Individual Convergence Rates in Empirical Vector Quantizer Design

András Antos, László Györfi, *Fellow, IEEE,* and András György, *Member, IEEE*

Abstract

We consider the rate of convergence of the expected distortion redundancy of empirically optimal vector quantizers. Earlier results show that the mean-squared distortion of an empirically optimal quantizer designed from n independent and identically distributed source samples converges uniformly to the optimum at a rate of $O(1/\sqrt{n})$, and that this rate is sharp in the minimax sense. We prove that for any fixed distribution supported on a given finite set the convergence rate is $O(1/n)$ (faster than the minimax lower bound), where the corresponding constant depends on the source distribution. For more general source distributions we provide conditions implying a little bit worse $O(\log n/n)$ rate of convergence. Although these conditions, in general, are hard to verify, we show that sources with continuous densities satisfying certain regularity properties (similar to the ones in [22] that were used to prove a central limit theorem for the code points of the empirically optimal quantizers) are included in the scope of this result. In particular, scalar distributions with strictly log-concave densities with bounded support (such as the truncated Gaussian distribution) satisfy these conditions.

Index Terms

convergence rates, fixed rate quantization, empirical design, individual convergence rate, log-concave densities

I. Introduction

The problem of empirical vector quantizer design is an important issue in data compression, since in many practical situations good source models are not available, but it is possible to collect source samples, called the training data, to gain information about the source statistics. Then the goal is to design a quantizer of a given rate, based on this data, whose average distortion on the source is as close to the distortion of the optimal quantizer (that is, one with minimum distortion) of the same rate as possible.

The usual, quite intuitive approach to this problem is empirical error minimization, which is based on the concept that if the training data describes the source statistics accurately, then a quantizer that performs well on the training samples should also have a good performance on the real source. Most existing design algorithms employ this principle, and search for an empirically optimal quantizer, i.e., a quantizer minimizing the empirical error on the training data, expecting that it will have near optimal performance when applied to the real source. (The reader is referred to Gersho and Gray [8] for a good summary of such algorithms.) Indeed, under general conditions on the source distribution Pollard [20], [21] showed that this method is consistent when the training data consists of n consecutive elements of a stationary and ergodic sequence drawn according to the source distribution: he proved that the mean-squared error (MSE) distortion $D(Q_n^*)$ of the empirically optimal quantizer Q_n^* (when applied to the real source) converges with probability one to the minimum MSE D^* achieved by an optimal quantizer.

Obviously, the above consistency result does not provide any information on how many training samples are needed to ensure that the distortion of the empirically optimal quantizer is close to the optimum. This question can be answered by analyzing the rate of convergence in $D(Q_n^*) \to D^*$, that is, by giving finite sample upper bounds for the distortion redundancy $D(Q_n^*) - D^*$. Linder *et al.* [16] showed that the expected distortion redundancy (with respect to the training data) can be bounded as $\mathbf{E}D(Q_n^*) - D^* \le c/\sqrt{n}$ for some appropriate constant c for all source distributions over a given bounded region. More precisely, in [16] only $O(\sqrt{\log n/n})$ rate was shown, supported with a discussion on how to improve the convergence rate to $O(1/\sqrt{n})$, but in the latter case the resulting constant was impractically large. A practically applicable constant can be obtained by combining the results of [16] with recent results of Linder [14]. (See also [15] for a summary.) This result has been extended in many ways. An extension to vector quantizers designed for noisy channels or for "noisy" sources was given by Linder *et al.* [17], an extension to unbounded sources was provided by Merhav and Ziv [19], while the case of dependent (mixing) training data was examined by Zeevi [24].

Bartlett *et al.* [3] showed that the $O(1/\sqrt{n})$ bound on the expected distortion redundancy is tight in the minimax sense. They proved that (for at least three quantization levels) for any empirical quantizer design method, that is, when the resulting quantizer Q_n is an arbitrary function of the training data, and for any n large enough, there is a distribution in the class of distributions over a bounded region such that $\mathbf{E}D(Q_n) - D^* > c/\sqrt{n}$. These "bad" distributions are quite simple, e.g., the

Manuscript received October 2, 2003; revised June 15, 2005. This research was supported in part by the NATO Science Fellowship, a research grant from the Research Group for Informatics and Electronics of the Hungarian Academy of Sciences, and NKFP-2/0017/2002 project Data Riddle.

A. Antos is with the Informatics Laboratory, Computer and Automation Research Institute of the Hungarian Academy of Sciences, Lágymányosi u. 11, 1111 Budapest, Hungary (email: antos@szit.bme.hu).

L. Györfi is with the Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Magyar tudósok körútja 2, 1117 Budapest, Hungary (email: gyorfi@szit.bme.hu).

A. György is with the Informatics Laboratory, Computer and Automation Research Institute of the Hungarian Academy of Sciences, Lágymányosi u. 11, 1111 Budapest, Hungary (email: gya@szit.bme.hu).

During part of this research A. Antos and A. György were also with the Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6.

distributions used in the proof are concentrated on finitely many atoms. However, the minimax lower bound gives information about the maximum distortion within the class, but not about the behavior of the distortion for a single fixed source distribution, as the sample size n increases. Moreover, the chosen "bad" distributions in the proof of the above result are different for all n, allowing the possibility that the upper bound may be improved in an individual sense, that is, a faster rate of convergence may be achievable, where the constant in the bound also depends on the (fixed) source distribution. Finding the best such individual rate (or weak rate) for the class of sources over a bounded region was labeled in [3] as an interesting and challenging problem.

There are some results suggesting that the convergence rate can be improved to $O(1/n)$: In the special case of a one-level quantizer the code point of the empirically (MSE) optimal quantizer is simply the average of the training samples, and it is easy to see that in this case $\mathbf{E}D(Q_n^*) - D^* = c/n$, where c is the variance of the source. Also, based on another result of Pollard [22] showing that for sources with continuous densities satisfying certain regularity properties the suitably scaled difference of the code points of the optimal and the empirically optimal quantizers has asymptotically multidimensional normal distribution, Chou [4] pointed out that for such sources the distortion redundancy decreases as $O(1/n)$ in probability.

In this paper we provide improved upper bounds for the convergence rate of the expected distortion redundancy individually for source distributions within the class of distributions over a bounded region. In Theorem 1 we show that $\mathbf{E}D(Q_n^*) - D^* \le c(\mu, N)/n$ for all source distributions μ concentrated on a finite set, where the constant $c(\mu, N)$ depends on the actual source distribution and the number of quantization levels N. The convergence rate for general source distributions is considered in Theorem 2. It is shown that for source distributions over a bounded region satisfying a certain regularity condition, the expected distortion redundancy can be upper bounded by $c(\mu, N)\log n/n$, where the actual value of the constant again depends on the actual source distribution. In Corollary 1 we prove that source distributions with bounded support satisfying essentially the same conditions as in [22] satisfy the requirements in Theorem 2, and in Corollary 2 we show that the conditions of Corollary 1 hold for scalar sources having strictly log-concave densities with bounded support (such as the truncated Gaussian distribution), and for the uniform distribution, implying $O(\log n/n)$ expected distortion redundancy. To give more insight to the problem we also illustrate that the similar conditions of Hartigan [10] (that are valid only in one-dimension) also imply the condition in Theorem 2.

Comparing our results with [3], it follows that the problem of empirical quantizer design is an interesting example of the unusual event when the orders of the minimax lower bound and the individual upper bound are different.

II. Empirical vector quantizer design

A d-dimensional N-level *vector quantizer* is a measurable mapping $Q : \mathbb{R}^d \to \mathcal{C}$, where the *codebook* $\mathcal{C} = \{y_1, \ldots, y_N\} \subset \mathbb{R}^d$ is a collection of N distinct d-vectors, called the *code points*. The quantizer is completely characterized by its codebook and the sets

$$S_i = \{x \in \mathbb{R}^d : Q(x) = y_i\} \qquad i = 1, \ldots, N,$$

called the *cells* or *partition cells* (as they form a partition of $\mathbb{R}^d$) via the rule

$$Q(x) = y_i \qquad \text{if } x \in S_i.$$

The set $\{S_1, \ldots, S_N\}$ is called the partition of Q. Throughout this paper, unless explicitly stated otherwise, all quantizers are assumed to be d-dimensional with N code points.

