A New Compensated Criterion in Testing Trained Codebooks

Dong Sik Kim* *Regular Member*

ABSTRACT

In designing the quantizer of a coding scheme using a training sequence (TS), the training algorithm tries to find a quantizer that minimizes the distortion measured in the TS. In order to evaluate the trained quantizer or compare the coding schemes, we can observe the minimized distortion. However, the minimized distortion is a biased estimate of the minimal distortion for the input distribution. Hence, we could often overestimate a quantizer or make a wrong comparison even if we use a validating sequence. In this paper, by compensating the minimized distortion for the TS, a new estimate is proposed. Compensating term is a function of the training ratio, the TS size to the codebook size. Several numerical results are also introduced for the proposed estimate.

I. Introduction

Quantization (or vector quantization) can effectively reduce the huge amount of data with possibly small error, which is called *quantizer distortion*, and is the basis for the lossy video and audio compression[8]. In pattern recognition/classification, including speech recognition, part of the unsupervised learning and clustering step can be regarded as vector quantization or the vector quantizer(VQ) design problem[2],[7],[9].

Since the codewords of the VQ codebook are elements of the k-dimensional Euclidean space R^k, deriving an explicit relation between an optimal quantizer and the input distribution is very difficult. Hence, explicitly designing a VQ codebook is difficult even if the input distribution is known. Clustering algorithms, however, can effectively design a codebook using a training sequence (TS) by minimizing the distortion measured in the given TS. The underlying distribution of the TS is equal to the input distribution. TS could be a part of the video or the speech spectral vectors. We call the distortion measured in the given TS the *training sequence distortion*(TSD). The well known clustering algorithms for designing VQ are the K-means algorithm and the Kohonen learning algorithm[22]. Such inductive methods are based on the *empirical risk minimization*(ERM) principle[6]. Abaya and Wise[19] studied the consistency problem of the sequence of the trained quantizers. They have shown that, as the TS size increases, the trained quantizers yield a distortion sequence that converges to the minimal distortion for the input distribution and that TSD also converges to the minimal distortion. Hence, if using a large TS is allowed, then the trained codebook for the TS will produce a distortion that is very close to the minimal distortion. However, training using a large TS is very complex, and acquiring a large TS is often difficult. When we develop a lossy coding scheme based on quantization, if the quantizer is designed using a TS, then we may encounter with the following questions:

1. How to evaluate a trained codebook,
2. How to compare the coding schemes.

Since TSD is easy of access in training the quantizer, TSD is usually employed to answer the questions. However, the well known effect of TS on TSD is that TSD decreases as the TS size decreases and eventually goes to zero. In other words, TSD is a biased estimate of the minimal distortion for the input distribution and the bias

* School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, U. S. A. (dskim@ecn.purdue.edu)
논문번호 : 98355-0814, 접수일자 : 1999년 8월 14일

increases as the TS size decreases (this fact will be shown in this paper).

We now consider the problems caused by the effect of TS size based on the questions. The first question is concerned with evaluating the optimality of a trained codebook by calculating the distortion using the trained codebook for the input distribution and comparing this distortion to the minimal distortion for the input distribution. The underlying distribution of the TS and the input distribution may be chosen to be different in order to evaluate the robustness of the trained codebook; this problem is referred as the codebook robustness or the quantizer mismatch problem[21]. However, in this paper, we will confine the problem to the case that the underlying distribution of TS is equal to the input distribution. Since the input distribution is usually unknown, we should use the validating sequence(VS) to calculate the distortion produced by the trained codebook for the input distribution. It is assumed that the underlying distribution of VS is equal to the input distribution. Let the distortion measured using the VS be denoted by VSD. Furthermore, since we do not know the minimal distortion for the input distribution, the biased TSD is employed instead of the minimal distortion. Hence, the following difference distortion

$$\mathrm{VSD} - \mathrm{TSD} \qquad (1)$$

is usually employed to observe the optimality of a trained codebook. However, since TSD is a biased estimate, this difference distortion is much greater than the real difference.

The second question is concerned with comparing the minimal distortions that can be achieved by each coding scheme for the same input distribution. A simple way to compare the coding schemes is comparing TSDs of each coding scheme. If the TS size is large enough, then this approach may be appropriate. However, since the TSD is a biased estimate of the minimal distortion for the input distribution, comparing TSDs could lead a codebook designer to make an erroneous design decision. For example, even though the minimal distortions of two coding schemes are the same, their TSDs could be different depending on certain conditions related with TS. A better coding scheme may reduce the TSD. However, decreasing the TS size could also reduce the TSD. It follows that we often overestimate the performance of a coding scheme. To avoid such a mistake, it is general procedure to compare each coding scheme's VSD in a similar manner as in the first case. However, in real applications, since the input often has a time-varying statistics, obtaining an appropriate VS is difficult. Hence, a careless choice of VS could also yield a wrong comparison.

In this paper, focusing on the mistakes that could be caused by the bias of TSD, we try to alleviate the bias of TSD. To do this, the TSD is analyzed by proposing several upper bounds. An estimate of the minimal distortion for the input distribution is then proposed by compensating the TSD. The bias due to the finite TS size will be significantly compensated by this estimate. Hence, when we evaluate a trained codebook or compare the coding schemes, employing the new estimate in place of TSD will reduce the effect caused by using TS.

This paper is organized as follows. In Section II, VQ and TSD are defined with mathematical notations and the bias of the TSD will be observed. In order to alleviate the bias, an asymptotic bound for the TSD will be introduced in Section III. A new difference distortion using the compensated estimate is followed in Section IV. In Section V, numerical results on the new estimate will be illustrated. Section VI concludes this paper.

