

Universités de Paris 6 & Paris 7 - CNRS (UMR 7599)

**PRÉPUBLICATIONS DU LABORATOIRE
DE PROBABILITÉS & MODÈLES ALÉATOIRES**

4, place Jussieu - Case 188 - 75 252 Paris cedex 05

<http://www.proba.jussieu.fr>

**Quantization of probability distributions
under norm-based distortion measures**

S. DELATTRE, S. GRAF, H. LUSCHGY & G. PAGÈS

SEPTEMBRE 2004

Prépublication n° 924

S. Delattre & G. Pagès : Laboratoire de Probabilités et Modèles Aléatoires,
CNRS-UMR 7599, Université Paris VI & Université Paris VII, 4 place Jussieu,
Case 188, F-75252 Paris Cedex 05.

S. Graf : Universität Passau, Fakultät für Mathematik und Informatik,
D-94030 Passau.

H. Luschgy : Universität Trier, FB IV-Mathematik, D-54286 Trier.

Quantization of probability distributions under norm-based distortion measures

SYLVAIN DELATTRE* SIEGFRIED GRAF†
HARALD LUSCHGY‡ AND GILLES PAGÈS §

June 04

Abstract

For a probability measure P on \mathbb{R}^d and $n \in \mathbb{N}$ consider $e_n = \inf \int \min_{a \in \alpha} V(\|x - a\|) dP(x)$ where the infimum is taken over all subsets α of \mathbb{R}^d with $\text{card}(\alpha) \leq n$ and V is a nondecreasing function. Under certain conditions on V , we derive the precise n -asymptotics of e_n for nonsingular and for (singular) self-similar distributions P and we find the asymptotic performance of optimal quantizers using weighted empirical measures.

Key words: High-rate vector quantization, norm-difference distortion, empirical measure, weak convergence, local distortion, point density measure.

2001 AMS classification: 60E99, 94A29, 28A80.

1 Introduction

Consider a random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ with distribution $\mathbb{P}^X = P$ and let $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a nondecreasing function. For $n \in \mathbb{N}$ and any norm $\|\cdot\|$ on \mathbb{R}^d , the n -optimal V -quantization is the global minimization of

$$E \min_{a \in \alpha} V(\|X - a\|)$$

over all subsets $\alpha \subset \mathbb{R}^d$ with $\text{card}(\alpha) \leq n$. Such a set α is called n -codebook or n -quantizer. So the resulting error by using $a \in \alpha$ instead of X is measured by the norm-difference distortion based on the loss function V . The minimal n th V -quantization error is then defined by

$$e_{n,V}(X) = e_{n,V}(P) := \inf \{ \mathbb{E} \min_{a \in \alpha} V(\|X - a\|) : \alpha \subset \mathbb{R}^d, 1 \leq \text{card}(\alpha) \leq n \}. \quad (1.1)$$

This quantity is finite provided

$$\mathbb{E} V(\|X\|) < \infty. \quad (1.2)$$

*Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Paris 7, case 2, 4, pl. Jussieu, F-75252 Paris Cedex 5. E-mail: delattre@math.jussieu.fr

†Universität Passau, Fakultät für Mathematik und Informatik, D-94030 Passau, BR Deutschland. Email: Graf@univ.passau.de

‡Universität Trier, FB IV-Mathematik, D-54286 Trier, BR Deutschland. E-mail: luschgy@uni-trier.de

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

For a given n -codebook α one defines an associated closest neighbour projection

$$\pi_\alpha := \sum_{a \in \alpha} a \mathbf{1}_{C_a(\alpha)}$$

and the induced α -quantized version (or α -quantization) of X by

$$\widehat{X}^\alpha := \pi_\alpha(X), \tag{1.3}$$

where $\{C_a(\alpha) : a \in \alpha\}$ is a Voronoi partition of \mathbb{R}^d w.r.t. α , that is

$$C_a(\alpha) \subset \{x \in \mathbb{R}^d : \|x - a\| = \min_{b \in \alpha} \|x - b\|\}$$

for every $a \in \alpha$. Then one easily checks that for any random variable $Y : \Omega \rightarrow \alpha \subset \mathbb{R}^d$,

$$\mathbb{E} V(\|X - Y\|) \geq \mathbb{E} V(\|X - \widehat{X}^\alpha\|) = \mathbb{E} \min_{a \in \alpha} V(\|X - a\|)$$

so that

$$\begin{aligned} e_{n,V}(X) &= \inf\{\mathbb{E} V(\|X - \widehat{X}\|) : \widehat{X} = f(X), f : \mathbb{R}^d \rightarrow \mathbb{R}^d \text{ measurable, } \text{card}f(\mathbb{R}^d) \leq n\} \\ &= \inf\{\mathbb{E} V(\|X - Y\|) : Y : \Omega \rightarrow \mathbb{R}^d \text{ measurable, } \text{card}(Y(\Omega)) \leq n\}. \end{aligned} \tag{1.4}$$

In electrical engineering this problem arises in the context of coding signals effectively. For these applications in information theory we refer to Gersho and Gray [10]. In statistics quantizers may be used as models for the grouping of data. More recently, quantization appears as promising tool for multidimensional nonlinear problems in numerical probability (see e.g. Pagès et al. [25]).

Much of the previous work is for r -quantization where $V(t) = t^r$ for some $r \in (0, \infty)$. For the mathematical aspects of r -quantization one may consult Graf and Luschgy [12]. The need of more general loss functions for applications in speech and image compression has been emphasized e.g. by Gardner and Rao [9] and Li et al. [20]. See also the investigation of Linder et al. [21]. In fact, the emphasis in these papers is on input weighted error measures of the type $\mathbb{E}(X - \widehat{X})^T M(X)(X - \widehat{X})$ for some continuous matrix-valued function M . By localization, our results are related to the special case of real input weights in that we provide an asymptotic evaluation of $\mathbb{E}g(X)V(\|X - \widehat{X}\|)$ for V -optimal quantizers \widehat{X} where g is real-valued (see Theorem 5).

Section 2 of this paper presents the basic features of the V -quantization problem including existence and uniqueness of optimal quantizers, necessary conditions for optimality and an application to numerical integration. Moreover, a portmanteaux-proposition about the different types of convergence of α_n -quantizations is established. Section 3 contains the n -asymptotics of the quantization error $e_{n,V}(P)$ for nonsingular probability distributions P and loss functions which are regularly varying at zero “without slowly varying part” that is, $V(t)$ behaves locally at zero as t^r for some $r \in (0, \infty)$. The result applies, for instance, to exponential quantization with $V(t) = \exp(t^r) - 1$, log-quantization with $V(t) = (\log(1+t))^r$ and exponentially weighted r -quantization with $V(t) = t^r e^t$. Related results for absolutely continuous distributions with compact support are contained in Gruber ([16],[17]). See also the references in these papers.

In Section 4 again for nonsingular distributions we establish the weak convergence of empirical and other measures induced by (asymptotically) V -optimal n -quantizers. In particular, the asymptotics of localized V -quantization errors and the point density measure are derived. In Section 5 we investigate the n -asymptotics of $e_{n,V}(P)$ and the asymptotic behaviour of V -optimal quantizers for self-similar probabilities P which provide an interesting class of singular distributions. In the nonarithmetic case and under distribution dependent rates one can achieve similar results as for nonsingular distributions.

We emphasize that some of the results are new even in the r -quantization framework (Theorems 6, 8 and 9) while others still provide improvements of known results for $V(t) = t^r$ (Theorems 4 and 5).

NOTATIONS: $a_n \sim b_n$ means $a_n = b_n + o(b_n)$, $a_n \approx b_n$ for positive real numbers means $\liminf a_n/b_n > 0$ and $\limsup a_n/b_n < \infty$, $a_n \asymp b_n$ means $a_n = O(b_n)$ and $b_n = O(a_n)$, λ^d denotes the d -dimensional Lebesgue measure on \mathbb{R}^d , \Rightarrow denotes weak convergence of finite measures on \mathbb{R}^d and $d(x, A) := \inf_{y \in A} \|x - y\|$ for $A \subset \mathbb{R}^d$.

2 Basic facts

2.1 r -Quantization

For the r -quantization problem where $V(t) = t^r$ for some $r \in (0, \infty)$ set

$$e_{n,r}(P) := \inf \{ \mathbb{E} \min_{a \in \alpha} \|X - a\|^r : \alpha \subset \mathbb{R}^d, 1 \leq \text{card}(\alpha) \leq n \}. \quad (2.1)$$

(Notice that we do not take the r th root). Let $P = P^a + P^s$ be the Lebesgue decomposition of P with respect to λ^d , where P^a denotes the absolutely continuous part and P^s the singular part of P . Furthermore, let

$$\|h\|_s := \left(\int |h|^s d\lambda^d \right)^{1/s}$$

for every Borel measurable function $h : \mathbb{R}^d \rightarrow \mathbb{R}$ and $s \in (0, \infty)$.

The rate of convergence of $e_{n,r}$ to zero is ruled by the following theorem.

Theorem 1 (Zador [27],[28]), Bucklew and Wise [3], Graf and Luschgy [12]) *Assume that $\int \|x\|^{r+\delta} dP(x) < \infty$ for some $\delta > 0$. Then*

$$\lim_{n \rightarrow \infty} n^{r/d} e_{n,r}(P) = J_{r,d} \|h\|_{d/(d+r)} < \infty,$$

where

$$J_{r,d} := \inf_{n \geq 1} n^{r/d} e_{n,r}(U([0,1]^d)) > 0$$

with $U([0,1]^d)$ denoting the uniform distribution on $[0,1]^d$ and h the λ^d -density of P^a .

Notice that $J_{r,d}$ depends on the underlying norm on \mathbb{R}^d . With a few exceptions the constant $J_{r,d}$ is unknown.

In case $P^a \neq 0$, set

$$Q_r(P) := J_{r,d} \|h\|_{d/(d+r)} \in (0, \infty) \quad (2.2)$$

The (r, s) -problem concerns the performance of quantizers under increasing powers $r < s$. For this, the following information about α -quantizations is useful.

Proposition 1 *Let $(\alpha_n)_n$ be a sequence of finite codebooks. The following statements are equivalent.*

(i) $\hat{X}^{\alpha_n} \Rightarrow X$.

(ii) $\hat{X}^{\alpha_n} \rightarrow X$ a.e.

(iii) $\mathbf{1}_K(X) \|X - \widehat{X}^{\alpha_n}\| \xrightarrow{L^\infty(\mathbb{P})} 0$ for every compact set $K \subset \mathbb{R}^d$.

(iv) $\lim_{n \rightarrow \infty} d(x, \alpha_n) = 0$ for every $x \in \text{supp}(P)$

(v) $\lim_{n \rightarrow \infty} \max_{x \in K} d(x, \alpha_n) = 0$ for every compact subset K of $\text{supp}(P)$.

If $\|X\| \in L^r(\mathbb{P})$ for some $r \in (0, \infty]$, then (i) is also equivalent to

(vi) $\|X - \widehat{X}^{\alpha_n}\| \xrightarrow{L^r(\mathbb{P})} 0$.

Proof. (i) \Rightarrow (iv). Let $x \in \text{supp}(P)$. For $\varepsilon > 0$, we obtain

$$\liminf_{n \rightarrow \infty} \mathbb{P}(\|\widehat{X}^{\alpha_n} - x\| \leq \varepsilon) \geq \mathbb{P}(\|X - x\| < \varepsilon) > 0.$$

Consequently, $B(x, \varepsilon) \cap \alpha_n \neq \emptyset$ and thus $d(x, \alpha_n) \leq \varepsilon$ for all large enough $n \in \mathbb{N}$, where $B(x, \varepsilon)$ is the closed ball centered at x with radius ε .

