Available online at www.sciencedirect.com

ScienceDirect

Journal of Approximation Theory 198 (2015) 111–131

Journal of Approximation Theory

www.elsevier.com/locate/jat

Full length article

Greedy vector quantization

Harald Luschgy[a], Gilles Pagès[b,*]

[a] Universität Trier, FB IV-Mathematik, D-54286 Trier, Germany

[b] Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5, France

Received 19 August 2014; received in revised form 15 April 2015; accepted 22 May 2015
Available online 16 June 2015

Communicated by Gitta Kutyniok

Abstract

We investigate the greedy version of the L^p-optimal vector quantization problem for an $\mathbb{R}^d$-valued random vector $X \in L^p$. We show the existence of a sequence $(a_N)_{N\geq 1}$ such that a_N minimizes $a \mapsto \big\|\min_{1\leq i\leq N-1}|X-a_i| \wedge |X-a|\big\|_{L^p}$ (L^p-mean quantization error at level N induced by $(a_1,\dots,a_{N-1},a)$). We show that this sequence produces L^p-rate optimal N-tuples $a^{(N)} = (a_1,\dots,a_N)$ (*i.e.* the L^p-mean quantization error at level N induced by $a^{(N)}$ goes to 0 at rate $N^{-\frac{1}{d}}$). Greedy optimal sequences also satisfy, under natural additional assumptions, the distortion mismatch property: the N-tuples $a^{(N)}$ remain rate optimal with respect to the L^q-norms, $p \leq q < p+d$.

MSC: 60G15; 60G35; 41A25

Keywords: Optimal vector quantization; Greedy optimization; Distortion mismatch; Lloyd's method I; Competitive Learning Vector Quantization; k-means

* Corresponding author.
E-mail addresses: luschgy@uni-trier.de (H. Luschgy), gilles.pages@upmc.fr (G. Pagès).

http://dx.doi.org/10.1016/j.jat.2015.05.005
0021-9045/

1. Introduction and definition of greedy quantization sequences

Let $p \in (0, +\infty)$ and $L^p_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P}) = \{Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$, measurable, $\|Y\|_p = \big(\mathbb{E}|Y|^p\big)^{\frac{1}{p}} < +\infty\}$ where $|\,.\,|$ denotes a norm on $\mathbb{R}^d$. We consider $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ an L^p-integrable random vector. For every finite $\Gamma \subset \mathbb{R}^d$, we define the L^p-mean quantization error induced by Γ as the L^p-mean of the distance of the random vector X to the subset Γ (with respect to the norm $|\,.\,|$), namely

$$e_p(\Gamma, X) = \big\|d(X, \Gamma)\big\|_p$$

where $d(\xi, A) = \inf_{a\in A} |\xi - a|$, $\xi \in \mathbb{R}^d$, $A \subset \mathbb{R}^d$, denotes the distance of ξ to A. This quantity is always finite when $X \in L^p(\mathbb{P})$ since $e_p(\Gamma, X) \leq \|X\|_p + \min_{a\in\Gamma} |a| < +\infty$ owing to Minkowski's inequality when $p \geq 1$. When $p \in (0, 1)$, one has likewise $e_p(\Gamma, X)^p \leq \|X\|_p^p + \min_{a\in\Gamma} |a|^p < +\infty$. The usual L^p-optimal quantization problem *at level* $N \geq 1$ is to solve the following minimization problem

$$e_{p,N}(X) = \min_{\Gamma\subset\mathbb{R}^d, |\Gamma|\leq N} e_p(\Gamma, X) \tag{1.1}$$

where $|\Gamma|$ denotes the cardinality of the subset Γ, sometimes called *grid* in Numerical Probability or *codebook* in Signal processing. The use of "min" instead of "inf" is justified by the fact (see Proposition 4.12 in [16], p. 47 or [21]) that this infimum is always attained by an *optimal quantization* grid $\Gamma^{(N)}$ (of full size N if the support of the distribution $\mu = \mathbb{P}_X$ of X has at least N elements).

The above optimal vector quantization problem is clearly related to the approximation rate of an $\mathbb{R}^d$-valued random vectors $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ by random vectors taking at most N values ($N \in \mathbb{N}$). One shows (see e.g. Theorem 4.12 in [16] combined with comments, Section 3.3, p. 33) that, for very $p \in (0, +\infty)$,

$$\begin{aligned} e_{p,N}(X) &= \min\Big\{\|X - q(X)\|_p, q : \mathbb{R}^d \to \mathbb{R}^d,\ \text{Borel},\ |q(\mathbb{R}^d)| \leq N\Big\} \\ &= \min\Big\{\|X - Y\|_p,\ Y : \Omega \to \mathbb{R}^d, \text{measurable},\ |Y(\Omega)| \leq N\Big\}, \end{aligned}$$

both minima being attained by random vectors of the form

$$Y^{(N)} = \widehat{X}^{(N)} := \pi_{\Gamma^{(N)}}(X) \tag{1.2}$$

where $\pi_{\Gamma^{(N)}}$ denotes a *Borel projection on* $\Gamma^{(N)}$ *following the nearest neighbor rule* where $\Gamma^{(N)} \subset \mathbb{R}^d$ has size at most N.

This modulus is also related to the Wasserstein (pseudo-)distance $\mathcal{W}_p$, $p \in (0, 1]$ on the space of Borel probability measure on $\mathbb{R}^d$: let $\mathcal{P}_N$ be the set of distributions whose support has at most N elements. Let μ be a Borel distribution on $\mathbb{R}^d$ and let $\nu \in \mathcal{P}_N$ that we can associate to random vectors X and Y respectively; then for every p-Hölder function $f : \mathbb{R}^d \to \mathbb{R}$, with p-Hölder ratio $[f]_{p,Hol} < +\infty$ and every $\nu \in \mathcal{P}_N$,

$$\left|\int_{\mathbb{R}^d} f\, d\mu - \int_{\mathbb{R}^d} f\, d\nu\right| = \Big|\mathbb{E}\, f(X) - \mathbb{E}\, f(Y)\Big| \leq [f]_{p,Hol} \|X - Y\|_p. \tag{1.3}$$