Asymptotic Performance of Vector Quantizers with a Perceptual Distortion Measure

Jia Li, Navin Chaddha, and Robert M. Gray, *Fellow, IEEE*

***Abstract*—Gersho's bounds on the asymptotic performance of vector quantizers are valid for vector distortions which are powers of the Euclidean norm. Yamada, Tazaki, and Gray generalized the results to distortion measures that are increasing functions of the norm of their argument. In both cases, the distortion is uniquely determined by the vector quantization error, i.e., the Euclidean difference between the original vector and the codeword into which it is quantized. We generalize these asymptotic bounds to input-weighted quadratic distortion measures and measures that are approximately output-weighted-quadratic when the distortion is small, a class of distortion measures often claimed to be perceptually meaningful. An approximation of the asymptotic distortion based on Gersho's conjecture is derived as well. We also consider the problem of source mismatch, where the quantizer is designed using a probability density different from the true source density. The resulting asymptotic performance in terms of distortion increase in decibels is shown to be linear in the relative entropy between the true and estimated probability densities.**

***Index Terms*—Asymptotic bounds, perceptual distortion, relative entropy, source coding, source mismatch.**

I. Introduction

In image processing, mean-squared error is the most commonly used distortion measure for evaluating the performance of compression algorithms because of rich theory and its mathematical tractability. In particular, for quantization or source coding it is simpler to design good encoders and decoders and faster to run them using the mean-squared error distortion. It has often been shown, however, that mean-squared error does not reflect subjective (human) dissatisfaction well [1], [2]. A fair amount of work has been done for developing objective distortion measures which are in accord with the evaluation of human observers [1]–[4]. With the rapid increase of image distribution over communication systems, e.g., the World Wide Web, the standards for image quality become more demanding. Therefore, perceptual distortion measures are receiving greater attention as another dimension of freedom for improving image compression systems. Data compression algorithms aimed at minimizing perceptually meaningful distortion measures are studied [1], [4]. Although perceptual distortion measures result in higher complexity for compression systems, algorithmic speed is less of an issue since the encoding is often off-line and computing power is improving steadily.

Since good perceptual distortion measures take the properties of the human vision system, which is generally nonlinear, into account, they cannot be accurately modeled by simple difference distortion measures. It is well know that human eyes are far more sensitive to changes in a particular range of intensity. When an area is too bright or too dark, the eyes will not notice intensity variations as much as in a grey area. To reflect such a fact, a perceptual distortion measure should be higher for the same quantization error if the original intensity value is in the sensitive range of the eyes. Other factors, such as spatial frequency sensitivity and color sensitivity, play roles in perceptual distortion measures as well. We propose a generalization of difference distortion measures which includes as special cases several nondifference distortion measures proposed as perceptually meaningful measures. Instead of constraining a measure to a difference distortion measure, we only require the measure to be approximately quadratic when distortion is small. This type of distortion measure includes the input-weighted quadratic distortion measure and the measure defined by Nill [4], which is evaluated in the cosine domain incorporating a model of the human vision system.

This paper is concerned with the approximations and bounds in the case of asymptotic quantization, i.e., high-rate quantization, when perceptually motivated distortion measures are used. Although in many practical cases, the quantization rate is far from the high rate required in the asymptotic analysis, the results can be useful for providing benchmarks for comparison and insight into quantizer design. Consequently, high-rate quantization theory has been developed extensively. For the continuity of the theory, we give a brief review of some of the important work in the field (see also Gray and Neuhoff [5]). The pioneering work of Bennett [6] analyzed the asymptotic performance of a scalar quantizer. Zador [7] and Elias [8] extended the bounds on average distortion to powers of the Euclidean norm for vector quantizers. Gersho [9] made a widely accepted conjecture about optimal quantizers, which yields widely used approximations and bounds for the asymptotic L_p distortion. A more direct extension of Bennett's result to vector quantizer was provided by Na and Neuhoff [10]. By introducing the concept of inertial profile, a form of the Bennett integral for vector quantizers was proven to be the limit of the asymptotic distortion under moderate constraints.

Manuscript received May 1, 1997; revised September 1, 1998. This work was supported in part by the National Science Foundation under Grant NSF MIP-9016974. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Ulm, Germany, June 1997.
J. Li and R. M. Gray are with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, CA 94305 USA (e-mail: jiali@isl.stanford.edu; rmgray@stanford.edu).

N. Chaddha is with Microsoft Corporation, Redmond, WA 98052 USA (e-mail: navinc@microsoft.com)
Communicated by R. Laroia, Associate Editor for Source Coding.

Publisher Item Identifier S 0018-9448(99)03552-X.

The vector form of the Bennett integral was also shown to be a good guide for viewing the cause of performance loss in the case of structured vector quantization. Yamada *et al.* [11] generalized the lower bounds of Gersho [9] to difference distortion measures that are increasing functions of the norm of their argument. Gardner and Rao [12] extended the fixed-rate coding results in [11] to a larger class of distortion measures $d(\boldsymbol{x}, \boldsymbol{y})$, defined as nonnegative functions with continuous derivatives, which they argued well modeled perceptual speech distortion. We use a distortion measure $d(\boldsymbol{x}, \boldsymbol{y})$ similar to Gardner and Rao's, but with more complete regularity constraints to permit more formal analysis. Standard asymptotic quantization analysis methods are applied to prove both fixed-rate and variable-rate performance bounds, extending the results of Yamada *et al.* [11] to our version of the distortion introduced by Gardner and Rao [12]. We also apply a variable-rate coding result to several popular perceptual distortion measures. A final issue of theoretical and practical importance in quantization is the loss of performance when the statistics of the source are not accurately known. In the last section, an asymptotic relation is derived which characterizes the performance loss due to source mismatch in terms of the relative entropy between the true source distribution and the estimated one.

In Section II, we provide preliminaries in which basic notation and prerequisite results are introduced. In Section III, Gardner and Rao's [12] bounds on asymptotic average distortion of fixed-rate codes are reviewed and a formal proof is provided. The approximation of the asymptotic distortion is also derived based on Gersho's conjecture [9]. The results are extended to variable-rate coding in Section IV. In Section V, the variable-rate coding results are applied to two examples of perceptual distortion measures. The issue of source mismatch is addressed in Section VI. The technique used in deriving the bounds in Sections III and IV is similar to that used by Yamada *et al.* [11] and Na and Neuhoff [10]. Hence the notation parallels that of [10] and [11] to facilitate reference.

We note that the results generalizing the Bennett integral to input-dependent quadratic distortion measures complement and are consistent with recent results for the same distortion measure by Linder and Zamir [13] on Shannon lower bounds to the rate-distortion function (which provide an approximation to the rate-distortion function for asymptotically small distortion, corresponding to our asymptotically high rate), and Linder, Zamir, and Zeger [14] on multidimensional companding with lattice codes for similar distortion measures.

II. Preliminaries

Let $\boldsymbol{X}$ be a k-dimensional random vector taking sample values $\boldsymbol{x}$ as described by a joint probability density function $p(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, \cdots, x_k) \in \Re^k$, k-dimensional Euclidean space. Suppose the support set of $p(\boldsymbol{x})$ is G. A k-dimensional vector quantizer Q is described by a collection of N reproduction vectors $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N \in \Re^k$, called the reproduction alphabet, and a partition $S_1, \cdots, S_N$ of $\Re^k$. The quantizer Q is defined by

$$Q(\boldsymbol{x}) = \boldsymbol{y}_i, \qquad \text{if } \boldsymbol{x} \in S_i.$$

The distortion between two vectors is generally denoted by $d(\boldsymbol{x}, \boldsymbol{y})$.