The source to be quantized is a random vector $X \in \mathbb{R}^d$ with distribution μ. We assume $\mathbf{E}\{\|X\|^2\} < \infty$, where $\|\cdot\|$ denotes the Euclidean norm. The performance of the quantizer Q in quantizing X is measured by the *average (mean-squared) distortion*

$$D(Q) = \mathbf{E}\{\|X - Q(X)\|^2\}.$$

A quantizer Q^* achieving the minimum distortion D^* is called *optimal*. Thus, in this case

$$D^* = D(Q^*) \le D(Q) \qquad \text{for all } Q \in \mathcal{Q}_N$$

where $\mathcal{Q}_N$ denotes the set of all N-level quantizers. It is well known (see, e.g., [13],[8]) that any optimal quantizer satisfies the centroid and nearest neighbor conditions, also known as the Lloyd-Max conditions. The quantizer Q satisfies the centroid condition if each code point is chosen to minimize the distortion over its associated cell, that is

$$\mathbf{E}\{\|X - y_i\|^2 | X \in S_i\} = \min_y \mathbf{E}\{\|X - y\| | X \in S_i\}, \tag{1}$$

and so

$$y_i = \mathbf{E}\{X | X \in S_i\}$$

for all $i = 1, \ldots, N$. A partition $\{S_1, \ldots, S_N\}$ is optimal if the quantizer Q with cells $S_1, \ldots, S_N$ satisfying the centroid condition is optimal. Q is called a *nearest neighbor* quantizer, if it satisfies

$$\|x - Q(x)\| = \min_i \|x - y_i\| \qquad \text{for all } x \in \mathbb{R}^d. \tag{2}$$

Note that
(i) a nearest neighbor quantizer is determined by its codebook $\{y_1, \ldots, y_N\}$ with ties arbitrarily broken;

(ii) for any non-nearest neighbor quantizer Q', a nearest neighbor quantizer Q with the same codebook has at most the same distortion as Q', that is, $D(Q) \leq D(Q')$, regardless of the distribution of X.

Thus any optimal quantizer can be assumed to be nearest neighbor, and so finding an optimal quantizer is equivalent to finding its codebook. Using this observation, Pollard [21] proved that if $\mathbf{E}\{\|X\|^2\} < \infty$, then there exists an optimal quantizer (which may not be unique).

In many situations the distribution μ is unknown, and the only available information about it is given in the form of *training data*, that is, a sequence $X_1^n = X_1, \ldots, X_n$ of n independent and identically distributed (i.i.d.) copies of X. The sequence X_1^n is also assumed to be independent of X. X_1^n is used to construct an *empirically designed* quantizer $Q_n(\cdot) = Q_n(\cdot, X_1, \ldots, X_n)$, which is a random function depending on the training data. The goal is to produce such quantizers with performance near D^*. The performance of Q_n in quantizing X is measured by the *test distortion*

$$D(Q_n) = \mathbf{E}\{\|X - Q_n(X)\|^2 | X_1^n\} = \int_{\mathbb{R}^d} \|x - Q_n(x)\|^2 \, \mu(dx).$$

Note that $D(Q_n)$ is a random variable.

The *empirical distortion* (or training distortion) of any Q is given by its mean-squared error in quantizing the training data

$$D_n(Q) = \frac{1}{n} \sum_{i=1}^{n} \|X_i - Q(X_i)\|^2.$$

Note that although Q is a deterministic mapping, the empirical distortion $D_n(Q)$ is also a random variable depending on the training data X_1^n.

Assume that Q_n^* minimizes the empirical distortion, that is,

$$D_n(Q_n^*) = \min_{Q \in \mathcal{Q}_N} D_n(Q).$$

Then Q_n^* (which is a specific empirically designed quantizer) is called an *empirically optimal vector quantizer*. Q_n^* is an optimal quantizer for the empirical distribution μ_n of the training data given as

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \in A\}}$$

for every Borel set $A \subset \mathbb{R}^d$, where I_E denotes the indicator function of the event E. Note that Q_n^* always exists (although it is not necessarily unique), and we can assume that it is a nearest neighbor quantizer. Using Q_n^* as an approximation of the optimal Q^* is consistent in the sense that its test distortion converges to the optimal distortion, that is,

$$\lim_{n \to \infty} D(Q_n^*) = D^*$$

almost surely for any $N \geq 1$ if $E\{\|X\|^2\} < \infty$, see [20],[21].

III. Rate of convergence

To determine the number of training samples necessary to achieve a preassigned level of distortion, the finite sample behavior of the *expected distortion redundancy*

$$\mathbf{E}D(Q_n^*) - D^*$$

has to be analyzed. To do this we assume the peak power constraint

$$\mathbf{P}\{\|X\| \leq B\} = 1 \tag{3}$$

and this assumption will be in effect throughout the paper. In other words, the distribution μ of the source X is an element of $\mathcal{P}(B)$, the family of distributions supported on the sphere

$$S(B) = \{x \in \mathbb{R}^d : \|x\| \leq B\}.$$

An important consequence of (3) is that it is sufficient for our purpose to consider only quantizers with code points in the sphere $S(B)$, since otherwise projecting a code point that is not in $S(B)$ to the surface of $S(B)$ clearly reduces the distortion.

It is of interest how fast $\mathbf{E}D(Q_n^*)$ converges to D^*. To our knowledge the best accessible upper bound can be obtained by combining results of [16] with recent developments of [14], implying:

$$0 \leq \sup_{\mu \in \mathcal{P}(B)} (\mathbf{E}D(Q_n^*) - D^*) \leq c_u B^2 \sqrt{\frac{Nd}{n}} \tag{4}$$

for all $n \geq 1$, where $c_u = 192$. A natural question is whether there exists a method, perhaps different from empirical distortion minimization, which provides an empirically designed quantizer with substantially smaller test distortion. In case of $N = 1$, it is easy to see that

$$\mathbf{E}D(Q_n^*) - D^* = \frac{\mathbf{Var}(X)}{n}.$$

Thus, the convergence rate is $O(1/n)$, substantially faster than the $O(1/\sqrt{n})$ rate above. However, for $N \geq 3$ the lower bound in [3] shows that the $O(1/\sqrt{n})$ convergence rate above cannot be improved in the minimax sense: There it is proved that if $N \geq 3$, then for any empirically designed quantizer Q_n trained on $n \geq n_0 \approx 10^4 N$ samples, we have

$$\sup_{\mu \in \mathcal{P}(B)} (\mathbf{E}D(Q_n) - D^*) \geq c_l B^2 \sqrt{\frac{N^{1-4/d}}{n}} \tag{5}$$

where $c_l \approx 2.67 \cdot 10^{-11}$. This result has recently been improved in [1], extending the results to the case $N = 2$, and improving the constant c_l to approximately $1.68 \cdot 10^{-4}$ and the constant n_0 to $8N$.

The results (4) and (5) imply that there exist positive constants $c_l(N,B,d)$ and $c_u(N,B,d)$ depending on N, B, and d such that

$$\frac{c_l(N,B,d)}{\sqrt{n}} \leq \inf_{Q_n} \sup_{\mu \in \mathcal{P}(B)} (\mathbf{E}D(Q_n) - D^*) \leq \frac{c_u(N,B,d)}{\sqrt{n}}$$

where the infimum is taken over all empirically designed quantizers Q_n (recall that $Q_n = Q_n(\cdot, X_1, \ldots, X_n)$ is a function $Q_n : \mathbb{R}^d \times \mathbb{R}^{dn} \to \mathbb{R}^d$). That is, the *minimax bounds* on the rate of convergence are $\Theta(1/\sqrt{n})$. The minimax lower bound expresses the minimum achievable worst-case error which is achievable for a given sample size n for the distribution class $\mathcal{P}(B)$, by describing the behavior of any quantizer design method for a source distribution which is the least suitable for the given n and the given design method.

The "bad" distributions achieving the supremum of the expected distortion redundancy in (5) may be different for each n. Indeed, in the construction of [3] (or [1]), although the bad distributions are concentrated on the same finitely many atoms for each n, the exact probability mass function of the bad distribution depends on n. Thus, the bound does not tell the behavior of the distortion redundancy for a *single fixed source distribution* μ. For example, it does not exclude the possibility that for some sequence of empirical quantizers $\{Q_n\}$, $\mathbf{E}D(Q_n) - D^*$ converges to 0 at $O(1/n)$ rate for *every fixed* μ, that is, it is possible to get faster *individual upper bounds* of the form

$$\mathbf{E}D(Q_n) - D^* \leq \frac{c(\mu, N)}{n} \tag{6}$$

for each $\mu \in \mathcal{P}(B)$ and $n \geq 1$, where the constant $c(\mu, N)$ *depends on the source distribution.* This type of upper bounds is the main purpose of this paper. Next we show (6) for discrete distribution μ, and with an additional factor $\log n$ for general distribution μ satisfying some regularity condition. The proofs are deferred to the next section.

Our first result shows that the expected distortion redundancy converges to 0 at a rate $O(1/n)$ for any fixed source distribution concentrated on a finite set of points in $\mathbb{R}^d$.

Theorem 1: Assume that the source distribution μ is concentrated on finitely many atoms. Then

$$\mathbf{E}D(Q_n^*) - D^* \leq \frac{c(\mu, N)}{n},$$

where the constant $c(\mu, N)$ depends on μ and N.