(iv) \Rightarrow (v). Let $\varepsilon > 0$. Choose a finite set $\beta \subset K$ such that $\max\{d(x, \beta) : x \in K\} \leq \varepsilon/2$ and then choose $n_0 \in \mathbb{N}$ with $\max\{d(y, \alpha_n) : y \in \beta\} \leq \varepsilon/2$ for every $n \geq n_0$. Consequently, for $x \in K$ there exists $y \in \beta$ such that $\|x - y\| = d(x, \beta) \leq \varepsilon/2$ so that for $n \geq n_0$,

$$d(x, \alpha_n) \leq \|x - y\| + d(y, \alpha_n) \leq \varepsilon.$$

(v) \Leftrightarrow (iii) follows from

$$\text{ess sup} \mathbf{1}_K(X) \|X - \widehat{X}^{\alpha_n}\| = \max\{d(x, \alpha_n) : x \in K \cap \text{supp}(P)\}.$$

and (iii) \Rightarrow (ii) \Rightarrow (i) is clear.

Now assume $\|X\| \in L^r(\mathbb{P})$ with $r \in (0, \infty)$. The implication (vi) \Rightarrow (i) is clear.

(ii) \Rightarrow (vi). Fix $y \in \text{supp}(P)$. Using (iv) one obtains for all large enough n (such that $d(y, \alpha_n) \leq 1$),

$$\begin{aligned} \|X - \widehat{X}^{\alpha_n}\| &= d(X, \alpha_n) \leq \|X - y\| + d(y, \alpha_n) \\ &\leq \|X - y\| + 1 \leq \|X\| + \|y\| + 1. \end{aligned}$$

Consequently, one may apply the Lebesgue dominated convergence theorem and $\mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^r \rightarrow 0$.

If $\|X\| \in L^\infty(\mathbb{P})$, then $\text{supp}(P)$ is compact and hence (iii) and (vi) are equivalent. \square

A weak solution of the (r, s) -problem is as follows.

Corollary 1 Let $s, r \in (0, \infty]$ with $s > r$ and let $(\alpha_n)_n$ be a sequence of n -quantizers.

(a) Assume $\|X\| \in L^s(\mathbb{P})$. If

$$\|X - \widehat{X}^{\alpha_n}\| \xrightarrow{L^r(\mathbb{P})} 0$$

then

$$\|X - \widehat{X}^{\alpha_n}\| \xrightarrow{L^s(\mathbb{P})} 0.$$

(b) Assume $\|X\| \in L^p(\mathbb{P})$ for every $p \in (0, \infty)$ and $s < \infty$. If

$$\mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^r = O(n^{-\gamma})$$

for some $\gamma > 0$ then

$$\mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^s = O(n^{-\gamma\rho}) \quad \text{for every } \rho \in (0, 1). \quad (2.3)$$

If $\|X\| \in L^\infty(\mathbb{P})$, one may take $\rho = 1$.

Proof. (a) is an immediate consequence of Proposition 1.

(b) Using Hölder's inequality, one obtains

$$\begin{aligned}\mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^s &= \mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^{r\rho+(s-r\rho)} \\ &\leq (\mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^r)^\rho (\mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^{\frac{s-r\rho}{1-\rho}})^{1-\rho} \\ &\leq C(\rho)(\mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^r)^\rho\end{aligned}$$

where $C(\rho) = \sup_n (\mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^{\frac{s-r\rho}{1-\rho}})^{1-\rho} < +\infty$ due to the equivalence (i) \Leftrightarrow (vi) in Proposition 1. \square

In situation (b), for instance, asymptotic r -optimality of $(\alpha_n)_n$ which means in case $P_a \neq 0$

$$\lim_{n \rightarrow \infty} n^{\frac{r}{d}} \mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^r = Q_r(P)$$

implies for every $s > r$ that

$$\limsup_{n \rightarrow \infty} (n^{\frac{s}{d}})^{\rho^{\frac{r}{s}}} \mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^s < \infty \quad (2.4)$$

for every $\rho \in (0, 1)$. However,

$$\limsup_{n \rightarrow \infty} n^{\frac{s}{d}} \mathbb{E}\|X - \widehat{X}^{\alpha_n}\|^s = \infty$$

may happen as is illustrated by the following example.

Example 1 Let $\mathbb{P}^X = P = U([0, 1])$ and for $n \geq 2$ and $\vartheta \in (0, \infty)$ set

$$\alpha_n = \alpha_n(\vartheta) := \left\{ \frac{1}{2n^\vartheta} \right\} \cup \left\{ \frac{1}{n^\vartheta} + \left(1 - \frac{1}{n^\vartheta}\right) \frac{2(k-1) - 1}{2(n-1)} : k = 2, \dots, n \right\}.$$

Let $r \in (0, \infty)$ and assume $\vartheta > r/(1+r)$. Using a non-Voronoi partition gives the upper estimate

$$\begin{aligned}\mathbb{E}|X - \widehat{X}^{\alpha_n}|^r &\leq \int \mathbf{1}_{[0, n^{-\vartheta}]}(x) \left| x - \frac{1}{2n^\vartheta} \right|^r dx \\ &\quad + \sum_{k=2}^n \int \mathbf{1}_{[\frac{1}{n^\vartheta} + (1 - \frac{1}{n^\vartheta}) \frac{k-2}{n-1}, \frac{1}{n^\vartheta} + (1 - \frac{1}{n^\vartheta}) \frac{k-1}{n-1}]}(x) \left| x - \left(\frac{1}{n^\vartheta} + \left(1 - \frac{1}{n^\vartheta}\right) \frac{2(k-1) - 1}{2(n-1)} \right) \right|^r dx \\ &= \frac{2}{r+1} \left(\frac{1}{2n^\vartheta} \right)^{r+1} + \left(1 - \frac{1}{n^\vartheta}\right)^{r+1} \frac{1}{(r+1)2^r(n-1)^r}.\end{aligned}$$

Hence

$$\begin{aligned}\limsup_{n \rightarrow \infty} n^r \mathbb{E}|X - \widehat{X}^{\alpha_n}|^r &\leq \limsup_{n \rightarrow \infty} \left(\frac{1}{2^r(r+1)} \frac{1}{n^{(r+1)\vartheta-r}} + \frac{1}{(r+1)2^r} \right) \\ &= \frac{1}{2^r(r+1)} = J_{r,1} = Q_r(P)\end{aligned}$$

so that in fact

$$\lim_{n \rightarrow \infty} n^r \mathbb{E}|X - \widehat{X}^{\alpha_n}|^r = Q_r(P)$$

(see Theorem 1). It follows that the sequence $(\alpha_n(\vartheta))_n$ is an asymptotically r -optimal n -quantizer for every $\vartheta \in (r/(r+1), \infty)$. Now, let $s > r$ and $\vartheta \in (r/(r+1), s/(s+1))$. Then

$$\mathbb{E}|X - \widehat{X}^{\alpha_n}|^s \geq \int_0^{1/2n^\vartheta} \left| x - \frac{1}{2n^\vartheta} \right|^s dx = \frac{1}{2^{s+1}(s+1)} n^{-\vartheta(s+1)}$$

so that

$$n^s \mathbb{E} |X - \widehat{X}^{\alpha_n}|^s \geq \frac{1}{2^{s+1}(s+1)} n^{s-\vartheta(s+1)}.$$

Consequently,

$$\lim_{n \rightarrow \infty} n^s \mathbb{E} |X - \widehat{X}^{\alpha_n}|^s = \infty.$$

Moreover, one shows that for $\eta > 0$, $\eta < \frac{s}{r} - \frac{1+s}{1+r}$ and $\vartheta = \vartheta(\eta)$ such that

$$\frac{r}{r+1} < \vartheta(\eta) < r \left(\frac{1}{1+r} + \frac{\eta}{1+s} \right),$$

$$\lim_{n \rightarrow \infty} (n^s)^{\tilde{\rho}r/s} \mathbb{E} |X - \widehat{X}^{\alpha_n(\vartheta)}|^s = \infty$$

with $\tilde{\rho} = (1+s)/(1+r) + \eta$ and $\tilde{\rho}r/s < 1$. It remains a gap between this limiting relation and (2.3).

Example 2 Let $\mathbb{P}^X = P = C_p e^{-x^p} \mathbf{1}_{\{x \geq 0\}}$ for some $p \in (0, 1)$ and let $\gamma \in (0, r)$. For $n \geq 2$ set

$$\alpha_n = \alpha_n(\gamma) := \left\{ k \frac{u_n}{n} : k = 1, \dots, n \right\}$$

where (u_n) is defined by the implicit equation $e^{-u_n^p u_n^{(1-p)(1+r)}} = n^{-\gamma}$. Then $u_n \rightarrow +\infty$ and $(\frac{u_n}{n})^r = o(n^{-\gamma})$. Then, one derives after some standard computations that, for every $s > r$,

$$n^\gamma \mathbb{E} |X - \widehat{X}^{\alpha_n}|^s \asymp u_n^{(1-p)(s-r)} \longrightarrow +\infty \quad \text{as } n \rightarrow +\infty.$$

On the other hand

$$\mathbb{E} |X - \widehat{X}^{\alpha_n}|^r = \mathbb{E} (|X - \widehat{X}^{\alpha_n}| \mathbf{1}_{\{X \leq u_n\}})^r + C_p \int_{u_n}^{+\infty} (x - u_n)^r e^{-x^p} dx.$$

One checks as above that $\mathbb{E} (|X - \widehat{X}^{\alpha_n}| \mathbf{1}_{\{X \leq u_n\}})^r \asymp (\frac{u_n}{n})^r$ and then using elementary changes of variable that

$$e^{u_n^p} u_n^{(p-1)(1+r)} \int_{u_n}^{+\infty} (x - u_n)^r e^{-x^p} dx \longrightarrow p^{-(1+r)} \int_0^{+\infty} e^{-\xi} d\xi = p^{-(1+r)} \quad \text{as } n \rightarrow +\infty.$$

Consequently, using the implicit equation satisfied by u_n implies $\mathbb{E} |X - \widehat{X}^{\alpha_n}|^r \asymp \max \left((\frac{u_n}{n})^r, n^{-\gamma} \right) \asymp n^{-\gamma}$. This example shows that the rate in (2.3) of Corollary 1(b) may be optimal. Note however that this occurs with a sequence of quantizers (α_n) which is not rate optimal since $\gamma < r$.

2.2 V-quantization

Now let $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a nondecreasing function with $V(0) = 0$ and assume (1.2). (If $V(0) > 0$ one may turn to $V - V(0)$.)

A set $\alpha \subset \mathbb{R}^d$ with $1 \leq \text{card}(\alpha) \leq n$ is called V -optimal n -quantizer for P if

$$\mathbb{E} \min_{a \in \alpha} V(\|X - a\|) = e_{n,V}(P).$$

Let $\mathcal{C}_{n,V}(P)$ denote the set of all these optimal quantizers. We provide two properties of V -optimal quantizers. The first proposition shows that V -stationarity is a necessary condition for V -optimality. Its proof is a straightforward adaptation of Proposition 2.2 a) in [13]. For any Borel subset A of \mathbb{R}^d with $P(A) > 0$ let $P(\cdot | A) := P(\cdot \cap A)/P(A)$.

Proposition 2 Assume that $V_{[0,t_0]}$ is (strictly) increasing for some $t_0 > 0$ and $\text{card}(\text{supp}(P)) \geq n$. Let $\alpha \in \mathcal{C}_{n,V}(P)$ and let $\{C_a : a \in \alpha\}$ be a Voronoi partition of \mathbb{R}^d w.r.t. α . Then

$$\text{card}(\alpha) = n, P(C_a) > 0,$$

$$\{a\} \in \mathcal{C}_{1,V}(P(\cdot | C_a)) \quad \text{for every } a \in \alpha.$$

It is clear that the above condition on V cannot be dropped. Consider, for instance,

$$V(t) = \begin{cases} 0, & t \leq t_0 \\ 1, & t > t_0 \end{cases}, t_0 > 0$$

and a probability P with $\text{supp}(P) = [-t_0, t_0]$. Then $e_{n,V}(P) = 0$ and $\{0\} \in \mathcal{C}_{n,V}(P)$ for every $n \geq 1$.