For a vector (i.e., a block space of pixels or a subimage) based perceptual distortion measure, the average distortion of an image is the mean of the distortions contributed by all of the blocks in the image. The general form of the distortion for every vector is a nonnegative function $d(\boldsymbol{x}, \boldsymbol{y})$, zero if and only if $\boldsymbol{x} = \boldsymbol{y}$, where $\boldsymbol{x}$ is the original vector and $\boldsymbol{y}$ is the quantized vector. For the derivation, we require regularity constraints on $d(\boldsymbol{x}, \boldsymbol{y})$, which are listed below.

1) $d(\boldsymbol{x}, \boldsymbol{y})$ has continuous partial derivatives of third order almost everywhere.
2) The matrix $B(\boldsymbol{y})$ defined as a $k \times k$-dimensional matrix with the j, nth element

$$B_{j,n}(\boldsymbol{y}) = \frac{1}{2} \left. \frac{\partial^2 d(\boldsymbol{x}, \boldsymbol{y})}{\partial x_j \, \partial x_n} \right|_{\boldsymbol{x}=\boldsymbol{y}} \tag{1}$$

is positive definite almost everywhere.

Actually, combined with condition 1), $d(\boldsymbol{x}, \boldsymbol{y})$ being nonnegative and being zero if and only if $\boldsymbol{x} = \boldsymbol{y}$ implies that $B(\boldsymbol{y})$ is semipositive-definite. Gardner *et al.* [12] used a similar $d(\boldsymbol{x}, \boldsymbol{y})$ to model a perceptual distortion measure for speech. The matrix $B(\boldsymbol{y})$ is the "sensitivity matrix" in [12]. Here, we clarify the constraints on $d(\boldsymbol{x}, \boldsymbol{y})$. In [12], it is not explicitly stated that $B(\boldsymbol{y})$ must be positive-definite and the order of continuous partial derivatives of $d(\boldsymbol{x}, \boldsymbol{y})$ must be three.

The requirement that $B(\boldsymbol{y})$ be positive-definite may appear to be a very strong constraint because it means that the distortion is dominated by the quadratic terms when $\boldsymbol{x}$ and $\boldsymbol{y}$ are very close. Nevertheless, these conditions are usually satisfied by perceptual distortion measures. Examples can be found in Eskicioglu and Fisher [15]. Also Nill's [4] definition of quality measure and the input-weighted quadratic distortion measure analyzed in detail later in this paper satisfy these conditions.

For the high-rate optimum quantizer Q, $\boldsymbol{x}$ and $Q(\boldsymbol{x})$ are close enough so that $d(\boldsymbol{x}, \boldsymbol{y}_i)$, $\boldsymbol{x} \in S_i$ can be approximated by

$$d(\boldsymbol{x}, \boldsymbol{y}_i) \cong (\boldsymbol{x} - \boldsymbol{y}_i)^t B(\boldsymbol{y}_i)(\boldsymbol{x} - \boldsymbol{y}_i). \tag{2}$$

For simplicity of notation, define $B_i = B(\boldsymbol{y}_i)$. The derivation of (2) is given in the Appendix. In the case when $d(\boldsymbol{x}, \boldsymbol{y}_i)$ exactly equals the right-hand side of the above expression, all the results below hold.

A useful quantity is the volume of the unit sphere with respect to the quadratic norm $\|\boldsymbol{x}\|^2_{(i)} = \boldsymbol{x}^t B_i \boldsymbol{x}$, in k-dimensional space, denoted as V_i

$$V_i = V\left(\left\{\boldsymbol{u}\colon \sqrt{\boldsymbol{u}^t B_i \boldsymbol{u}} \le 1\right\}\right).$$

From [11]

$$V_i = [\det(B_i)]^{-(1/2)} C_k \tag{3}$$

where $C_k = (2\Gamma(1/2)^k)/(k\Gamma(k/2))$ is the volume of the k-dimensional unit sphere with respect to the Euclidean norm. Recall the scaling property

$$V(\{\boldsymbol{x}\colon \|\boldsymbol{x} - \boldsymbol{y}_i\|_{(i)} \le a\}) = a^k V_i. \tag{4}$$

III. Bounds for Asymptotically Optimal Performance with Fixed-Rate Coding

We first consider the case of fixed-rate coding, that is, the rate is measured by $\log N$ where N is the number of codewords. A lower bound on the asymptotic average distortion D and the corresponding optimal limiting density function $\lambda(\boldsymbol{x})$ are obtained for the distortion of Section II.

The performance of a quantizer Q is measured by the average distortion

$$\begin{aligned} D &= Ed(\boldsymbol{x}, Q(\boldsymbol{x})) \\ &= \int p(\boldsymbol{x})\, d(\boldsymbol{x}, Q(\boldsymbol{x}))\, d\boldsymbol{x} \\ &= \sum_{i=1}^{N} \int_{S_i} p(\boldsymbol{x})\, d(\boldsymbol{x}, \boldsymbol{y}_i)\, d\boldsymbol{x} \\ &\cong \sum_{i=1}^{N} p(\boldsymbol{y}_i) \int_{S_i} (\boldsymbol{x}-\boldsymbol{y}_i)^t B_i (\boldsymbol{x}-\boldsymbol{y}_i)\, d\boldsymbol{x}. \end{aligned}$$

Just as in all studies of asymptotic quantization, to obtain the above approximation, we assume the probability density $p(\boldsymbol{x})$ is sufficiently "smooth" to ensure that $p(\boldsymbol{x})$ is effectively constant over small bounded sets. It is also assumed that the contribution of overloaded regions to the distortion is negligible when N is large. We define the *normalized moment of inertia* of the cell S_i similarly to [10] as

$$M(S_i) = \frac{\int_{S_i} (\boldsymbol{x}-\boldsymbol{y}_i)^t B_i (\boldsymbol{x}-\boldsymbol{y}_i)\, d\boldsymbol{x}}{V(S_i)^{1+(2/k)}}$$

where $V(S_i)$ denotes the volume of the cell S_i. The approximation of D can thus be written as

$$D \cong \sum_{i=1}^{N} p(\boldsymbol{y}_i) M(S_i) V(S_i)^{1+(2/k)}.$$

Following Gersho [9], define the k-dimensional reproduction vector density by

$$g_N(\boldsymbol{x}) = (NV(S_i))^{-1}, \qquad \text{if } \boldsymbol{x} \in S_i,\ i = 1, 2, \cdots, N.$$

We assume that as $N \to \infty$, there is a limiting density $\lambda(\boldsymbol{x})$ having unit integral, and hence

$$V(S_i) \cong (N\lambda(\boldsymbol{y}_i))^{-1}$$

for every bounded region S_i. We also assume as in [10] that the limit of $M(S_i)$ exists, which is denoted as $m(\boldsymbol{x})$, the inertial profile. When N is sufficiently large, the normalized moment of inertia $M(S_i)$ of the cell S_i which contains the point $\boldsymbol{x}$ is approximately $m(\boldsymbol{x})$. The approximation of D is further expressed as a function of $\lambda(\boldsymbol{x})$ and $m(\boldsymbol{x})$ as

$$\begin{aligned} D &\cong \sum_{i=1}^{N} p(\boldsymbol{y}_i) m(\boldsymbol{y}_i) (N\lambda(\boldsymbol{y}_i))^{-(2/k)} V(S_i) \\ &\cong N^{-(2/k)} \int p(\boldsymbol{x}) \frac{m(\boldsymbol{x})}{\lambda(\boldsymbol{x})^{2/k}}\, d\boldsymbol{x}. \end{aligned} \tag{5}$$