II. Theoretical Observation on TSD

In this section, the TSD will be defined and theoretically observed. It will be shown that the TSD is a biased estimate of the minimal distortion for the input distribution. Consider a discrete-time, memoryless, stationary source, $X_1, X_2, \cdots$ with source symbols that are independent and identically distributed(i.i.d.) random vectors taking values in R^k. Let F be defined as the distribution function of the random vector X_1. Suppose that $E\|X_1\|^2 < \infty$, where $\|x\| := \sqrt{x_1^2 + \cdots + x_k^2}$, for $x = (x_1, \cdots, x_k)$.

Let C_n denote the class of sets that have n points in R^k, and let the sets in C_n be called the n-level codebooks, where each codebook has n codewords. For a codebook C ($\in C_n$), a VQ is then described as a map Q_C defined as

$$Q_C : R^k \to C,$$

$$Q_C(x) \mapsto \arg\min_{y \in C} \|x - y\|^2. \quad (2)$$

Furthermore, the average distortion $D(C, F)$ resulting when the random vector X_1 is quantized using codebook C, is given by

$$D(C, F) := \int \|x - Q_C(x)\|^2 dF(x). \quad (3)$$

The quantity $\inf_{C \in C_n} D(C, F)$ is called the (n-*level*) *F-optimal distortion*, and a code C^* that yields the F-optimal distortion is called an (n-*level*) *F-optimal codebook* (i.e., $D(C^*, F) = \inf_{C \in C_n} D(C, F)$ if C^* exists[16]). Note that the optimal codebook design problem is that of finding an F-optimal codebook.

We define the TSD quantitatively as follows. Let $X_1^\omega, \cdots, X_m^\omega$ be a given TS, where ω is a sample point in the underlying sample space Ω and m is the TS size. For a codebook C ($\in C_n$), TSD, the quantizer distortion measured in the TS, is defined as

$$\text{TSD} := D(C, F_m^\omega) = \int \|x - Q_C(x)\|^2 dF_m^\omega(x), \quad (4)$$

where F_m^ω is the empirical distribution function of the TS[10,p.268]. Thus the TSD is a random variable that is defined on the underlying sample space and is a function of ω. However, for simplicity, we omit ω in this notation. The number of possible ways in grouping the m points $X_1^\omega, \cdots, X_m^\omega$ into the n (or less than n) groups is n^m, which is finite[1]. Hence, there exists at least one codebook that minimizes the TSD for a given F_m^ω. Let C_X^ω, which satisfies

$$D(C_X^\omega, F_m^\omega) = \inf_{C \in C_n} D(C, F_m^\omega), \quad (5)$$

be called an *F_m-optimal codebook* and $D(C_X^\omega, F_m^\omega)$ called the *F_m-optimal distortion*[27]. Note that $\inf_{C \in C_n} D(C, F_m^\omega) = 0$ every ω if $\beta = 1$. Through this paper, we assume that the codebook in TSD is F_m-optimal, i.e., $\text{TSD} = \inf_{C \in C_n} D(C, F_m^\omega)$ from the codebook design algorithms that is based on the ERM principle. Otherwise the codebook in TSD will be specified.

Now, we observe the bias of TSD. For any fixed codebook B, it is clear from [25, Appendix] that $E\{D(B, F_m)\} = D(B, F)$. It follows that

$$\inf_{C \in C_n} E\{D(C, F_m)\} = \inf_{C \in C_n} D(C, F). \quad (6)$$

Hence, if C^* ($\in C_n$) is an F-optimal codebook, then the TSD for C^* will be an unbiased estimate of the F-optimal distortion, since $E\{D(C^*, F_m)\} = D(C^*, F)$. On the other hand, from Appendix A,

$$E\{\inf_{C \in C_n} D(C, F_m)\} \le \inf_{C \in C_n} D(C, F)\left(1 - \frac{1}{m}\right). \quad (7)$$

Hence for finite m,

$$E\{\inf_{C \in C_n} D(C, F_m)\} < \inf_{C \in C_n} D(C, F), \quad (8)$$

which implies that the F_m-optimal distortion is a biased estimate of the F-optimal distortion. Since the TSD is equal to the F_m-optimal distortion from the assumption, the TSD is biased obviously. Note that a similar result can also be derived from a tighter bound derived in [26, Proposition 1].

Ⅲ. A Compensated TSD

In this section, in order to compensate for the bias of TSD, an asymptotic result on the F_m-optimal distortion is introduced for an absolutely continuous F.

Let the training ratio β be defined as the ratio of the TS size to the codebook size, i.e., $\beta := m/n$. Let $\rho := k/(k+2)$, f be the density function of X_1, and

$$\|f\|_\rho := \left[\int f^\rho(x)dx\right]^{1/\rho}. \tag{9}$$

If f is bounded, then, from [26,Theorem 2], we have an asymptotic bound for the F_m-optimal distortion given by

$$\begin{aligned} \limsup_{n\to\infty} \; & n^{2/k}E\{\inf_{C\in C_n} D(C, F_{m_n})\} \\ & \le J_k\|f\|_\rho\left[1-\frac{\zeta-\xi(\beta)}{\beta}\right] \\ & \le J_k\|f\|_\rho\left(1-\frac{1-e^{-\beta}}{\beta}\right), \end{aligned} \tag{10}$$

for $\beta\ (\ge \zeta-\xi(\beta))$, where β is a constant and (m_n) is a sequence of n such that $m_n/n \to \beta$ as $n\to\infty$,

$$\begin{aligned} \zeta \; &:= \left[\int f^{2\rho-1}(x)dx\right], \\ \xi(\beta) \; &:= \left[\int f^{2\rho-1}(x)e^{-f^{1-\rho}(x)\|f\|_\rho^\rho\beta}dx\right] / (\|f\|_\rho)^{2\rho}, \end{aligned} \tag{11}$$

and J_k a constant only depending on the vector dimension[14]. For example, it is known that $J_1 = 0.0833\cdots$, $J_2/2 = 0.0801\cdots$, and $J_3/3 = 0.0785\cdots$ [18,Table I],[17]. Several bounds for J_k/k are introduced in [12], [13], and [20]. It is clear from [15] and [26] that

$$\limsup_{n\to\infty} n^{2/k}\inf_{C\in C_n} D(C, F) \le J_k\|f\|_\rho. \tag{12}$$

However, it is conjectured that the asymptotically optimal quantizer is a function of J_k[12],[15]. In other words,

$$\lim_{n\to\infty} n^{2/k}\inf_{C\in C_n} D(C, F) = J_k\|f\|_\rho. \tag{13}$$

In (11), we assume that $\int f^{2\rho-1}(x)dx < \infty$. Note that the minimum of $\zeta-\xi(\beta)$, i.e., $1-e^{-\beta}$ in (10) is obtained when F is a uniform distribution. For the uniform distribution case, the upper bound is simple as shown in (10).