The existence of V -optimal n -quantizers for lower semi-continuous loss functions is well known (cf. [1]). The proof is based on the observation that by the Fatou lemma, the distortion function

$$D_n : (\mathbb{R}^d)^n \rightarrow \mathbb{R}_+, D_n(a) := \mathbb{E} \min_{1 \leq i \leq n} V(\|X - a_i\|) \quad (2.5)$$

is lower semi-continuous for every n . It is to be noticed that lower semi-continuity of the nondecreasing function V simply means continuity on the left. On the other hand, since V has only countably many discontinuities one obtains the same property for the distortion function if P vanishes on spheres.

Proposition 3 Assume that V is continuous on the left or that P vanishes on spheres. Then for every $n \in \mathbb{N}$,

$$\mathcal{C}_{n,V}(P) \neq \emptyset.$$

Proof. If P vanishes on spheres, set $V_-(t) := V(t-)$ for $t > 0$ and $V_-(0) := 0$. Then V_- is nondecreasing, and continuous on the left and $V(\|X - a\|) = V_-(\|X - a\|)$ a.e. for every $a \in \mathbb{R}^d$. In particular, we have

$$\mathcal{C}_{n;V}(P) = \mathcal{C}_{n;V_-}(P) \text{ for every } n \in \mathbb{N}.$$

Therefore, we may assume without loss of generality that V is continuous on the left.

Let $a^k \rightarrow a$ in $(\mathbb{R}^d)^n$. Then

$$\min_{1 \leq i \leq n} \|x - a_i^k\| \rightarrow \min_{1 \leq i \leq n} \|x - a_i\|$$

and hence

$$\liminf_{k \rightarrow \infty} V(\min_{1 \leq i \leq n} \|x - a_i^k\|) \geq V(\min_{1 \leq i \leq n} \|x - a_i\|) \text{ for every } x \in \mathbb{R}^d.$$

It follows from the Fatou lemma that

$$\liminf_{k \rightarrow \infty} D_n(a^k) \geq D_n(a)$$

so that D_n is lower semi-continuous for every n .

Next consider a fixed $n \geq 1$ satisfying $e_n < e_{n-1}$, where $e_n := e_{n,V}(P)$ and $e_0 := \sup V$. Choose a constant c such that $e_n < c < e_{n-1}$. Choose s and S with $0 < s < S$ such that

$$V(S - s)P(B(0, s)) > c.$$

This is possible since $c < \sup V$. Then the level set $\{D_n \leq c\}$ is bounded. Assume the contrary. Then there exists a sequence $(b^k)_k$ in $\{D_n \leq c\}$ such that

$$\max_{1 \leq i \leq n} \|b_i^k\| \rightarrow \infty \text{ as } k \rightarrow \infty.$$

Assume without loss of generality $\|b_1^k\| \leq \dots \leq \|b_n^k\|$ for every k . Then

$$\|b_1^k\| \leq S \text{ for every } k$$

since otherwise

$$c \geq \int_{B(0,s)} V(\min_{1 \leq i \leq n} \|x - b_i^k\|) dP(x) \geq V(S - s)P(B(0, s)),$$

a contradiction. Therefore, $n \geq 2$ and taking a subsequence (also indexed by k) we may assume that there exists $m \in \{1, \dots, n-1\}$ and $b \in (\mathbb{R}^d)^m$ such that

$$\begin{aligned} b_i^k &\rightarrow b_i & , & \quad i \in \{1, \dots, m\} \\ \|b_i^k\| &\rightarrow \infty & , & \quad i \in \{m+1, \dots, n\}. \end{aligned}$$

Consequently,

$$\min_{1 \leq i \leq n} \|x - b_i^k\| \rightarrow \min_{1 \leq i \leq m} \|x - b_i\|$$

so that

$$\liminf_{k \rightarrow \infty} V(\min_{1 \leq i \leq n} \|x - b_i^k\|) \geq V(\min_{1 \leq i \leq m} \|x - b_i\|)$$

for every $x \in \mathbb{R}^d$. Fatou's lemma implies that

$$c \geq \liminf_{k \rightarrow \infty} D_n(b^k) \geq D_m(b) \geq e_{n-1},$$

hence a contradiction. Thus the level set $\{D_n \leq c\}$ is compact. Since a lower semi-continuous function on a compact set takes a minimum value on this set, $\mathcal{C}_{n,V}(P) \neq \emptyset$.

Now proceed inductively. For $n = 1$, if $e_1 = e_0$, then every $a \in \mathbb{R}^d$ is a V -optimal 1-quantizer. If $e_1 < e_0$, then by the above reasoning $\mathcal{C}_{1,V}(P) \neq \emptyset$. Assume $\mathcal{C}_{n-1,V}(P) \neq \emptyset$ for some $n \geq 2$. If $e_n = e_{n-1}$, then $\emptyset \neq \mathcal{C}_{n-1,V}(P) \subset \mathcal{C}_{n,V}(P)$ and if $e_n < e_{n-1}$, then again it follows from the above reasoning that $\mathcal{C}_{n,V}(P) \neq \emptyset$. \square

Abaya and Wise [2] observed that without any condition imposed on V or P , $\mathcal{C}_{n,V}(P)$ may be empty.

Finally, for the nonincreasing sequence of quantization errors we have:

Proposition 4 (a) Assume $V(0+) = 0$. Then

$$\lim_{n \rightarrow \infty} e_{n,V}(P) = 0.$$

(b) Assume that $V \upharpoonright [0, t_0]$ is (strictly) increasing for some $t_0 > 0$ and that $\text{supp}(P)$ is not finite. If $\mathcal{C}_{n,V}(P) \neq \emptyset$ for every n , then the sequence $(e_{n,V}(P))_n$ is (strictly) decreasing.

Proof. (a) Let $\{a_1, a_2, \dots\}$ be a countable dense subset of \mathbb{R}^d with $a_1 = 0$. Then

$$e_{n,V}(P) \leq \mathbb{E} V(\min_{1 \leq i \leq n} \|X - a_i\|) \text{ for every } n$$

and

$$\min_{1 \leq i \leq n} \|X - a_i\| \rightarrow 0$$

so that

$$V(\min_{1 \leq i \leq n} \|X - a_i\|) \rightarrow 0 \quad \text{everywhere.}$$

Since $V(\min_{1 \leq i \leq n} \|X - a_i\|) \leq V(\|X\|) \in L^1(\mathbb{P})$ it follows from the Lebesgue dominated convergence theorem that

$$\mathbb{E} V(\min_{1 \leq i \leq n} \|X - a_i\|) \rightarrow 0.$$

(b) follows immediately from Proposition 2. \square

2.3 Application to numerical integration

Emphasizing the aspect of numerical integration, the case $V(t) = t^r$ has been investigated by Pagès in [23] and [24]. For general loss functions V , consider the Hölder class H^V of measurable functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $|f(x) - f(y)| \leq V(\|x - y\|)$ for every $x, y \in \mathbb{R}^d$. By (1.2), $H^V \subset L^1(P)$. For every $f \in H^V$ and every finite set $\alpha \subset \mathbb{R}^d$ one obtains

$$\begin{aligned} \left| \int f dP - \sum_{a \in \alpha} P(C_a) f(a) \right| &= | \mathbb{E} f(X) - \mathbb{E} f(\widehat{X}^\alpha) | \\ &\leq \mathbb{E} | f(X) - f(\widehat{X}^\alpha) | \\ &\leq \mathbb{E} V(\|X - \widehat{X}^\alpha\|) = \mathbb{E} \min_{a \in \alpha} V(\|X - a\|) \end{aligned} \quad (2.6)$$

so that

$$\inf \left\{ \sup_{f \in H^V} \left| \int f dP - \sum_{a \in \alpha} P(C_a(\alpha)) f(a) \right| : \alpha \subset \mathbb{R}^d, \text{card}(\alpha) \leq n \right\} \leq e_{n,V}(P). \quad (2.7)$$

Thus $e_{n,V}(P)$ rules the rate of convergence of the error when approximating $\int f dP$ by $\sum_{a \in \alpha} P(C_a) f(a)$ uniformly over $f \in H^V$. If V satisfies $V(s + t) \leq V(s) + V(t)$ for every $s, t \in \mathbb{R}_+$, then $V(d(\cdot, \alpha)) \in H^V$ and thus we have equality in (2.6). For such loss functions V and probabilities P with compact Jordan-measurable support the n -asymptotics of $e_{n,V}(P)$ has been investigated in Chornaya [5] (cf. also [16], [17]).

2.4 Uniqueness of the V -optimal quantizer in one dimension

The uniqueness problem consists in finding conditions on V and P which ensure that $\mathcal{C}_{n,V}(P)$ is reduced to a single quantizer for every $n \in \mathbb{N}$. In higher dimension, this essentially never occurs so we will assume throughout this paragraph that $d = 1$. The earliest result in that direction is due to Fleischer [8] in 1964 for that standard loss function $V(t) = t^2$ and absolutely continuous distributions with strictly log-concave density function. This was successively extended by Kieffer and Trushkin (see [18], [26]) to convex loss functions V , still under some log-concavity assumption for the density function. More recently, Lamberton and Pagès proposed for quadratic loss functions a more geometric approach to uniqueness based on the so-called Mountain Pass Lemma (see [19]). This approach was then developed by Cohort in [6] in a more general setting by the use of Lagrangian techniques. This leads to the following result.

Assume that $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfies

$$\begin{aligned} V(t) &= \int_0^t v(s) ds \quad \forall t \in \mathbb{R}_+, \text{ where } v : \mathbb{R}_+ \rightarrow \mathbb{R} \text{ is continuous on the right} \\ &\text{with left-hand limits and } v(t) > 0 \text{ for every } t > 0. \end{aligned} \quad (2.8)$$

Then, $V(0) = 0$, V is continuous, increasing on \mathbb{R}_+ , right and left differentiable. Assume that $P = h.\lambda$, where the density function h satisfies

$$\text{supp } h = [m, M] \text{ in } \overline{\mathbb{R}} \text{ and } h \text{ is log-concave on } (m, M). \quad (2.9)$$

In particular, h has a right derivative denoted by h' . Furthermore one needs some joint assumption on P and V , namely

$$\forall a \in \mathbb{R}, \quad \int_{\mathbb{R}} V(|x - a|)h(x)dx < +\infty, \quad (2.10)$$

$$\forall a, b \in \mathbb{R}, a \leq b, \quad \int_{a \leq u \leq b} \sup v(|x - u|)(h(x) + |h'(x)|)dx < +\infty. \quad (2.11)$$

Then the following uniqueness result holds.

Theorem 2 (Cohort [6]) *Assume that V and the distribution P satisfy (2.8)-(2.11). Then, for every $n \in \mathbb{N}$, $\text{card } \mathcal{C}_{n,V}(P) = 1$.*

In fact, one proves a bit more than that: if the above assumptions hold, the distortion function D_n is differentiable and admits a unique critical point (zero of its gradient) in the set $\{(a_1, \dots, a_n) \in \mathbb{R}^n : m < a_1 < \dots < a_n < M\}$. This in turns implies that the V -quantization error function has no other local minima or local maxima.

One derives that uniqueness holds true if the loss function V is convex and if P satisfies the log-concavity assumption (2.8), provided that the integrability assumption (2.9) holds. But it also holds for any non-convex loss function V like piecewise affine increasing functions or antiderivative of positive functions like $V(t) = \int_0^t \exp(-1/s)ds$ (which is bounded).

