2.3. We have

\[
\int g_\varepsilon(x)^{-\frac{r}{r+\delta}} dP(x) \leq K_{\delta,r}^{-\frac{r}{r+\delta}} \int (1 \vee (1 + \varepsilon)\|x - a_1\|)^{r+\delta} dP(x)
\]

so that, applying \( L^{r+\delta} \)-Minkowski inequality, one obtains

\[
\left( \int g_\varepsilon(x)^{-\frac{r}{r+\delta}} dP(x) \right)^{\frac{1}{r}} \leq K_{\delta,r}^{-\frac{1}{3}} (1 + (1 + \varepsilon)\sigma_r)^{1+\frac{\delta}{r}}.
\]
Consequently, by Proposition 2.3, for $\varepsilon \in (0, 1/3)$,

$$e_r(a(n), P) \leq V_d^{-\frac{1}{d}} \left( \frac{r}{d} \right)^{\frac{1}{d}} K_{\delta,r}^{-\frac{1}{d}} (1 + (1 + \varepsilon)\sigma_{r+\delta})^{1+\frac{\delta}{r}} \varphi_r(\varepsilon)^{-\frac{1}{d}}(n-1)^{-\frac{1}{d}} \tag{2.6}$$

Now, we introduce an equivariance argument. For $\lambda > 0$, let $X_\lambda := \lambda (X - a_1) + a_1$ and $(a_{\lambda,n})_{n \geq 1} := (\lambda(a_{n-1} - V) + a_1)_{n \geq 1}$. It is clear that $(a_{\lambda,n})_{n \geq 1}$ is an $L^r$-optimal greedy sequence for $X_\lambda$ and $e_r(a^{(n)}, X) = \frac{1}{\lambda} e_r(a^{(n)}, X_\lambda)$. Plugging this in inequality (2.6) yields

$$e_r(a(n), P) \leq V_d^{-\frac{1}{d}} \left( \frac{r}{d} \right)^{\frac{1}{d}} K_{\delta,r}^{-\frac{1}{d}} \left( (1 + (1 + \varepsilon)\lambda \sigma_{r+\delta})^{1+\frac{\delta}{r}} \varphi_r(\varepsilon)^{-\frac{1}{d}}(n-1)^{-\frac{1}{d}} \right)$$

$$= V_d^{-\frac{1}{d}} \left( \frac{r}{d} \right)^{\frac{1}{d}} K_{\delta,r}^{-\frac{1}{d}} \left( (1 + (1 + \varepsilon)\lambda \sigma_{r+\delta})^{1+\frac{\delta}{r}} \varphi_r(\varepsilon)^{-\frac{1}{d}}(n-1)^{-\frac{1}{d}} \right)$$.  

Finally, one deduces the result by setting $\lambda = \frac{1}{\delta (1 + \varepsilon)\sigma_{r+\delta}}$.

(b) Let $\delta > 0$ be fixed. We set $\nu(dx) = \gamma_{r,\delta}(x)\lambda_{d}(dx)$ where

$$\gamma_{r,\delta}(x) = \frac{K_{\delta,r}}{(1 \vee \|x - a_1\|)^d (\log(\|x - a_1\| \vee e))^{1+\frac{\delta}{r}}} \tag{2.7}$$

with $K_{\delta,r} = \left( \int \frac{dx}{(1 \vee \|x\|^{d(\log(\|x\| \vee e))^{1+\frac{\delta}{r}}})} \right)^{-1} < +\infty$, is a probability density w.r.t. $\lambda_d$. Let $\varepsilon \in (0, 1)$ and $t > 0$. For every $x \in \mathbb{R}^d$ such that $\varepsilon \|x - a_1\| \geq t$ and every $y \in B(x,t)$, $\|y - a_1\| \leq \|y - x\| + \|x - a_1\| \leq (1 + \varepsilon)\|x - a_1\|$ so that

$$\nu(B(x,t)) \geq \frac{K_{\delta,r} V_d d}{(1 \vee (1 + \varepsilon)\|x - a_1\|)^d (\log((1 + \varepsilon)\|x - a_1\| \vee e))^{1+\frac{\delta}{r}}}$$

$$\geq \frac{K_{\delta,r} V_d d}{(1 + \varepsilon)^d (1 \vee \|x - a_1\|)^d (\log(\|x - a_1\| \vee e))^{1+\frac{\delta}{r}}}$$

since $\log(1 + \varepsilon) \leq e$.

$$\geq \frac{K_{\delta,r} V_d d}{2^{1+\frac{\delta}{r}} (1 + \varepsilon)^d (1 \vee \|x - a_1\|)^d (\log(\|x - a_1\| \vee e))^{1+\frac{\delta}{r}}}$$

so we can apply Proposition 2.3. We have

$$\left( \int g_\varepsilon(x)^{-\frac{\delta}{r}} dP(x)^{\frac{1}{2}} \right)^{\frac{1}{2}} \leq \frac{(1 + \varepsilon)^{\frac{1+\frac{\delta}{r}}{2}}}{K_{\delta,r}^{\frac{1}{2}}} \left( \int (1 \vee \|x - a_1\|)^r (\log(\|x - a_1\| \vee e))^{\delta+\frac{\delta}{r}} dP(x)^{\frac{1}{2}} \right)^{\frac{1}{2}}.$$
Consequently, one applies Proposition 2.3 to deduce the first part. For the second part of the proposition, we start by noticing that

\[(1 \lor \|x - a_1\|)^r \leq (1 + \|x\| + \|a_1\|)^r \leq 2^{(r-1)\log(\|x\| + (1 + \|a_1\|)^r)}\]

and

\[\log(\|x - a_1\| \lor e) \leq \log(\|x\| \lor e) + \frac{\|a_1\| \lor e}{\|x\| \lor e} \leq \log_+ \|x\| + 1 + \frac{\|a_1\| \lor e}{e}
\]

where \(\log_+ u = \log u \mathbb{1}_{u \geq 1}\), so that

\[(1 \lor \|x - a_1\|)^r (\log(\|x - a_1\| \lor e))^{\delta + \frac{2}{r}} \leq 2^{(r-1)\log_+(\|x\|^{\delta + \frac{2}{r}} + A_2 \|x\|^r)} + A_1 \log_+ \|x\|^{\frac{2}{r} + \delta} + A_1 A_2\]

where \(A_1 = (1 + \|a_1\|)^r\) and \(A_2 = \left(1 + \frac{\|a_1\| \lor e}{e}\right)^{\frac{2}{r} + \delta}\). Since \(\log \|x\|^{\delta + \frac{2}{r}} = \frac{1}{r} \left(\frac{\|x\|}{d} + \delta\right) \log_+ \|x\|^r\), then \(\log_+ \|x\|^{\delta + \frac{2}{r}} = \frac{1}{r} \left(\frac{\|x\|}{d} + \delta\right) \log_+ \|x\|^r\). Moreover, \(\log_+ \|x\|^r \leq \|x\|^r - 1\) if \(\|x\|^r \geq 1\) and equal to zero otherwise so

\[\log_+ \|x\|^{\delta + \frac{2}{r}} \leq \frac{1}{r} \left(\frac{\|x\|}{d} + \delta\right) \left(\|x\|^r - 1\right) + \frac{1}{r} \left(\frac{\|x\|}{d} + \delta\right) (1 + \|x\|^r).
\]

Consequently,

\[(1 \lor \|x - a_1\|)^r (\log(\|x - a_1\| \lor e))^{\delta + \frac{2}{r}} \leq 2^{\beta \log_+(\|x\|^{\delta + \frac{2}{r}} + A'_1 \|x\|^r + A'_2)}\]

where \(\beta = (r-1) + (\frac{\|x\|}{d} + \delta - 1) +, A'_1 = A_2 + \frac{1}{r} \left(\frac{\|x\|}{d} + \delta\right) A_1\) and \(A'_2 = \frac{1}{r} \left(\frac{\|x\|}{d} + \delta\right) A_1 + A_1 A_2\). The result is deduced from the fact that \(\sup\{\|a_1\| : a_1 \in \arg\min_{\xi \in [P]} R_{\xi}(\xi, P)\} < +\infty\) (see (Graf S. and Luschgy H., 2000, Lemma 2.2)) and \(\kappa_{d, \epsilon, \rho}\) does not depend on \(a_1\). \(\square\)

**Remark 2.1.** One checks that \(\varphi_\epsilon\) attains its maximum at \(\frac{1}{3} \left(\frac{d}{\pi + r}\right)^{\frac{1}{2\epsilon}}\) on \((0, \frac{1}{3})\), so one concludes that \(\min_{\epsilon \in (0, \frac{1}{3})} \left((1 + \epsilon)\varphi_\epsilon(\epsilon)^{-\frac{1}{2}}\right) \leq \left(1 + \frac{1}{3} \left(\frac{d}{\pi + r}\right)^{\frac{1}{2}}\right)^{\frac{3}{2}} \left(1 + \frac{d}{\pi + r}\right)^{\frac{3}{2}}\).

At this stage, one could wonder whether it is possible to have a kind of hybrid Zador-Pierce result (i.e. a result which is both non-asymptotic like in Pierce’s Lemma, and controlled by \(\|h\|_{L_{\pi + r}}\) like in Zador’s Theorem) where, if \(P = h.\lambda_d\), one has

\[e_\epsilon(a^{(n)}(\lambda_d), P) \leq C \|h\|_{L_{\pi + r}}^{\frac{d}{\pi + r}} n^{-\frac{1}{d}}\]

for some real constant \(C\). To this end, we will not directly rely on the criteria (2.2). Instead, we have to consider a specific auxiliary distribution \(\nu\) defined by

\[\nu = \frac{h_{\pi + r}^{\frac{d}{\pi + r}}}{\int h_{\pi + r}^{\frac{d}{\pi + r}} d\lambda_d} \lambda_d.
\]

This is related to the following local growth control condition of densities.
Definition 2.4. Let $A \subset \mathbb{R}^d$. A function $f : \mathbb{R}^d \to \mathbb{R}_+$ is said to be almost radial non-increasing on $A$ w.r.t. $a \in A$ if there exists a norm $\| \cdot \|_0$ on $\mathbb{R}^d$ and real constant $M \in (0, 1]$ such that

$$\forall x, y \in A \setminus \{a\}; \|y - a\|_0 \leq \|x - a\|_0, \quad f(y) \geq M f(x).$$

(2.8)

If (2.8) holds for $M = 1$, then $f$ is called radial non-increasing on $A$ w.r.t. $a$.