On the other hand, for the non-uniform distribution case, explicitly deriving ζ and $\xi(\beta)$ are difficult.

First, we observe the constant ζ. Since

$$\int f^{\rho-1/2}f^{1/2}dx \le \left(\int f^{2\rho-1}dx\right)^{1/2}\left(\int f dx\right)^{1/2}, \tag{14}$$

by the Schwarz inequality, $\|f\|_\rho < \infty$ and $\zeta \ge 1$. Suppose that J is a non-singular $n\times n$ matrix. Define a random vector $Z = JX$. If X has a density function f_X, then Z has the density function as

$$f_Z(z) = f_X(J^{-1}z)/|\det J|. \tag{15}$$

Hence, ζ_Z for Z satisfies the following relation

$$\begin{aligned} \zeta_Z &= \left[\int f_Z^{2\rho-1}(z)dz\right] / \left[\int f_Z^{2\rho}(z)dz\right]^2 \\ &= \left[\int f_X^{2\rho-1}(x)|\det J|^{-2\rho+2}dx\right] / \left[\int f_X^{2\rho}(x)|\det J|^{-\rho+1}dx\right]^2 \\ &= \zeta_X. \end{aligned} \tag{16}$$

Hence, ζ is invariant under a linear transformation of X_1, e.g., independent of the variance of the input distribution or the correlation inside the vector X_1. For the uniform distribution case, $\zeta = \zeta^U := 1$. For $k \ge 3$, it is known from [26] that the Gaussian distribution has

$$\zeta = \zeta_k^N := \left(\frac{k^2}{k^2-4}\right)^{k/2}, \tag{17}$$

$$\zeta = \zeta_k^L := \left(\frac{k^2}{k^2-4}\right)^{k}. \tag{18}$$

Next, we consider $\xi(\beta)$. It is clear that $e^{-\beta} \le \xi(\beta) \le \zeta$ from (11) and [26,Proposition 1]. For the uniform distribution, $\xi(\beta) = e^{-\beta}$. However, calculating $\xi(\beta)$ for an arbitrary f is difficult. Hence, an approximation on $\xi(\beta)$ is introduced as follows. Since

$$f^{2\rho-1}(x)e^{f^{1-\rho}(x)\|f\|_\rho^\rho\beta} \to 0, \quad \text{a.e., as } \beta\to\infty, \tag{19}$$

and $f^{2\rho-1}e^{-f^{1-\rho}(x)\|f\|_\rho^\rho\beta} \le f^{2\rho-1}(x)$, where $\int f^{2\rho-1}$

$(x)dx < \infty$, $\lim_{\beta\to\infty}\xi(\beta)=0$ by Lebesgue's dominated convergence theorem[3,p.110]. In other words, $\xi(\beta)/\beta$ converges to zero at a rate faster than β^{-1}. It follows that $[\zeta-\xi(\beta)]/\beta \approx \zeta/\beta$ for relatively large β. For example, when the input has a uniform distribution, if $\beta=5$, then $\beta^{-1}=0.2$, which is approximately equal to $(1-e^{-\beta})/\beta = 0.198\cdots$. Conclusively, for relatively large β, we can obtain an approximate relation as follows.

$$\limsup_{n\to\infty} n^{2/k} E\{\inf_{C\in C_n} D(C, F_{m_n})\} \le J_k \|f\|_\rho \left(1-\frac{\zeta}{\beta}\right). \tag{20}$$

In Fig. 1, several examples of the term $(1-\zeta/\beta)$ in (20) are depicted. We can see the effect of the input distribution on the codebook training. A uniform distribution may require the smallest TS size, and a larger vector dimension k may require a less TS size for the Gaussian and Laplacian distribution in training a codebook, since $\zeta^U < \zeta_k^N < \zeta_k^L$, for $k=3,4,\cdots$. We should notice that maintaining a high training ratio β is important for good codebook. Moreover, increasing the VQ codebook size without increasing TS does not guarantee the expected improvement, since the training performance is degenerated if we decrease the training ratio. Furthermore, a usual mistake is overestimating the performance of the trained quantizer by observing the TSD.

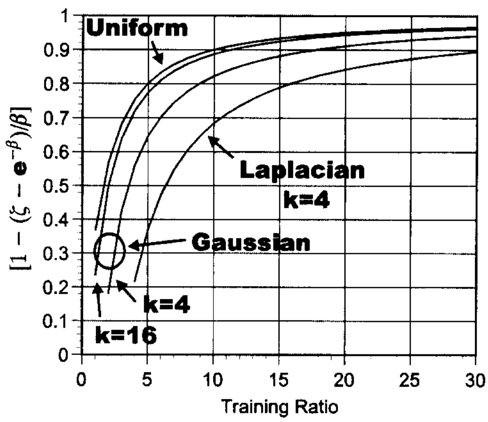


Fig 1. Example of (20).