3 Asymptotics of the quantization errors

In this section let $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a nondecreasing function satisfying $V(0) = 0$. Assume (1.2) where $\|\cdot\|$ denotes any norm on \mathbb{R}^d . We show that if $V(t)$ behaves locally at zero as t^r for some $r \in (0, \infty)$ then under additional moment conditions the V -quantization errors $e_{n,V}$ exhibit the same asymptotic behaviour as the r -quantization errors $e_{n,r}$ (see Theorem 1). More precisely, the condition is as follows . Let $r \in (0, \infty)$.

(A_r) $V(t) \sim t^r$ as $t \rightarrow 0+$ and there exists a nondecreasing function $W : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that

$$V(t) \leq t^r W(t) \text{ for every } t \in \mathbb{R}_+$$

and

$$\int \|x\|^{pr+\delta} dP(x) < \infty, \quad \int W(\|x\|)^{p/(p-1)} dP(x) < \infty \quad (3.1)$$

for some $p > 1$ and $\delta > 0$.

Notice that condition (A_r) implies $V(0+) = V(0) = 0$, $V(t) > 0$ for every $t > 0$ and the function W must satisfy $W(0) \geq 1$.

Theorem 3 *Assume (A_r) . Then*

$$\lim_{n \rightarrow \infty} n^{r/d} e_{n,V}(P) = J_{r,d} \|h\|_{d/(d+r)}, \quad (3.2)$$

where $h = dP^a/d\lambda^d$.

The preceding theorem contains Theorem 1 as a special case.

Example 3 Typical examples are exponential quantization with $V(t) = \exp(t^r) - 1$, where $W(t) = \exp(t^r)$ and $V(t) = (\log(1+t))^r$, where $W(t) = 1$. More generally, for $r = 1$, if $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is nondecreasing, continuously differentiable and $f(0) = 0$, $f'(0) > 0$, then $f(t) \sim f'(0)t$ and the choice $V = f/f'(0)$ and $W(t) = \sup_{0 \leq s \leq t} f'(s)/f'(0)$ is possible. Of course, loss functions of the type $V(t) = t^r \tilde{W}(t)$ with $\tilde{W} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ nondecreasing and $\tilde{W}(t) \rightarrow 1$ as $t \rightarrow 0+$ are included. An example is exponentially weighted r -quantization with $V(t) = t^r e^t$. Notice that the use of the loss function $V(t) = \log(t)$ (instead of $\log(1+t)$) leads by a suitable transformation to the quantization problem

$$\inf\{\exp(\mathbb{E} \log(\min_{a \in \alpha} \|X - a\|)) : \alpha \subset \mathbb{R}^d, 1 \leq \text{card}(\alpha) \leq n\}$$

which occurs as the limiting case of r -quantization as $r \rightarrow 0$. This $\log(t)$ -problem is essentially different from the $\log(1+t)$ -case and outside of the present setting (cf. Graf and Luschgy [14]).

We will use the following lemma for the proof of Theorem 3.

Lemma 1 Assume that $\text{supp}(P)$ is compact and $V(t) \sim t^r$ as $t \rightarrow 0+$. Let $(s_n)_n$ be a sequence in $(0, \infty)$ such that $\lim_{n \rightarrow \infty} s_n/s_{n+k} = 1$ for every $k \in \mathbb{N}$. Then

$$\limsup_{n \rightarrow \infty} s_n e_{n,V}(P) \leq \limsup_{n \rightarrow \infty} s_n e_{n,r}(P)$$

and

$$\liminf_{n \rightarrow \infty} s_n e_{n,V}(P) \geq \liminf_{n \rightarrow \infty} s_n e_{n,r}(P).$$

Proof. Set $K := \text{supp}(P)$. Let $c \in (0, 1)$. Choose $t_0 \in (0, \infty)$ such that $V(t) \geq ct^r$ for every $t \in [0, t_0]$ and then choose a finite set $\beta \subset \mathbb{R}^d$ with $\max\{d(x, \beta) : x \in K\} \leq t_0$. Let $m := \text{card}(\beta)$. For $\alpha \subset \mathbb{R}^d$ with $\text{card}(\alpha) \leq n$ one obtains

$$\begin{aligned} \int V(d(x, \alpha)) dP(x) &\geq \int_K V(d(x, \alpha \cup \beta)) dP(x) \\ &\geq c \int_K d(x, \alpha \cup \beta)^r dP(x) \\ &\geq c e_{n+m,r}(P) \end{aligned}$$

and hence

$$e_{n,V}(P) \geq c e_{n+m,r}(P).$$

Consequently,

$$\begin{aligned} \liminf_{n \rightarrow \infty} s_n e_{n,V}(P) &\geq c \liminf_{n \rightarrow \infty} \frac{s_n}{s_{n+m}} s_{n+m} e_{n+m,r}(P) \\ &\geq c \liminf_{n \rightarrow \infty} s_n e_{n,r}(P). \end{aligned}$$

Letting $c \rightarrow 1$ yields the lower estimate. As for the upper estimate, let $c > 1$ and choose $t_0 \in (0, \infty)$ such that $V(t) \leq ct^r$ for every $t \in [0, t_0]$. Choose β as above. One obtains for $\alpha \subset \mathbb{R}^d$ with $\text{card}(\alpha) \leq n - m, n > m$,

$$\begin{aligned} e_{n,V}(P) &\leq \int_K V(d(x, \alpha \cup \beta)) dP(x) \\ &\leq c \int_K d(x, \alpha \cup \beta)^r dP(x) \\ &\leq c \int d(x, \alpha)^r dP(x) \end{aligned}$$

and hence

$$e_{n,V}(P) \leq c e_{n-m,r}(P).$$

Consequently,

$$\begin{aligned} \limsup_{n \rightarrow \infty} s_n e_{n,V}(P) &\leq \limsup_{n \rightarrow \infty} \frac{s_n}{s_{n-m}} s_{n-m} e_{n-m,r}(P) \\ &\leq c \limsup_{n \rightarrow \infty} s_n e_{n,r}(P) \end{aligned}$$

and letting $c \rightarrow 1$ gives the upper estimate. \square

In particular, if $\lim_{n \rightarrow \infty} s_n e_{n,r}(P)$ exists (in $[0, \infty]$), then $\lim_{n \rightarrow \infty} s_n e_{n,V}(P)$ exists and

$$\lim_{n \rightarrow \infty} s_n e_{n,V}(P) = \lim_{n \rightarrow \infty} s_n e_{n,r}(P).$$

Proof of Theorem 3. The proof is given in three steps.

STEP 1. Assume that P has compact support. Then the limiting assertion (3.2) follows immediately from Theorem 1 and Lemma 1.

STEP 2. Let P be arbitrary. For a compact set $K \subset \mathbb{R}^d$ with $P(K) > 0$, the λ^d -density of the absolutely continuous part of $P(\cdot | K)$ is given by $\mathbf{1}_K h / P(K)$. Since for every $n \geq 1$,

$$e_{n,V}(P) \geq P(K) e_{n,V}(P(\cdot | K))$$

it follows from Step 1 that

$$\begin{aligned} \liminf_{n \rightarrow \infty} n^{r/d} e_{n,V}(P) &\geq P(K) J_{r,d} \|\mathbf{1}_K h / P(K)\|_{d/(d+r)} \\ &= J_{r,d} \|\mathbf{1}_K h\|_{d/(d+r)}. \end{aligned}$$

Letting $K \uparrow \mathbb{R}^d$ ($K = [-k, k]^d$ and $k \rightarrow \infty$, say) yields

$$\liminf_{n \rightarrow \infty} n^{r/d} e_{n,V}(P) \geq J_{r,d} \|h\|_{d/(d+r)} \quad (3.3)$$

STEP 3. Now assume (A_r) and that $\text{supp}(P)$ is not compact. Let $K \subset \mathbb{R}^d$ be a compact set with $0 < P(K) < 1$ and consider the decomposition $P = \sum_{i=1}^2 P(K_i) P(\cdot | K_i)$, where $K_1 = K$ and $K_2 = K^c$. First observe that for $n_1, n_2 \in \mathbb{N}$,

$$e_{n_1+n_2,V}(P) \leq \sum_{i=1}^2 P(K_i) e_{n_i,V}(P(\cdot | K_i)). \quad (3.4)$$

In fact, let $\varepsilon > 0$ and choose $\alpha_i \subset \mathbb{R}^d$ with $\text{card}(\alpha_i) = n_i$ such that

$$\int V(d(x, \alpha_i)) dP(\cdot | K_i)(x) \leq e_{n_i,V}(P(\cdot | K_i)) + \varepsilon.$$

Setting $\alpha := \alpha_1 \cup \alpha_2$ yields

$$\begin{aligned} e_{n_1+n_2,V}(P) &\leq \int V(d(x, \alpha)) dP(x) \leq \sum_{i=1}^2 P(K_i) \int V(d(x, \alpha_i)) dP(\cdot | K_i)(x) \\ &\leq \sum_{i=1}^2 P(K_i) e_{n_i,V}(P(\cdot | K_i)) + \varepsilon \end{aligned}$$

and hence (3.4).

Now for $\varepsilon \in (0, 1)$ and $n \geq \max\{\frac{1}{\varepsilon}, \frac{1}{1-\varepsilon}\}$, let $n_1 := [(1 - \varepsilon)n]$ and $n_2 := [\varepsilon n]$, where $[y]$ denotes the integer part of $y \in \mathbb{R}$. Then by (3.4),

$$n^{r/d} e_{n,V}(P) \leq \sum_{i=1}^2 P(K_i) \left(\frac{n}{n_i}\right)^{r/d} n_i^{r/d} e_{n_i,V}(P(\cdot | K_i)).$$

Consequently, using Step 1,

$$\limsup_{n \rightarrow \infty} n^{r/d} e_{n,V}(P) \leq (1 - \varepsilon)^{-r/d} J_{r,d} \|\mathbf{1}_K h\|_{d/(d+r)} + \varepsilon^{-r/d} P(K^c) \limsup_{n \rightarrow \infty} n^{r/d} e_{n,V}(P(\cdot | K^c)). \quad (3.5)$$

If α denotes an $(n - 1)$ -optimal pr -quantizer for P with p from (A_r) (that is, α is $(n - 1)$ -optimal for P with respect to the loss function $V(t) = t^{pr}$) and $\beta := \alpha \cup \{0\}$, then using the Hölder inequality

$$\begin{aligned} P(K^c) e_{n,V}(P(\cdot | K^c)) &\leq \int_{K^c} V(d(x, \beta)) dP(x) \\ &\leq \int_{K^c} d(x, \beta)^r W(d(x, \beta)) dP(x) \\ &\leq \left(\int_{K^c} d(x, \alpha)^{pr} dP(x) \right)^{1/p} \left(\int_{K^c} W(\|x\|)^q dP(x) \right)^{1/q} \\ &= e_{n-1,pr}(P)^{1/p} \left(\int_{K^c} W(\|x\|)^q dP(x) \right)^{1/q} \end{aligned}$$

for every $n \geq 2$, where $q = p/(p - 1)$. Consequently, by (3.1) and Theorem 1

$$P(K^c) \limsup_{n \rightarrow \infty} n^{r/d} e_{n,V}(P(\cdot | K^c)) \leq J_{pr,d}^{1/p} \|h\|_{d/(d+pr)}^{1/p} \left(\int_{K^c} W(\|x\|)^q dP(x) \right)^{1/q} < \infty.$$

The moment condition (3.1) implies

$$\int_{K^c} W(\|x\|)^q dP(x) \rightarrow 0 \text{ as } K \uparrow \mathbb{R}^d.$$

Therefore, letting $K \uparrow \mathbb{R}^d$ and then letting ε tend to zero yields.

$$\limsup_{n \rightarrow \infty} n^{r/d} e_{n,V}(P) \leq J_{r,d} \|h\|_{d/(d+r)}. \quad (3.6)$$

In view of (3.3) the proof is complete. \square

4 Asymptotic behaviour of optimal quantizers

In this section $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ denotes a nondecreasing function with $V(0) = 0$ satisfying $V(t) > 0$ for every $t > 0$. This is to exclude degenerate cases and holds under condition (A_r) . As for P , assume (1.2) and $\text{supp}(P)$ is not finite. Then $e_{n,V}(P) > 0$ for every $n \in \mathbb{N}$.