The main idea in the proof is that with high probability the empirical and the real source distributions are so close that the corresponding optimal quantizer partitions coincide, and in this case we only need to find the centroid of these partitions. Proving similar rate of convergence results for more general source distributions is significantly harder, since the partitions of optimal quantizers for "close" distributions are different in general.

Next we give conditions on the source distribution μ which ensure that the expected distortion redundancy converges to 0 at rate $O(\log n/n)$. For a nearest neighbor quantizer Q let

$$\Delta_Q(x) = \|x - Q(x)\|^2 - \|x - Q_Q^*(x)\|^2$$

for all $x \in S(B)$, where Q_Q^* is the "closest" optimal nearest neighbor quantizer to Q in the sense that it achieves the minimum

$$\min_{\widehat{Q}: D(\widehat{Q}) = D^*} \mathbf{Var}\{\|X - Q(X)\|^2 - \|X - \widehat{Q}(X)\|^2\}. \tag{7}$$

Note that the minimum can always be achieved as can be seen by a continuity-compactness type argument. This minimization is introduced to avoid problems that occur if the optimal quantizer Q^* is not unique.

Theorem 2: Assume that $\mathbf{P}\{\|X\| \leq B\} = 1$, and let Q_n^* be an empirically optimal quantizer. Assume furthermore that

$$A = \inf_{Q: D(Q) > D^*} \frac{\mathbf{E}\{\Delta_Q(X)\}}{\mathbf{Var}\{\Delta_Q(X)\}} > 0 \tag{8}$$

where the infimum is taken over all non-optimal nearest neighbor quantizers having all their code points in the sphere $S(B)$. Then

$$\mathbf{E}D(Q_n^*) - D^* \leq \frac{c_1 \log n}{n} + \frac{c_2}{n} \tag{9}$$

with constants

$$c_1 = 4dN \max\left\{\frac{e-2}{A}, 4B^2\right\}$$

and

$$c_2 = 4N \max\left\{\frac{e-2}{A}, 4B^2\right\} \log\left(V\left(\frac{3eB}{N\sqrt{d}\max\left\{\frac{e-2}{A}, 4B^2\right\}}\right)^d\right)$$

where log denotes the natural logarithm and V denotes the volume of the sphere $S(B)$.

The proof of the theorem is based on a proof of [17] combined with a technique developed by Barron [2] and Lee *et al.* [12] (see also, e.g., [9, Chapter 16]). The essence of the latter, which is an interesting result itself, is formulated in Lemma 1 in the next section.

Remark 1: It is expected that at the expense of a more complicated analysis, the $\log n$ term can be removed from the upper bound (9), giving the desired $O(1/n)$ rate.

Remark 2: The constants in the above theorem can slightly be improved to

$$c_1 = \frac{16dNB^2}{G^{-1}(4AB^2)}$$

and

$$c_2 = \frac{16NB^2}{G^{-1}(4AB^2)} \log\left(V\left(\frac{3eG^{-1}(4AB^2)}{4BN\sqrt{d}}\right)^d\right)$$

where G^{-1} is the inverse of the function G given as

$$G(c) = \frac{e^c - c - 1}{c} \qquad \text{for } c > 0. \tag{10}$$

This is shown at the end of the proof of the theorem.

Condition (8) is hard to check for general source distributions, therefore the scope of Theorem 2 is not clear. The next corollary shows that the theorem is valid for sources with continuous densities satisfying certain regularity properties.

Let $\{y_1^*, \ldots, y_N^*\}$ be the code points of an optimal quantizer for μ and let $\{S_1^*, \ldots, S_N^*\}$ denote the corresponding nearest neighbor cells. It is known (see [22, Lemma C and Theorem]) that if μ has a continuous density f with bounded support, then the distortion $D(y_1, \ldots, y_N) = D(Q)$ of the nearest neighbor quantizer Q with code points $\{y_1, \ldots, y_N\}$ is a continuous function of the vector $(y_1, \ldots, y_N)$ which has a second derivative block matrix $\Gamma(y_1^*, \ldots, y_N^*) = [\Gamma_{ij}(y_1^*, \ldots, y_N^*)]$ at $(y_1^*, \ldots, y_N^*)$ made up of $d \times d$ blocks

$$\Gamma_{ij}(y_1^*, \ldots, y_N^*) = \begin{cases} 2\mu(S_i^*)I_d - 2\sum_{l \neq i} \frac{\int_{F_{il}} f(x)(x-y_i^*)(x-y_i^*)^T \, d\lambda_{d-1}(x)}{\|y_i^* - y_l^*\|} & \text{for } j = i, \\ 2\frac{\int_{F_{ij}} f(x)(x-y_i^*)(x-y_j^*)^T \, d\lambda_{d-1}(x)}{\|y_i^* - y_j^*\|} & \text{for } j \neq i, \end{cases} \quad (1 \le i, j \le N)$$

where F_{ij} is the (possibly empty) common face of S_i^* and S_j^* (it is a convex set in a $(d-1)$-dimensional hyperplane), I_d is the $d \times d$ identity matrix, and λ_{d-1} is the $(d-1)$-dimensional Lebesgue measure. [1] It is clear that since $\{y_1^*, \ldots, y_N^*\}$ is an optimal codebook, the matrix $\Gamma(y_1^*, \ldots, y_N^*)$ is positive semidefinite. The next corollary shows that if $\Gamma(y_1^*, \ldots, y_N^*)$ is also positive definite, then the desired $O(\log n/n)$ convergence rate can be established.

Corollary 1: Assume that the random variable X has a continuous density supported in $S(B)$, and the matrix $\Gamma(y_1^*, \ldots, y_N^*)$ is positive definite for all optimal codebooks. Then

$$\mathbf{E}D(Q_n^*) - D^* = O\left(\frac{\log n}{n}\right).$$

The conditions of Corollary 1 on the distribution are essentially the same as those of Pollard [22] (and of Chou [4]). The conditions in [22] are weaker in the sense that there the usual assumption of X having a bounded support is replaced by a tail condition. This extension might be possible for Corollary 1 and Theorem 2 at the expense of some complication in the proof. On the other hand, while Pollard assumes the uniqueness of an optimal quantizer, we allow multiple optima. However, if the set of optimal quantizers (each represented by the N-vector of its codebook) has an accumulation point, then usually Γ is not positive definite for all optimal codebooks. Thus, Corollary 1 is not applicable, for example, for a multidimensional truncated Gaussian distribution (although we suspect that the results can be sharpened to include such cases, as well). Nevertheless, there are cases when the optimal quantizer is not unique and the set of optimal codebooks does not have an accumulation point. For example, for scalar sources with symmetric, not log-concave densities with a large spike in the middle usually two asymmetric optimal two-level quantizers exist (if the density is log-concave, then the optimal quantizer is unique [7],[23]).

[1]Note that Pollard [22] made a slight error in the derivation of Γ_{ij} for $i \neq j$, and arrived to a formula with an incorrect sign.

Note that Corollary 1 implies Chou's result for bounded distributions with the additional $\log n$ factor. However, the other direction requires some kind of uniform integrability of the random variables $\{n(D(Q_n^*) - D^*)\}_{n\geq 1}$, which can be arbitrary large as n goes to infinity, even for bounded source distributions.

Although it is not easy to determine in general whether or not the matrix $\Gamma(y_1^*, \ldots, y_N^*)$ is positive definite, sufficient conditions can be obtained easily in the scalar case. For example, it is easy to show, that $\Gamma(y_1^*, \ldots, y_N^*)$ is positive definite for the uniform distribution, where the unique optimal quantizer is the N-level uniform quantizer. More general sufficient conditions can be given based on a result of Fleischer [7], who, while proving the uniqueness of an optimal quantizer, also showed that if the source has a density f for which the derivative $d^2 \log f(x)/dx^2$ is negative over its total support, then the matrix $\Gamma(y_1^*, \ldots, y_N^*)$ is positive definite for all optimal codebooks, and hence for the unique optimal codebook. A slight modification of Fleischer's original proof allows us to replace the condition on the second derivative by the assumption that $\log f(x)$ is a strictly concave function (then f is called a strictly log-concave function). Thus we obtain the following result.

Corollary 2: Assume that the scalar random variable X has a strictly log-concave density f supported in the interval $[-B, B]$ (that is, $\log f(x)$ is strictly concave over its support), or X is uniformly distributed in $[-B, B]$. Then

$$\mathbf{E}D(Q_n^*) - D^* = O\left(\frac{\log n}{n}\right).$$

We note here that the result of Fleischer proving that the matrix $\Gamma(y_1^*, \ldots, y_N^*)$ is positive definite proves that scalar sources with strictly log-concave densities and sufficiently light tails (such as one with Gaussian distribution) satisfy the conditions of Pollard's central limit theorem [22] (this fact escaped Pollard's attention), which also implies that for such sources the distortion redundancy is $O(1/n)$ in probability by [4].