Remark 2.2. (a) (2.8) reads $f(B_{\| \cdot \|_0}(a, \|x - a\|_0) \cap A \setminus \{a\}) \geq M f(x)$ for all $x \in A \setminus \{a\}$.

(b) If $f$ is radial non-increasing on $\mathbb{R}^d$ w.r.t. $a \in \mathbb{R}^d$ with parameter $\| \cdot \|_0$, then there exists a non-increasing measurable function $g : (0, +\infty) \to \mathbb{R}_+$ satisfying $f(x) = g(\|x - a\|_0)$ for every $x \neq a$.

(c) From a practical point of view, many classes of distributions satisfy (2.8), e.g. the $d$-dimensional normal distribution $\mathcal{N}(\mu, \Sigma)$ for $h(y) = (2\pi)^{-\frac{d}{2}} \det(\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2} y^T \Sigma^{-1} y}$ and density $f(x) = h(||x - \mu||_2)$ where $\|x\|_0 = \|\Sigma^{-\frac{1}{2}} x\|_2$, and the family of distributions defined by $f(x) \propto e^{-a|x|^b}$, for every $x \in \mathbb{R}^d, a, b > 0$ and $c > -d$, for which one considers $h(u) = u^c e^{-ax^b}$. In the one-dimensional case, we can mention the Gamma distribution, the Weibull distributions, the Pareto distributions and the log-normal distributions.

Theorem 2.2. Assume $P = h, \lambda_d = h \in L^1_{\mathcal{F}, \pi}(\lambda_d)$ and $\int_{\mathbb{R}^d} \|x\|^r dP(x) < +\infty$. Let $a_1$ denote the $L^r$-median of $P$. Assume that $\text{supp}(P) \subset A$ and $a_1 \in A$ for some $A$ star-shaped and peakless with respect to $a_1$ in the sense that

$$p(A, \| \cdot - a_1 \|) := \inf \left\{ \frac{\lambda_d(B(x, t) \cap A)}{\lambda_d(B(x, t))} : x \in A, 0 < t \leq \|x - a_1\| \right\} > 0.$$  (2.9)

Assume $h$ is almost radial non-increasing on $A$ w.r.t. $a_1$ in the sense of (2.8). Then,

$$\forall n \geq 2, \quad e_r(a^{(n)}, P) \leq \kappa \|\| \cdot - a_1\| \|^\frac{r}{2} \frac{\lambda_d}{L^1_{\mathcal{F}, \pi}(\lambda_d)} (n - 1)^{\frac{1}{2}},$$

where $\kappa = \kappa_{d, r, M, C_0, p(A, \| \cdot - a_1 \|)} \leq \frac{2c_0^{\frac{r}{2}}}{d^r M^{d+r} V_d^2 p(A, \| \cdot - a_1 \|)^\frac{r}{2}} \min_{\varepsilon \in (0, \frac{1}{2})} \left[ \varphi_r(\varepsilon)^{-\frac{3}{2}} \right]$ with $C_0 \in [1, +\infty)$ satisfying, for every $x \in \mathbb{R}^d$, $\frac{1}{C_0} \|x\|_0 \leq \|x\| \leq C_0 \|x\|_0$.

Remark 2.3. (a) If $A = \mathbb{R}^d$, then $p(A, \| \cdot - a_1 \|) = 1$ for every $a \in \mathbb{R}^d$.

(b) The most typical unbounded sets satisfying (2.9) are convex cones that is cones $K \subset \mathbb{R}^d$ of vertex 0 with $0 \in K$ ($K \neq \emptyset$) and such that $\lambda x \in K$ for every $x \in K$ and $\lambda \geq 0$. For such convex cones $K$ with $\lambda_d(K) > 0$, we even have that the lower bound

$$p(K) := \inf \left\{ \frac{\lambda_d(B(x, t) \cap K)}{\lambda_d(B(x, t))} : x \in K, t > 0 \right\} = \frac{\lambda_d(B(0, 1) \cap K)}{V_d} > 0.$$

Thus if $K = \mathbb{R}^d$, then $p(K) = 2^{-d}$.

The proof of Theorem 2.2 is based on the following lemma.
Lemma 2.5. Let $\nu = f.\lambda_d$ be a probability measure on $\mathbb{R}^d$ where $f$ is almost radial non-increasing on $A \in B(\mathbb{R}^d)$ w.r.t. $a_1 \in A$, $A$ being star-shaped relative to $a_1$ and satisfying (2.9). Then, for every $x \in A$ and positive $t \in (0, \|x - a_1\|)$,

$$\nu(B(x, t)) \geq M\rho(A, \| - a_1\|)(2C_0^2)^{-d}V_d f(x)t^d$$

where $C_0 \in [1, +\infty)$ satisfies, for every $x \in \mathbb{R}^d$, $\frac{1}{C_0}\|x\|_0 \leq \|x\| \leq C_0\|x\|_0$.

**Proof.** For every $x \in A$ and $t > 0$,

$$\nu(B(x, t)) \geq \int_{B(x, t) \cap A \cap \{f \geq Mf(x)\}} fd\lambda_d \geq Mf(x)\lambda_d(B(x, t) \cap A \cap \{f \geq Mf(x)\})$$

and $B(x, t) \cap (A \setminus \{a_1\}) \cap B_{\|\|_0}(a_1, \|x - a_1\|_0) \subset B(x, t) \cap A \cap \{f \geq Mf(x)\}$. Now, assume $0 < t \leq \|x - a_1\| \leq C_0\|x - a_1\|_0$. Setting $x' := \left(1 - \frac{t}{2C_0\|x - a_1\|_0}\right)x + \frac{t}{2C_0\|x - a_1\|_0}a_1 \in A$ (since $A$ is star-shaped with respect to $a_1$), we notice that, for $y \in B \left(x', \frac{t}{2C_0}\right) \subset B_{\|\|_0}(x', \frac{t}{2C_0})$,

$$\|y - x\| \leq \|y - x'\| + C_0\|x' - x\| \leq \frac{t}{2C_0^2} + C_0\left\|\frac{t}{2C_0\|x - a_1\|_0}(x - a_1)\right\|_0 = \frac{t}{2C_0^2} + \frac{t}{2} \leq t$$

and $\|y - a_1\|_0 \leq \|y - x'\|_0 + \|x' - a_1\|_0 \leq \frac{t}{2C_0} + \left(1 - \frac{t}{2C_0\|x - a_1\|_0}\right)\|x - a_1\|_0 = \|x - a_1\|_0$, so that, $B \left(x', \frac{t}{2C_0}\right) \subset B(x, t) \cap B_{\|\|_0}(a_1, \|x - a_1\|_0)$. Consequently,

$$\nu(B(x, t)) \geq Mf(x)\lambda_d \left( B \left(x', \frac{t}{2C_0}\right) \cap A \right).$$

Moreover, $\frac{t}{2C_0^2} \leq \frac{t}{2} \leq \frac{1}{2}\|x - a_1\| \leq \|x' - a_1\|$. Hence, we have

$$\lambda_d \left( B \left(x', \frac{t}{2C_0}\right) \cap A \right) \geq \rho(A, \| - a_1\|)\lambda_d \left( B \left(x', \frac{t}{2C_0}\right) \right) = \rho(A, \| - a_1\|)(2C_0^2)^{-d}V_d \lambda_d(B(0, 1)).$$

□

**Proof of Theorem 2.2.** Consider $\nu = h_r \lambda_d := \frac{h_r^d}{\int h_r^d d\lambda_d} \lambda_d$. Notice that $h_r$ is almost radial non-increasing on $A$ w.r.t. $a_1$ with parameter $M^d \pi^d r^d$ so that Lemma 2.5 yields for every $x \in A$ and $t \in (0, \|x - a_1\|)$

$$\nu(B(x, t)) \geq M^d \pi^d \rho(A, \| - a_1\|)(2C_0^2)^{-d}V_d h_r(x)t^d.$$  

Consequently, using the fact that

$$\int_{\mathbb{R}^d} h_r^{-\frac{t}{d}}dP = \int_{\mathbb{R}^d} h_r^{-\frac{r}{\pi^d}}hd\lambda_d \times \left( \int_{\mathbb{R}^d} h_r^d d\lambda_d \right)^{\frac{t}{d}} = \left( \int_{\mathbb{R}^d} h_r^d d\lambda_d \right)^{1 + \frac{t}{d}} = \|h\|_{L^d}^{d}(\lambda_d),$$

the assertion follows from Proposition 2.1 with the same reasoning as in the proof of Proposition 2.3. □
Remark 2.4. Note that, by applying Hölder inequality with the conjugate exponents \( p = 1 + \frac{q}{r} \) and \( q = 1 + \frac{d}{r} \), one has

\[
\int_{\mathbb{R}^d} h(\xi)^{\frac{d}{p-r}} d\xi \leq \left( \int_{\mathbb{R}^d} h(\xi)(1 + |\xi|)^{r+\delta} d\xi \right)^{\frac{d}{p-r}} \left( \int_{\mathbb{R}^d} \frac{d\xi}{(1 + |\xi|)^{d(1+\frac{q}{r})}} \right)^{\frac{r}{p-r}} .
\]

Consequently, since \( \int_{\mathbb{R}^d} \frac{d\xi}{(1 + |\xi|)^{d(1+\frac{q}{r})}} < +\infty \), one deduces that \( \| h \|_{\frac{d}{p-r}} = \mathcal{O}(1 + \frac{\delta}{r}) \).

We note that Zador’s Theorem 1.1 implies \( \liminf_n n^{\frac{1}{2}} e_r(a^{(n)}, P) \geq \liminf_n n^{\frac{1}{2}} e_{r,n}(P, \mathbb{R}^d) \geq Q_r(P)^{\frac{1}{2}} \). The next proposition may appear as a refinement of Pierce’s Lemma and Theorem 2.2 in the sense that it gives a lower convergence rate for the discrete derivative of the quantization error, that is its increment.