A sequence $(\alpha_n)_{n \in \mathbb{N}}$ of quantizers is called asymptotically V -optimal n -quantizer for P if $1 \leq \text{card}(\alpha_n) \leq n$ and $\mathbb{E} \min_{a \in \alpha_n} V(\|X - a\|) < \infty$ for every n and

$$\mathbb{E} \min_{a \in \alpha_n} V(\|X - a\|) \sim e_{n,V}(P) \text{ as } n \rightarrow \infty. \quad (4.1)$$

Under condition (A_r) and $P^a \neq 0$ this reads using Theorem 3

$$\lim_{n \rightarrow \infty} n^{r/d} \mathbb{E} \min_{a \in \alpha_n} V(\|X - a\|) = Q_r(P). \quad (4.2)$$

In particular, the size $\text{card}(\alpha_n)$ necessarily satisfies $\text{card}(\alpha_n)/n \rightarrow 1$ (see Theorem 4).

Mainly for such sequences of quantizers we investigate the weak convergence and the weak limits of the standard empirical measures

$$\frac{1}{n} \sum_{a \in \alpha_n} \delta_a,$$

the weighted empirical measures

$$\sum_{a \in \alpha_n} P(C_a(\alpha_n)) \delta_a,$$

$$n^{r/d} \sum_{a \in \alpha_n} \int_{C_a(\alpha_n)} V(\|x - a\|) dP(x) \delta_a$$

and the finite measures

$$n^{r/d} \int_{(\cdot)} \min_{a \in \alpha_n} V(\|x - a\|) dP(x),$$

where $\{C_a(\alpha_n) : a \in \alpha_n\}$ is a Voronoi partition of \mathbb{R}^d w.r.t. α_n .

The following simple observation turns out to be useful in order to deal with nonsingular but not absolutely continuous distributions.

Lemma 2 *Assume (A_r) and $P^a \neq 0$. If $(\alpha_n)_n$ is an asymptotically V -optimal n -quantizer for P then also for the probability distribution $P^a/P^a(\mathbb{R}^d)$.*

Proof. Let $\tilde{P}^a := P^a/P^a(\mathbb{R}^d)$. It follows from Theorem 3 applied to \tilde{P}^a that

$$\lim_{n \rightarrow \infty} n^{r/d} e_{n,V}(\tilde{P}^a) = Q_r(\tilde{P}^a) = Q_r(P)/P^a(\mathbb{R}^d).$$

Thus one obtains

$$\begin{aligned} Q_r(P) &= \lim_{n \rightarrow \infty} n^{r/d} P^a(\mathbb{R}^d) e_{n,V}(\tilde{P}^a) \\ &\leq \liminf_{n \rightarrow \infty} n^{r/d} P^a(\mathbb{R}^d) \int \min_{a \in \alpha_n} V(\|x - a\|) d\tilde{P}^a(x) \\ &\leq \limsup_{n \rightarrow \infty} n^{r/d} P^a(\mathbb{R}^d) \int \min_{a \in \alpha_n} V(\|x - a\|) d\tilde{P}^a(x) \\ &\leq \lim_{n \rightarrow \infty} n^{r/d} \int \min_{a \in \alpha_n} V(\|x - a\|) dP(x) \\ &= Q_r(P). \end{aligned}$$

Consequently,

$$\lim_{n \rightarrow \infty} n^{r/d} \int \min_{a \in \alpha_n} V(\|x - a\|) d\tilde{P}^a = Q_r(\tilde{P}^a). \quad \square$$

For P with $P^a = h\lambda^d \neq 0$ and $\|h\|_{d/(d+r)} < \infty$ (or what is the same, $Q_r(P) < \infty$) define the point density probability measure of P with respect to $r \in (0, \infty)$ by

$$P_r := h_r \lambda^d, h_r := \frac{h^{d/(d+r)}}{\int h^{d/(d+r)} d\lambda^d}. \quad (4.3)$$

For instance, if $P = N(0, 1)$ then $P_r = N(0, 1+r)$. Recall that $\|h\|_{d/(d+r)} < \infty$ is satisfied if $\int \|x\|^{r+\delta} dP(x) < \infty$ for some $\delta > 0$ (cf. Theorem 1).

The empirical measure problem concerns the weak convergence of $\sum_{a \in \alpha_n} \delta_a/n$ for asymptotically optimal quantizers $(\alpha_n)_n$ and has been solved for r -quantization and absolutely continuous distributions P by McClure [22] (for 1-dimensional distributions with compact support) and Bucklew [4]. See Graf and Luschgy [12] for a rigorous formulation and proof.

Theorem 4 Assume (A_r) and $P^a \neq 0$. Let $(\alpha_n)_n$ be an asymptotically V -optimal n -quantizer for P . Then

$$\frac{1}{n} \sum_{a \in \alpha_n} \delta_a \Rightarrow P_r \text{ as } n \rightarrow \infty.$$

This simply means that $\text{card}(\alpha_n \cap A)/n \rightarrow P_r(A)$ for every Borel set $A \subset \mathbb{R}^d$ with $P_r(\partial A) = 0$ where ∂A denotes the boundary of A and explains the notion "point density measure" for P_r . On the other hand, the standard empirical measure is not the right object to reconstruct P as soon as $P \neq P_r$.

As for the proof we will rely on the following lemma and a simple observation concerning weak convergence.

Lemma 3 (The "firewall" construction). Let $A \subset \mathbb{R}^d$ be a bounded set and $\varepsilon > 0$. Let

$$A_\varepsilon := \{x \in \mathbb{R}^d : d(x, A^c) > \varepsilon\} \quad \text{and} \quad A^\varepsilon := \{x \in \mathbb{R}^d : d(x, A) \leq \varepsilon\}.$$

Then there exists a finite set $\beta \subset \mathbb{R}^d$ such that

$$d(\cdot, \beta) < d(\cdot, A^c) \text{ on } A_\varepsilon \quad \text{and} \quad d(\cdot, \beta) < d(\cdot, A) \text{ on } (A^\varepsilon)^c.$$

In particular, for every finite codebook $\alpha \subset \mathbb{R}^d$,

$$d(\cdot, \alpha \cup \beta) = d(\cdot, (\alpha \cup \beta) \cap A) \text{ on } A_\varepsilon \quad \text{and} \quad d(\cdot, \alpha \cup \beta) = d(\cdot, (\alpha \cup \beta) \cap A^c) \text{ on } (A^\varepsilon)^c.$$

Proof. Since A_ε has compact closure and $\{d(\cdot, A) = \varepsilon\}$ is compact, there are finite sets $\beta_1 \subset A_\varepsilon$ and $\beta_2 \subset \{d(\cdot, A) = \varepsilon\}$ such that $d(x, \beta_1) \leq \varepsilon/2$ for every $x \in A_\varepsilon$ and $d(x, \beta_2) \leq \varepsilon/2$ for every $x \in \{d(\cdot, A) = \varepsilon\}$. Set $\beta := \beta_1 \cup \beta_2$. Clearly, for every $x \in A_\varepsilon$,

$$d(x, \beta) \leq \varepsilon/2 < \varepsilon < d(x, A^c).$$

Now let $x \in (A^\varepsilon)^c = \{d(\cdot, A) > \varepsilon\}$. There exists $y \in \text{cl}(A)$ (closure of A) such that

$$d(x, A) = \|x - y\| > \varepsilon.$$

Consider the line segment $\{z_s : s \in [0, 1]\}$ joining x and y , where $z_s := sx + (1 - s)y$. Since $s \mapsto d(z_s, A)$ is continuous, $d(z_0, A) = 0$ and $d(z_1, A) > \varepsilon$, the intermediate value theorem yields the existence of an $t \in (0, 1)$ such that $d(z_t, A) = \varepsilon$. We have

$$\begin{aligned} \|x - y\| &= \|x - z_t\| + \|z_t - y\|, \\ \|z_t - y\| &\geq d(z_t, A) = \varepsilon \end{aligned}$$

and

$$d(z_t, \beta) \leq \varepsilon/2.$$

Consequently,

$$\begin{aligned} d(x, \beta) &\leq \|x - z_t\| + d(z_t, \beta) \leq \|x - z_t\| + \varepsilon/2 \\ &\leq \|x - z_t\| + \varepsilon \leq \|x - z_t\| + \|z_t - y\| \\ &= \|x - y\| = d(x, A). \end{aligned}$$

□

Lemma 4 Let $\nu_n, n \in \mathbb{N}$ and ν be finite Borel measures on \mathbb{R}^d . Then

$$\nu_n \Rightarrow \nu$$

if (and only if)

$$\limsup_{n \rightarrow \infty} \nu_n(\mathbb{R}^d) \leq \nu(\mathbb{R}^d)$$

and

$$\liminf_{n \rightarrow \infty} \nu_n(O) \geq \nu(O)$$

for every bounded open subset $O \subset \mathbb{R}^d$ with $\nu(\partial O) = 0$.

Proof. Let $O \subset \mathbb{R}^d$ be a bounded open set. The boundaries of the bounded open subsets $O_\varepsilon = \{d(\cdot, O^c) > \varepsilon\}$ of O are disjoint for different values of $\varepsilon > 0$, so that at most countably many of them can have nonzero ν -measure. Choose a sequence $\varepsilon_k \downarrow 0$ with $\nu(\partial O_{\varepsilon_k}) = 0$ for every k . Consequently, by assumption for every k ,

$$\liminf_{n \rightarrow \infty} \nu_n(O) \geq \liminf_{n \rightarrow \infty} \nu_n(O_{\varepsilon_k}) \geq \nu(O_{\varepsilon_k}).$$

Since $O_{\varepsilon_k} \uparrow \{d(\cdot, O^c) > 0\} = O$ and hence $\nu(O_{\varepsilon_k}) \rightarrow \nu(O)$ as $k \rightarrow \infty$, this yields

$$\liminf_{n \rightarrow \infty} \nu_n(O) \geq \nu(O).$$

Clearly, this relation then holds for every (not necessarily bounded) open set O and in particular

$$\liminf_{n \rightarrow \infty} \nu_n(\mathbb{R}^d) \geq \nu(\mathbb{R}^d).$$

One obtains

$$\lim_{n \rightarrow \infty} \nu_n(\mathbb{R}^d) = \nu(\mathbb{R}^d)$$

and thus $\nu_n \Rightarrow \nu$. □

Proof of Theorem 4. Using Lemma 2 and since $(P^a/P^a(\mathbb{R}^d))_r = P_r$, we may assume without loss of generality that $P = P^a$. Then P and P_r are mutually absolutely continuous. Set

$$\mu_n := \frac{1}{n} \sum_{a \in \alpha_n} \delta_a.$$

Clearly, we have

$$\limsup_{n \rightarrow \infty} \mu_n(\mathbb{R}^d) = \limsup_{n \rightarrow \infty} \frac{\text{card}(\alpha_n)}{n} \leq 1. \quad (4.4)$$

Let $O \subset \mathbb{R}^d$ be an arbitrary bounded open set with $P_r(\partial O) = 0$. By Lemma 4, it is enough to show that

$$M(O) := \liminf_{n \rightarrow \infty} \mu_n(O) \geq P_r(O). \quad (4.5)$$

First assume $0 < P(O) < 1$. For $\varepsilon > 0$ such that $P(O_\varepsilon) > 0$ and $P(O^\varepsilon) < 1$ choose a finite codebook $\beta = \beta(\varepsilon, O) \subset \mathbb{R}^d$ according to Lemma 3 so that for every n ,

$$d(\cdot, \alpha_n \cup \beta) = d(\cdot, (\alpha_n \cup \beta) \cap O) \text{ on } O_\varepsilon$$

and

$$d(\cdot, \alpha_n \cup \beta) = d(\cdot, (\alpha_n \cup \beta) \cap O^c) \text{ on } (O^\varepsilon)^c.$$