The second step of the approximation above was obtained by recognizing that the left-hand side in the first step is an approximation of the Riemann integral. This is simply the extension of Bennett's integral for vector quantizers in Na and Neuhoff [10] to our distortion measure, the only difference being the definition of $m(\boldsymbol{x})$. For mean-squared error in [10], $M(S_i)$ depends on the shape of the Voronoi cell around $\boldsymbol{y}_i$. In our case, $M(S_i)$ depends on the shape of S_i as well as B_i, which varies with $\boldsymbol{y}_i$. $M(S_i)$ is lower-bounded and the minimum is achieved if S_i is a sphere with respect to the norm $\|\boldsymbol{x}\|_{(i)} = \sqrt{\boldsymbol{x}^t B_i \boldsymbol{x}}$. To show this, we introduce two concepts as in [11]: the effective radius $R(S_i)$ and effective region $T(S_i)$. The effective radius $R(S_i)$ of S_i is the radius of the sphere with the same volume as the region S_i. The corresponding norm is defined as $\|\boldsymbol{x}\|_{(i)}^2 = \boldsymbol{x}^t B_i \boldsymbol{x}$. By the scaling property stated in (4), it is easy to see that

$$R(S_i) = (V(S_i)/V_i)^{1/k}.$$

The effective region $T(S_i)$ of S_i centered at $\boldsymbol{y}_i$ is defined by

$$\begin{aligned} T(S_i) &= \left\{\boldsymbol{x}\colon \sqrt{(\boldsymbol{x}-\boldsymbol{y}_i)^t B_i (\boldsymbol{x}-\boldsymbol{y}_i)} \le R(S_i)\right\} \\ &= \{\boldsymbol{x}\colon \|\boldsymbol{x}-\boldsymbol{y}_i\|_{(i)} \le R(S_i)\}. \end{aligned}$$

$T(S_i)$ is essentially a sphere centered at $\boldsymbol{y}_i$ with radius $R(S_i)$. Obviously,

$$V(T(S_i)) = R(S_i)^k V_i = V(S_i).$$

A crucial inequality [11] is

$$\int_S d(\boldsymbol{x}, \mathrm{y})\, d\boldsymbol{x} \ge \int_{T(S)} d(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{x}$$

where $d(\boldsymbol{x}, \boldsymbol{y})$ is a difference distortion measure. In our special case, $d(\boldsymbol{x}, \boldsymbol{y}_i) = (\boldsymbol{x}-\boldsymbol{y}_i)^t B_i (\boldsymbol{x}-\boldsymbol{y}_i)$, and hence

$$\int_{S_i} (\boldsymbol{x}-\boldsymbol{y}_i)^t B_i (\boldsymbol{x}-\boldsymbol{y}_i)\, d\boldsymbol{x} \ge \int_{T(S_i)} (\boldsymbol{x}-\boldsymbol{y}_i)^t B_i (\boldsymbol{x}-\boldsymbol{y}_i)\, d\boldsymbol{x}$$

equality holds if and only if $T(S_i) = S_i$. This inequality is equivalent to the statement that $M(S_i)$ reaches its lower bound when S_i is a sphere under the norm $\|\boldsymbol{x}\|_{(i)} = \sqrt{\boldsymbol{x}^t B_i \boldsymbol{x}}$. The minimum normalized moment of inertia is thus

$$\begin{aligned} M(T(S_i)) &= \frac{\int_{T(S_i)} (\boldsymbol{x}-\boldsymbol{y}_i)^t B_i (\boldsymbol{x}-\boldsymbol{y}_i)\, d\boldsymbol{x}}{(R(S_i)^k V_i)^{1+(2/k)}} \\ &= \frac{k}{k+2} C_k^{-(2/k)} (\det(B_i))^{1/k}. \end{aligned}$$

Correspondingly, the inertial profile satisfies

$$m(\boldsymbol{x}) \ge \frac{k}{k+2} C_k^{-(2/k)} (\det(B(\boldsymbol{x})))^{1/k}.$$

As a result, a lower bound for D is given by

$$\begin{aligned} D \ge{}& \frac{k}{k+2} C_k^{-(2/k)} N^{-(2/k)} \\ &\cdot \int p(\boldsymbol{x}) (\det(B(\boldsymbol{x})))^{1/k} \lambda(\boldsymbol{x})^{-(2/k)}\, d\boldsymbol{x}. \end{aligned}$$

The right-hand side of the above inequality is minimized by properly choosing $\lambda(\boldsymbol{x})$. Since $\int \lambda(\boldsymbol{x})\, d\boldsymbol{x} = 1$, the above inequality for D is equivalent to

$$\begin{aligned} D \geq & \frac{k}{k+2} C_k^{-(2/k)} N^{-(2/k)} \\ & \cdot \left[\int p(\boldsymbol{x}) (\det(B(\boldsymbol{x})))^{1/k} \lambda(\boldsymbol{x})^{-(2/k)} \, d\boldsymbol{x} \right] \\ & \cdot \left[\int \lambda(\boldsymbol{x}) \, d\boldsymbol{x} \right]^{2/k} \\ \geq & \frac{k}{k+2} C_k^{-(2/k)} N^{-(2/k)} \\ & \cdot \left\{ \int \left[p(\boldsymbol{x}) (\det(B(\boldsymbol{x})))^{1/k} \right]^{k/(k+2)} d\boldsymbol{x} \right\}^{(k+2)/k}. \end{aligned}$$

The last inequality follows from Hölder's inequality and the equality is achieved if and only if

$$\lambda(\boldsymbol{x}) \propto \left[p(\boldsymbol{x}) (\det(B(\boldsymbol{x})))^{1/k} \right]^{k/(k+2)}.$$

Subject to the unit integral constraint,

$$\lambda_{\text{opt}}(\boldsymbol{x}) = \frac{\left[p(\boldsymbol{x}) (\det(B(\boldsymbol{x})))^{1/k} \right]^{k/(k+2)}}{\int \left(p(\boldsymbol{x}') (\det(B(\boldsymbol{x}')))^{1/k} \right)^{k/(k+2)} d\boldsymbol{x}'}. \quad (6)$$

Define $D_L(Q_{\text{opt}})$ as the lower bound to D with the optimal $\lambda(\boldsymbol{x})$

$$\begin{aligned} D_L(Q_{\text{opt}}) = & \frac{k}{k+2} C_k^{-(2/k)} N^{-(2/k)} \int p(\boldsymbol{x}) \\ & \cdot (\det(B(\boldsymbol{x})))^{1/k} (\lambda_{\text{opt}}(\boldsymbol{x}))^{-(2/k)} \, d\boldsymbol{x}. \end{aligned} \quad (7)$$

Hence, the lower bound for the asymptotic distortion is

$$\begin{aligned} D \geq & D_L(Q_{\text{opt}}) \\ = & \frac{k}{k+2} C_k^{-(2/k)} N^{-(2/k)} \\ & \cdot \left\{ \int \left[p(\boldsymbol{x}) (\det(B(\boldsymbol{x})))^{1/k} \right]^{k/(k+2)} d\boldsymbol{x} \right\}^{(k+2)/k} \\ = & \frac{k}{k+2} C_k^{-(2/k)} N^{-(2/k)} \cdot \left\| p(\boldsymbol{x}) (\det(B(\boldsymbol{x})))^{1/k} \right\|_{k/(k+2)} \end{aligned} \quad (8)$$

where $\|\cdot\|_\alpha$ denotes the L_α norm. If we specialize to the mean-square error (MSE) distortion, $\det(B(\boldsymbol{x})) = 1$. Hence

$$D_L(Q_{\text{opt}}) = k/(k+2) C_k^{-(2/k)} N^{-(2/k)} \cdot \|p(\boldsymbol{x})\|_{k/(k+2)}$$

as proved in Gersho [9].