**Proposition 2.6.** Assume \( \int_{\mathbb{R}^d} \| x \|^r dP(x) < +\infty \). Then,

\[
\liminf_n n^{\frac{1}{2}} \min_{1 \leq i \leq n} \left( e_r(a^{(i)}, P)^r - e_r(a^{(i+1)}, P)^r \right) > 0.
\]

**Proof.** We start by choosing \( N > 0 \) such that \( P(B(0, N)) > 0 \). Proposition 2.1 yields, for every probability measure \( \nu \) on \( \mathbb{R}^d \), for every \( n \geq n_0 \) and \( c \in (0, \frac{1}{2}) \),

\[
e_r(a^{(n)}, P)^r - e_r(a^{(n+1)}, P)^r \geq \frac{(1-c)^r - c^r}{(c+1)^r} \int_{B(0,N) \cap \text{supp}(P)} \nu \left( B \left( x, \frac{c}{c+1} d(x, a^{(n)}) \right) \right) d(x, a^{(n)})^r dP.
\]

We choose \( \nu = U(B(0,N)) \). Then, for every \( x \in B(0, N) \), \( t \leq N \) and \( x' = (1 - \frac{t}{2N}) x \), one has \( B \left( x', \frac{t}{2} \right) \subset B(x, t) \cap B(0, N) \) since, for every \( y \in B \left( x', \frac{t}{2} \right) \),

\[
\| y - x \| \leq \| y - x' \| + \| x' - x \| \leq \frac{t}{2} + \frac{t}{2N} \| x \| \leq t
\]

and

\[
\| y \| \leq \| y - x' \| + \| x' \| \leq \frac{t}{2} + \left( 1 - \frac{t}{2N} \right) \| x \| \leq \frac{t}{2} + \left( 1 - \frac{t}{2N} \right) N = N.
\]

Consequently,

\[
\nu(B(x, t)) \geq \frac{\lambda_d(B(x', \frac{t}{2}))}{\lambda_d(B(0, N))} = (2N)^{-d} t^d.
\]

Moreover, we denote \( C := \sup_{n \geq 1} \max_{x \in B(0,N) \cap \text{supp}(P)} d(x, a^{(n)}) \) which is finite because \( a^{(n)} \in \text{conv}(\text{supp}(P)) \). Consequently, for every \( c \in (0, \frac{1}{2}) \) such that \( \frac{c}{c+1} C \leq N \) and every \( n \geq n_0 \),

\[
e_r(a^{(n)}, P)^r - e_r(a^{(n+1)}, P)^r \geq \frac{(1-c)^r - c^r}{(c+1)^r} \left( \frac{c}{c+1} \right)^d (2N)^{-d} d(x, a^{(n)})^d dP(x)
\]

\[
\geq \frac{\varphi \left( \frac{c}{c+1} \right)}{c+1} (2N)^{-d} P(B(0, N)) e^{d+r}(a^{(n)}, P(\cdot | B(0, N)) ).
\]
where $P(.|A) = \frac{P(\cdot \cap A)}{P(A)}$ denotes the conditional distribution of $P$ given $A$ when $P(A) > 0$. Finally, one deduces the result using that $(\epsilon_n^{d+r}(\mathcal{A}(n), P(.|B(0, N))))_{n \geq 1}$ is nonincreasing and relying on Zador’s Theorem (Theorem 1.1).

3. Distortion mismatch

We address now the problem of distortion mismatch, i.e. the property that the rate optimal decay property of $L^r$-quantizers remains true for $L^s(P)$-quantization error for $s \in (0, +\infty)$. This problem was originally investigated in Graf S. et al. (2008) for optimal quantizers. If $s \leq r$, the monotonicity of the $L^s$-norm as a function of $s$ ensures that any $L^r$-optimal greedy sequence remains $L^s$-rate optimal for the $L^s$-norm. The challenge is when $s$ is larger than $r$. The problem is solved in Luschgy H. and Pagès G. (2015) for $s \in (0, +\infty)$ relying on an integrability assumption of the $b$-maximal function $\Psi_b$. Here, we give an additional nonasymptotic result for $s \in (r, d + r)$, in the same settings as for Proposition 2.3, considering auxiliary probability distributions $\nu$ satisfying (2.2).

**Theorem 3.1.** Let $P$ be such that $\int_{\mathbb{R}^d} \|x\|^{r} dP(x) < +\infty$. Let $s \in (r, d + r)$. Let $(\mathcal{A}_n)$ be an $L^s$-optimal greedy sequence for $P$. For any distribution $\nu$ and Borel function $g : \mathbb{R}^d \to (0, +\infty)$, $\varepsilon \in (0, \frac{1}{3})$, satisfying (2.2), for every $n \geq 3$,

$$e_s(\mathcal{A}_n, P) \leq \kappa \left( \int g \frac{r}{2} d\nu \right)^{\frac{d+r-s}{r}} + \left( \int g \frac{r}{2} d\nu \right)^{\frac{1}{2}} (n - 2)^{\frac{-1}{2}}$$

where $\kappa = \kappa_r, \epsilon = 2 \frac{1}{\eta} \left( \frac{r}{a} \right)^{\frac{d+r-s}{r}} \left( \frac{r}{a} \right)^{\frac{1}{2}} \min_{\varepsilon \in (0, \frac{1}{3})}[\Phi_r(\varepsilon)]^{\frac{1}{2}}$.

**Proof.** We assume $g \in L^{\frac{d}{d+r-s}}(P)$ so that $g \in L^{\frac{d}{d+r-s}}(P)$ since $\frac{d}{d+r-s} \geq \frac{d}{d} = \frac{d}{d}$. Inequality (2.5) from the proof of Proposition 2.3 still holds, i.e.

$$e_r(\mathcal{A}_n, P)^r - e_r(\mathcal{A}_{n+1}, P)^r \geq K_e \int g \frac{r}{2} d\nu \left( \frac{r}{a} \right)^{\frac{d+r-s}{r}} dP(x),$$

with, for every $c \in (0, \frac{r}{1+c}] \cap (0, 1/2)$, $K_e = V_d \Phi_r \left( \frac{c}{1+c} \right)$ where $\Phi_r(u) = \left( \frac{1}{3^r} - u \right) u^d$. The reverse Hölder inequality applied with $p = \frac{d}{d+r-s} \in (0, 1)$ and $q = -\frac{s}{d+r-s} \in (-\infty, 0)$ yields that

$$e_r(\mathcal{A}_n, P)^r - e_r(\mathcal{A}_{n+1}, P)^r \geq C_1 e_s(\mathcal{A}_n, P)^{d+r}$$

where $C_1 = K_e \left( \int g \frac{r}{2} d\nu \right)^{-\frac{d+r-s}{r}}$. Hence, knowing that $k \mapsto e_s(\mathcal{A}_k, P)$ is non-increasing and summing between $n$ and $2n - 1$, we obtain for $n \geq 1$

$$n e_s(a(2n-1), P)^{d+r} \leq \sum_{k=n}^{2n-1} e_s(a(k), P)^{d+r} \leq \frac{1}{C_1} \sum_{k=n}^{2n-1} e_r(a(k), P)^r - e_r(a(k+1), P)^r \leq \frac{1}{C_1} e_r(a(n), P)^r.$$

Finally, since $2 \left[ \frac{n}{2} \right] - 1 \leq n$, we have $e_s(a(n), P) \leq e_s \left( a^2 \left[ \frac{n}{2} \right], P \right)$ and we derive that

$$\frac{n}{2} e_s(a(n), P)^{d+r} \leq \left[ \frac{n}{2} \right] e_s(a(n), P)^{d+r} \leq \left[ \frac{n}{2} \right] e_s \left( a^2 \left[ \frac{n}{2} \right], P \right)^{d+r} \leq \frac{1}{C_1} e_r \left( a^2 \left[ \frac{n}{2} \right], P \right)^r.$$
Consequently, plugging in $C_1$, 
\[ e_s(a(n), P) \leq 2^{\frac{1}{d+r+s}} V_d^{-\frac{1}{d+r+s}} \varphi_r \left( \frac{c}{c+1} \right)^{\frac{1}{d+r+s}} \left( \int g_\epsilon \frac{e}{(d+x)^{\frac{1}{d+r+s}}} dP \right)^{\frac{d+r-s}{d(d+r+s)}} \left( \int g_\epsilon \frac{e}{(d+x)^{\frac{1}{d+r+s}}} dP \right)^{\frac{1}{d+r+s}} \epsilon_r \left( \frac{c}{c+1} \right) P. \]

Consequently, one can deduce from Proposition 2.3, for $n \geq 3$, 
\[ e_s(a(n), P) \leq \frac{2^{\frac{1}{d+r+s}} V_d^{-\frac{1}{d+r+s}}}{V_d^2 d^{\frac{1}{d+r+s}}} \left( \int g_\epsilon \frac{e}{(d+x)^{\frac{1}{d+r+s}}} dP \right)^{\frac{d+r-s}{d(d+r+s)}} \left( \int g_\epsilon \frac{e}{(d+x)^{\frac{1}{d+r+s}}} dP \right)^{\frac{1}{d+r+s}} \varphi_r \left( \frac{c}{c+1} \right) P. \]

Hence, the result follows from the fact that $\varphi_r \left( \frac{c}{c+1} \right) \leq \varphi_r (\epsilon)$ for $c \in (0, \frac{\epsilon}{1-\epsilon}]$.