Now, by compensating the bias in TSD based on (20), a new estimate is proposed as

$$\mathrm{TSD}' := \mathrm{TSD}/\left(1-\frac{\zeta}{\beta}\right). \tag{21}$$

This estimate can substitute for TSD in evaluating the trained codebooks or comparing the coding schemes. Instead of TSD, for relatively large $\beta\ (>\zeta)$. Using this compensated estimate TSD′, we may alleviate the problem that overestimating the performance of the trained codebook due to the TS size. If the input distribution is unknown, we can simply set $\zeta=1$ in (21).

Ⅳ. A New Difference Criterion for Testing Codebooks

In this section, a new difference distortion is proposed using the estimate TSD′ in (21).

In order to validate the optimality of a codebook $B\ (\in C_n)$ for F, we should calculate $D(B,F)$ and compare it with the F-optimal distortion $\inf_{C\in C_n} D(C,F)$. Let Δ denote the difference distortion as

$$\Delta := D(B,F) - \inf_{C\in C_n} D(C,F). \tag{22}$$

Note that $\Delta \ge 0$. If this difference distortion is small enough, then the codebook can be regarded as a good codebook for F. However, since the F-optimal distortion is usually unknown, the TSD is employed instead of the F-optimal distortion [25]. If a VS is available, then using the VS and the trained codebook, we can estimate the difference distortion Δ in (22). First, we consider the traditional difference distortion VSD − TSD.

Consider a sequence of i.i.d. random vectors $\overline{X_1}, \overline{X_2}, \cdots$. Suppose that $\overline{X_1}$ has the same distribution function as X_1 has. In a similar manner in (4), let $\overline{X_1^{\omega'}}, \cdots, \overline{X_{\overline{m}}^{\omega'}}$ be a given VS, where $\omega' \in \Omega$ and $\overline{m}$ is the VS size. For an F-optimal codebook C_X^{ω}, VSD is defined as

$$\mathrm{VSD} := D(C_X^\omega, \overline{F_m^{\omega'}}), \tag{23}$$

where $\overline{F_m^{\omega'}}$ is the empirical distribution function of the VS. In general, the following difference distortion

$$\Delta_1 := \mathrm{VSD} - \mathrm{TSD} \tag{24}$$

is employed for validating a trained codebook instead of Δ, since the F-optimal distortion is usually unknown, and

$$\begin{aligned} E\{\Delta\} &= E\{D(C_X, \overline{F_m^{\omega'}})\} - \inf_{C\in C_n} D(C,F) \\ &\le E\{D(C_X, \overline{F_m^{\omega'}})\} - E\{\inf_{C\in C_n} D(C,F_m)\} \\ &= E\{\Delta_1\}, \end{aligned} \tag{25}$$

where C_X^ω is an F_m-optimal codebook. Note that, from the *Strong Law of Large Numbers*[3], $D(B,F_m) \to D(B,F)$ almost surely (a.s.), as $m\to\infty$, for any codebook B. Therefore, the TSD and VSD converge to the same quantity $D(B,F)$ a.s. if C^ω is replaced by a fixed codebook B. In other words, the difference Δ_1 goes to zero independently of the optimality of B as m gets large; Δ_1 is meaningless in this case. However, since C_X^ω is F_m-optimal from [19, Theorem 1], there exists a sequence of codebooks such that $(D(C_X^\omega, F))$ and $(D(C_X^\omega, F_m^\omega))$ converge to the F-optimal distortion $\inf_{C\in C_n} D(C,F)$. Δ_1 also approaches zero as m gets large under several assumptions. Therefore, Δ_1 is traditionally employed for codebook validation especially in algorithms based on the ERM principle.

Now, we make several approximations based on the asymptotic results in Section III. From (13), we can obtain an approximation $\inf_{C\in C_n} D(C,F) \approx n^{-2/k} J_k \|f\|_\rho$. Therefore, Δ can then have an approximate relation:

$$\begin{aligned} E\{\Delta\} &\approx E\{\mathrm{VSD} - n^{-2/k} J_k \|f\|_\rho\} \\ &\le E\left\{\mathrm{VSD} - \frac{\mathrm{TSD}}{1-\zeta/\beta}\right\}, \end{aligned} \tag{26}$$

for relatively large β, where $\beta > \zeta$. Let Δ_2^ζ denote the new difference distortion given by

$$\Delta_2^\zeta := \mathrm{VSD} - \mathrm{TSD}'. \tag{27}$$

Since $\Delta_2^\zeta < \Delta_1$ $(= \mathrm{VSD} - \mathrm{TSD})$, Δ_2^ζ is tighter than the widely used difference distortion Δ_1. Note that ζ is dependent on the input distribution as shown in (11).

It is convenient to consider the signal to noise ratios (SNRs) of the TSD and VSD. Let $\overline{\mathrm{TSD}}$ and $\overline{\mathrm{VSD}}$ denote the SNRs of the TSD and VSD, respectively, where
$\overline{\mathrm{TSD}} := -10\log(\mathrm{TSD}/\overline{\sigma}^2)$,
$\overline{\mathrm{VSD}} := -10\log(\mathrm{VSD}/\overline{\sigma}^2)$, and $\overline{\sigma}^2$ is the signal variance. In VQ, $\overline{\sigma}^2$ is defined as $\overline{\sigma}^2 := (\det S)^{1/k}$, where S is the auto-covariance matrix of X_1 [5,p.473],[23].