While Corollary 1 uses the nearest neighbor condition to capture optimality of quantizers, another approach is to use the centroid condition instead. This approach was used by Hartigan [10], an ancestor of [22] for one-dimension, who applied differentiation with respect to the quantization thresholds instead of differentiation with respect to the code points. This method is illustrated in the next example for the scalar case when $N = 2$.

Example 1: For $d = 1$ and $N = 2$, define the split function

$$D(t) = \mathbf{Var}\{X|X < t\}\mathbf{P}\{X < t\} + \mathbf{Var}\{X|X \geq t\}\mathbf{P}\{X \geq t\},$$

that is, the minimal distortion corresponding to the partition $\{(-\infty, t), [t, \infty)\}$, and let $t^* = (y_1^* + y_2^*)/2$ denote the cell boundary (or threshold) of an optimal quantizer with codebook $\{y_1^*, y_2^*\}$. Then if the scalar random variable X has a continuous density in the interval $[-B, B]$, and $\frac{d^2 D}{dt^2}(t^*)$ is positive for any optimal threshold t^*, then

$$\mathbf{E}D(Q_n^*) - D^* = O\left(\frac{\log n}{n}\right).$$

To see this, rewrite $D(t)$ as

$$D(t) = \mathbf{E}\{X^2\} - \mathbf{E}^2\{X|X < t\}\mathbf{P}\{X < t\} - \mathbf{E}^2\{X|X > t\}\mathbf{P}\{X \geq t\}.$$

Then

$$\frac{d^2 D}{dt^2}(t^*) = f(t^*)(y_2^* - y_1^*)\left(2 - \frac{f(t^*)(y_2^* - y_1^*)}{2\mathbf{P}\{X < t^*\}\mathbf{P}\{X \geq t^*\}}\right),$$

and the assumption $\frac{d^2 D}{dt^2}(t^*) > 0$ implies that

$$\begin{array}{rcl} 2\Gamma_{1,1}(y_1^*, y_2^*) & = & 4\mathbf{P}\{X < t^*\} - f(t^*)(y_2^* - y_1^*) \\ & \geq & 4\mathbf{P}\{X < t^*\}\mathbf{P}\{X \geq t^*\} - f(t^*)(y_2^* - y_1^*) \\ & = & \det\Gamma(y_1^*, y_2^*) \\ & > & 0, \end{array}$$

and thus $\Gamma(y_1^*, y_2^*)$ is positive definite for any optimal codebook, thus the result follows by Corollary 1. □

Finally, a comparison of the results in this section with [3] shows that the problem of empirical quantizer design is an interesting example of the unusual event when the orders of the minimax lower bound and the individual upper bound are different (the latter being smaller).

IV. Proofs

Proof of Theorem 1: Let $S = \{x_1, \ldots, x_m\} \subset \mathbb{R}^d$ denote the support of μ with corresponding probability mass $p_i = \mathbf{P}\{X = x_i\} > 0$, $i = 1, \ldots, m$. Assume $m \geq N + 1$, since otherwise

$$\mathbf{E}D(Q_n^*) \leq 4B^2 \sum_{i=1}^{m} (1 - p_i)^n$$

by the power constraint (3).

Let Π_μ be the set of optimal partitions for μ. Let $P_n^* = \{S_{n,1}^*, \ldots, S_{n,N}^*\}$ denote the partition of an empirically optimal quantizer Q_n^*. Clearly $P_n^* \in \Pi_{\mu_n}$. By the centroid rule (1) the code point $y_{n,i}^*$ associated with the cell $S_{n,i}^*$ is the average of samples falling into this cell. This can be computed whenever $\mu_n(S_{n,i}^*)$, the ratio of samples falling into the cell is positive. Without loss of generality we can assume that otherwise $S_{n,i}^* = \emptyset$ in which case the definition of $y_{n,i}^*$ is immaterial, since keeping only the code points corresponding to the nonempty cells can only increase the test distortion of Q_n^*, and hence increase the expected distortion redundancy.

Decompose the expected distortion redundancy of Q_n^* in the following way:

$$\begin{aligned}
&\mathbf{E}D(Q_n^*) - D^* \\
&\quad = \mathbf{E}\{I_{\{P_n^* \in \Pi_\mu\}}(D(Q_n^*) - D^*)\} \\
&\qquad + \mathbf{E}\{I_{\{P_n^* \notin \Pi_\mu\}}(D(Q_n^*) - D^*)\}.
\end{aligned} \tag{11}$$

Let $\bar{y}_{n,i}^* = \mathbf{E}\{X | X \in S_{n,i}^*\}$ for all i; now if $P_n^* \in \Pi_\mu$ then the quantizer with partition P_n^* and codebook $\{\bar{y}_{n,1}^*, \ldots, \bar{y}_{n,N}^*\}$ is optimal for μ, and so for the first term of (11) we have

$$\begin{aligned}
&\mathbf{E}\{I_{\{P_n^* \in \Pi_\mu\}}(D(Q_n^*) - D^*)\} \\
&\quad = \mathbf{E}\left\{ I_{\{P_n^* \in \Pi_\mu\}} \sum_{i=1}^N (y_{n,i}^* - \bar{y}_{n,i}^*)^2 \mu(S_{n,i}^*) \right\} \\
&\quad = \mathbf{E}\left\{ I_{\{P_n^* \in \Pi_\mu\}} \sum_{i=1}^N I_{\{\mu_n(S_{n,i}^*)>0\}} (y_{n,i}^* - \bar{y}_{n,i}^*)^2 \mu(S_{n,i}^*) \right\}.
\end{aligned} \tag{12}$$

For any partition $P = \{S_1, \ldots, S_N\} \in \Pi_\mu$, let $y_{n,i}$ be the average of samples falling into S_i (it can be defined arbitrarily if $\mu_n(S_i) = 0$) and let $\bar{y}_i = \mathbf{E}\{X | X \in S_i\}$. Then (12) can be continued as

$$\begin{aligned}
&\mathbf{E}\left\{ I_{\{P_n^* \in \Pi_\mu\}} \sum_{i=1}^N I_{\{\mu_n(S_{n,i}^*)>0\}} (y_{n,i}^* - \bar{y}_{n,i}^*)^2 \mu(S_{n,i}^*) \right\} \\
&\quad \le \mathbf{E}\left\{ \sum_{P: P \in \Pi_\mu} \sum_{i=1}^N I_{\{\mu_n(S_i)>0\}} (y_{n,i} - \bar{y}_i)^2 \mu(S_i) \right\} \\
&\quad = \sum_{P: P \in \Pi_\mu} \sum_{i=1}^N \mathbf{E}\left\{ I_{\{\mu_n(S_i)>0\}} \mathbf{E}\{(y_{n,i} - \bar{y}_i)^2 | I_{\{X_1 \in S_i\}}, \ldots, I_{\{X_n \in S_i\}}\} \right\} \mu(S_i) \\
&\quad = \sum_{P: P \in \Pi_\mu} \sum_{i=1}^N \mathbf{E}\left\{ I_{\{\mu_n(S_i)>0\}} \frac{\mathbf{Var}\{X | X \in S_i\}}{n \mu_n(S_i)} \right\} \mu(S_i) \\
&\quad = \sum_{P: P \in \Pi_\mu} \sum_{i=1}^N \mathbf{E}\left\{ \frac{I_{\{B_i>0\}}}{B_i} \right\} \mathbf{Var}\{X | X \in S_i\} \mu(S_i) \\
&\quad \le \sum_{P: P \in \Pi_\mu} \sum_{i=1}^N \frac{2\,\mathbf{Var}\{X | X \in S_i\}}{n+1} \qquad (13) \\
&\quad \le \frac{2B^2 |\Pi_\mu| N}{n+1} \qquad (14)
\end{aligned}$$

where B_i has a binomial distribution with parameters n and $\mu(S_i)$, and (13) holds by Devroye *et al.* [6, Lemma A.2].