**Corollary 3.1.** Let $s \in (r, d+r)$. Assume that \( \int \|x\|^{\frac{d}{d+r-s}} \epsilon \|x\|^{\frac{s}{d+r-s}} (1+\frac{d}{d+r}) dP(x) < +\infty \) for some $\delta > 0$, then 
\[ \limsup_n n^{\frac{s}{d+r-s}} \sup \{ e_s(a(n), P) : (a_n)_{L^r} \text{-optimal greedy sequence for } P \} < +\infty. \]

**Proof.** The proof is divided in two steps.

**Step 1:** Let $\delta > 0$ be fixed. We consider $\nu(dx) = \gamma_{r,\delta}(x) \lambda_d(dx)$ where $\gamma_{r,\delta}(x)$ is a probability density with respect to the Lebesgue measure on $\mathbb{R}^d$ defined by (2.7) in the proof of Theorem 2.1(b) and verifying (2.2) with 
\[ g_\epsilon(x) = \frac{K_{\delta,r}}{(1+\epsilon)^d 2^{\delta+r} (1 \vee \|x-a_1\|)^d (\log(\|x-a_1\| \vee \epsilon))^{1+\frac{d}{d+r}}. \]

Consequently, Theorem 3.1 yields, for $n \geq 3$, 
\[ e_s(a(n), P) \leq C_{d,r,\delta} \left( \int (1 \vee \|x-a_1\|)^\tau (\log(\|x-a_1\| \vee \epsilon))^{\delta+r} dP(x) \right)^{\frac{1}{d+r+s}} \times \left( \int (1 \vee \|x-a_1\|)^{\frac{d}{d+r-s}} (\log(\|x-a_1\| \vee \epsilon))^{1+\frac{d}{d+r}} dP(x) \right)^{\frac{d+r-s}{d(d+r+s)}} (n-2)^{-\frac{1}{d+r+s}} \]

where $C_{d,r,\delta} \leq 2^{\frac{\delta}{d+r}} V_d^{-\frac{1}{d+r}} (\frac{2}{d})^{\frac{1}{d+r-s}} K_{\delta,r}^{-\frac{1}{d+r}} \min_{\epsilon \in (0,\frac{1}{2})} [1+\epsilon] \varphi_r (\epsilon)^{-\frac{1}{d+r}} \left[ (1+\epsilon) \varphi_r (\epsilon)^{-\frac{1}{d+r}} \right]$.

**Step 2:** Just as in the proof of Theorem 2.1(b), we have 
\[ (1 \vee \|x-a_1\|)^\tau (\log(\|x-a_1\| \vee \epsilon))^{\delta+r} \leq 2^{\delta} \left( \left\| x \right\|^\tau \log_+ \left\| x \right\|^{\delta+r} + A_1 \left\| x \right\|^\tau + A_2 \right) \]
and, denoting $\beta' = (1 + \frac{d}{d+r}) \frac{\delta+r}{d+r-s}$,
\[ (1 \vee \|x-a_1\|)^{\frac{d}{d+r-s}} (\log(\|x-a_1\| \vee \epsilon))^{\beta'} \leq 2^{\delta} \left( \frac{d}{d+r-s} - 1 \right)_{+} (\beta'-1)_{+} \]
\[ \times \left( \left\| x \right\|^{\frac{d}{d+r-s}} \log_+ \left\| x \right\|^{\beta'} + B_1 \left\| x \right\|^\tau + B_2 \right) \]
where $A_1, A_2, B_1$ and $B_2$ are constants depending only on $r, d, s, \delta$ and $a_1$. Since, $\frac{d}{d+r-s} \geq \frac{\delta}{d+r}$, one has $\frac{d}{d+r-s} > r$ and $\beta' \geq \delta + \frac{\tau}{d+r}$, so that the two above quantities are finite (by the
assumption made in the theorem). The result is deduced from the fact that \( \sup \{ ||a_1|| : a_1 \in \arg \min_{\xi \in \mathbb{R}^d} \epsilon_r(\{\xi\}, P) \} < +\infty. \)

\[ \]  

4. Algorithmics

An important application of quantization is numerical integration. Let us consider the quadratic case \( r = 2 \) and an \( L^2 \)-optimal greedy quantization sequence \( a^{(n)} \) for a random variable \( X \) with distribution \( \mathbb{P}_X = P \). Since we know that \( \epsilon_2(a^{(n)}), X = ||X - \hat{X}^{(n)}||_{L^2(P)} \) converges to 0 when \( n \) goes to infinity, this means that the Voronoï quantization \( \hat{X}^{(n)} \) of \( X \) converges towards \( X \) in \( L^2 \) and hence in distribution. So, one can approximate \( \mathbb{E}[f(X)] \), for every continuous function \( f : \mathbb{R}^d \to \mathbb{R} \), by the following cubature formula

\[
I(f) := \mathbb{E}[f(X)] \approx \sum_{i=1}^{n} p^n_i f(a^{(n)}_i) \tag{4.1}
\]

where, for every \( i \in \{1, \ldots, n\} \), \( p^n_i = P(X \in W_i(a^{(n)})) \) represents the weight of the \( i^{th} \) Voronoï cell corresponding to the increasing reordering \( \{a^{(n)}_1, \ldots, a^{(n)}_n\} \) of the greedy quantization sequence \( a^{(n)} = \{a_1, \ldots, a_n\} \). When the function \( f \) satisfies certain regularities, one establishes error bounds for this quantization-based cubature formula and obtains an \( \mathcal{O}(n^{-\frac{1}{2}}) \) rate of convergence, we refer to Pagès G. (2018) for details. For example, if \( f \) is \( |f|_{\text{Lip}} \)-Lipschitz continuous,

\[
\left| \sum_{i=1}^{n} p^n_i f(a^{(n)}_i) - \mathbb{E}[f(X)] \right| \leq |f|_{\text{Lip}} \mathbb{E}[|X - \hat{X}^{(n)}|^2] = \mathcal{O}(n^{-\frac{1}{2}}).
\]

When working on the unit cube \( [0,1]^d \), it is natural to compare an optimal greedy sequence of the uniform distribution \( \mathcal{U}([0,1]^d) \) and a uniformly distributed sequence with low discrepancy used in the quasi-Monte Carlo method (QMC). A \( [0,1]^d \)-valued sequence \( \xi = (\xi_n)_{n \geq 1} \) is uniformly distributed if \( \mu_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{\xi_k} \) converges weakly to \( \lambda_d \) (where \( \lambda_d \) denotes the Lebesgue measure on \( (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \)). It is well known (see Kuipers L. and Niederreiter H. (1974) for example) that \( (\xi_n)_{n \geq 1} \) is uniformly distributed if and only if

\[
D^*_n(\xi) = \sup_{u \in [0,1]^d} \left| \frac{1}{n} \sum_{i=1}^{n} 1_{\xi_i \in [0,u]^d} - \lambda_d([0,u]^d) \right| \to 0 \quad \text{as} \quad n \to +\infty.
\]

The above modulus is known as the star-discrepancy of \( \xi \) at order \( n \) and can be defined, for fixed \( n \in \mathbb{N} \), for any \( n \)-tuple \( (\xi_1, \ldots, \xi_n) \) whose components \( \xi_k \) lie in \( [0,1]^d \). There exists many sequences (Halton, Kakutani, Faure, Niederreiter, Sobol', see Bouleau N. and Lépine D. (1994); Pagès G. (2018) for example) achieving a \( \mathcal{O}\left(\frac{\log(n+1)^{d-1}}{n}\right) \) rate of decay for their star-discrepancy and it is a commonly shared conjecture that this rate is optimal, such sequences are called \textit{sequences with low discrepancy}. By a standard so-called Hammersley argument, one shows that if a \( [0,1]^d \)-valued sequence \( \xi = (\xi_n)_{n \geq 1} \) has low discrepancy i.e. there exists a real constant \( C(\xi) \in (0, +\infty) \) such that \( D^*_n(\xi) \leq C(\xi) \left( \frac{\log(n+1)}{n} \right)^{d-1} \), for every \( n \geq 1 \), then, for
every \( n \geq 1 \), the \([0,1]^d\)-valued \( n \)-tuple \( \left( (\zeta_k, \frac{k}{n}) \right)_{1 \leq k \leq n} \) satisfies

\[
D^*_{n} \left( \left( (\zeta_k, \frac{k}{n}) \right)_{1 \leq k \leq n} \right) \leq C(\zeta) \left( \frac{\log(n+1)}{n} \right)^{d-1}.
\]

The QMC method finds its gain in the following error bound for numerical integration. Let \((\xi_1, \ldots, \xi_n)\) be a fixed \( n \)-tuple in \([0,1]^d\), then, for every \( f : [0,1]^d \to \mathbb{R} \) with finite variation (in the Hardy and Krause sense, see Niederreiter H. (1992) or in the measure sense see Bouleau N. and Lepingle D. (1994); Pagès G. (2018)),

\[
\left| \frac{1}{n} \sum_{i=1}^{n} f(\xi_i) - \int_{[0,1]^d} f(u) du \right| \leq V(f)D^*_{n}(\xi_1, \ldots, \xi_n).
\] (4.2)

where \( V(f) \) denotes the (finite) variation of \( f \). So a \( O\left( \frac{(\log(n+1))^d}{n} \right) \) or \( O\left( \frac{(\log(n+1))^{d-1}}{n} \right) \) rate of convergence can be achieved, for this class of functions, whether we use a \( n \)-tuple or the \( n \) first terms of a sequence. However, the class of functions with finite variation becomes sparser in the space of functions defined from \([0,1]^d\) to \( \mathbb{R} \) (for example, the function \((x_1, \ldots, x_d) \mapsto (x_1 + \ldots + x_d) \wedge 1\) has finite variation for \( d = 1 \) and 2 but not for \( d \geq 3 \) (see Pagès G. (2018)) and it seems natural to evaluate the performance of the low-discrepancy sequences or \( n \)-tuples on a more natural space of test functions like the Lipschitz functions that do not necessarily have finite variations, especially in high dimensions. This is the purpose of Proinov’s theorem reproduced below.