Set $m := \text{card}(\beta)$, $\gamma_n := (\alpha_n \cup \beta) \cap O$ and $\delta_n := (\alpha_n \cup \beta) \cap O^c$. One obtains for every n

$$\begin{aligned} & \int \min_{a \in \alpha_n} V(\|x - a\|) dP(x) \\ & \geq \int_{O_\varepsilon} V(d(x, \alpha_n \cup \beta)) dP(x) + \int_{(O^\varepsilon)^c} V(d(x, \alpha_n \cup \beta)) dP(x) \\ & = \int_{O_\varepsilon} V(d(x, \gamma_n)) dP(x) + \int_{(O^\varepsilon)^c} V(d(x, \delta_n)) dP(x) \\ & \geq P(O_\varepsilon) e_{n\mu_n(O)+m, V}(P(\cdot | O_\varepsilon)) + P((O^\varepsilon)^c) e_{n\mu_n(O^c)+m, V}(P(\cdot | (O^\varepsilon)^c)). \end{aligned}$$

Choose a subsequence (also indexed by n) such that

$$\lim_{n \rightarrow \infty} \mu_n(O) = M(O)$$

and that

$$M(O^c) := \lim_{n \rightarrow \infty} \mu_n(O^c)$$

exists. Consequently, using Theorem 3 twice,

$$\begin{aligned} Q_r(P) &= \lim_{n \rightarrow \infty} n^{r/d} \int \min_{a \in \alpha_n} V(\|x - a\|) dP(x) \\ &\geq P(O_\varepsilon) M(O)^{-r/d} Q_r(P(\cdot | O_\varepsilon)) + P((O^\varepsilon)^c) M(O^c)^{-r/d} Q_r(P(\cdot | (O^\varepsilon)^c)). \end{aligned}$$

Letting $\varepsilon \downarrow 0$ yields $O_\varepsilon \uparrow O$ and $O^\varepsilon \downarrow O^c$. Since $P(O) = P(O^c)$, this implies

$$P(O_\varepsilon) Q_r(P(\cdot | O_\varepsilon)) \rightarrow P(O) Q_r(P(\cdot | O))$$

and

$$P((O^\varepsilon)^c) Q_r(P(\cdot | (O^\varepsilon)^c)) \rightarrow P(O^c) Q_r(P(\cdot | O^c)).$$

One obtains

$$Q_r(P) \geq P(O) (P(\cdot | O)) M(O)^{-r/d} + P(O^c) Q_r(P(\cdot | O^c)) M(O^c)^{-r/d}. \quad (4.6)$$

In particular, $\min\{M(O), M(O^c)\} > 0$ and moreover, by (4.4) we have $M(O) + M(O^c) \leq 1$. Now it follows from Hölder's inequality (for exponents less than 1) and the equality case of this inequality that $M(O) = P_r(O)$, that is, (4.5) (see [12], p. 98).

If $P(O) = 0$, then $P_r(O) = 0$ and (4.5) is obvious. If $P(O) = 1$, then omit the second summands. One gets

$$Q_r(P) \geq Q_r(P) M(O)^{-r/d}$$

and hence $M(O) \geq 1 = P_r(O)$. □

If, for a codebook α , one weights the Dirac mass δ_a for $a \in \alpha$ with the P -measure of its Voronoi regions then one arrives at the probability measure

$$\sum_{a \in \alpha} P(C_a(\alpha)) \delta_a.$$

Under suitable conditions this weighted empirical measure provides a reconstruction of P .

Proposition 5 *Let $(\alpha_n)_n$ be a sequence of quantizers satisfying $\limsup \text{card}(\alpha_n)/n \leq 1$.*

(a) *If $\lim_{n \rightarrow \infty} \mathbb{E} \min_{a \in \alpha_n} V(\|X - a\|) = 0$, then*

$$\widehat{X}^{\alpha_n} \Rightarrow X, \text{ where } \mathbb{P}^{\widehat{X}^{\alpha_n}} = \sum_{a \in \alpha_n} P(C_a(\alpha_n)) \delta_a$$

and hence

$$\widehat{X}^{\alpha_n} \rightarrow X \text{ a.e.}$$

(b) Assume (A_r) . If $(\alpha_n)_n$ is rate-optimal i.e. $\mathbb{E} \min_{a \in \alpha_n} V(\|X - a\|) = O(e_{n,V}(P))$, then the sequence $(n^{1/d}(X - \widehat{X}^{\alpha_n}))_n$ is uniformly tight. Moreover, if $r > d$

$$n^\vartheta \|X - \widehat{X}^{\alpha_n}\| \rightarrow 0 \text{ a.e. as } n \rightarrow \infty$$

for every $\vartheta \in (0, \frac{1}{d} - \frac{1}{r})$ and .

Proof. (a) For every $\varepsilon > 0$,

$$\mathbb{P}(\|\widehat{X}^{\alpha_n} - X\| \geq \varepsilon) \leq \frac{1}{V(\varepsilon)} \mathbb{E} V(\|\widehat{X}^{\alpha_n} - X\|) = \frac{1}{V(\varepsilon)} \mathbb{E} \min_{a \in \alpha_n} V(\|X - a\|) \rightarrow 0.$$

The a.e. assertion follows from Proposition 1.

(b) Using Theorem 3, we have for $\varepsilon > 0$ and some constant $c \in (0, \infty)$,

$$\sum_{n=1}^{\infty} \mathbb{P}(n^\vartheta \|\widehat{X}^{\alpha_n} - X\| \geq \varepsilon) \leq \sum_{n=1}^{\infty} \frac{1}{V(\varepsilon/n^\vartheta)} \mathbb{E} \min_{a \in \alpha_n} V(\|X - a\|) \leq c\varepsilon^{-r} \sum_{n=1}^{\infty} n^{r\vartheta} n^{-r/d} < \infty.$$

Uniform tightness of $(n^{1/d}(X - \widehat{X}^{\alpha_n}))_n$ follows from similar estimates. \square

Next we investigate the local quantization errors

$$\int_A \min_{a \in \alpha_n} V(\|x - a\|) dP(x), A \subset \mathbb{R}^d$$

for asymptotically V -optimal quantizers. For r -quantization and absolutely continuous measures a (probably not quite correct version of the) statement of the following local error asymptotics is contained in Bucklew [4].

Theorem 5 Assume (A_r) and $P^a \neq 0$. Let $(\alpha_n)_n$ be an asymptotically V -optimal n -quantizer for P . For $n \in \mathbb{N}$, define a finite measure on \mathbb{R}^d by

$$\frac{d\nu_n}{dP}(x) := n^{r/d} \min_{a \in \alpha_n} V(\|x - a\|), x \in \mathbb{R}^d.$$

Then

$$\nu_n \implies Q_r(P)P_r \quad \text{as } n \rightarrow \infty.$$

Thus we see that the limiting measure is the $Q_r(P)$ -weighted point density measure.

In terms of the α_n -quantization of X the preceding theorem reads

$$\lim_{n \rightarrow \infty} n^{r/d} \mathbb{E} g(X) V(\|X - \widehat{X}^{\alpha_n}\|) = Q_r(P) \int g dP_r$$

for every bounded λ^d -a.e. continuous function $g : \mathbb{R}^d \rightarrow \mathbb{R}$.

Proof. By asymptotic optimality of $(\alpha_n)_n$,

$$\lim_{n \rightarrow \infty} \nu_n(\mathbb{R}^d) = Q_r(P).$$

According to Lemma 4 it remains to show that for every bounded open set $O \subset \mathbb{R}^d$ with $P_r(\partial O) = 0$,

$$\liminf_{n \rightarrow \infty} \nu_n(O) \geq Q_r(P)P_r(O). \quad (4.7)$$

Let O be such a set. Notice that P^a and P_r are mutually absolutely continuous. So if $P^a(O) = 0$, then $P_r(O) = 0$ and (4.7) is clearly true. Now assume $P^a(O) > 0$. For $\varepsilon > 0$ such that $P^a(O_\varepsilon) > 0$ choose a finite codebook $\beta = \beta(\varepsilon, O)$ according to Lemma 3 so that for every n ,

$$d(\cdot, \alpha_n \cup \beta) = d(\cdot, (\alpha_n \cup \beta) \cap O) \text{ on } O_\varepsilon.$$

Set $m := \text{card}(\beta)$ and $\gamma_n := (\alpha_n \cup \beta) \cap O$. One obtains for every n ,

$$\begin{aligned} \nu_n(O) &= n^{r/d} \int_O \min_{a \in \alpha_n} V(\|x - a\|) dP(x) \\ &\geq n^{r/d} \int_{O_\varepsilon} V(d(x, \alpha_n \cup \beta)) dP(x) \\ &= n^{r/d} \int_{O_\varepsilon} V(d(x, \gamma_n)) dP(x) \\ &\geq n^{r/d} P(O_\varepsilon) e_{n\mu_n(O)+m, V}(P(\cdot | O_\varepsilon)), \end{aligned}$$

where $\mu_n := \sum_{a \in \alpha_n} \delta_a/n$ denotes the standard empirical measure. It follows from Theorem 4 that

$$\lim_{n \rightarrow \infty} \mu_n(O) = P_r(O) > 0.$$

Consequently, by Theorem 3,

$$\liminf_{n \rightarrow \infty} \nu_n(O) \geq P(O_\varepsilon) P_r(O)^{-r/d} Q_r(P(\cdot | O_\varepsilon)).$$

Letting $\varepsilon \downarrow 0$ yields $O_\varepsilon \uparrow O$ and hence

$$\liminf_{n \rightarrow \infty} \nu_n(O) \geq P(O) P_r(O)^{-r/d} Q_r(P(\cdot | O)) = Q_r(P) P_r(O).$$

This is (4.7). □

The same limiting measure occurs when considering n -dependent localization of the quantization errors at Voronoi regions. This is made precise in the following theorem which is seemingly a new result even in the r -quantization framework.

Theorem 6 *Assume (A_r) and $P^a \neq 0$. Let $(\alpha_n)_n$ be an asymptotically V -optimal n -quantizer for P . For $n \in \mathbb{N}$, let $\{C_a(\alpha_n) : a \in \alpha_n\}$ be a Voronoi partition of \mathbb{R}^d w.r.t. α_n . Then*

$$n^{r/d} \sum_{a \in \alpha_n} \int_{C_a(\alpha_n)} V(\|x - a\|) dP(x) \delta_a \Rightarrow Q_r(P) P_r \text{ as } n \rightarrow \infty.$$

In terms of the α_n -quantization of X this reads

$$\lim_{n \rightarrow \infty} n^{r/d} \mathbb{E} g(\hat{X}^{\alpha_n}) V(\|X - \hat{X}^{\alpha_n}\|) = Q_r(P) \int g dP_r$$

for every bounded λ^d -a.e. continuous function $g : \mathbb{R}^d \rightarrow \mathbb{R}$.

Combining the preceding theorem and Theorem 4 provides an indication for the uniformity feature of local distortion

$$\int_{C_{a_n}(\alpha_n)} V(\|x - a_n\|) dP(x) \sim \frac{e_{n, V}(P)}{n}.$$

However, rigorous proofs are available only in dimension 1 for the r -quantization problem (see Delattre et al. [7]).

If α_n is (exactly) n -optimal for P for every n and $V_{|[0,t_0]}$ is (strictly) increasing for some $t_0 > 0$ then the above result reads

$$n^{r/d} \sum_{a \in \alpha_n} e_{1,V}(P(\cdot | C_a(\alpha_n))P(C_a(\alpha_n))\delta_a \Rightarrow Q_r(P)P_r$$

(see Proposition 2).