We also need to give a bound on the second term of (11). Pollard [21] proved that the set of optimal partitions is continuous with respect to the L_2 Wasserstein distance of the distributions (defined by $\rho_W(\mu, \mu') = \inf_{(X,Y)} (\mathbf{E}\|X - Y\|^2)^{1/2}$, where the infimum is taken over all joint distributions of (X, Y) such that X and Y have distributions μ and μ', respectively), where a sequence of sets Π_k of partitions converges to Π if the corresponding characteristic functions converge, that is, $I_{\{P \in \Pi_k\}} \to I_{\{P \in \Pi\}}$ as $k \to \infty$ for each partition P of S. Moreover, it follows from [18] that the L_2 metric

$$\rho(\mu, \mu') = \left(\sum_{i=1}^m (\mu(x_i) - \mu'(x_i))^2 \right)^{1/2}$$

for the family of distributions concentrated on S is as strong as the Wasserstein distance, that is, $\rho(\mu_n, \mu) \to 0$ if and only if $\rho_W(\mu_n, \mu) \to 0$. Therefore, if $\rho(\mu_n, \mu) \to 0$ for the sequence $\{\mu_n\}$ of the empirical distributions, then $\Pi_{\mu_n} \to \Pi_\mu$. Since the number of possible partitions of S is finite, this implies that there is a $\delta_\mu > 0$ such that

$$\Pi_{\mu_n} = \Pi_\mu \qquad \text{for } \rho(\mu_n, \mu) < \delta_\mu.$$

Since $P_n^* \in \Pi_{\mu_n}$, this and the power constraint (3) imply

$$\begin{aligned}\mathbf{E}\{I_{\{P_n^*\notin\Pi_\mu\}}(D(Q_n^*)-D^*)\} &\leq 4B^2\mathbf{P}\{P_n^*\notin\Pi_\mu\}\\ &\leq 4B^2\mathbf{P}\{\rho(\mu_n,\mu)\geq\delta_\mu\}.\end{aligned}$$

Applying Markov's inequality we obtain

$$\begin{aligned}\mathbf{P}\{\rho(\mu_n,\mu)\geq\delta_\mu\} &= \mathbf{P}\left\{\sum_{i=1}^m(\mu_n(x_i)-p_i)^2\geq\delta_\mu^2\right\}\\ &\leq \frac{\mathbf{E}\{\sum_{i=1}^m(\mu_n(x_i)-p_i)^2\}}{\delta_\mu^2}\\ &= \frac{\sum_{i=1}^m p_i(1-p_i)}{n\delta_\mu^2}\\ &\leq \frac{1}{n\delta_\mu^2}.\end{aligned}$$

Thus from (11), (12), and (14) we obtain

$$\mathbf{E}D(Q_n^*)-D^*\leq 2B^2\left(\frac{|\Pi_\mu|N}{n+1}+\frac{2}{n\delta_\mu^2}\right)\leq\frac{c(\mu,N)}{n},$$

where

$$c(\mu,N)=2B^2\left(|\Pi_\mu|N+2/\delta_\mu^2\right).$$

(Note that $|\Pi_\mu|$ is bounded above by some function of N and m.) □

The proof of Theorem 2 is based on the following lemma:

Lemma 1: Let X_{ij}, $i=1,\ldots,n$, $j=1,\ldots N$ be random variables such that for each fixed j, $X_{1j},\ldots,X_{nj}$ are independent and identically distributed such that for each $s_0\geq s>0$

$$\mathbf{E}\{e^{sX_{ij}}\}\leq e^{s^2\sigma_j^2}.$$

For $\delta_j>0$, put

$$\alpha=\min_{j\leq N}\frac{\delta_j}{\sigma_j^2}.$$

Then

$$\mathbf{E}\left\{\max_{j\leq N}\left(\frac{1}{n}\sum_{i=1}^n X_{ij}-\delta_j\right)\right\}\leq\frac{\log N}{\min\{\alpha,s_0\}n}. \tag{15}$$

If

$$\mathbf{E}\{X_{ij}\}=0$$

and

$$|X_{ij}|\leq K,$$

then, for any $L>0$

$$\mathbf{E}\left\{\max_{j\leq N}\left(\frac{1}{n}\sum_{i=1}^n X_{ij}-\delta_j\right)\right\}\leq\frac{K\log N}{\min\left\{K\alpha^*\frac{L^2}{e^L-1-L},L\right\}n}, \tag{16}$$

where

$$\alpha^*=\min_{j\leq N}\frac{\delta_j}{\mathbf{Var}(X_{ij})}.$$

Proof: For the notation

$$Y_j=\frac{1}{n}\sum_{i=1}^n X_{ij}-\delta_j$$

we have that for any $s_0\geq s>0$

$$\begin{aligned}\mathbf{E}\{e^{snY_j}\} &= \mathbf{E}\{e^{sn\left(\frac{1}{n}\sum_{i=1}^n X_{ij}-\delta_j\right)}\}\\ &= e^{-sn\delta_j}\left(\mathbf{E}\{e^{sX_{1j}}\}\right)^n\\ &\leq e^{-sn\delta_j}e^{ns^2\sigma_j^2}\\ &\leq e^{-sn\alpha\sigma_j^2+s^2n\sigma_j^2}.\end{aligned}$$

Thus

$$\begin{aligned} e^{sn\mathbf{E}\{\max_{j\le N} Y_j\}} &\le \mathbf{E}\{e^{sn\max_{j\le N} Y_j}\} \\ &= \mathbf{E}\{\max_{j\le N} e^{snY_j}\} \\ &\le \sum_{j\le N} \mathbf{E}\{e^{snY_j}\} \\ &\le \sum_{j\le N} e^{-sn\sigma_j^2(\alpha-s)}. \end{aligned}$$

For $s=\min\{\alpha, s_0\}$ it implies that

$$\mathbf{E}\{\max_{j\le N} Y_j\} \le \frac{1}{sn}\log\left(\sum_{j\le N} e^{-sn\sigma_j^2(\alpha-s)}\right) \le \frac{\log N}{\min\{\alpha, s_0\}n}.$$

In order to prove the second half of the lemma, notice that, for any $L>0$ and $|x|\le L$ we have the inequality

$$\begin{aligned} e^x &= 1+x+x^2\sum_{i=2}^{\infty}\frac{x^{i-2}}{i!} \\ &\le 1+x+x^2\sum_{i=2}^{\infty}\frac{L^{i-2}}{i!} \\ &= 1+x+x^2\frac{e^L-1-L}{L^2}, \end{aligned}$$

therefore $0<s\le s_0=L/K$ implies that $s|X_{ij}|\le L$, so

$$e^{sX_{ij}} \le 1+sX_{ij}+(sX_{ij})^2\frac{e^L-1-L}{L^2}.$$

Thus,

$$\mathbf{E}\{e^{sX_{ij}}\} \le 1+s^2\mathbf{Var}(X_{ij})\frac{e^L-1-L}{L^2} \le e^{s^2\mathbf{Var}(X_{ij})\frac{e^L-1-L}{L^2}},$$

so (16) follows from (15). □

Remark 3: (16) implies different bounds for different choice of L. If $L=1$ then

$$\begin{aligned} \frac{K\log N}{\min\{K\alpha^*\frac{L^2}{e^L-1-L}, L\}n} &= \frac{K}{\min\{K\alpha^*/(e-2), 1\}}\frac{\log N}{n} \\ &= \max\left\{\frac{e-2}{\alpha^*}, K\right\}\frac{\log N}{n}. \end{aligned}$$

Hamers and Kohler [11] derived a similar bound:

$$\left(\frac{1}{2\alpha^*}+\frac{2K}{3}\right)\frac{\log N}{n}.$$

There is a better choice of L. Introduce the function

$$G(L)=\frac{e^L-L-1}{L} \qquad \text{for } L>0,$$

and choose L such that

$$K\alpha^*\frac{L^2}{e^L-1-L}=L,$$

i.e.,

$$K\alpha^*=G(L),$$

i.e.,

$$L=G^{-1}(K\alpha^*),$$

then (16) implies the bound

$$\frac{K\log N}{G^{-1}(K\alpha^*)n}. \tag{17}$$

Remark 4: In its spirit, Lemma 1 is similar to the inequality of Devroye and Lugosi [5] with the essential difference that they considered the maximum of zero mean random variables:

$$\mathbf{E}\left\{\max_{j\le N}\left(\frac{1}{n}\sum_{i=1}^{n} X_{ij}\right)\right\} \le 2\max_{j\le N}\sigma_j\sqrt{\frac{\log N}{n}}.$$

Proof of Theorem 2: Following the proof of [17] consider a cubic grid of width δ in $S(B)$ with the minimum number of grid-points. It can be seen that if the origin of the grid is uniformly distributed in the d-dimensional unit cube centered at the origin of the sphere, then the expected number of grid points within the ball is V/δ^d. Thus the grid in $S(B)$ with minimum number of points has at most V/δ^d points. Let $\mathcal{Q}'_N$ denote the set of N-level nearest neighbor quantizers which have all their code points on this grid. Since for any $y \in S(B)$ there is an y' on the grid such that $\|y - y'\| \le \delta\sqrt{d}$, for any quantizer Q with code points inside $S(B)$ there is a $Q' \in \mathcal{Q}'_N$ such that

$$\sup_{x\in S(B)} \left| \|x - Q(x)\|^2 - \|x - Q'(x)\|^2 \right| \le 4\delta B\sqrt{d}.$$

Letting $\epsilon = 4\delta B\sqrt{d}$, specifically there is a quantizer $Q'_n \in \mathcal{Q}'_N$ satisfying

$$\sup_{x\in S(B)} \left| \|x - Q^*_n(x)\|^2 - \|x - Q'_n(x)\|^2 \right| \le \epsilon$$

with probability 1, since the code points of an empirically optimal quantizer are almost surely concentrated in $S(B)$. Concerning the cardinality of $\mathcal{Q}'_N$ we have

$$|\mathcal{Q}'_N| \le \left(\frac{V}{\delta^d}\right)^N = V^N(4B\sqrt{d})^{dN}\epsilon^{-dN}. \tag{18}$$