When the dimension k is high, Gardner and Rao [12] argued that in high-rate quantization, the normalized moment of inertia of S_i approaches that of $T(S_i)$ arbitrarily closely. The reason is that the slopes of the Voronoi regions S_i will approach the relative dimensions of the hyper-ellipsoidal regions $T(S_i)$ although the hyper-ellipsoids cannot be formed into a lattice as required by the quantizers. The error incurred in this approximation was investigated in [16] for spaces up to dimension 10. Consequently, the distortion D of the optimal quantizer is approximately equal to the lower bound, i.e.,

$$\begin{aligned} D(Q_{\text{opt}}) \cong & \frac{k}{k+2} C_k^{-(2/k)} N^{-(2/k)} \\ & \cdot \left\| p(\boldsymbol{x}) (\det(B(\boldsymbol{x})))^{1/k} \right\|_{k/(k+2)}. \end{aligned} \quad (9)$$

When k is not large, in order to estimate $m(\boldsymbol{x})$ for an optimal quantizer, we heuristically proceed as follows. For notational simplicity, we denote the transformation of a region S by matrix B as $B[S]$, which is defined below

$$B[S] = \{\boldsymbol{x}\colon \exists \boldsymbol{y} \in S \text{ st. } \boldsymbol{x} = B\boldsymbol{y}, \, \boldsymbol{x} \in \Re^k\}.$$

Based on Gersho's conjecture for optimal quantizers with mean-squared error, we make the same assumption about the geometry of the partition of an optimal quantizer with distortion measure $(\boldsymbol{x} - Q(\boldsymbol{x}))^t B (\boldsymbol{x} - Q(\boldsymbol{x}))$, where B is a fixed positive-definite matrix. Recall that Gersho [9] postulated that for N sufficiently large, the optimal quantizer for a random vector uniformly distributed on some convex set S will have a partition whose regions are all congruent to some polytope H, with the possible exception of regions touching the boundary of S. The conjecture was proved for $k = 2$ by Fejes Toth [17]. To see that Gersho's conjecture implies its extension to the weighted distortion case, consider the fact

$$\begin{aligned} & \int_S (\boldsymbol{x} - Q(\boldsymbol{x}))^t B (\boldsymbol{x} - Q(\boldsymbol{x})) \, d\boldsymbol{x} \\ & \quad = (\det(B))^{-(1/2)} \int_{B^{1/2}[S]} (\boldsymbol{u} - Q(\boldsymbol{u}))^t (\boldsymbol{u} - Q(\boldsymbol{u})) \, d\boldsymbol{u} \end{aligned}$$

by changing variable $\boldsymbol{u} = B^{1/2}\boldsymbol{x}$ and setting $Q(\boldsymbol{u}) = B^{1/2}Q(\boldsymbol{x})$. Thus the optimal quantization of $\boldsymbol{u}$ leads to the optimal quantization of $\boldsymbol{x}$ and *vice versa*. Since $\boldsymbol{u}$ is uniformly distributed on $B^{1/2}[S]$, which is convex by the convexity of S, the optimal quantizer for $\boldsymbol{u}$ on $B^{1/2}[S]$ has a partition with regions congruent to some polytope H. Consequently, the optimal partition for $\boldsymbol{x}$ in S should have regions congruent to $B^{-(1/2)}[H]$, which is also a polytope since the new region is just a linear transform of H.

We will show that if $H_k^\star$ is the optimal admissible polytope with respect to MSE, i.e., the polytope with minimum normalized moment of inertia which generates a tessellation, the optimal admissible polytope for distortion weighted by B is $B^{-(1/2)}[H_k^\star]$ and the minimum normalized moment of inertia is $(\det(B))^{1/k} I(H_k^\star)$, where $I(H_k^\star)$ is the normalized inertia defined by Gersho [9]

$$I(H) = \frac{\int_H (\boldsymbol{x} - Q(\boldsymbol{x}))^t (\boldsymbol{x} - Q(\boldsymbol{x})) \, d\boldsymbol{x}}{(V(H))^{1+(2/k)}}.$$

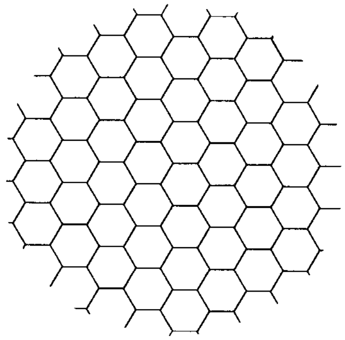
Without loss of generality, we assume $Q(\boldsymbol{x}) = \boldsymbol{0}$, for $\boldsymbol{x} \in H$. By definition

$$M(S) = \frac{\int_S \boldsymbol{x}^t B \boldsymbol{x} \, d\boldsymbol{x}}{V(S)^{1+(2/k)}}$$

changing variable, $\boldsymbol{u} = B^{1/2}\boldsymbol{x}$,

$$\begin{aligned} M(S) & = \frac{(\det(B))^{-(1/2)} \int_{B^{1/2}[S]} \boldsymbol{u}^t \boldsymbol{u} \, d\boldsymbol{u}}{\left((\det(B))^{-(1/2)} V\left(B^{1/2}[S]\right) \right)^{1+(2/k)}} \\ & = (\det(B))^{1/k} \frac{\int_{B^{1/2}[S]} \boldsymbol{u}^t \boldsymbol{u} \, d\boldsymbol{u}}{(V(B^{1/2}[S]))^{1+(2/k)}} \\ & \geq (\det(B))^{1/k} I(H_k^\star) \end{aligned}$$

where equality is achieved if and only if $B^{1/2}[S] = H_k^\star$, equivalent to $S = B^{-(1/2)}[H_k^\star]$. For $k = 2$, an example of a



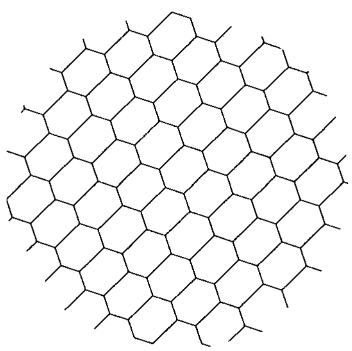


Fig. 1. Tessellation of optimal polytopes. Top: mean-square error distortion measure, the optimal admissible polytope is the regular hexagon, denoted by H. Bottom: weighted quadratic distortion measure with weighting matrix B, the optimal admissible polytope is the transformed polytope of the regular hexagon, i.e., $B^{-(1/2)}[H]$.

tessellation by the optimal admissible polytopes for both mean-squared error and the weighted quadratic error with weighting matrix

$$B = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$$

are shown in Fig. 1. For mean-squared error, the optimal polytope is the regular hexagon. For the weighted quadratic distortion, the optimal polytope is the transformed regular hexagon with transforming matrix

$$B^{-(1/2)} = \begin{pmatrix} -1 & \sqrt{1/3} \\ 1 & \sqrt{1/3} \end{pmatrix}.$$

As shown in the figure, the optimal admissible polytope in the case of weighted quadratic distortion is a nonregular hexagon, a rotated and stretched version of the regular hexagon.