The next natural question in view of Theorems 5 and 6 is to elucidate the weak asymptotics of the uniformly tight sequence $(n^{1/d}(X - \widehat{X}^{\alpha_n}))_n$ as $n \rightarrow \infty$ (see Proposition 5). For instance, in case $\mathbb{P}^X = U([0, 1])$ we have (at least for certain $(\alpha_n)_n$ like the sequence of optimal n -codebooks $\alpha_n = \{(2i - 1)/(2n), i = 1, \dots, n\}$)

$$n(X - \widehat{X}^{\alpha_n}) \Rightarrow U([-1/2, 1/2]).$$

However, one must be aware that a simple result is hopeless in higher dimension due to the usual “geometric” non-uniqueness of the optimal quantizers (which occurs *e.g.* for the uniform distribution on the unit square $[-1, 1]^2$ since one easily checks that an optimal quantizer cannot be invariant by a rotation $\mathcal{R}(0; \pi/2)$ of angle $\pi/2$).

Theorem 6 can be deduced from Theorem 5 and the following lemma which provides an improvement of one part of Lemma 3 under a mild assumption on the quantizers α_n in that no extra codebook is needed.

Lemma 5 *Let $A \subset \mathbb{R}^d$ be a bounded set and $\varepsilon > 0$. Let $(\alpha_n)_n$ be a sequence of finite codebooks such that*

$$\widehat{X}^{\alpha_n} \Rightarrow X.$$

Then for all large enough n ,

$$d(\cdot, \alpha_n) = d(\cdot, \alpha_n \cap A) \text{ on } A_\varepsilon \cap \text{supp}(P)$$

where $A_\varepsilon = \{d(\cdot, A^c) > \varepsilon\}$ or what is the same,

$$A_\varepsilon \cap \text{supp}(P) \subset \bigcup_{a \in \alpha_n \cap A} C_a(\alpha_n).$$

Proof. Since $A \cap \text{supp}(P)$ has a compact closure contained in $\text{supp}(P)$, it follows from Proposition 1 that

$$\sup\{d(x, \alpha_n) : x \in A \cap \text{supp}(P)\} \leq \varepsilon$$

for all large enough n , $n \geq n_0$ say. For $x \in A_\varepsilon \cap \text{supp}(P)$ and $n \geq n_0$, choose $a_n \in \alpha_n$ such that

$$\|x - a_n\| = d(x, \alpha_n) \leq \varepsilon.$$

Then $a_n \in A$. Consequently,

$$d(x, \alpha_n) = d(x, \alpha_n \cap A).$$

□

Proof of Theorem 6. Let μ_n denote the finite discrete measure of the left hand side of the limiting assertion. Then

$$\lim_{n \rightarrow \infty} \mu_n(\mathbb{R}^d) = \lim_{n \rightarrow \infty} n^{r/d} \int \min_{a \in \alpha_n} V(\|x - a\|) dP(x) = Q_r(P).$$

Let $O \subset \mathbb{R}^d$ be a bounded open set, $C(\alpha_n, O) := \cup\{C_a(\alpha_n) : a \in \alpha_n \cap O\}$ and $\varepsilon > 0$. By Lemma 5 and Proposition 5, for all large enough n ,

$$\begin{aligned}\mu_n(O) &= n^{r/d} \sum_{a \in \alpha_n \cap O} \int_{C_a(\alpha_n)} V(d(x, \alpha_n)) dP(x) \\ &= n^{r/d} \int_{C(\alpha_n, O)} V(d(x, \alpha_n)) dP(x) \\ &\geq n^{r/d} \int_{O_\varepsilon} V(d(x, \alpha_n)) dP(x) \\ &= \nu_n(O_\varepsilon)\end{aligned}$$

with the finite measure ν_n from Theorem 4. Since O_ε is open, it follows from Theorem 5 that

$$\liminf_{n \rightarrow \infty} \mu_n(O) \geq \liminf_{n \rightarrow \infty} \nu_n(O_\varepsilon) \geq Q_r(P)P_r(O_\varepsilon).$$

Letting $\varepsilon \downarrow 0$ yields

$$\liminf_{n \rightarrow \infty} \mu_n(O) \geq Q_r(P)P_r(O)$$

and the assertion follows from Lemma 4. \square

Finally, we comment on probabilities P with compact support. In this case, all results of this section hold with (A_r) replaced by the condition $V(t) \sim t^r$ as $t \rightarrow 0+$. Moreover, Theorem 4 then follows immediately from the corresponding result for r -quantization. This is a consequence of the following proposition.

Proposition 6 *Assume that $\text{supp}(P)$ is compact and $V(t) \sim t^r$ as $t \rightarrow 0+$. Assume that $\lim_{n \rightarrow \infty} e_{n,r}(P)/e_{n+k,r}(P) = 1$ for every k . Let (α_n) be an asymptotically V -optimal n -quantizer for P . Then (α_n) is asymptotically r -optimal n -quantizer for P .*

Proof. An application of Lemma 1 with $s_n = 1/e_{n,r}(P)$ yields

$$e_{n,V}(P) \sim e_{n,r}(P) \text{ as } n \rightarrow \infty.$$

Let $K := \text{supp}(P)$. Since $\lim e_{n,V}(P) = 0$, one obtains $\lim \int V(d(x, \alpha_n)) dP(x) = 0$. Therefore, by Proposition 1,

$$\lim \max_{x \in K} d(x, \alpha_n) = 0.$$

Let $c \in (0, 1)$ and choose $t_0 \in (0, \infty)$ with $V(t) \geq ct^r$ for every $t \in [0, t_0]$. Then for all large enough n (such that $\max_{x \in K} d(x, \alpha_n) \leq t_0$),

$$e_{n,r}(P) \leq \int d(x, \alpha_n)^r dP(x) \leq \frac{1}{c} \int V(d(x, \alpha_n)) dP(x).$$

Letting $c \rightarrow 1$ yields

$$e_{n,r}(P) \sim \int d(x, \alpha_n)^r dP(x).$$

\square

5 Asymptotic quantization for self-similar probability measures

For singular probability measures P the main result of Section 3 only yields

$$e_{n,V}(P) = o(n^{-r/d}) \text{ as } n \rightarrow \infty \quad (5.1)$$

and consequently, Theorems 5 and 6 read

$$n^{r/d} \int_{(\cdot)} \min_{a \in \alpha_n} V(\|x - a\|) dP(x) \Rightarrow 0$$

and

$$n^{r/d} \sum_{a \in \alpha_n} \int_{C_a(\alpha_n)} V(\|x - a\|) dP(x) \delta_a \Rightarrow 0$$

while Theorem 4 does not apply.

In this section we investigate the precise asymptotics of the V -quantization errors and the point density measure for self-similar probabilities on \mathbb{R}^d which provide an interesting class of (continuous) probability measures with compact support. Most of these probability measures are singular.

Let $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be nondecreasing with $V(0) = 0$ and let $\|\cdot\|$ denote any norm on \mathbb{R}^d . The condition (A_r) is replaced by

$$(B_r) \quad V(t) \sim t^r \text{ as } t \rightarrow 0+.$$

Notice that (B_r) implies $V(t) > 0$ for every $t > 0$.

In what follows N is a natural number with $N \geq 2$ and $S_1, \dots, S_N : \mathbb{R}^d \rightarrow \mathbb{R}^d$ are contractive similitudes. Let s_i be the contraction number of S_i , i.e. $s_i \in (0, 1)$ and $\|S_i(x) - S_i(y)\| = s_i \|x - y\|$ for all $x, y \in \mathbb{R}^d$. Sometimes the N -tuple (S_1, \dots, S_N) is called an iterated function system. Its attractor A is the unique nonempty compact subset A of \mathbb{R}^d with

$$A = \bigcup_{i=1}^N S_i(A). \quad (5.2)$$

For every probability vector (p_1, \dots, p_N) there exists a unique Borel probability P on \mathbb{R}^d which satisfies

$$P = \sum_{i=1}^N p_i P^{S_i}, \quad (5.3)$$

where P^{S_i} denotes the image measure of P under S_i . P is called the self-similar probability measure corresponding to $(S_1, \dots, S_N; p_1, \dots, p_N)$.

We will always assume that $p_i > 0$ for every i so that A equals the support $\text{supp}(P)$ of P . (S_1, \dots, S_N) is said to satisfy the open-set-condition (OSC) if there exists a nonempty open set $U \subset \mathbb{R}^d$ with $S_i(U) \subset U$ and $S_i(U) \cap S_j(U) = \emptyset$ for all i, j with $i \neq j$. From now on let (S_1, \dots, S_N) satisfy the OSC. For $r \in (0, \infty)$ there is a unique number $D_r = D_r(P) \in (0, \infty)$ satisfying

$$\sum_{i=1}^N (p_i s_i^r)^{D_r/(D_r+r)} = 1 \quad (5.4)$$

(cf. [12], Lemma 14.4.). We will see (Theorem 7 below) that under condition (B_r) the number D_r equals the V -quantization dimension of P defined by

$$\lim_{n \rightarrow \infty} \frac{\log n}{-\frac{1}{r} \log e_{n,V}(P)}$$

which in view of (5.1) is bounded above by the space dimension d . In the nonsingular setting, under condition (A_r) , the V -quantization dimension is simply the space dimension itself (see Theorem 3). In the sequel let P be the self-similar probability corresponding to $(S_1, \dots, S_N; p_1, \dots, p_N)$.

Let $\{1, \dots, N\}^*$ denote the set of all finite words (sequences) on the alphabet $1, \dots, N$ including the empty word \emptyset . For $\sigma \in \{1, \dots, N\}^*$ set

$$S_\sigma := \begin{cases} id_{\mathbb{R}^d}, & \sigma = \emptyset \\ S_{\sigma_1} \circ \dots \circ S_{\sigma_n}, & \sigma = \sigma_1 \dots \sigma_n \end{cases}$$

and

$$s_\sigma := \begin{cases} 1, & \sigma = \emptyset \\ \prod_{i=1}^n s_{\sigma_i}, & \sigma = \sigma_1 \dots \sigma_n. \end{cases}$$

p_σ is defined analogously.

Observe first that the existence of V -optimal n -quantizers for P is ensured if V is continuous on the left (see Proposition 3) and without any condition on V if the underlying norm on \mathbb{R}^d is the l_2 -norm. This follows again from Proposition 3 and the fact that P vanishes on l_2 -spheres (see [9]).

The precise asymptotic behaviour of the quantization errors $e_{n,V}$ and the point density measure of P w.r.t. V -quantization can be deduced from recent results on the r -quantization problem. Define P_r as the self-similar probability corresponding to $(S_1, \dots, S_N, q_1, \dots, q_N)$ where

$$q_i = (p_i s_i^r)^{D_r/(D_r+r)}. \quad (5.5)$$

A vector $(a_1, \dots, a_N) \in (\mathbb{R} \setminus \{0\})^N$ is called arithmetic if $(a_1, \dots, a_N) \in a\mathbb{Z}^N$ for some $a \in \mathbb{R}$. We need the following condition:

(C_r) $(\log(p_1 s_1^r), \dots, \log(p_N s_N^r))$ is not arithmetic.

Theorem 7 *Assume (B_r) .*

(a) $e_{n,V}(P) \approx n^{-r/D_r}$ as $n \rightarrow \infty$.

(b) *Assume (C_r) . Then*

$$Q_r(P) := \lim_{n \rightarrow \infty} n^{r/D_r} e_{n,V}(P)$$

exists in $(0, \infty)$.

(c) *Assume (C_r) . Let $(\alpha_n)_n$ be an asymptotically V -optimal n -quantizer for P . Then*

$$\frac{1}{n} \sum_{a \in \alpha_n} \delta_a \Rightarrow P_r \text{ as } n \rightarrow \infty.$$

Proof. It is known that

$$e_{n,r}(P) \approx n^{-r/D_r}$$

and under (C_r) ,

$$\lim_{n \rightarrow \infty} n^{r/D_r} e_{n,r}(P) \text{ exists in } (0, \infty)$$

and

$$\frac{1}{n} \sum_{b \in \beta_n} \delta_b \Rightarrow P_r$$

for any asymptotically r -optimal n -quantizer $(\beta_n)_n$ (cf. Graf and Luschgy [15]). Therefore, the assertions (a) and (b) follow from Lemma 1 and (c) follows from Proposition 6. \square

It is known that without the condition (C_r) parts (b) and (c) of the preceding theorem are not true. An example is the classical Cantor distribution (see [15]).