For $Q = Q'_n$ let $\widehat{Q}_n$ denote the optimal quantizer achieving the minimum in (7). Using the empirical optimality of Q^*_n we proceed with the following decomposition.

$$\begin{aligned}
&\mathbf{E}\{\|X - Q^*_n(X)\|^2|X_1^n\} - D^* \\
&\le \mathbf{E}\{\|X - Q^*_n(X)\|^2|X_1^n\} - \mathbf{E}\{\|X - \widehat{Q}_n(X)\|^2|X_1^n\} \\
&\quad -\frac{2}{n}\sum_{i=1}^n \left(\|X_i - Q^*_n(X_i)\|^2 - \|X_i - \widehat{Q}_n(X_i)\|^2\right) \\
&\le 3\epsilon + \mathbf{E}\{\|X - Q'_n(X)\|^2|X_1^n\} - \mathbf{E}\{\|X - \widehat{Q}_n(X)\|^2|X_1^n\} \\
&\quad -\frac{2}{n}\sum_{i=1}^n \left(\|X_i - Q'_n(X_i)\|^2 - \|X_i - \widehat{Q}_n(X_i)\|^2\right)
\end{aligned} \tag{19}$$

with probability 1. Then

$$\begin{aligned}
&\mathbf{E}\{\|X - Q^*_n(X)\|^2\} - D^* \\
&\le \mathbf{E}\Big\{\mathbf{E}\{\|X - Q'_n(X)\|^2|X_1^n\} - \mathbf{E}\{\|X - \widehat{Q}_n(X)\|^2|X_1^n\} \\
&\quad -\frac{2}{n}\sum_{i=1}^n \left(\|X_i - Q'_n(X_i)\|^2 - \|X_i - \widehat{Q}_n(X_i)\|^2\right)\Big\} + 3\epsilon \\
&= \mathbf{E}\left\{\mathbf{E}\{\Delta_{Q'_n}(X)|X_1^n\} - \frac{2}{n}\sum_{i=1}^n \Delta_{Q'_n}(X_i)\right\} + 3\epsilon \\
&\le \mathbf{E}\left\{\max_{Q\in\mathcal{Q}'_N}\left(\mathbf{E}\{\Delta_Q(X)\} - \frac{2}{n}\sum_{i=1}^n \Delta_Q(X_i)\right)\right\} + 3\epsilon \\
&= \mathbf{E}\left\{\max_{Q\in\mathcal{Q}'_N}\left(\frac{1}{n}\sum_{i=1}^n 2\left(\mathbf{E}\{\Delta_Q(X)\} - \Delta_Q(X_i)\right) - \mathbf{E}\{\Delta_Q(X)\}\right)\right\} \\
&\quad + 3\epsilon \\
&\le 4\max\left\{\frac{e-2}{A}, 4B^2\right\}\frac{\log|\mathcal{Q}'_N|}{n} + 3\epsilon &(20) \\
&\le 4\max\left\{\frac{e-2}{A}, 4B^2\right\}\frac{\log\left(V^N(4B\sqrt{d})^{dN}\epsilon^{-dN}\right)}{n} + 3\epsilon, &(21)
\end{aligned}$$

where (20) follows from Lemma 1 with $L = 1$ and $X_{i,Q} = 2(\mathbf{E}\{\Delta_Q(X)\} - \Delta_Q(X_i))$ and $\delta_Q = \mathbf{E}\{\Delta_Q(X)\}$; and (21) follows form (18). Choosing

$$\epsilon = 4\max\left\{\frac{e-2}{A}, 4B^2\right\}\frac{dN}{3n}$$

completes the proof of the theorem.

The improved constants of Remark 2 can be obtained if one uses the improved version of Lemma 1 with bound (17) instead of with $L = 1$ in (20). □

Proof of Corollary 1: We only need to show that condition (8) is satisfied under the assumptions of the corollary; then the result follows by Theorem 2. It is easy to see that if (8) does not hold, then there is a sequence of strictly suboptimal quantizers $Q_n \in \mathcal{Q}_N$ converging to an optimal quantizer Q^* in the sense that $y_{n,i} \to y_i^*$ for all i, where $\{y_{n,1}, \ldots, y_{n,N}\}$ denotes the codebook of Q_n and $\{y_1^*, \ldots, y_N^*\}$ denotes the codebook of Q^*, such that

$$\lim_{n\to\infty} \frac{\mathbf{E}\{\Delta_{Q_n}(X)\}}{\mathbf{Var}\{\Delta_{Q_n}(X)\}} = 0. \tag{22}$$

In what follows we will show that

$$\liminf_{n\to\infty} \frac{\mathbf{E}\{\|X - Q_n(X)\|^2 - \|X - Q^*(X)\|^2\}}{\mathbf{Var}\{\|X - Q_n(X)\|^2 - \|X - Q^*(X)\|^2\}} > 0 \tag{23}$$

which readily implies that (22) cannot hold, proving the corollary.

Since $D(y_1, \ldots, y_N) = \mathbf{E}\{\|X - Q(X)\|^2\}$ is twice differentiable according to the vector $\mathbf{y} = (y_1, \ldots, y_N) \in \mathbb{R}^{dN}$ in a neighborhood of $\mathbf{y}^* = (y_1^*, \ldots, y_N^*)$, where Q is a nearest neighbor quantizer with codebook $\{y_1, \ldots, y_N\}$ [22], $D(y_1, \ldots, y_N)$ has the following Taylor expansion:

$$\begin{aligned} D(y_1, \ldots, y_N) &= D(y_1^*, \ldots, y_N^*) + \frac{dD(\mathbf{y}^*)}{d\mathbf{y}}(\mathbf{y} - \mathbf{y}^*) \\ &\quad + \frac{1}{2}(\mathbf{y} - \mathbf{y}^*)^T \Gamma(\mathbf{y}^*)(\mathbf{y} - \mathbf{y}^*) + o(\|\mathbf{y} - \mathbf{y}^*\|^2) \end{aligned}$$

where $dD(\mathbf{y}^*)/d\mathbf{y}$ denotes the vector formed by the partial derivatives of D according to its variables at $\mathbf{y}^*$. It is also shown in [22] that the derivative $dD(\mathbf{y})/d\mathbf{y}$ is made up of the d-vectors $\partial D(y_1, \ldots, y_N)/\partial y_i = -2\mathbf{E}\{I_{\{X\in S_i\}}(X - y_i)\}$. However, since $\mathbf{y}^*$ minimizes $D(\mathbf{y})$, $dD(\mathbf{y}^*)/d\mathbf{y} = \mathbf{0}$ (here $\mathbf{0}$ denotes the zero vector), and so

$$\begin{aligned} &\mathbf{E}\{\|X - Q(X)\|^2 - \|X - Q^*(X)\|^2\} \\ &\quad = \quad D(y_1, \ldots, y_N) - D(y_1^*, \ldots, y_N^*) \\ &\quad = \quad \frac{1}{2}(\mathbf{y} - \mathbf{y}^*)^T \Gamma(\mathbf{y}^*)(\mathbf{y} - \mathbf{y}^*) + o(\|\mathbf{y} - \mathbf{y}^*\|^2). \end{aligned}$$

Furthermore, since $\Gamma(\mathbf{y}^*)$ is positive definite by assumption, its smallest eigenvalue λ is positive, and for any $\mathbf{a} \in \mathbb{R}^{dN}$ we have

$$\mathbf{a}^T \Gamma(\mathbf{y}^*)\mathbf{a} \geq \lambda \|\mathbf{a}\|^2$$

which in turn implies that

$$\mathbf{E}\{\|X - Q(X)\|^2 - \|X - Q^*(X)\|^2\} \geq \frac{\lambda}{2}\|\mathbf{y} - \mathbf{y}^*\|^2 + o(\|\mathbf{y} - \mathbf{y}^*\|^2). \tag{24}$$