We are now ready to estimate $m(\boldsymbol{x})$. We argue that when the rate is high, the quantization of a small region around $\boldsymbol{x}$ is more and more like the quantization of a uniform density, with distortion measure $(\boldsymbol{x}-Q(\boldsymbol{x}))^t B(\boldsymbol{x})(\boldsymbol{x}-Q(\boldsymbol{x}))$, provided that the probability density function $p(\boldsymbol{x})$, the point density $\lambda(\boldsymbol{x})$, and $B(\boldsymbol{x})$ are sufficiently smooth. Thus the optimal $m(\boldsymbol{x})$ is approximately $(\det(B(\boldsymbol{x})))^{1/k} I(H_k^\star)$.

Substituting $m(\boldsymbol{x})$ to (5), we obtain an approximation for D of the optimal quantizer

$$D \cong N^{-(2/k)} I(H_k^\star) \int p(\boldsymbol{x})(\det(B(\boldsymbol{x})))^{1/k} \lambda(\boldsymbol{x})^{-(2/k)}\, d\boldsymbol{x}. \tag{10}$$

This approximation differs from the lower bound by only a constant ratio. Thus the optimal point density $\lambda(\boldsymbol{x})$ is also

$$\lambda_{\rm opt}(\boldsymbol{x}) = \frac{\left(p(\boldsymbol{x})(\det(B(\boldsymbol{x})))^{1/k}\right)^{k/(k+2)}}{\int \left(p(\boldsymbol{x})(\det(B(\boldsymbol{x})))^{1/k}\right)^{k/(k+2)} d\boldsymbol{x}}$$

and the optimal distortion is

$$D \cong I(H_k^\star) N^{-(2/k)} \cdot \left\| p(\boldsymbol{x})(\det(B(\boldsymbol{x})))^{1/k} \right\|_{k/(k+2)}. \tag{11}$$

The values and bounds for $I(H_k^\star)$ with different dimension k are listed in Gersho [9].

IV. Bounds for Asymptotically Optimal Performance with Variable-Rate Coding

In this section, we consider variable-rate coding. In this case, the rate is approximately the entropy of the encoded source. We constrain the number of reproduction vectors, N, to be finite. As in the previous section, we derive a lower bound on the asymptotic average distortion D and the corresponding optimal limiting density function $\lambda(\boldsymbol{x})$. An approximation of D can be similarly obtained by using the approximation (10) derived before.

To start the derivation, express $D_L(Q_{\rm opt})$ as a function of $\lambda_{\rm opt}(\boldsymbol{x})$. From (7), we obtain

$$D_L(Q_{\rm opt}) = \frac{k}{k+2} C_k^{-(2/k)} \int p(\boldsymbol{x})(\det(B(\boldsymbol{x})))^{1/k} \cdot [N\lambda_{\rm opt}(\boldsymbol{x})]^{-(2/k)}\, d\boldsymbol{x}. \tag{12}$$

For the convenience of derivation, we define

$$f(\boldsymbol{x}) = \frac{(\det(B(\boldsymbol{x})))^{1/k} p(\boldsymbol{x})}{\int (\det(B(\boldsymbol{x}')))^{1/k} p(\boldsymbol{x}')\, d\boldsymbol{x}'}.$$

The density $f(\boldsymbol{x})$ is the original probability density function weighted by $(\det(B(\boldsymbol{x})))^{1/k}$. In the special case of a difference distortion measure, $\det(B(\boldsymbol{x}))$ is a constant and $f(\boldsymbol{x})$ is the same as $p(\boldsymbol{x})$.

Substituting $f(\boldsymbol{x})$ into (12) simplifies the equation to

$$D_L(Q_{\rm opt}) = \frac{k}{k+2} C_k^{-(2/k)} \left[\int (\det(B(\boldsymbol{x})))^{1/k} p(\boldsymbol{x})\, d\boldsymbol{x}\right] \cdot \int [N\lambda_{\rm opt}(\boldsymbol{x})]^{-(2/k)} f(\boldsymbol{x})\, d\boldsymbol{x}.$$

From [11, eq. (28)], the entropy of the encoded source is approximately

$$H_Q \cong h(p) - E\left\{\log \frac{1}{N\lambda(\boldsymbol{x})}\right\}.$$

We take the base of the logarithm as e. Hence, the dimension of H_Q is nats here. We denote $g(\boldsymbol{x}) = N\lambda(\boldsymbol{x})$.

Now, finding $Q_{\rm opt}$ with fixed entropy H_Q becomes a constrained optimization problem

$$\begin{aligned} &\text{given} \quad h(p) - E\left[\log\left(\frac{1}{g(\boldsymbol{x})}\right)\right] \leq \text{constant} \\ &\text{i.e.} \quad \int \log(g(\boldsymbol{x}))p(\boldsymbol{x})\,d\boldsymbol{x} \leq C \\ &\text{where} \quad C \text{ is a constant} \\ &\text{find} \quad \min_{g(\boldsymbol{x})} D_L(Q_{\rm opt}) \\ &\text{i.e.} \quad \min_{g(\boldsymbol{x})} \int g(\boldsymbol{x})^{-(2/k)} f(\boldsymbol{x})\,d\boldsymbol{x}. \end{aligned}$$

As $f(\boldsymbol{x})$ and $p(\boldsymbol{x})$ are fixed functions in the minimization, the constraint

$$\int \log(g(\boldsymbol{x}))p(\boldsymbol{x})\,d\boldsymbol{x} \leq C$$

is equivalent to

$$\int \log(g(\boldsymbol{x})^{-(2/k)}(f(\boldsymbol{x})/p(\boldsymbol{x})))p(\boldsymbol{x})\,d\boldsymbol{x} \geq C'$$

where C' is another constant. By Jensen's inequality

$$\log \int \left(g(\boldsymbol{x})^{-(2/k)}\,\frac{f(\boldsymbol{x})}{p(\boldsymbol{x})}\right)p(\boldsymbol{x})\,d\boldsymbol{x} \geq \int \log\left(g(\boldsymbol{x})^{-(2/k)}\,\frac{f(\boldsymbol{x})}{p(\boldsymbol{x})}\right)p(\boldsymbol{x})\,d\boldsymbol{x}.$$

The above inequality can be rewritten as

$$\begin{aligned} &\int \left(g(\boldsymbol{x})^{-(2/k)}\,\frac{f(\boldsymbol{x})}{p(\boldsymbol{x})}\right)p(\boldsymbol{x})\,d\boldsymbol{x} \\ &= \int g(\boldsymbol{x})^{-(2/k)} f(\boldsymbol{x})\,d\boldsymbol{x} \\ &\geq \exp\left(\int \log\left(g(\boldsymbol{x})^{-(2/k)}\,\frac{f(\boldsymbol{x})}{p(\boldsymbol{x})}\right)p(\boldsymbol{x})\,d\boldsymbol{x}\right) \\ &\geq e^{C'} \end{aligned} \tag{13}$$

where the equality holds if and only if $g(\boldsymbol{x})^{-(2/k)}(f(\boldsymbol{x})/p(\boldsymbol{x}))$ is a constant. With $g(\boldsymbol{x})$ and $f(\boldsymbol{x})$ substituted in, we can conclude that the lower bound is achieved if and only if $\lambda_{\rm opt}(\boldsymbol{x}) \propto (\det(B(\boldsymbol{x})))^{1/2}$, i.e.,

$$\lambda_{\rm opt}(\boldsymbol{x}) = \frac{(\det(B(\boldsymbol{x})))^{1/2}}{\int_{\boldsymbol{x}'\in G}(\det(B(\boldsymbol{x}')))^{1/2}\,d\boldsymbol{x}'}. \tag{14}$$

The above conclusion about the optimal point density shows that $\lambda_{\rm opt}(\boldsymbol{x})$ increases when $\det(B(\boldsymbol{x}))$ increases. This accords with intuition because the area with larger $\det(B(\boldsymbol{x}))$ receives more penalty in terms of distortion for the same quantization error, which consequently requires denser codewords in it to keep the total distortion low. Recall in the case of a difference distortion measure, $\lambda_{\rm opt}(\boldsymbol{x})$ is shown to be constant in [11], which follows here by setting $\det(B(\boldsymbol{x}))$ to a constant.