**Theorem 4.1.** (see Proinov P.D. (1988)) (a) Let \( \mathbb{R}^d \) be equipped with the \( \ell^\infty \)-norm \( |(u^1, \ldots, u^d)|_\infty = \max_{1 \leq i \leq d} |u^i| \). For every continuous function \( f : [0,1]^d \to (\mathbb{R}, |.|_\infty) \), we define the uniform continuity modulus of \( f \) by \( w(f, \delta) = \sup_{\xi, \xi' \in [0,1]^d, |\xi - \xi'| \leq \delta} |f(\xi) - f(\xi')| \). Then, there exists a constant \( C_d \in (0,4) \) (only depending on \( d \)) such that, for every \( n \geq 1 \) and every \( n \)-tuple \((\xi_1, \ldots, \xi_n)\),

\[
\left| \frac{1}{n} \sum_{i=1}^{n} f(\xi_i) - \int_{[0,1]^d} f(x) dx \right| \leq C_d w(f, D^*_{n}(\xi_1, \ldots, \xi_n) \frac{1}{n}^\frac{1}{d}).
\]

(b) Lipschitz setting In particular, if \( f \) is \( [f]_{\text{Lip}} \)-Lipschitz continuous (for the \( |.|_\infty \)-norm) and \((\xi_n)_{n \geq 1} \) is a \([0,1]^d\)-valued sequence with low discrepancy, then

\[
\left| \frac{1}{n} \sum_{i=1}^{n} f(\xi_i) - \int_{[0,1]^d} f(x) dx \right| \leq C_d [f]_{\text{Lip}} D^*_{n}(\xi_1, \ldots, \xi_n) \frac{1}{n}^\frac{1}{d} \leq C_d [f]_{\text{Lip}} C(\xi) \frac{\log(n+1)}{n^\frac{1}{d}}.
\]

Finally, if \( d \geq 2 \), for any sequence of “Hammersley” \( n \)-tuples of the form \( \xi^{(n)} = \left( \frac{k}{n}, \xi_n \right), n \geq 1 \), where \((\xi_n)_{n \geq 1} \) is a \([0,1]^{d-1}\)-valued sequence with low discrepancy, then for every \( n \geq 1 \),

\[
\left| \frac{1}{n} \sum_{i=1}^{n} f(\xi_i) - \int_{[0,1]^d} f(x) dx \right| \leq C_d [f]_{\text{Lip}} C(\xi)^\frac{1}{d-1} \frac{\log(n+1)}{n^\frac{1}{d}}.
\]

This suggests that, at least for a commonly encountered class of regular functions, the curse of dimensionality is more severe with QMC than with quantization due to the extra
(log(n + 1))^{1 - \frac{1}{4}} factor in QMC. This is the price paid by QMC for considering uniform weights
\( p_i = \frac{1}{n}, i = 1, \ldots, n. \)

With greedy quantization sequences, we will show that it is possible to keep the \( n^{-\frac{1}{2}} \) rate of decay for numerical integration but also keep the asset of a sequence which is a recursive formula for cubatures.

4.1. Optimization of the algorithm and the numerical integration in the 1-dimensional case

Quadratic optimal greedy quantization sequences are obtained by implementing, in a recursive way, variants of algorithms such as Lloyd’s I algorithm, also known as \( k \)-means algorithm, or the Competitive Learning Vector Quantization (CLVQ) algorithm, which is a stochastic gradient descent algorithm associated to the \( L^2 \)-distortion function \( G^2_{\| \cdot \|_2} \) defined on \( \mathbb{R}^n \) by \( G^2_{\| \cdot \|_2}(a., X)^2 \). We refer to an extended version of Luschgy H. and Pagès G. (2015) on ArXiv where greedy variants of these procedures are explained in detail and give here a brief idea of one-dimensional greedy Lloyd’s algorithm. The implementation is as follows: at the \( n \)-th iteration, we freeze the \( n - 1 \) points of \( a_{n-1} = \{a_1, \ldots, a_{n-1}\} \) which have been already computed and we sort them in an increasing order \( a_{n-1}^{(n-1)} < \ldots < a_{n-1}^{(n-1)} \). Then, we add a new point \( \bar{a}_0 \) which will constitute the starting point of the optimization procedure. In other words, we compute the inter-point local inertia given by

\[
\sigma_i^{2(n-1)} := \int_{a_i^{(n-1)}}^{a_{i+\frac{1}{2}}^{(n-1)}} |a_i^{(n-1)} - \xi|^2 P(d\xi) + \int_{a_{i+\frac{1}{2}}^{(n-1)}}^{a_{i+1}^{(n-1)}} |a_{i+1}^{(n-1)} - \xi|^2 P(d\xi), \quad i = 0, \ldots, n - 1
\]

(where \( a_{i+\frac{1}{2}}^{(n-1)} = \frac{a_i^{(n-1)} + a_{i+1}^{(n-1)}}{2} \) with \( a_{\frac{1}{2}}^{(n-1)} = a_0^{(n-1)} = -\infty \) and \( a_{\frac{n-1}{2}}^{(n-1)} = a_n^{(n-1)} = +\infty \)) and add a random point \( \bar{a}_0 \) in the inter-point zone with the maximal local inertia \( (a_{i_0}^{(n-1)}, a_{i_0+1}^{(n-1)}) \) where \( i_0 \) is the index such that \( \sigma_i^{2(n-1)} = \max_{1 \leq i \leq n-1} \sigma_i^{2(n-1)} \). This point \( \bar{a}_0 \) is the starting point of the optimization procedure which converges to the \( n \)-th point \( a_n \) of the sequence.

Passing to the \( n + 1 \)-th iteration, we notice that, since all the points are frozen, the local inter-point inertia \( \sigma_i^{2(n-1)} \) remain unchanged for every \( i \in \{0, \ldots, n - 1\} \) except \( \sigma_{i_0}^{2(n-1)} \) which is now replaced by two local inter-point inerti\( a_{i_0}^{(n)} \)\( a_{i_0+1}^{(n)} \) which corresponds to \( a_{i_0}^{(n-1)} \) and \( a_{i_0}^{(n)} \) which corresponds to the point \( a_n \) added at the \( n \)-th iteration), and \( \sigma_{i_0}^{2(n)} \), the inertia between \( a_{i_0}^{(n)} \) and \( a_{i_0+1}^{(n)} \) which corresponds to \( a_{i_0}^{(n-1)} \). Thus, at each iteration, the computation of \( n \) inerti\( a_{i_0}^{(n)} \)\( a_{i_0+1}^{(n)} \) can be reduced to the computation of only 2, thereby reducing the cost of the procedure.

Likewise, the weights \( p_i^2 = P(W_i(a^{(n)})) \) of the Voronoï cells remain mostly unaffected. The only cells that change from one step to another are the cell \( W_{i_0}(a^{(n)}) \) having for centroid the new point \( a_{i_0}^{(n)} \) and the two neighboring cells \( W_{i_0-1}(a^{(n)}) \) and \( W_{i_0+1}(a^{(n)}) \). Thus, the online computation of cell weights just needs 3 calculations instead of \( n \) (or 2 in case the added point is the first or last point in the reordered sequence). The utility of the weights of the Voronoï cells is featured in the approximation of \( \mathbb{E}f(X) \) for \( f : \mathbb{R}^d \rightarrow \mathbb{R} \) by the quadrature
New approach to greedy vector quantization

formula (4.1) using the reordered sequence $a^{(n)}$. Thus, based on the fact that only 3 Voronoï cells are modified at each iteration, one can deduce an iterative formula for the approximation of $I(f)$ by $I_n(f)$, as follows

$$I_n(f) = I_{n-1}(f) - p^n f(a_{i_0}^{(n)}) - p^n f(a_{i_0+1}^{(n)}) + (p^n + p^n) f(a_{i_0}^{(n)}),$$

where

- $a_{i_0}^{(n)}$ is the point added to the greedy sequence at the $n$-th iteration, i.e. the point $a_n$,
- $a_{i_0-1}^{(n)}$ and $a_{i_0+1}^{(n)}$ are the points lower and greater than $a_{i_0}^{(n)}$, i.e. $a_{i_0-1}^{(n)} < a_{i_0}^{(n)} < a_{i_0+1}^{(n)}$,
- $p^n = P([a_{i_0-1}^{(n)}, a_{i_0}^{(n)}])$ and $p^n_{i_0} = P([a_{i_0}^{(n)}, a_{i_0+1}^{(n)}])$.

In conclusion, this iterative formula for numerical integration can be applied without computing the weights of the Voronoï cells and storing them, which usually appear as significant drawbacks for quantization. Instead, one only need to compute 2 weights $p^n$ and $p^n_{i_0}$ which are not Voronoï weights, using only 3 points of the sequence $a_{i_0}^{(n)}$, $a_{i_0-1}^{(n)}$ and $a_{i_0+1}^{(n)}$.

4.2. Product greedy quantization ($d > 1$)

In higher dimensions, greedy quantization has always the recursive properties, so it gets interesting to apply the same numerical improvements as in the one-dimensional case. However, when the dimension $d$ grows, the construction of multidimensional greedy quantization sequences becomes more and more complex and expensive since it relies on complicated stochastic optimization algorithms. In the two-dimensional case, a deterministic variant of greedy algorithms is established in El Nmeir R. (2020) (in collaboration with V. Lemaire) and allows to obtain deterministic two-dimensional greedy quantization sequences which also share the recursive character of greedy sequences. However, this deterministic variant is not extended to the high-dimensional cases when $d \geq 3$ so, as an alternative, one can use one-dimensional greedy quantization grids as tools to obtain multidimensional greedy quantization sequences in some cases.

4.2.1. How to build multi-dimensional greedy product grids

Multidimensional quantization sequences can be obtained as a result of the tensor product of one-dimensional sequences, when the target law is a tensor product of its independent marginal laws. These grids are, of course, not optimal nor asymptotically optimal but they allow to approach the multidimensional law.