Next we bound the variance $\mathbf{Var}\{\|X - Q(X)\|^2 - \|X - Q^*(X)\|^2\}$. Notice that $\|x - Q(x)\|^2$ can be decomposed as

$$\begin{aligned} &\|x - Q(x)\|^2 \\ &\quad = \quad \sum_{i=1}^N \|x - y_i\|^2 I_{\{x\in S_i^*\}} \\ &\qquad\quad + \sum_{i=1}^N \sum_{1\leq j\leq N, j\neq i} \left(\|x - y_j\|^2 - \|x - y_i\|^2\right) I_{\{x\in S_j \cap S_i^*\}} \end{aligned}$$

where, as usual, S_i denotes the cell of Q corresponding to the code point y_i. Therefore,

$$\begin{aligned} &\|x - Q(x)\|^2 - \|x - Q^*(x)\|^2 \\ &\quad = \quad \sum_{i=1}^N \left(\|x - y_i\|^2 - \|x - y_i^*\|^2\right) I_{\{x\in S_i^*\}} \\ &\qquad\quad + \sum_{i=1}^N \sum_{1\leq j\leq N, j\neq i} \left(\|x - y_j\|^2 - \|x - y_i\|^2\right) I_{\{x\in S_j \cap S_i^*\}} \\ &\quad = \quad \sum_{i=1}^N J_i(x) + \sum_{i=1}^N \sum_{1\leq j\leq N, j\neq i} K_{i,j}(x), \end{aligned}$$

and so by the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathbf{Var}\{\|X - Q(X)\|^2 - \|X - Q^*(X)\|^2\} \\
&= \mathbf{Var}\left\{\sum_{i=1}^N J_i(X) + \sum_{i=1}^N \sum_{1\le j\le N, j\ne i} K_{i,j}(X)\right\} \\
&\le N^2\left(\sum_{i=1}^N \mathbf{Var}\{J_i(X)\} + \sum_{i=1}^N \sum_{1\le j\le N, j\ne i} \mathbf{Var}\{K_{i,j}(X)\}\right). \qquad (25)
\end{aligned}$$

First we bound $\mathbf{Var}\{J_i(X)\}$:

$$\begin{aligned}
&\mathbf{Var}\{J_i(X)\} \\
&\le \mathbf{E}\{J_i^2(X)\} \\
&\le \mathbf{E}\left\{\Big((X-y_i)^T(X-y_i) - (X-y_i^*)^T(X-y_i^*)\Big)^2\right\} \\
&= \mathbf{E}\left\{\Big((2X - y_i - y_i^*)^T(y_i^* - y_i)\Big)^2\right\} \\
&\le \mathbf{E}\left\{\|2X - y_i - y_i^*\|^2\|y_i^* - y_i\|^2\right\} \qquad (26) \\
&\le 16B^2\|y_i^* - y_i\|^2 \qquad (27)
\end{aligned}$$

where (26) follows from the Cauchy-Schwarz inequality, and (27) by the fact that $\|2X - y_i - y_i^*\| \le 4B$ by the triangle inequality if $\|X\| \le B$.

Bounding the terms $\mathbf{Var}\{K_{i,j}(X)\}$ is somewhat more complicated. Let $X_{i,j}$ denote the orthogonal projection of X to the line connecting y_i and y_j. Then for $k = i, j$, $\|X - y_k\|^2 = \|X_{i,j} - y_k\|^2 + \|X - X_{i,j}\|^2$, and so

$$\begin{aligned}
&\mathbf{Var}\{K_{i,j}(X)\} \\
&\le \mathbf{E}\{K_{i,j}^2(X)\} \\
&= \mathbf{E}\left\{\left(\|X - y_j\|^2 - \|X - y_i\|^2\right)^2 I_{\{X\in S_j\cap S_i^*\}}\right\} \\
&= \mathbf{E}\left\{\left(\|X_{i,j} - y_j\|^2 - \|X_{i,j} - y_i\|^2\right)^2 I_{\{X\in S_j\cap S_i^*\}}\right\} \\
&= \mathbf{E}\left\{\|2X_{i,j} - y_i - y_j\|^2\|y_i - y_j\|^2 I_{\{X\in S_j\cap S_i^*\}}\right\} \\
&\le 4B^2\mathbf{E}\left\{\|2X_{i,j} - y_i - y_j\|^2 I_{\{X\in S_j\cap S_i^*\}}\right\} \qquad (28)
\end{aligned}$$

where the last equality holds because y_i, y_j and $X_{i,j}$ lie in the same line (otherwise it would be an inequality as in (26)).

Next we prove that the expectation in (28) is $O(\|y_i - y_i^*\|^2 + \|y_j - y_j^*\|^2)$. Let

$$H_{j,i} = \{x \in \mathbb{R}^d : \|x - y_j\| \le \|x - y_i\|\}$$

and

$$H_{i,j}^* = \{x \in \mathbb{R}^d : \|x - y_i^*\| \le \|x - y_j^*\|\}.$$

First we show that if $X \in H_{j,i} \cap H_{i,j}^*$, $\|X\| \le B$, and $\|y_i - y_i^*\|$ and $\|y_j - y_j^*\|$ are sufficiently small (relative to $\|y_i^* - y_j^*\|$), then

$$\left\|X_{i,j} - \frac{y_i + y_j}{2}\right\| \le \left(\frac{2B}{\|y_i^* - y_j^*\|} + \frac{1}{2}\right)(\|y_i - y_i^*\| + \|y_j - y_j^*\|). \qquad (29)$$

To see this, let $h_{j,i}$ and $h_{i,j}^*$ denote the hyperplanes corresponding to $H_{j,i}$ and $H_{i,j}^*$, respectively, let $h_{j,i}'$ denote the hyperplane containing the point $(y_i^* + y_j^*)/2$ parallel to $h_{j,i}$, and let $H_{j,i}'$ denote the corresponding shifted version of the halfspace $H_{j,i}$. Furthermore, let $\rho(x, h)$ denote the Eucledian distance of the point $x \in \mathbb{R}^d$ from the hyperplane $h \subset \mathbb{R}^d$. Then, since $(y_i + y_j)/2 \in h_{j,i}$, for any x between the hyperplanes $h_{j,i}$ and $h_{j,i}'$ we have

$$\rho(x, h_{j,i}) \le \left\|\frac{y_i^* + y_j^*}{2} - \frac{y_i + y_j}{2}\right\|, \qquad (30)$$

and for any (other) $x \in \mathbb{R}^d$ we have

$$\rho(x, h_{j,i}) \le \left\|\frac{y_i^* + y_j^*}{2} - \frac{y_i + y_j}{2}\right\| + \rho(x, h_{j,i}'). \qquad (31)$$

Let α denote the angle of the vectors $(y_i - y_j)$ and $(y_i^* - y_j^*)$. It is easy to see that if $\|y_i - y_i^*\|$ and $\|y_j - y_j^*\|$ are small enough (relative to $\|y_i^* - y_j^*\|$), then $0 \le \alpha \le \pi/2$, and so for all $x \in H'_{j,i} \cap H^*_{i,j}$,$\|x\| \le B$ we have

$$\rho(x, h'_{j,i}) \le \left\| x - \frac{y_i^* + y_j^*}{2} \right\| \sin\alpha \le 2B \sin\alpha$$

since $\|(y_i^* + y_j^*)/2\| \le B$. Furthermore, in this case,

$$\sin\alpha \le \frac{\|y_i - y_i^*\| + \|y_j - y_j^*\|}{\|y_i^* - y_j^*\|},$$

and so for any $x \in H'_{j,i} \cap H^*_{i,j}$, $\|x\| \le B$, and $\|y_i - y_i^*\|$ and $\|y_j - y_j^*\|$ sufficiently small, by (31) we have

$$\begin{aligned} \left\| x_{i,j} - \frac{y_i + y_j}{2} \right\| &= \rho(x, h_{j,i}) \\ &\le \left\| \frac{y_i^* + y_j^*}{2} - \frac{y_i + y_j}{2} \right\| + \frac{2B(\|y_i - y_i^*\| + \|y_j - y_j^*\|)}{\|y_i^* - y_j^*\|} \\ &\le \left(\frac{2B}{\|y_i^* - y_j^*\|} + \frac{1}{2} \right) (\|y_i - y_i^*\| + \|y_j - y_j^*\|) \end{aligned} \tag{32}$$

where $x_{i,j}$ denotes the orthogonal projection of x to the line connecting y_i and y_j. Now since if $x \in H_{j,i}$ then it is either between $h_{j,i}$ and $h'_{j,i}$, or $x \in H'_{j,i}$ (or both), (30) implies that (32) is valid for all $x \in H_{j,i} \cap H^*_{i,j}$, $\|x\| \le B$. This implies (29).