Finally, substituting $\lambda_{\rm opt}(\boldsymbol{x})$ yields

$$\begin{aligned} H_Q = h(p) + \log N + \tfrac{1}{2}\int \log(\det(B(\boldsymbol{x})))p(\boldsymbol{x})\,d\boldsymbol{x} \\ - \log\left(\int (\det(B(\boldsymbol{x})))^{1/2}\,d\boldsymbol{x}\right) \end{aligned} \tag{15}$$

and

$$\begin{aligned} D_L(Q_{\rm opt}) = \frac{k}{k+2}\,C_k^{-(2/k)} N^{-(2/k)} \\ \cdot \left[\int (\det(B(\boldsymbol{x})))^{1/2}\,d\boldsymbol{x}\right]^{2/k}. \end{aligned} \tag{16}$$

To relate $D_L(Q_{\rm opt})$ directly to H_Q, apply (15) to replace N to obtain the following lower bound for the asymptotic distortion:

$$\begin{aligned} D &\geq D_L(Q_{\rm opt}) \\ &= \frac{k}{k+2}\,C_k^{-(2/k)} \cdot \left[\int (\det(B(\boldsymbol{x})))^{1/2}\,d\boldsymbol{x}\right]^{2/k} \\ &\quad \cdot \exp\left(-\frac{2}{k}\left(H_Q - h(p) + \log\left(\int (\det(B(\boldsymbol{x})))^{1/2}\,d\boldsymbol{x}\right)\right.\right. \\ &\quad \left.\left. - \frac{1}{2}\int \log(\det(B(\boldsymbol{x})))p(\boldsymbol{x})\,d\boldsymbol{x}\right)\right). \end{aligned} \tag{17}$$

Further calculation simplifies (17) to

$$\begin{aligned} D &\geq D_L(Q_{\rm opt}) \\ &= \frac{k}{k+2}\,C_k^{-(2/k)} \cdot \exp\left(-\frac{2}{k}\left(H_Q - h(p) - \frac{1}{2}\right.\right. \\ &\quad \left.\left. \cdot \int \log(\det(B(\boldsymbol{x})))p(\boldsymbol{x})\,d\boldsymbol{x}\right)\right). \end{aligned} \tag{18}$$

The above inequality shows how a nonconstant $\det(B(\boldsymbol{x}))$ affects the lower bound of asymptotic distortion. In the case of MSE distortion measure, $\det(B(\boldsymbol{x}))$ is a constant and (16) can be simplified to the same form as in [9]. An approximation of D can be similarly derived by using the approximation (10) in previous section. The difference is that the constant multiplier is $I(H_k^\star)$ instead of $k/(k+2)C_k^{-(2/k)}$ as in $D_L(Q_{\rm opt})$. The optimal $\lambda(\boldsymbol{x})$ with respect to the approximation is the same as $\lambda_{\rm opt}(\boldsymbol{x})$ derived for the lower bound $D_L(Q_{\rm opt})$.

V. Examples

We give two examples in this section to show how the variable-rate results affect quantization strategy for different perceptual distortion measures. The optimal limiting density functions $\lambda(\boldsymbol{x})$ for the perceptual distortion developed by Nill [4] and the input-weighted squared-error measure [18] are derived. Eskicioglu and Fisher [15] compared many distortion measures with subjective ratings. They found that Nill's definition of the distortion measure correlates with subjective ratings consistently better than mean-squared error, L_1 and L_3 distortion measures for compressed images obtained from several popular quantization algorithms.

Nill [4] defined a distortion measure in the cosine transform domain incorporating a human visual model. The distortion for every vector or subimage is

$$d(\boldsymbol{x}, \hat{\boldsymbol{x}}) = W(\boldsymbol{x}) \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} H^2(r)[F(u, v) - \hat{F}(u, v)]^2$$

where

$H(r)$ rotationally symmetric spatial frequency response of human visual system, $r = \sqrt{u^2 + v^2}$;

F, $\hat{F}$ transform coefficients of original and quantized subimage, respectively;
u, v coordinates in the transform domain;
M, N number of coefficients in orthogonal u, v directions;
$\boldsymbol{x}$ $[F(0,0), \cdots, F(0, N-1), F(1,0), \cdots, F(1, N-1), \cdots, \cdots, F(M-1, N-1)]$;
$\hat{\boldsymbol{x}}$ $[\hat{F}(0,0), \cdots, \hat{F}(0, N-1), \hat{F}(1,0), \cdots, \hat{F}(1, N-1), \cdots, \cdots, \hat{F}(M-1, N-1)]$;
$W(\boldsymbol{x})$ weighting factor proportional to subimage's intensity level variance. It is a quadratic function of x_i's.

We suppose the quantization is done in the transform domain, i.e., the vector sequence $\boldsymbol{x}_i$ to be quantized is the cosine transformed data of the original vector. To calculate $\lambda(\boldsymbol{x}_i)$, we need $\det(B(\boldsymbol{x}))$. By definition (1)

$$B_{j,n}(\boldsymbol{x}) = \frac{1}{2} \frac{\partial^2 d(\boldsymbol{y}, \boldsymbol{x})}{\partial y_j \, \partial y_n}\bigg|_{\boldsymbol{y}=\boldsymbol{x}}$$

when

$$j \neq n, \qquad B_{j,n}(\boldsymbol{x}) = 0$$

when $j = n$, suppose

$$F(u, v) = x_n$$
$$B_{j,n}(\boldsymbol{x}) = H^2(r) W(\boldsymbol{x}), \qquad \text{where } r = \sqrt{u^2 + v^2}.$$

Hence

$$\det(B(\boldsymbol{x})) = \left(\prod_{u=0}^{M-1} \prod_{v=0}^{N-1} H^2\left(\sqrt{u^2+v^2}\right)\right) W(\boldsymbol{x})^{M \cdot N}.$$

As

$$\prod_{u=0}^{M-1} \prod_{v=0}^{N-1} H^2(\sqrt{u^2+v^2})$$

is a constant with respect to $\boldsymbol{x}$

$$\lambda(\boldsymbol{x}) \propto W(\boldsymbol{x})^{(M \cdot N)/2}.$$

Hence according to Nill's definition of distortion, the limiting density of the codewords depends on the sub-image's intensity level variance. It increases as the $(M \cdot N)/2$th power of the intensity variance.