The traditional $\Delta_1 = \mathrm{VSD} - \mathrm{TSD}$ can then be changed to the difference SNR $\overline{\Delta_1}$ defined by

$$\overline{\Delta_1} := \overline{\mathrm{TSD}} - \overline{\mathrm{VSD}} \ \ (\mathrm{dB}). \tag{28}$$

Note that $\overline{\Delta}^2$ is independent of the signal variance. In a similar manner to (26), we can obtain a relation for

$$\overline{\Delta} := 10\log D(C^{\omega^*}, F) - \log[\inf_{C\in C_n} D(C,F)] \tag{29}$$

as

$$\begin{aligned} E\{\overline{\Delta}\} &\approx E\{n^{-2/k} J_k \|f\|_\rho - \overline{\mathrm{VSD}}\} \\ &\le E\{\overline{\Delta_1}\} + 10\log(1-\zeta/\beta), \end{aligned} \tag{30}$$

for $\zeta/\beta < 1$. Hence, $\overline{\Delta_2}^\zeta$, the difference SNR, is defined as

$$\overline{\Delta_2}^\zeta := \overline{\Delta_1} + 10\log(1-\zeta/\beta), \ \ (\mathrm{dB}) \tag{31}$$

where $10\log(1-\zeta/\beta)$ is the compensation term for the biased estimate TSD.

V. Numerical Results

For the numerical simulation, synthetic inputs that represent the uniform distribution and the Gaussian distribution were used.

Note that TSD is a random quantity but, the randomness is reduced as the codebook size increases for arbitrary fixed β. Under several conditions, we can prove that a sequence of the upper bounds for $\inf_{C \in C_n} D(C, F_{m_n})$ [26,(2.3)] converges to its expectation in mean square as $n \to \infty$ [24]. Thus, there exists a subsequence that converges a.s. [4,p.117]. In fact, $n = 2, 4, 8, \cdots$ specifies an example of such a subsequence and the expectation can be the bound in (10). Hence, from (20), we have the approximation

$$\mathrm{TSD} \le n^{-2/k} J_k \|f\|_\rho \left(1 - \frac{\zeta}{\beta}\right), \quad (32)$$

depending on the sample point $\omega \in \Omega$, for large n and β. This fact is demonstrated in Fig. 2, where the TSD curves for three different TSs overlap. Fig. 2 shows that each of the TSD curve from the different TS are very similar. This implies that we can expect similar TSD curves even if TS is changed, i.e., the sample point is changed.

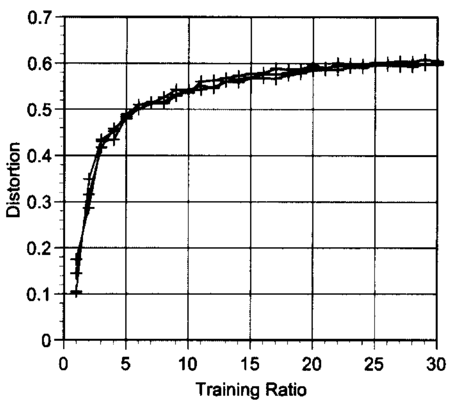


Fig 2. TSD curves for three different TSs (k=16, n=64, and a Gaussian i.i.d. with unit variance).

The first simulation was performed to observe the tightness of the bound in (10) for various types of inputs. Consider a uniform density function $f_U(x)$, which is defined as $f_U(x) := a^{-k}$, if $x \in \times_{i=1}^{k} [0, a]$ and 0, otherwise, where a is a positive constant. For this uniform distribution or its linearly transformed distributions, the bound can be rewritten as

$$\mathrm{TSD} \le n^{-2/k} J_k (12\ \overline{\sigma}^2) \left(1 - \frac{1 - e^{-\beta}}{\beta}\right). \quad (33)$$

Fig. 3 is an example of the uniform i.i.d. input case, where $k = 3$, $n = 512$, and the signal variance $\overline{\sigma}^2 = 1$. We can see that the training ratio β dominates the TSD curve, and the bound curve is very close to the TSD curve. However, for $\beta \ge 15$, the TSD values are greater than the corresponding bounds. This is probably due to the assumption that $\mathrm{TSD} = \min_{C \in C_n} D(C, F_m^{\omega})$ in the codebook design algorithm. In this case, it seems that the trained codebook is close to optimal, since the TSD is very close to $n^{-2/k} J_k\ (12\ \overline{\sigma}^2)$ (≈ 0.0442). However, the codebook training algorithms often get stuck in a local minimum. For illustration, the VSD for the trained codebook at $\beta = 30$ is 0.0513; yet there is a gap between VSD and TSD. Similar simulations were also conducted using Gaussian i.i.d. inputs. In this case, the upper bound is given by

$$\mathrm{TSD} \le n^{-2/k} J_k [2\pi\rho^{-(k+2)/2}\ \overline{\sigma}^2] \left(1 - \frac{\zeta_k^N}{\beta}\right) \quad (34)$$

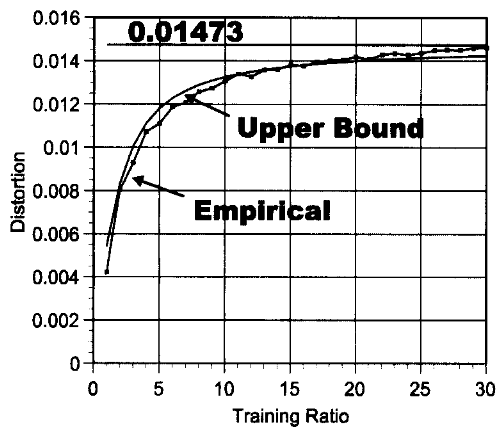


Fig 3. The upper bound in (33) and empirical result versus the training ratio (k=3, n=512, uniform i.i.d. input without correlation inside the vector).

from (31) for relatively large $\beta > \zeta_k^N$. Fig. 4 was obtained for a Gaussian i.i.d. input with $\overline{\sigma}^2 = 1$ and $k = 3$. In this simulation, at $\beta = 5120$, the TSD is 0.081 and is approximately equal to the

VSD. In Figs. 3 and 4, there is no correlation within the random vector X_1. However, further simulation was conducted for a correlated input inside the input vector. Fig. 5 was obtained for a Gauss-Markov input with $k=3$, where the auto-covariance matrix is given by

$$S=\begin{pmatrix} 1 & 0.9 & 0.81 \\ 0.9 & 1 & 0.9 \\ 0.81 & 0.9 & 1 \end{pmatrix}, \tag{35}$$

hence $\overline{\sigma}^2=(\det S)^{1/k}=0.330\cdots$. This Gauss-Markov input is generated using a method based on conditional distributions[11]. The constant ζ is independent of the variance of the input or the correlation inside the vector[26]. Thus, Fig. 5 has the same $\zeta=\zeta_k^N$ that of Fig. 4.