Therefore, since $S_j \subset H_{j,i}$ and $S_i^* \subset H^*_{i,j}$, for small enough $\|y_i - y_i^*\|$, $\|y_j - y_j^*\|$, we have

$$\begin{aligned} &\mathbf{E}\left\{ \|2X_{i,j} - y_i - y_j\|^2 I_{\{X \in S_j \cap S_i^*\}} \right\} \\ &\le \mathbf{E}\left\{ \|2X_{i,j} - y_i - y_j\|^2 I_{\{X \in H_{j,i} \cap H^*_{i,j}\}} \right\} \\ &\le 4\left(\frac{2B}{\|y_i^* - y_j^*\|} + \frac{1}{2} \right)^2 (\|y_i - y_i^*\| + \|y_j - y_j^*\|)^2 \\ &\le 8\left(\frac{2B}{\|y_i^* - y_j^*\|} + \frac{1}{2} \right)^2 (\|y_i - y_i^*\|^2 + \|y_j - y_j^*\|^2). \end{aligned}$$

Thus from (28)

$$\begin{aligned} &\mathbf{Var}\{K_{i,j}(X)\} \\ &\le 32B^2 \left(\frac{2B}{\|y_i^* - y_j^*\|} + \frac{1}{2} \right)^2 (\|y_i - y_i^*\|^2 + \|y_j - y_j^*\|^2) \\ &\le 32B^2 \left(\frac{2B}{\min_{u \ne v} \|y_u^* - y_v^*\|} + \frac{1}{2} \right)^2 (\|y_i - y_i^*\|^2 + \|y_j - y_j^*\|^2), \end{aligned}$$

which, together with (25) and (27), implies

$$\begin{aligned} &\mathbf{Var}\{\|X - Q(X)\|^2 - \|X - Q^*(X)\|^2\} \\ &\le B^2 N^2 \left(16 + 64(N-1) \left(\frac{2B}{\min_{u \ne v} \|y_u^* - y_v^*\|^2} + \frac{1}{2} \right)^2 \right) \sum_{i=1}^N \|y_i - y_i^*\|^2 \\ &= C\|\mathbf{y} - \mathbf{y}^*\|^2 \end{aligned}$$

if $\|\mathbf{y} - \mathbf{y}^*\|$ is sufficiently small, where

$$C = B^2 N^2 \left(16 + 64(N-1) \left(\frac{2B}{\min_{u \ne v} \|y_u^* - y_v^*\|^2} + \frac{1}{2} \right)^2 \right)$$

is a positive constant. Now for such $\mathbf{y}$'s (24) yields

$$\frac{\mathbf{E}\{\|X - Q(X)\|^2 - \|X - Q^*(X)\|^2\}}{\mathbf{Var}\{\|X - Q(X)\|^2 - \|X - Q^*(X)\|^2\}} \ge \frac{\lambda}{2C} + \frac{o(\|\mathbf{y} - \mathbf{y}^*\|^2)}{\|\mathbf{y} - \mathbf{y}^*\|^2}$$

and hence (23) holds, completing the proof. Note that carrying out the same proof in one-dimension, that is, when $d = 1$, the constant C has the much simpler form $C = 32(2N-1)B^2$, thus in that case C does not depend on Q^*. □

ACKNOWLEDGEMENTS

The authors wish to thank Zsolt Bihary and Tamás Linder for useful discussions.

REFERENCES

[1] A. Antos, "Improved minimax bounds on the test and training distortion of empirical designed vector quantizers," *IEEE Transactions on Information Theory*, under revision.

[2] A. R. Barron, "Complexity regularization with application to artificial neural networks," in *Nonparametric Functional Estimation and Related Topics* (G. Roussas, ed.), pp. 561–576, NATO ASI Series, Kluwer Academic Publishers, Dordrecht, 1991.

[3] P. L. Bartlett, T. Linder, and G. Lugosi, "The minimax distortion redundancy in empirical quantizer design," *IEEE Transactions on Information Theory*, vol. IT-44, pp. 1802–1813, Sept. 1998.

[4] P. Chou, "The distortion of vector quantizers trained on n vectors decreases to the optimum as $O_p(1/n)$," in *Proc. IEEE International Symposium on Information Theory*, (Trondheim, Norway), p. 457, 1994.

[5] L. Devroye and G. Lugosi, *Combinatorial Methods in Density Estimation*. New York: Springer-Verlag, 2001.

[6] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. New York: Springer-Verlag, 1996.

[7] P. E. Fleischer, "Sufficient conditions for achieving minimum distortion in a quantizer," IEEE Int. Conv. Rec., Part I, pp. 104-111, 1964.

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

[9] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*. New York: Springer, 2002.

[10] J. A. Hartigan, "Asymptotic distributions for clustering criteria," *Annals of Statistics*, vol. 6, pp. 117–131, 1978.

[11] M. Hamers and M. Kohler, "A bound on the expected maximal deviation of averages from their means," *Statistics and Probability Letters*, vol. 62, pp. 137–144, 2003.

[12] W. S. Lee, P. L. Bartlett, and R. C. Williamson, "Efficient agnostic learning of neural networks with bounded fan-in," *IEEE Transactions on Information Theory*, vol. IT-42, pp. 2118–2132, Nov. 1996.

[13] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. COM-28, pp. 84–95, Jan. 1980.

[14] T. Linder, "On the training distortion of vector quantizers," *IEEE Transactions on Information Theory*, vol. IT-46, pp. 1617–1623, Jul. 2000.

[15] T. Linder, "Learning-theoretic methods in vector quantization," in *Principles of Nonparametric Learning* (Ed.: L. Györfi) Springer-Verlag, Wien, New York, 2002.

[16] T. Linder, G. Lugosi, and K. Zeger, "Rates of convergence in the source coding theorem, in empirical quantizer design, and in universal lossy source coding," *IEEE Transactions on Information Theory*, vol. IT-40, pp. 1728–1740, Nov. 1994.

[17] T. Linder, G. Lugosi, and K. Zeger, "Empirical quantizer design in the presence of source noise or channel noise," *IEEE Transactions on Information Theory*, vol. IT-43, pp. 612–623, Mar. 1997.

[18] C. L. Mallows, "A note on asymptotic joint normality," *Annals Math. Statistics*, vol. 43, pp. 508–515, 1972.

[19] N. Merhav and J. Ziv, "On the amount of side information required for lossy data compression," *IEEE Transactions on Information Theory*, vol. IT-43, pp. 1112–1121, July 1997.

[20] D. Pollard, "Strong consistency of k-means clustering," *Annals of Statistics*, vol. 9, no. 1, pp. 135–140, 1981.

[21] D. Pollard, "Quantization and the method of k-means," *IEEE Transactions on Information Theory*, vol. IT-28, pp. 199–205, Mar. 1982.

[22] D. Pollard, "A central limit theorem for k-means clustering," *Annals of Probability*, vol. 10, no. 4, pp. 919–926, 1982.

[23] A. V. Trushkin, "Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting function," *IEEE Transactions on Information Theory*, vol. IT-28, pp. 187–198, Mar. 1982.

[24] A. J. Zeevi, "On the performance of vector quantizers empirically designed from dependent sources," in *Proc. Data Compression Conference, DCC'98* (J. Storer and M. Cohn, eds.), (Los Alamitos, CA), pp. 73–82, IEEE Computer Society Press, 1998.

András Antos was born in Budapest, Hungary in 1972. He received the M.Sc. degree (with distinction) in engineering informatics from the Technical University of Budapest in 1995, and the Ph.D. degree in technical informatics from the Budapest University of Technology and Economics in 2000.

Since 1998, he has been with the Informatics Laboratory, Computer and Automation Research Institute of the Hungarian Academy of Sciences, where currently he is a senior researcher. He was a NATO Science Postdoctoral Fellow at the Department of Mathematics and Statistics, Queen's University in 2001–2003. His research interests include nonparametric statistics, information theory, source coding, vector quantization, machine learning, and random trees.

Dr. Antos received the Student Author Award of the IFIP Working Group 1.4 in 1999.

László Györfi (M'87-SM'94-F'97) was born in Hercegfalva, Hungary, on July 6,1947. He received a degree in mathematics and physics in 1970 from the Loránd Eötvös University, Budapest, Hungary

He coauthored, with L. Devroye the book *Nonparametric Density Estimation: the L_1 View*, (New York: Wiley, 1985), with L. Devroye and G. Lugosi the book *Probabilistic Theory of Pattern Recognition*, (New York: Springer, 1996), and with M. Kohler, A. Krzyżak, and H. Walk the book *A Distribution-Free Theory of Nonparametric Regression*, (New York: Springer, 2002). His main interests are nonparametric statistics and multiple access communication. He is now with the Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Budapest, Hungary.

Dr. Györfi became an Ordinary Member of the Hungarian Academy of Sciences in 2001.

András György (S'01-A'03-M'04) was born in Budapest, Hungary, in 1976. He received the M.Sc. degree (with distinction) in technical informatics from the Technical University of Budapest, in 1999, the M.Sc. degree in mathematics and engineering from Queen's University, Kingston, ON, Canada, in 2001, and the Ph.D. degree in technical informatics from the Budapest University of Technology and Economics in 2003.

He was a Visiting Research Scholar in the Department of Electrical and Computer Engineering, University of California, San Diego, in spring of 1998. Since 2002, he has been a researcher in the Informatics Laboratory of the Computer and Automation Research Institute of the Hungarian Academy of Sciences. In 2003–2004 he was a NATO Science Fellow in the Department of Mathematics and Statistics, Queen's University. His research interests include information theory, source coding, vector quantization, communication networks, and nonparametric statistics.

Dr. György received the Gyula Farkas prize of the János Bolyai Mathematical Society in 2001 and the Academic Golden Ring of the President of the Hungarian Republic in 2003.