The second example is the input-weighted squared-error measure [18], which is defined as

$$d(\boldsymbol{y}, \boldsymbol{x}) = (\boldsymbol{y} - \boldsymbol{x})^t W(\boldsymbol{y})(\boldsymbol{y} - \boldsymbol{x}).$$

With the requirement that every element of $W(\boldsymbol{x})$ has continuous partial derivatives of third order, it can be shown that

$$B(\boldsymbol{x}) = \tfrac{1}{2}[W(\boldsymbol{x})^t + W(\boldsymbol{x})].$$

Thus the positive definiteness of $W(\boldsymbol{x})$ assures that of $B(\boldsymbol{x})$. The optimal point density satisfies

$$\lambda(\boldsymbol{x}) \propto (\det(W(\boldsymbol{x})^t + W(\boldsymbol{x})))^{1/2}.$$

If $W(\boldsymbol{x})$ is symmetric, then

$$\lambda(\boldsymbol{x}) \propto (\det(W(\boldsymbol{x})))^{1/2}.$$

An example for $W(\boldsymbol{x})$ being symmetric is the Itakura–Saito distortion measure [19], [20] which sets $W(\boldsymbol{x})$ to be an autocorrelation matrix.

VI. Source Mismatch

The previous analysis was based on the assumption that the probability density function $p(\boldsymbol{x})$ of the source is known. However, this is usually not the situation in practice. In real life, $p(\boldsymbol{x})$ must be estimated. For high-rate variable-rate coding, $p(\boldsymbol{x})$ is not required since $\lambda_{\text{opt}}(\boldsymbol{x})$ is unrelated to $p(\boldsymbol{x})$. For fixed-rate coding, $\lambda_{\text{opt}}(\boldsymbol{x})$ does depend on $p(\boldsymbol{x})$ and it is of interest to quantify the possible change in performance due to mismatch. This section addresses this problem.

We constrain our interest to the case that for vector $\boldsymbol{x} = (x_1, \cdots, x_k)$, x_i's are independent and identically distributed (i.i.d.) random variables. We analyze the asymptotic case when $k \to \infty$. The result can be easily generalized to $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k)$ where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{ir})$, $\boldsymbol{x}_i$'s are i.i.d. random vectors. The only reason to consider the scalar case is to simplify the mathematical notation. Since we consider the limit case $k \to \infty$, we must put some constraint on how $\det(B(\boldsymbol{x}))$ changes with k to get a reasonable result. The assumption we make here is that

$$\det(B(\boldsymbol{x})) = \prod_{i=1}^{k} \det(B(\boldsymbol{x}_i))$$

where $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k)$, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{ir})$. This assumption is essentially saying that all the subvectors $\boldsymbol{x}_i$ play an equal and independent role in determining the distortion. In the following derivation, $\boldsymbol{x}_i$ is a scalar, hence $\det(B(x_i)) = B(x_i)$. Consequently, the assumption becomes

$$\det(B(\boldsymbol{x})) = \prod_{i=1}^{k} B(x_i).$$

Recall the optimal $\lambda(\boldsymbol{x})$ of (16). If we estimate $p(\boldsymbol{x})$ by $\hat{p}(\boldsymbol{x})$, the optimal $\lambda(\boldsymbol{x})$ is

$$\hat{\lambda}_{\text{opt}}(\boldsymbol{x}) = \frac{\left[\hat{p}(\boldsymbol{x})(\det(B(\boldsymbol{x})))^{1/k}\right]^{k/(k+2)}}{\displaystyle\int \left[\hat{p}(\boldsymbol{x}')(\det(B(\boldsymbol{x}')))^{1/k}\right]^{k/(k+2)} d\boldsymbol{x}'}.$$

As a result of $\hat{p}(\boldsymbol{x}) = \prod_{i=1}^{k} \hat{p}(x_i)$ and $\det(B(\boldsymbol{x})) = \prod_{i=1}^{k} B(x_i)$

$$\hat{\lambda}_{\text{opt}}(\boldsymbol{x}) = \prod_{i=1}^{k} \hat{\lambda}_{\text{opt}}(x_i)$$

where

$$\hat{\lambda}_{\text{opt}}(x_i) = \frac{\left[\hat{p}(x_i)(B(x_i))^{1/k}\right]^{k/(k+2)}}{\displaystyle\int \left[\hat{p}(x_i)(B(x_i))^{1/k}\right]^{k/(k+2)} dx_i}.$$

We are being sloppy here with the notation $\hat{p}$, B, and $\hat{\lambda}_{\text{opt}}$, but it should be clear from the context whether we mean the one-dimensional form or the vector form.

The asymptotic distortion becomes

$$D(\hat{Q}_{\rm opt}) \cong \hat{D}_L = \frac{k}{k+2} C_k^{-(2/k)} N^{-(2/k)} \cdot \int p(\boldsymbol{x})(\det(B(\boldsymbol{x})))^{1/k} \hat{\lambda}_{\rm opt}(\boldsymbol{x})^{-(2/k)}\, d\boldsymbol{x}.$$

As the true optimal distortion $D(Q_{\rm opt})$ is given by (9), the increase of distortion in decibels is $10\log(D(\hat{Q}_{\rm opt})/D(Q_{\rm opt}))$. We choose decibels as a measure of performance loss because in practice, the signal-to-noise ratio (SNR) or the peak signal-to-noise ratio (PSNR) is common evaluation of performance, and the decrease of SNR or PSNR is equal to the increase of distortion in decibels. With $D(\hat{Q}_{\rm opt})$ and $D(Q_{\rm opt})$ substituted in and some algebra, we get

$$\begin{aligned}\ln\frac{D(\hat{Q}_{\rm opt})}{D(Q_{\rm opt})} &\cong \ln\frac{\hat{D}_L}{D_L(Q_{\rm opt})}\\ &= \ln\left(\frac{\int p(x)B(x)^{1/k}\hat{\lambda}_{\rm opt}(x)^{-(2/k)}\,dx}{\int p(x)B(x)^{1/k}\lambda_{\rm opt}(x)^{-(2/k)}\,dx}\right)^k\\ &= 2\ln\int \hat{p}(x)^{k/(k+2)}B(x)^{1/(k+2)}\,dx\\ &\quad + k\ln\int p(x)\hat{p}(x)^{-(2/(k+2))}B(x)^{1/(k+2)}\,dx\\ &\quad - (k+2)\ln\int p(x)^{k/(k+2)}B(x)^{1/(k+2)}\,dx. \end{aligned} \tag{19}$$

The following limits are proved in the Appendix under assumptions (21)–(24):

$$\begin{aligned}&\lim_{k\to\infty}\ln\int \hat{p}(x)^{k/(k+2)}B(x)^{1/(k+2)}\,dx = 0\\ &\lim_{k\to\infty} k\ln\int p(x)\hat{p}(x)^{-(2/(k+2))}B(x)^{1/(k+2)}\,dx\\ &\qquad = \int p(x)\ln\left(\hat{p}(x)^{-2}B(x)\right)dx\\ &\lim_{k\to\infty} -(k+2)\ln\int p(x)^{k/(k+2)}B(x)^{1/(k+2)}\,dx\\ &\qquad = -\int p(x)\ln\left(p(x)^{-2}B(x)\right)dx. \end{aligned} \tag{20}$$

The following conditions for these limits to hold are developed in the derivation in the Appendix.

1)
$$E[\ln^2(\hat{p}(x)^{-2}B(x))] < \infty \tag{21}$$
with respect to both $\hat{p}$ and p.

2)
$$E_p[\ln^2(p(x)^{-2}B(x))] < \infty. \tag{22}$$
.

3)
$$\exists \epsilon > 0 \quad \text{st.} \quad E\left[(\hat{p}(x)^{-2}B(x))^{\epsilon}\right] < \infty \tag{23}$$
with respect to both $\hat{p}$ and p.

4)
$$E_p\left[(p(x)^{-2}B(x))^{\epsilon}\right] < \infty. \tag{24}$$

In many practical situations, the support of x is a bounded set. If this is the case and if on the support of x, $B(x)$ is a continuous function and $p(x)$ and $\hat{p}(x)$ do not drop to zero too fast, the above four conditions hold.