Let $X_1, \ldots, X_d$ be $d$ independent $L^2$-random variables taking values in $\mathbb{R}$ with respective distributions $\mu_1, \ldots, \mu_d$ and $a^{1,(n_1)}, \ldots, a^{d,(n_d)}$ the corresponding greedy quantization sequences. By computing the tensor product of these $d$ one-dimensional greedy sequences, we obtain the $d$-dimensional greedy quantization grid $a^{1,(n_1)} \otimes \ldots \otimes a^{d,(n_d)}$ of the product law $\mu = \mu_1 \otimes \ldots \otimes \mu_d$. 
given by \((a^{(n)}_j)_{1 \leq j \leq n} = (a^{1,(n_1)}_{j_1}, \ldots, a^{d,(n_d)}_{j_d})_{1 \leq j_1 \leq n_1, \ldots, 1 \leq j_d \leq n_d}\) of size \(n = \prod_{i=1}^d n_i\). The corresponding quantization error is given by

\[
e_r(a^{1,(n_1)} \otimes \cdots \otimes a^{d,(n_d)}, \mu)^r = \sum_{k=1}^d e_r(a^{k,(n_k)}, X_k)^r.
\] (4.6)

The weights \(p^{(n)}_j\) of the \(d\)-dimensional Voronoï cells \((W_j(a^{(n)}))_{1 \leq j \leq n}\) are deduced from the one-dimensional Voronoï weights \((p^{k,n_k}_j)_{1 \leq j \leq n_d}\), \(k = 1, \ldots, d\), corresponding to the one-dimensional greedy sequences, via

\[
p_j = p^{1,n_1}_j \times \cdots \times p^{d,n_d}_j \quad \forall j = (j_1, \ldots, j_d) \in \prod_{k=1}^d \{1, \ldots, n_k\}.
\]

Note that this is true for any norm \(p^r, p > 0\), on the product space \(\mathbb{R}^d\). The implementation of \(d\)-dimensional grids is not a point-by-point implementation. In fact, at each iteration \(n\), having the \(d\) one-dimensional sequences, one must add a point to one one-dimensional sequence, generating this way several points of the multidimensional sequence. One must choose between \(d\) possibilities: add one point to only one sequence \(a^{k,(n_k)}\) among the \(d\) marginal sequences to obtain \(a^{n_1 \times \cdots \times n_k-1 \times (n_k+1) \times n_{k+1} \times \cdots \times n_d}\). These \(d\) cases are not similar since each one produces a different error quantization, and one does not choose randomly. To make the right decision, one must compute in each case, using (4.6), the quantization error \(E_k\) obtained if we add a point to \(a^{k,(n_k)}\) for a \(k \in \{1, \ldots, d\}\). In other words, we compute, for \(k = 1, \ldots, d\)

\[
E_k = e_r(a^{k,(n_k+1)}, \mu_k)^r + \sum_{\ell \in \{1, \ldots, d\} \setminus \{k\}} e_r(a^{\ell,(n_\ell)}, \mu_\ell)^r.
\]

Then, one chooses the index \(i\) such that \(E_i = \min_{1 \leq k \leq d} E_k\), adds a point to the sequence \(a^{i,(n_i)}\) and obtains the grid \(a^{n_1 \times \cdots \times n_{i-1} \times (n_i+1) \times n_{i+1} \times \cdots \times n_d}\). We note that if the marginal laws \(\mu_1, \ldots, \mu_d\) are identical, this step is not necessary and the choice of the sequence to which a point is added, at each iteration, is systematically done in a periodic manner.

4.2.2. Numerical integration

Similarly to the 1-dimensional case, the majority of the Voronoï cells do not change while passing from an iteration \(n\) to an iteration \(n+1\). At the \(n\)-th iteration, having \(n_1 \times \cdots \times n_d\) points in the sequence, one adds a new point to \(a^{i,(n_i)}\). Hence, we will have \(n_1 \times \cdots \times n_{i-1} \times n_{i+1} \times \cdots \times n_d\) new created cells having for centroids the new points added to the \(d\)-dimensional sequence \(a^{(n)}\), and another \(2(n_1 \times \cdots \times n_{i-1} \times n_{i+1} \times \cdots \times n_d)\) modified cells, corresponding to all the neighboring cells of the new added cells. In total, there is \(3(n_1 \times \cdots \times n_{i-1} \times n_{i+1} \times \cdots \times n_d)\) new Voronoï cells, while the rest remains unchanged. This leads to an iterative formula for quantization-based numerical integration (where the same principle as in the one-dimensional case is applied) as follows: we denote, for the sake of simplicity \(f_{i_0} = f(a^{1,(n_1)}_{j_1}, \ldots, a^{i,(n_i)}_{j_i}, \ldots, a^{d,(n_d)}_{j_d})\), \(f_{i_0-1} = f(a^{1,(n_1)}_{j_1}, \ldots, a^{i,(n_i-1)}_{j_i}, \ldots, a^{d,(n_d)}_{j_d})\) and \(f_{i_0+1} = f(a^{1,(n_1)}_{j_1}, \ldots, a^{i,(n_i+1)}_{j_i}, \ldots, a^{d,(n_d)}_{j_d})\)

\[
I_{n+1}(f) = I_n(f) - \sum_{j_k = 1}^{n_k} \prod_{k=1}^d p^{k,(n_k)}_j (f_{i_0+1} - f_{i_0})
\]
New approach to greedy vector quantization

\begin{equation}
-p_i^{n_i+1} \sum_{j_k=1}^{n_k} \prod_{k \in \{1,...,d\} \setminus \{i\}} p_{jk}^{k(n_k)} (f_{i0+1} - f_{i0}) \tag{4.7}
\end{equation}

Note that in the \(d\)-dimensional case, the weights \(p_{jk}^{k(n_k)}\), \(k \in \{1,...,d\} \setminus \{i\}\) of the Voronoi cells of the other marginal sequences obtained at the previous iteration are needed, as well as the ordered one-dimensional greedy sequences \(a^{k(n_k)}\).

5. Numerical applications and examples

5.1. Greedy quantization of \(\mathcal{N}(0, I_d)\) via Box-Muller

The Box-Muller method allows to generate a random vector with normal distribution \(\mathcal{N}(0, I_d)\), actually two independent one-dimensional random variables \(Z_1\) and \(Z_2\) with distribution \(\mathcal{N}(0, 1)\) by considering two independent random variables \(E\) and \(U\) with respective distributions \(\mathcal{E}(1)\) (Exponential distribution with parameter 1) and \(\mathcal{U}([0, 1])\) (Uniform distribution on \([0, 1]\)). Then, \(2E \sim \mathcal{E}(\frac{1}{2})\) and \(2U \sim \mathcal{U}([0, 2\pi])\), so, the variables

\[Z_1 = \sqrt{2E} \cos(2\pi U) \quad \text{and} \quad Z_2 = \sqrt{2E} \sin(2\pi U)\]

are independent and with normal distribution \(\mathcal{N}(0, 1)\).

We use greedy quantization sequences \(\varepsilon^{(n_1)}\) and \(u^{(n_2)}\) of respective distributions \(\mathcal{E}(1)\) and \(\mathcal{U}([0, 1])\) to design two \(\mathcal{N}(0, 1)\)-distributed independent sequences \(\varepsilon_1^{(n_1)}\) and \(\varepsilon_2^{(n_2)}\), of size \(n = n_1 \times n_2\), via the previous formulas so we can get a greedy sequence \(\varepsilon^{(n)}\) of the two-dimensional normal distribution \(\mathcal{N}(0, I_d)\). The procedure is implemented as described in section 4.2. At each iteration, we must choose the one-dimensional distribution to which we should add a point. Thus, we compute the error induced if we add a point to \(u^{(n_2)}\)

\[E_u = e_2\left(u^{(n_2+1)}, \mathcal{U}([0, 2\pi])\right)^2 + e_2\left(\varepsilon^{(n_1)}, \mathcal{E}\left(\frac{1}{2}\right)\right)^2 = 4\pi^2 e_2\left(\varepsilon^{(n_1+1)}, \mathcal{E}(1)\right)^2 + 4\pi e_2\left(\varepsilon^{(n_1)}, \mathcal{E}(1)\right)^2\]

and the error induces if we add a point to \(\varepsilon^{(n_1)}\)

\[E_\varepsilon = e_2\left(\varepsilon^{(n_1+1)}, \mathcal{E}\left(\frac{1}{2}\right)\right)^2 + e_2\left(u^{(n_2)}, \mathcal{U}([0, 2\pi])\right)^2 = 4\pi^2 e_2\left(u^{(n_2+1)}, \mathcal{U}([0, 1])\right)^2 + 4\pi e_2\left(u^{(n_2)}, \mathcal{U}([0, 1])\right)^2\]

and we add a point to \(u^{(n_2)}\) if \(E_u < E_\varepsilon\) and a point to \(\varepsilon^{(n_1)}\) if \(E_\varepsilon < E_u\).

To design sequences in dimension \(d > 2\), one uses several couples \((E_i, U_i)\) to get several pairs \((Z_i, Z_j)\) and uses the wanted number of \((Z_k)k\) to obtain multidimensional sequences. In figure 1, we compare two greedy quantization sequences of the distribution \(\mathcal{N}(0, I_3)\) of size \(N = 15^3\), one is obtained using the Box-Muller method based on two greedy exponential sequences \(\mathcal{E}(1)\) and two greedy uniform sequences \(\mathcal{U}([0, 1])\), and the other obtained by greedy product quantization based on 3 one-dimensional Gaussian greedy sequences. The weights of the Voronoï cells in both cases are represented by a color scale (growing from blue to red).

Note that, even if the greedy product quantization of a Normal distribution takes the shape of a cube (which is unusual for such distribution), the low values of the Voronoï weights at the edges of this cube allow to consider such a sequence as a valid approximation of the Gaussian distribution.
5.2. Pricing of a 3-dimensional basket of European call options

We consider a Call option on a basket of 3 positive risky assets, with strike price $K$ and maturity $T$, with payoff $h_T = \left(\sum_{i=1}^{3} w_i X^i_T - K\right)^+$ where $(X^1, X^2, X^3)$ represent the prices of the 3 traded assets of the market and $(w_i)_{1 \leq i \leq 3}$ are positive weights such that $\sum_{i=1}^{3} w_i = 1$. 