Fig. 6 shows a numerical example of TSD′, the compensated TSD in (21). In this case, $k=16, n=64$, the input has a Gaussian distribution, and TSD* is obtained from the TSD when $\beta=5120$. A bias between TSD′ and TSD* remains as is shown in Fig. 5. However, the influence of the training ratio β on TSD′ is reduced significantly.

In Fig. 7, the new difference distortion Δ_2^ζ is compared with Δ_1. Fig. 6(a) shows the VSD and TSD curves for a Gaussian i.i.d. input when $k=16$ ($\zeta_{16}^N=1.134\cdots$). As β increases, VSD and TSD converge to a distortion. Note that when $\beta=5120$, TSD is 10.82 and VSD is 10.85. This convergence behavior is more clear in Fig. 6(b) when we observe Δ_1. However, the proposed difference distortion Δ_2^ζ is a better bound than Δ_1. In the new difference distortion, the biased quantity TSD is compensated for by $(1-\zeta/\beta)$. Thus, Δ_2^ζ can ensure a more accurate estimation of the trained codebook performance and the TS size effect than Δ_1 does. For the case of $\beta=10$ in Fig. 7(b), Δ_1 is 2.67, which is an upper bound for Δ. However, the proposed Δ_2^ζ is 1.57. In this case, the compensated amount is 0.52~dB from the difference SNR in (30). In Fig. 8, a similar simulation is performed for the $k=4$ case ($\zeta_4^N=1.777\cdots$). In this case the compensated SNR is 0.85 dB. If we do not know the type of the input distribution, then we can simply use

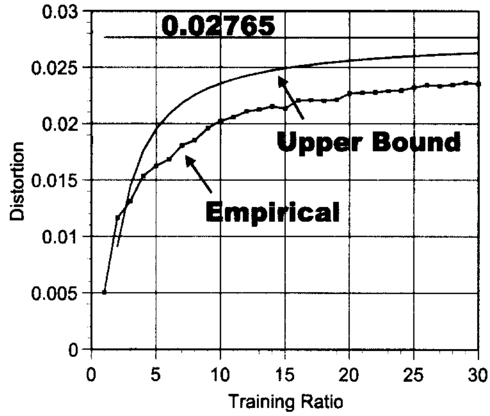


Fig 4. The upper bound and empirical result versus the training ratio (k=3, n=512, Gaussian i.i.d. input without correlation inside the input vector).

$$\Delta_2^1=\text{VSD}-\frac{\text{TSD}}{(1-\beta^{-1})}, \tag{36}$$

since $\Delta_2^1\le\Delta_2^1<\Delta_1$.

Ⅵ. Conclusion

In this paper, the bias in TSD is analyzed and asymptotic bounds for TSD are observed. A compensated estimate

$$\text{TSD}/\left(1-\frac{\zeta}{\beta}\right) \tag{37}$$

is proposed. This estimate can significantly alleviate the bias of the TSD and is convenient especially when a VS is unavailable. Furthermore, a new and simple difference distortion, $\text{VSD}-\text{TSD}/(1-\zeta/\beta)$, is proposed for the validation of the trained codebook when VS is available. As a result, we can avoid a wrong estimate due to the biased TSD which is used in the traditional difference distortion VSD − TSD without any compensation. The constant ζ depends on the input distribution. However, we can simply set $\zeta=1$ for unknown input distributions.

Appendix A

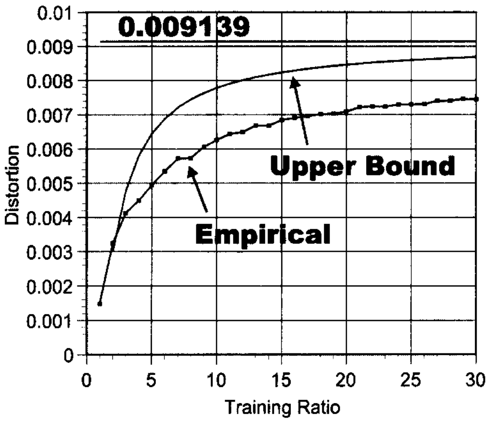


Fig 5. The upper bound and empirical results versus the training ratio (k=3, n=512, Gauss-Markov with correlation coefficient = 0.9 inside the input vector).

Consider a codebook $B=\{y_1,\cdots,y_n\}\in C_n$ and the corresponding Voronoi partition $\{S_i\}_{i=1}^{n}$. Then $D(B,F)$, the quantizer distortion for B, can be rewritten as

$$D(B,F)=\sum_{i=1}^{n}\int_{S_i}\|x-y_i\|^2dF(x), \qquad \text{(A1)}$$

for a distribution function F. If we consider a vector W_i^{ω} that is defined as

$$W_i^{\omega}=\frac{1}{m_i^{\omega}}\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})X_l^{\omega}, \qquad \text{(A2)}$$

if $m_i^{\omega}\neq 0$, and $W_i^{\omega}=(0,\cdots,0)$, otherwise, where $m_i^{\omega}:=\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})$, for each S_i, then it is clear that

$$\begin{aligned}\inf_{C\in C_n} D(C,F_m^{\omega}) &= \inf_{C\in C_n}\frac{1}{m}\sum_{l=1}^{m}\min_{y\in C}\|X_l^{\omega}-y\|^2 \qquad \text{(A3)}\\ &\leq \frac{1}{m}\sum_{i=1}^{n}\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})\|X_l^{\omega}-W_i^{\omega}\|^2,\end{aligned}$$

where $F_m^{\omega}(x):=m^{-1}\sum_{l=1}^{m}I_{(-\infty,x]}\|X_l^{\omega}-W_i^{\omega}\|^2$ and $-\infty:=(-\infty,\cdots,-\infty)$. From (A2), the last term in (A3) can be expanded as follows.