Notice that by Lemma 1, $Q_r(P)$ in fact depends on r and not on the exact form of V .

Next we will investigate the local quantization errors of asymptotically V -optimal n -quantizers $(\alpha_n)_n$ for the self-similar distribution P (in the sense of (4.1)).

Theorem 8 *Assume (B_r) and (C_r) . For $n \in \mathbb{N}$, define a finite measure on \mathbb{R}^d by*

$$\frac{d\nu_n}{dP}(x) := n^{r/D_r} \min_{a \in \alpha_n} V(\|x - a\|), x \in \mathbb{R}^d.$$

Then

$$\nu_n \Rightarrow Q_r(P)P_r \text{ as } n \rightarrow \infty$$

with $Q_r(P)$ from Theorem 7(b).

Proof. Let $O \subset \mathbb{R}^d$ be an arbitrary open set. By Lemma 5.4 in [15] there exists a finite or infinite (possibly empty) sequence $(\sigma^{(k)})_k$ in $\{1, \dots, N\}^*$ such that $(S_{\sigma^{(k)}}(A))$ is a sequence of pairwise disjoint subsets of O with $P_r(O) = \sum_k P_r(S_{\sigma^{(k)}}(A))$. For $\sigma \in \{1, \dots, N\}^*$ set $\alpha_n(\sigma) := \{a \in \alpha_n : W_a(\alpha_n) \cap S_\sigma(A) \neq \emptyset\}$ and $n(\sigma) := \text{card}(\alpha_n(\sigma))$ where $W_a(\alpha_n)$ is the closed Voronoi region $W_a(\alpha_n) = \{x \in \mathbb{R}^d : \|x - a\| = d(x, \alpha_n)\}$. It follows that

$$\begin{aligned} \nu_n(O) &= n^{r/D_r} \int_O V(d(x, \alpha_n)) dP(x) \\ &\geq n^{r/D_r} \sum_k \int_{S_{\sigma^{(k)}}(A)} V(d(x, \alpha_n(\sigma^{(k)}))) dP(x). \end{aligned}$$

By the self-similarity of P we obtain

$$\begin{aligned} \int_{S_{\sigma^{(k)}}(A)} V(d(x, \alpha_n(\sigma^{(k)}))) dP(x) &= p_{\sigma^{(k)}} \int V(d(S_{\sigma^{(k)}}(x), \alpha_n(\sigma^{(k)}))) dP(x) \\ &= p_{\sigma^{(k)}} \int V(s_{\sigma^{(k)}} d(x, S_{\sigma^{(k)}}^{-1} \alpha_n(\sigma^{(k)}))) dP(x). \end{aligned}$$

For $s > 0$ let $V_s : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be defined by

$$V_s(t) := s^{-r} V(st).$$

Then V_s is nondecreasing and we deduce

$$\int_{S_{\sigma^{(k)}}(A)} V(d(x, \alpha_n(\sigma^{(k)}))) dP(x) \geq p_{\sigma^{(k)}} s_{\sigma^{(k)}}^r e_{n(\sigma^{(k)}), V_s(\sigma^{(k)})}(P).$$

Since V_s satisfies (B_r) and since

$$\lim_{n \rightarrow \infty} \frac{n(\sigma)}{n} = q_\sigma = (p_\sigma s_\sigma^r)^{D_r/(r+D_r)}$$

for every $\sigma \in \{1, \dots, N\}^*$ by (24) in [15] we obtain from Theorem 7(b) that

$$\lim_{n \rightarrow \infty} n(\sigma^{(k)})^{r/D_r} e_{n(\sigma^{(k)}), V_s(\sigma^{(k)})}(P) = Q_r(P),$$

hence

$$\begin{aligned}
& \liminf_{n \rightarrow \infty} n^{r/D_r} \int_{S_{\sigma^{(k)}}(A)} V(d(x, \alpha_n(\sigma^{(k)})) dP(x) \\
& \geq \liminf_{n \rightarrow \infty} p_{\sigma^{(k)}} s_{\sigma^{(k)}}^r \left(\frac{n}{n(\sigma^{(k)})} \right)^{r/D_r} n(\sigma^{(k)})^{r/D_r} e_{n(\sigma^{(k)}), V_{S_{\sigma^{(k)}}}}(P) \\
& = p_{\sigma^{(k)}} s_{\sigma^{(k)}}^r ((p_{\sigma^{(k)}} s_{\sigma^{(k)}}^r)^{-D_r/(r+D_r)})^{r/D_r} Q_r(P) \\
& = Q_r(P) (p_{\sigma^{(k)}} s_{\sigma^{(k)}}^r)^{D_r/(r+D_r)}.
\end{aligned}$$

We conclude that

$$\liminf_{n \rightarrow \infty} \nu_n(O) \geq Q_r(P) \sum_k (p_{\sigma^{(k)}} s_{\sigma^{(k)}}^r)^{D_r/(r+D_r)}.$$

Since $P_r(S_{\sigma^{(k)}}(A)) = (p_{\sigma^{(k)}} s_{\sigma^{(k)}}^r)^{D_r/(r+D_r)}$ this implies

$$\liminf_{n \rightarrow \infty} \nu_n(O) \geq Q_r(P) P_r(O).$$

Since $\nu_n(\mathbb{R}^d) = n^{r/D_r} \int V(d(x, \alpha_n)) dP(x)$ and since (α_n) is asymptotically V -optimal we have $\lim_{n \rightarrow \infty} \nu_n(\mathbb{R}^d) = Q_r(P)$. Hence Lemma 4 yields the conclusion of the theorem. \square

Now the asymptotics for error localization at Voronoi regions can be deduced from Theorem 8 and Lemma 5 just as Theorem 6 follows from Theorem 5 and Lemma 5.

Theorem 9 *Assume (B_r) and (C_r) . For $n \in \mathbb{N}$, let $\{C_a(\alpha_n) : a \in \alpha_n\}$ be a Voronoi partition of \mathbb{R}^d w.r.t. α_n . Then*

$$n^{r/D_r} \sum_{a \in \alpha_n} \int_{C_a(\alpha_n)} V(\|x - a\|) dP(x) \delta_a \Rightarrow Q_r(P) P_r \text{ as } n \rightarrow \infty.$$

Combining the preceding theorem and Theorem 7 (c) provides an indication that the uniformity feature

$$\int_{C_{a_n}(\alpha_n)} V(\|x - a_n\|) dP(x) \sim \frac{e_{n,V}(P)}{n}$$

holds for self-similar probabilities P satisfying (C_r) . However, as yet no rigorous proof is available.

References

- [1] Abaya, E.F. and Wise, G.L. (1982). On the existence of optimal quantizers. *IEEE Trans. Inform. Theory*, **28**, 937-940.
- [2] Abaya, E.F. and Wise, G.L. (1984). Some remarks on the existence of optimal quantizers. *Statistics and Probab. Letters*, **2**, 349-351.
- [3] Bucklew, J.A. and Wise, G.L. (1982). Multidimensional asymptotic quantization theory with r th power distortion. *IEEE Trans. Inform. Theory*, **28**, 239-247.
- [4] Bucklew, J.A. (1984). Two results on the asymptotic performance of quantizers. *IEEE Trans. Inform. Theory*, **30** 341-348.
- [5] Chornaya, E.V. (1995). On the optimization of weighted cubature formulae on certain classes of continuous functions. *East J. on Approximations*, **1**, 47-60.
- [6] Cohort, P. (1998). A geometric method for uniqueness of locally optimal quantizer. Technical Report No. 464, LPMA and Ph.D. Thesis, *Sur quelques problèmes de quantification*, 2000, Univ. Paris 6.

- [7] Delattre, S., Fort, J.-C. and Pagès, G. (2004). Local distortion and μ -mass of the cells of one dimensional asymptotically optimal quantizers, *Communications in Statistics*, **33**, n⁰⁵, 1087-1118.
- [8] Fleischer, P.E. (1964). Sufficient conditions for achieving minimum distortion in a quantizer. *IEEE Int. Conv. Rec.*, part I, 104-111.
- [9] Gardner, W.R. and Rao, B.D. (1995). Theoretical analysis of the high-rate vector quantization of LPC parameters. *IEEE Trans. Speech Audio Processing*, **3**, 367-381.
- [10] Gersho, A. and Gray, R.M. (1992). *Vector Quantization and Signal Compression*. Kluwer, Boston.
- [11] Graf, S. and Luschgy, H. (1996). The quantization dimension of self-similar sets. Research Report No. 9, Dept. of Mathematics and Computer Science, Univ. of Passau.
- [12] Graf, S. and Luschgy, H. (2000). *Foundations of Quantization for Probability Distributions*. Lect. Notes in Math. 1730, Springer, Berlin.
- [13] Graf, S. and Luschgy, H. (2002). Rates of convergence for the empirical quantization error. *Ann. Probab.*, **30**, 874-897.
- [14] Graf, S. and Luschgy, H. (2004). Quantization for probability measures with respect to the geometric mean error. *Math. Proc. Cambridge Phil. Soc.*, **136**, 687-717.
- [15] Graf, S. and Luschgy, H. (2003). The point density measure in the quantization of self-similar probabilities. *Math. Proc. Cambridge Phil. Soc.* (to appear).
- [16] Gruber, P.M. (2002). Optimale Quantisierung. *Math. Semesterber.*, **49**, 227-251.
- [17] Gruber, P.M. (2004). Optimum quantization and its applications. *Adv. Math.* (to appear).
- [18] Kieffer, J.C. (1983). Uniqueness of locally optimal quantizer for log-concave density and convex error weighting functions, *IEEE Trans. Inform. Theory*, **29**, 42-47.
- [19] Lambertson, D. and Pagès, G. (1996). On the critical points of the 1-dimensional Competitive Learning Vector Quantization Algorithm. *Proceedings of the ESANN'96*, (ed., M. Verleysen), Editions D Facto, Bruxelles, 97-106.
- [20] Li, J. Chaddha, N. and Gray, R.M. (1999). Asymptotic performance of vector quantizers with a perceptual distortion measure. *IEEE Trans. Inform. Theory*, **45**, 1082-1091.
- [21] Linder, T., Zamir, R. and Zeger, K. (1999). High-resolution source coding for non-difference distortion measures: multidimensional companding. *IEEE Trans. Inform. Theory*, **45**, 548-561.
- [22] McClure, D.E. (1975). Nonlinear segmented function approximation and analysis of line patterns. *Quart. Appl. Math.*, **33**, 1-37.
- [23] Pagès, G. (1993). Voronoi tessellation, space quantization algorithm and numerical integration. *Proceedings of the ESANN'93*, (ed., M. Verleysen), Editions D Facto, Bruxelles, 221-228.
- [24] Pagès, G. (1997). A space quantization method for numerical integration. *J. Comput. Appl. Math.*, **89**, 1-38.
- [25] Pagès, G., Pham, H. and Printems, J. (2003). Optimal quantization methods and applications to numerical methods and applications in finance. Technical Report No. 813, LPMA, Univ. Paris 6, to appear in *Handbook of Numerical Methods in Finance* (ed., S. Rachev), Birkhäuser.
- [26] Trushkin, A.V. (1982). Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions, *IEEE Trans. Inform. Theory*, **28**, 187-198.
- [27] Zador, P.L. (1963). Development and evaluation of procedures for quantizing multivariate distributions. Ph.D. dissertation, Stanford Univ.
- [28] Zador, P.L. (1982). Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inform. Theory*, **28**, 139-149.