We consider a 3-dimensional correlated Black-Scholes model where the prices of the assets are given, for every $i \in \{1, 2, 3\}$, by

$$X^i_t = X^i_0 \exp \left( (r - \frac{\sigma^2_i}{2}) t + \sum_{j=1}^{q} \sigma_{ij} W^j_t \right), \quad t \in [0, T]$$

where $r$ is the interest rate, $\sigma = (\sigma_{ij})_{1 \leq i \leq 4, 1 \leq j \leq 3} \in \mathcal{M}(d,q,\mathbb{R})$ is a matrix representing the variance-correlation structure of the model and $(W^1, W^2, W^3)$ is a standard non-correlated 3-dimensional Brownian motion. The volatility $\sigma_i$ of $X^i$ is given by $\sigma_i^2 = \sum_{j=1}^{q} \sigma_{ij}^2$. First, we compute $V_0 = e^{-rT} E[h_T(X^1_T, X^2_T, X^3_T)]$ by a quadrature formula, according to (4.1), using a 3-dimensional greedy quantization sequences of $\mathcal{N}(0, I_3)$ obtained, on one hand, by the Box-Muller algorithm explained in the previous section and, on the other hand, by greedy product quantization of 3 one-dimensional sequences. Then, we estimate $V_0$ by the recursive formula (4.7) for $d = 3$ using the greedy product quantization sequence. We build sequences of size 32000 and consider

$$\sigma = \begin{pmatrix} 0.3 & 0 & 0 \\ 0.15 \frac{3\sqrt{3}}{\sqrt{2}} & 0 \\ 0.15 \frac{3\sqrt{3}}{\sqrt{10}} & \frac{\sqrt{5}}{\sqrt{10}} \end{pmatrix}$$

so that $\sigma_1 = \sigma_2 = \sigma_3 = 0.3$. Also, we consider the following parameters

$$r = 0.1, \quad X^i_0 = 100, \quad T = 1 \text{ and } K = 100.$$
Table 1. Approximation of a 3-dimensional basket of call options in a BS model by Box-Muller with quadrature formula (BM), greedy product quantization with quadrature formula (GPQ) and with recursive formula (GPI).

<table>
<thead>
<tr>
<th>n</th>
<th>BM</th>
<th>GPQ</th>
<th>GPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.1848</td>
<td>0.3019</td>
<td>0.4247</td>
</tr>
<tr>
<td>8000</td>
<td>0.0368</td>
<td>0.0989</td>
<td>0.0988</td>
</tr>
<tr>
<td>15000</td>
<td>0.0271</td>
<td>0.0815</td>
<td>0.0815</td>
</tr>
<tr>
<td>30000</td>
<td>0.0034</td>
<td>0.0049</td>
<td>0.0047</td>
</tr>
</tbody>
</table>

The reference price is given by a large Monte Carlo simulation with control variate of size \(M = 2 \times 10^7\). We consider the control variate

\[
k_T = \left( e^{\sum_{i=1}^{3} w_i \log(X^T_i)} - K \right)_+
\]

which is positive and lower than \(h_T\) (owing to the convexity of the exponential). Since \(e^{-rT} E k_t\) has a normal distribution with mean \((r - \frac{1}{2} \sum_{i=1}^{3} w_i \sigma_i^2)T\) and variance \(w_i \sigma_i \sigma^T w_i T\), it admits a closed form given by

\[
e^{-rT} E k_t = \text{Call}_{\text{BS}} \left( \prod_{i=1}^{3} X_{i,j} e^{-\frac{1}{2}T(\sum_{i=1}^{3} w_i \sigma_i^2 - w_i \sigma^T w)} \right) K, r, \sqrt{w_i \sigma_i \sigma^T w}, T\).
\]

We compare the three methods in table 1 where we expose the errors obtained by each method for particular number of points. We deduce that the recursive numerical integration gives the same results as the quadrature formula-based numerical integration making quantization-based numerical integration less expensive and more advantageous by reducing the cost in time and storage. Moreover, one notices that the Box-Muller algorithm is more accurate than the greedy product quantization, this can be explained by the fact that Box-Muller sequences fill the space in a way that resembles more to the normal distribution (see figure 1).

**Remark 5.1.** The matrix \(\sigma = (\sigma_{ij})_{1 \leq i,j \leq 3}\) can be written \(\sigma_{ij} = \sigma_i C_{ij}\) where \(C = (C_{ij})_{1 \leq i,j \leq 3}\) is a matrix such that \(CC^T = R\), \(R\) being the correlation matrix of a correlated Brownian motion \((B_i)_{1 \leq i \leq 3}\) given by \(B_i = \sum_{j=1}^{3} C_{ij} W_j\).

6. Further properties and numerical remarks

In this section, we present, based on numerical experiments, some properties of the one-dimensional quadratic greedy quantization sequences. We recall that \(a^{(n)}_i = \{a_1^{(n)}, \ldots, a_i^{(n)}\}\) denotes the reordered greedy sequence of the \(n\) first elements \(\{a_1, \ldots, a_n\}\) of \((a_n)_{n \geq 1}\).

6.1. Sub-optimality of greedy quantization sequences

The implementation of a greedy quantization sequence \((a_n)_{n \geq 1}\) of a distribution \(P\) and the computation of the corresponding weights \(p_i^n\) of the Voronoï cells \(W_i(a^{(n)})\) for \(i \in \{1, \ldots, n\}\) defined by (1.1) is, in general, not optimal. However, numerical implementations and graphs
Figure 2. Representation of $a_i \mapsto p_i^n$ where $(p_i^n)_{1 \leq i \leq n}$ denote the Voronoï weights of the greedy quantization sequence of $N(0,1)$ for $n = 255 = 2^8 - 1$ (left), $n = 400$ (right).

representing $a_i \mapsto p_i^n = P(X \in W_i(a^{(n)}))$ for different number of points $n$ show that, for certain distributions, the weights of the Voronoï cells converge towards the density curve of the corresponding distribution when the greedy sequence has a certain number of points.

For the Normal distribution $N(0,1)$ and Laplace distribution $L(0,1)$, this is observed when the sequence is of size $n = 2^k - 1$, for every integer $k \geq 1$. Regarding the Uniform distribution on $[0,1]$, the convergence of the Voronoï weights towards the density curve is observed for two sub-sequences $\alpha^{(n)} = a^{(k_i)}$ of $a^{(n)}$ for values of $k_i$ defined by

$$
\begin{align*}
k_0 &= 3 \\
k_i &= 2k_{i-1} + 1 \quad \text{if } i \equiv 1 \pmod{3} \\
k_i &= 2(k_i - 2) + 1 \quad \text{if } i \equiv 2 \pmod{3} \\
k_i &= 2(k_i + 2) + 1 \quad \text{if } i \equiv 0 \pmod{3}
\end{align*}
$$

This leads to the following conjecture

**Conjecture 6.1.** A greedy quantization sequence is sub-optimal in the sense that there exists one (or more) subsequence which is optimal itself. In the case of symmetrical distributions around 0, the subsequence is given by

$$
\alpha^{(n)} = a^{(2^k - 1)} \quad \text{s.t.} \quad n = 2^k - 1, \; k \in \mathbb{N}^*
$$

For the $U([0,1])$-distribution, there exists two subsequences of size defined by (6.1).

Some results for the normal distribution are represented in figure 2 where we observe the unimodal weights for $n = 255 = 2^8 - 1$ and non-unimodal weights for $n = 400$.

### 6.2. Convergence of standard and weighted empirical measures

Sequences of asymptotically optimal $n$-quantizers $(\Gamma_n)_{n \geq 1}$ of $P$ satisfy some empirical measure convergence theorems established in Graf S. and Luschgy H. (2000) (see Theorem 7.5 p. 96) and Delattre S. et al. (2004) and recalled below, where

$$
\hat{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i^n} \quad \text{and} \quad \tilde{P}_n = \sum_{i=1}^{n} P(W_i(\Gamma_n)) \delta_{x_i^n}
$$
New approach to greedy vector quantization

Figure 3. Comparison of the exact Voronoï weights (blue) and the limit weights (red) for the exponential distribution $E(1)$ for $n = 645$ (left) and $n = 1379$ (right).

designate, respectively, the empirical standard measure and the empirical weighted measure associated to $\Gamma_n = \{x_1^n, \ldots, x_n^n\}$.

Theorem 6.1 (Graf S. and Luschgy H. (2000)). Assume $P$ is absolutely continuous w.r.t the Lebesgue measure on $\mathbb{R}^d$ with density $f$. Let $(\Gamma_n)_{n \geq 1}$ be a sequence of asymptotically $L^r$-optimal $n$-quantizers of $P$. Then, denoting $C = \left( \int_{\mathbb{R}} f^{\frac{d}{r}}(u) du \right)^{-1}$, one has

$$
\tilde{P}_n \overset{n \to +\infty}{\Rightarrow} P \quad \text{and} \quad \hat{P}_n \overset{n \to +\infty}{\Rightarrow} C f^{\frac{d}{r}}(u) du. \quad (6.3)
$$

Due to the existence of suboptimal greedy quantization sequences, detailed previously, we hope to obtain such results for greedy sequences or, at least, for sub-optimal greedy sequences defined in the previous section. To this end, we implement the following numerical experience: First, we “divide” the two limits mentioned in (6.3), along the sequence $(W_i(a(n)))_{1 \leq i \leq n}$ to obtain, for every $i \in \{1, \ldots, n\}$, the limiting measure of the Voronoï cells of the greedy sequence

$$
P_l(W_i(a(n))) \simeq f^{\frac{d}{r}}(a_i^{(n)}) C n. \quad (6.4)
$$

Then, we test if the exact weights of the Voronoï cells, computed using the c.d.f of $P$, converge to the limit weights $P_l(W_i(a(N)))$ given in (6.4). If this is the case, we can conclude with the following conjecture.

Conjecture 6.2. Greedy quantization sequences satisfy the empirical measure Theorem 6.1. A faster convergence is satisfied by the optimal subsequences defined in Section 6.1.

Numerical experiments were established for the one-dimensional standard Normal, Uniform, Exponential and Laplace distribution. We observe that the exact weights of the Voronoï cells computed online get closer to the limit weights $P_l$ when $n$ increases. For the Gaussian distribution, we observe a much faster convergence for the subsequences $a^{(2k-1)}$ (as predicted). We present, in figure 3 the obtained results for the exponential distribution where we compare the exact weights (blue) and the limit weights (6.4) (red) for different number of points $n$. 
An interesting question is to see if the greedy sequences are stationary i.e. satisfy

$$a_i^{(n)} = \mathbb{E}(X|X \in W_i(a^{(n)})), i = 1, \ldots, n,$$

or can be close to stationarity, a property shared by quadratic optimal quantizers. Numerical experiments conducted for several probability distributions yield that, unfortunately, greedy sequences are not stationary in this sense. In fact, one can prove that the greedy quantization sequence $a^{(n)}$ of a symmetric unimodal distribution is not stationary, except for $n \in \{1, 3\}$. A proof is available in El Nmeir R. et al. (2020), an extended version of this paper posted on arXiv.