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^{n}\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})\|X_l^{\omega}-W_i^{\omega}\|^2\\ &= \frac{1}{m}\sum_{i=1}^{n}\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})\|(X_l^{\omega}-y_i)+(y_i-W_i^{\omega})\|^2\\ &= \frac{1}{m}\sum_{i=1}^{n}\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})\|X_l^{\omega}-y_i\|^2\\ &\quad -\frac{1}{m}\sum_{i=1}^{n}\sum_{l=1}^{m}m_i^{\omega}\|W_i^{\omega}-y_i\|^2.\end{aligned} \qquad \text{(A4)}$$

In (A4), if $m_i^{\omega}\neq 0$, then

$$\begin{aligned}m_i^{\omega}\|W_i^{\omega}-y_i\|^2 &= \frac{1}{m_i^{\omega}}\|\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})(X_l^{\omega}-y_i)\|^2\\ &= \frac{1}{m_i^{\omega}}\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})\|X_l^{\omega}-y_i\|^2+\frac{1}{m_i^{\omega}}a\end{aligned} \qquad \text{(A5)}$$

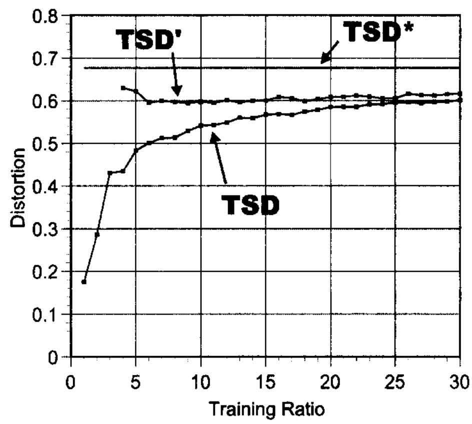


Fig 6. An example of compensated TSD (TSD': compensated TSD and TSD*: TSD at the training ratio 5120).

where

$$\begin{aligned}a_i^{\omega} := &\ 2I_{S_i}(X_1^{\omega})I_{S_i}(X_2^{\omega})(X_1^{\omega}-y_i)(X_2^{\omega}-y_i)+\\ &\cdots+2I_{S_i}(X_{m-1}^{\omega})I_{S_i}(X_m^{\omega})(X_{m-1}^{\omega}-y_i)(X_m^{\omega}-y_i),\end{aligned} \qquad \text{(A6)}$$

and $m_i^{\omega}\|W_i^{\omega}-y_i\|^2=0$ otherwise. Hence, by changing m_i^{ω} to m in the last term of (A5) and from (A4), we obtain a relation

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^{m}\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})\|X_l^{\omega}-y_i\|^2-\frac{1}{m}\sum_{i=1}^{n}m_i^{\omega}\|W_i^{\omega}-y_i\|^2\\ &\leq\frac{1}{m}\sum_{i=1}^{n}\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})\|X_l^{\omega}-y_i^2\|\\ &\quad -\Big[\frac{1}{m^2}\sum_{i=1}^{n}\sum_{l=1}^{m}I_{S_i}(X_l^{\omega})\|X_l^{\omega}-y_i\|^2+1\ m^2\sum_{i=1}^{n}a_i^{\omega}\Big].\end{aligned} \qquad \text{(A7)}$$

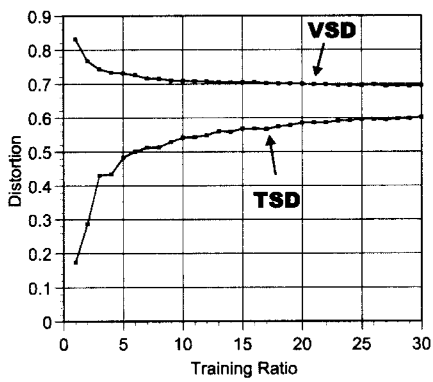
Note that $E\left\{\sum_{i=1}^{n}\alpha_i\right\}\ge 0$, where the equality holds when $y_i=\int_{S_i} x\,dF(x)/\int_{S_i} dF$, and

$$E\left\{\frac{1}{m}\sum_{i=1}^{m} I_{S_i}(X_i^{\omega})\|X_i^{\omega}-y_i\|^2\right\}=D(B,F). \qquad \text{(A8)}$$

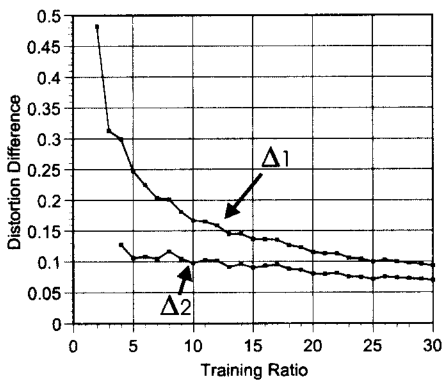
Therefore, from (A3), (A4), and (A7), we obtain (7).

Acknowledgment

The author would like to thank Prof. M.R. Bell and P. Salama of Purdue University, IN, USA, for their helpful comments.