However, further different numerical observations lead to the following conjecture.

**Conjecture 6.3.** *Greedy quantization sequences satisfy a $\rho$-quasi-stationarity, approaching the stationary property and defined, for $r \in \{1, 2\}$ and $\rho \in [0, 1]$, by*

$$\left\| \hat{X}^{a^{(n)}} - \mathbb{E}(X|\hat{X}^{a^{(n)}}) \right\| = o\left(\|\hat{X}^{a^{(n)}} - X\|^1_{1+\rho}\right) \; \text{or} \; \frac{\left\| \hat{X}^{a^{(n)}} - \mathbb{E}(X|\hat{X}^{a^{(n)}}) \right\|_r}{\|\hat{X}^{a^{(n)}} - X\|_{1+\rho}} \xrightarrow{n \rightarrow +\infty} 0. \quad (6.5)$$

It is satisfied by greedy sequences for $\rho$ lower than certain optimal values $\rho_l$ depending on $r$ and $P$. We expose, in table 2, these values of $\rho_l$ for $r \in \{1, 2\}$ for the Normal, Uniform and exponential distribution. This property is important because it brings improvements to quantization-based numerical integration. In fact, if $f$ is $C^1$ with $\rho$-holder gradient with coefficient $|\nabla f|_\rho$, the classical error bound is given by (see Pagès G. (2015))

$$|\mathbb{E}(f(X) - \mathbb{E}(f(\hat{X}^{a^{(n)}}))| \leq |\nabla f|_\rho \|X - \hat{X}^{a^{(n)}}\|_{1+\rho}. $$

And, based on the $\rho$-quasi-stationarity property deduced in this section, one can decrease this upper bound by a factor of $\frac{1}{1+\rho}$ for $\rho \in [0, 1]$. In fact, denoting $\langle \cdot, \cdot \rangle$ the inner product and $\mathbb{E}[\cdot, \cdot]$ the conditional expectation in this section, one has

$$\mathbb{E}(f(X) - \mathbb{E}(f(\hat{X}^{a^{(n)}})) \leq \mathbb{E}\langle \nabla f(\hat{X}^{a^{(n)}}) | X - \hat{X}^{a^{(n)}} \rangle \leq \mathbb{E}\left[ \int_0^1 \langle \nabla f(\hat{X}^{a^{(n)}} + t(X - \hat{X}^{a^{(n)}})) - \nabla f(\hat{X}^{a^{(n)}}))|X - \hat{X}^{a^{(n)}} \rangle dt \right]$$

where the second expectation in the right term of the above inequality is bounded by $|\nabla f|_\rho \mathbb{E}\|X - \hat{X}^{a^{(n)}}\|_{1+\rho} \int_0^1 t^{1+\rho} dt$ and

$$\mathbb{E}\langle \nabla f(\hat{X}^{a^{(n)}}) | X - \hat{X}^{a^{(n)}} \rangle = \mathbb{E}\langle \nabla f(\hat{X}^{a^{(n)}}) | X \rangle - \mathbb{E}\langle \nabla f(\hat{X}^{a^{(n)}}) | \hat{X}^{a^{(n)}} \rangle$$

**Table 2.** Values of optimal $\rho_l$ for different distributions and for $r \in \{1, 2\}$.

<table>
<thead>
<tr>
<th></th>
<th>$\mathcal{N}(0, 1)$</th>
<th>$\mathcal{U}([0, 1])$</th>
<th>$\mathcal{E}(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r = 1$</td>
<td>$\rho_l = 0.92$</td>
<td>$\rho_l = \frac{1}{3}$</td>
<td>$\rho_l = \frac{2}{3}$</td>
</tr>
<tr>
<td>$r = 2$</td>
<td>$\rho_l = 0.47$</td>
<td>$\rho_l = \frac{1}{5}$</td>
<td>$\rho_l = \frac{2}{5}$</td>
</tr>
</tbody>
</table>
New approach to greedy vector quantization

\[ \mathbb{E} \langle \nabla f(\tilde{X}^{a(n)}) | \mathbb{E}[X|\tilde{X}^{a(n)}] - \tilde{X}^{a(n)} \rangle. \]

So, if (6.5) is satisfied, then one uses the conditional Cauchy-Schwarz inequality to obtain

\[ |\mathbb{E} f(X) - \mathbb{E} f(\tilde{X}^{a(n)})| \leq \| \nabla f(\tilde{X}^{a(n)})\|_2 \| \mathbb{E}[X|\tilde{X}^{a(n)}] - \tilde{X}^{a(n)} \|_2 + \frac{1}{1+\rho} \| \nabla f \|_\rho \| X - \tilde{X}^{a(n)} \|^{1+\rho}_{1+\rho}. \]

6.4. Discrepancy of greedy sequences

The comparison established, in the beginning of section 4, between greedy quantization-based numerical integration and quasi-Monte Carlo methods, showing a gain of \( \log(n+1) \)-factor with greedy quantization in terms of convergence rate, drives us to build a relation, based on Proćnov’s Theorem 4.1, between the error quantization and the discrepancy. In fact, for every \( n \)-tuple \( \Xi = (\xi_1, \ldots, \xi_n) \in [0,1]^n \), noticing that a Lipschitz function \( f \) has always a finite variation and considering the function \( f : u \rightarrow \min_{1 \leq i \leq n} |u - \xi_i| \) which is 1-Lipschitz (since \( \min_i a_i - \min_i b_i \leq \max_i |a_i - b_i| \) and satisfies \( f(\xi_i) = 0 \) for every \( i \in \{1, \ldots, n\} \) and \( \int_0^1 f(u)du = e_1(\Xi, \mathcal{U}([0,1])) \), one applies the Koksma-Hlawka inequality (4.2) to \( f \) to deduce that

\[ e_1(\Xi, \mathcal{U}([0,1])) \leq D_n^*(\Xi). \quad (6.6) \]

This motivates us to study the discrepancy of greedy sequences hoping that they can be comparable to low discrepancy sequences. We compute this quantity for \( d \in \{1,2,3\} \), using formulas from Doerr C. et al. (2014) and deduce that, when \( d = 1 \), greedy sequences can be used as a low discrepancy sequence. But, when \( d > 1 \), the situation becomes less convincing: The discrepancy of pure greedy sequences, designed by implementing Lloyd’s algorithm, and that of low discrepancy sequences (Niederreiter sequences for example) are comparable, but the problem that arises is the complexity of the computations making greedy sequences less practical. On the other hand, if we build greedy product sequences, the computations will be less expensive but there is no gain in terms of discrepancy. Figure 4 shows a comparison of the discrepancy of a Niederreiter sequence in dimension 2 to that of a product greedy quantization sequence of \( \mathcal{U}([0,1]^2) \) on the one hand, and to that of pure greedy quantization sequence of \( \mathcal{U}([0,1]^2) \) on the other hand.

The positive results obtained for \( d = 1 \) encourage us to try and manipulate low discrepancy sequences, such as Van der Corput sequences, in order to use them as greedy quantization sequences. In other words, we will assign to them a Voronoï diagram, compute the weights of the corresponding Voronoï cells instead of considering uniform weights and observe the impact this brings to numerical integration. To this end, we consider a basic example where we compute the price of a European call \( C_0 = \mathbb{E}[(X_T - K)_+] \) for a maturity \( T \) and a strike price \( K \) where the price of the asset \( X_t \) at a time \( t \) is given by

\[ X_t = x_0 \exp \left( (r - \frac{\sigma^2}{2})t + \sigma W_t \right). \]
Figure 4. Comparisons of the star discrepancy of the Niederreiter sequence to a greedy product quantization sequence of the uniform distribution $\mathcal{U}([0,1]^2)$ (left) and to a pure greedy quantization sequence (right) for $d=2$.

where $r$ is the interest rate, $\sigma$ the volatility and $(W_t)_{t \in [0,T]}$ a standard Brownian motion (such that $W_t = \sqrt{t}Z$, $Z \sim N(0,1)$, for every $t \in [0,T]$). We consider

$$T = 1, K = 9, x_0 = 10, \mu = 0.06, \sigma = 0.1.$$  

The exact price is given by a closed formula known in the Black-Scholes case and is approximately equal to 1.5429. We compute the price $C_0$ first via a classical quadrature formula using the new weights $p^n_i$ assigned to the VdC sequence, then by a classical quasi-Monte Carlo simulation (using uniform weights of a VdC sequence) and finally by a quantization-based quadrature formula based on a greedy quantization sequence of $\mathcal{U}([0,1])$. We compare the errors induced by these three methods in figure 5 and deduce that the procedure using the greedy quantization sequence converges faster than the ones using the Van der Corput sequence. Consequently, one can say that, even when we assign non-uniform weights to low discrepancy sequences, they still are less advantageous than greedy sequences in terms of rate of convergence of the numerical integration error. One more general conclusion one can make is that greedy sequences are more advantageous than low discrepancy sequences, not only due to the $\log(n+1)$-factor gain but also because we showed that the drawback of vector quantization, which is the cost of the computation of non-uniform weights, is dramatically reduced due to the recursive character of the greedy quantization.

Acknowledgments

The authors would like to express a sincere gratitude to, Dr. Rami El Haddad, the co-advisor of R. El Nmeir, for his help and advice during this work. Also, they would like to acknowledge the National Council for Scientific Research of Lebanon (CNRS-L) for granting a doctoral fellowship to Rancy El Nmeir, in a joint program with Agence Universitaire de la Francophonie of the Middle East and the research council of Saint-Joseph University of Beirut.

References

New approach to greedy vector quantization

Figure 5. Price of a European call in a Black-Scholes model via a usual QMC method (blue), greedy quantization-based quadrature formula (red) and quadrature formula using VdC sequence with non-uniform weights (logarithmic scale).