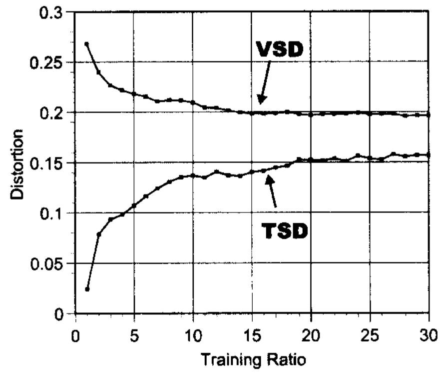


(a)

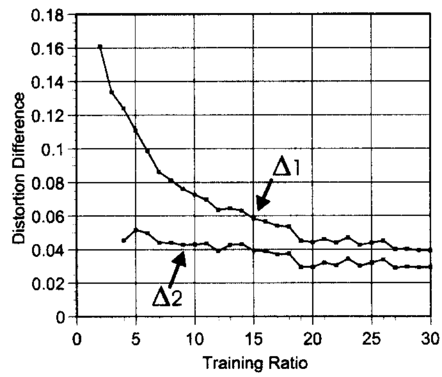


(b)

Fig 7. TSD, VSD, and the proposed difference distortion versus the training ratio (Gaussian i.i.d. with unit variance, k=16, and n=64). (a) TSD and VSD. (b) Δ1 = VSD-TSD and Δ2 = TSD/(1 - 1.134/β).



(a)



(b)

Fig 8. TSD, VSD, and the proposed difference distortion versus the training ratio (Gaussian i.i.d. with unit variance, k=4, and n=64). (a) TSD and VSD. (b) Δ1 =VSD-TSD and Δ2 = VSD-TSD/(1-1.778/β).

References

[1] M.R. Anderberg, *Clustering Analysis for Applications*. New York: Academic, 1973.

[2] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons, 1973.

[3] P.R. Halmos, *Measure Theory*. New York: Springer-Verlag, 1974.

[4] M. Loève, *Probability Theory I*. New York: Springer-Verlag, 1977, 4th. ed.

[5] A.J. Viterbi and J.K. Omura, *Principles of Digital Communication and Coding*. McGraw

Hill, 1979.

[6] V. Vapnik, *Estimation of Dependences Based on Empirical Data*. New York: Springer-Verlag, 1982.

[7] R. Hecht-Nielsen, *Neurocomputing*. New York: Addison-Wesley, 1990.

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

[9] L.R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. New Jersey: Prentice Hall, 1993.

[10] P. Billingsley, *Probability and Measure*. New York: John Wiley \& Sons, 1995, 3rd. ed.

[11] E.M. Scheuer and D.S. Stoller, "On the generation of normal random vectors," *Technometrics*, vol.4, pp. 278-281, 1962.

[12] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 4, pp. 373-380, July 1979.

[13] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT-28, no.2, pp.139-149, March 1982.

[14] J.H. Conway and N.J.A. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inform. Theory*, vol. IT-28, no.2, pp.211-226, March 1982.

[15] J.A. Bucklew and G.L. Wise, "Multidimensional asymptotic quantization theory with *r*th power distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-28, no.2, pp.239-247, March 1982.

[16] E.F. Abaya and G.L. Wise, "On the existence of optimal quantizers," *IEEE Trans. Inform. Theory*, vol. IT-28, no.6, pp.937-940, Nov. 1982.

[17] E.S. Barnes and N.J.A. Sloane, "The optimal lattice quantizer in three dimensions," *SIAM J. Alg. Discrete Methods*, vol.4, no.1, pp.30-41, March 1983.

[18] J.H. Conway and N.J.A. Sloane, "On the voronoi regions of certain lattices," *SIAM J. Alg. Discrete Methods*, vol.5, no.3, pp.294-305, Sep. 1984.

[19] E.A. Abaya and G.L. Wise, "Convergence of vector quantizers with applications to optimal quantization," *SIAM J. Appl. Math.*, vol.44, no.1, pp.183-189, Feb. 1984.

[20] J.H. Conway and N.J.A. Sloane, "A lower bound on the average error of vector quantizers," *IEEE Trans. Inform. Theory*, vol. IT-31, no.1, pp.106-109, Jan. 1985.

[21] D.L. Neuhoff and R. García-Muñoz, "Robust source coding of weakly compact classes," *IEEE Trans. Inform. Theory*, vol. IT-33, pp.522-529, July. 1987.

[22] T. Kohonen, "The self-organizing map," in *Proc. IEEE*, vol. 78, no. 9, pp. 1464-1480, 1990.

[23] D.S. Kim and S.U. Lee, "Image vector quantizer based on a classification in the DCT domain," *IEEE Trans. Commun.*, vol. COM-39, no.4, pp.549-556, April 1991.

[24] D.S. Kim, "Performance of vector quantizers and its relation to training-set size," Ph.D. dissertation, Dep. of Control and Instrumentation, Seoul National Univ. Seoul, Korea, Feb. 1994.

[25] D.S. Kim, T. Kim, and S.U. Lee, "On testing trained vector quantizer codebooks," *IEEE Trans. Image Processing*, vol. IP-6, no.3, pp.398-406, March 1997.

[26] D.S. Kim and M.R. Bell, "Bounds for the trained vector quantizer distortion measured in training data," Tech. Rep. TR-ECE 98-6, Purdue University, April 1998.

[27] D.S. Kim and N.B. Shroff, "Quantization based on a novel sample-adaptive product quantizer (SAPQ)," accepted as a regular paper in *IEEE Trans. Inform. Theory*, 1999.

김 동 식(Dong Sik Kim) 정회원

1986년 2월 : 서울대학교 제어 계측 공학과 졸업.
1988년 2월 : 서울대학교 제어 계측 공학과 석사.
1994년 2월 : 서울대학교 제어 계측 공학과 박사.
1986년부터~현재 : (주)오토만의 이사.
1996년 10월~1998년 5월 : 미국 Purdue 대학의 연구원
1998년 6월~현재 : 연구 조교수
<주관심 분야> 부호 이론, 양자화 이론, 영상처리, 영상 후처리, 영상 및 비디오 코딩, network 및 wireless 환경에서의 영상 전송. microprocessor 회로 및 프로그램, RF 회로 설계.