Back to the limit results, substituting the three terms into (19) yields

$$\begin{aligned}\lim_{k\to\infty}\ln\frac{\hat{D}_L}{D_L(Q_{\rm opt})} &= \int p(x)\ln\left(p(x)^2B(x)^{-1}\right)dx\\ &\quad - \int p(x)\ln\left(\hat{p}(x)^2B(x)^{-1}\right)dx\\ &= 2\int p(x)\ln\frac{p(x)}{\hat{p}(x)}\,dx\\ &= 2\ln 2\int p(x)\log_2\frac{p(x)}{\hat{p}(x)}\,dx\\ &= 2\ln 2 D(p(x)\|\hat{p}(x))\end{aligned}$$

where $D(p(x)\|\hat{p}(x))$ is the relative entropy of distributions $p(x)$ and $\hat{p}(x)$ in bits. Changing the base of logarithm, we finally obtain the loss in decibels when $k\to\infty$

$$10\log\frac{\hat{D}_L}{D_L(Q_{\rm opt})} = 6D(p(x)\|\hat{p}(x)). \tag{25}$$

It is interesting to notice that the limit loss is independent of $B(x)$, which means that the effect of $B(x)$ is washed out as $k\to\infty$. Equation (25) also shows the linear dependence of the performance loss on the relative entropy of the true distribution and the estimated one. This relation demonstrates again that the relative entropy is an effective measure of the closeness of two probability density functions. In this case, every 1-bit difference in the relative entropy indicates a 6-dB loss in performance asymptotically.

APPENDIX

We prove in the appendix the approximation relation stated in (2) and the three limits in (20).

In (2), we claimed that with high-rate optimum quantizer Q, the following approximation holds:

$$d(\boldsymbol{x},\boldsymbol{y}_i) \cong (\boldsymbol{x}-\boldsymbol{y}_i)^tB(\boldsymbol{y}_i)(\boldsymbol{x}-\boldsymbol{y}_i)$$

where $\boldsymbol{y}_i$ in our consideration here is a constant vector. Readers need to keep this in mind to avoid confusion. Since $d(\boldsymbol{x},\boldsymbol{y}_i)$ has continuous partials of third order, Taylor's theorem [21] implies that

$$\begin{aligned}d(\boldsymbol{x},\boldsymbol{y}_i) &= \nabla d(\boldsymbol{x},\boldsymbol{y}_i)|_{\boldsymbol{x}=\boldsymbol{y}_i}\cdot(\boldsymbol{x}-\boldsymbol{y}_i)\\ &\quad + (\boldsymbol{x}-\boldsymbol{y}_i)^tB(\boldsymbol{y}_i)(\boldsymbol{x}-\boldsymbol{y}_i) + R_2(\boldsymbol{x}-\boldsymbol{y}_i,\boldsymbol{y}_i).\end{aligned}$$

We discuss the three terms on the right-hand side of the above equation one by one. First, $\nabla d(\boldsymbol{x},\boldsymbol{y}_i)|_{\boldsymbol{x}=\boldsymbol{y}_i}$ is the gradient of $d(\boldsymbol{x},\boldsymbol{y}_i)$ calculated at $\boldsymbol{y}_i$. It is equivalent to

$$\nabla d(\boldsymbol{x},\boldsymbol{y}_i)|_{\boldsymbol{x}=\boldsymbol{y}_i} = \left.\left(\frac{\partial L}{\partial x_1},\frac{\partial L}{\partial x_2},\cdots,\frac{\partial L}{\partial x_k}\right)\right|_{\boldsymbol{x}=\boldsymbol{y}_i}.$$

As $d(\boldsymbol{x}, \boldsymbol{y}_i) \geq 0$ with equality if and only if $\boldsymbol{x} = \boldsymbol{y}_i$, $\boldsymbol{x} = \boldsymbol{y}_i$ is a local extremum. From the local extremum theorem [21], $\nabla d(\boldsymbol{x}, \boldsymbol{y}_i)|_{\boldsymbol{x}=\boldsymbol{y}_i} = 0$.

In the second term, $B(\boldsymbol{y}_i)$ is the square matrix of second-order partials of $d(\boldsymbol{x}, \boldsymbol{y}_i)$ multiplied by $\frac{1}{2}$. To write it explicitly, for the j, nth element of $B(\boldsymbol{y_i})$

$$B_{j,n}(\boldsymbol{y}_i) = \frac{1}{2} \frac{\partial^2 d(\boldsymbol{x}, \boldsymbol{y}_i)}{\partial x_j \, \partial x_n}\bigg|_{\boldsymbol{x}=\boldsymbol{y}_i}.$$

We assume that $B(\boldsymbol{y}_i)$ is positive-definite. As a result,

$$(\boldsymbol{x} - \boldsymbol{y}_i)^t B(\boldsymbol{y}_i)(\boldsymbol{x} - \boldsymbol{y}_i) \geq \mu_{\min} \|\boldsymbol{x} - \boldsymbol{y}_i\|^2$$

where $\mu_{\min}$ is the minimum eigenvalue of $B(\boldsymbol{y}_i)$.

The third term $R_2(\boldsymbol{x} - \boldsymbol{y}_i, \boldsymbol{y}_i)$ is the remainder which satisfies $R_2(\boldsymbol{x} - \boldsymbol{y}_i, \boldsymbol{y}_i)/\|\boldsymbol{x} - \boldsymbol{y}_i\|^2 \to 0$ as $\|\boldsymbol{x} - \boldsymbol{y}_i\| \to 0$, where the norm is a Euclidean norm. In the case of the high-rate optimum quantization, for bounded cells S_i, $\|\boldsymbol{x} - \boldsymbol{y}_i\|$ is sufficiently small so that $R_2(\boldsymbol{x} - \boldsymbol{y}_i, \boldsymbol{y}_i)$ is negligible compared to $(\boldsymbol{x} - \boldsymbol{y}_i)^t B(\boldsymbol{y}_i)(\boldsymbol{x} - \boldsymbol{y}_i)$. For the unbounded cells, we suppose the probability for $\boldsymbol{x}$ to be in these cells is so small that the contribution of these cells to the total distortion is negligible.

In conclusion, we get the approximation

$$d(\boldsymbol{x}, \boldsymbol{y}_i) \cong (\boldsymbol{x} - \boldsymbol{y}_i)^t B(\boldsymbol{y}_i)(\boldsymbol{x} - \boldsymbol{y}_i). \tag{26}$$

Now we prove the three limits in (20), which are crucial for the result of source mismatch. We also derive the required conditions for the limits to hold in the process.

First, let us rewrite the three terms in the right-hand side of (19)

$$2 \ln \int \hat{p}(x)^{k/(k+2)} B(x)^{1/(k+2)} \, dx = 2 \ln \int \hat{p}(x) \big(\hat{p}(x)^{-2} B(x)\big)^{1/(k+2)} \, dx$$

$$-(k+2) \ln \int p(x)^{k/(k+2)} B(x)^{1/(k+2)} \, dx = -(k+2) \ln \int p(x) \big(p(x)^{-2} B(x)\big)^{1/(k+2)} \, dx$$

$$k \ln \int p(x) \hat{p}(x)^{-(2/(k+2))} B(x)^{1/(k+2)} \, dx = k \ln \int p(x) \big(\hat{p}(x)^{-2} B(x)\big)^{1/(k+2)} \, dx.$$

By Taylor's formula with the remainder [22] for exponential function expanded at the origin

$$\big(\hat{p}(x)^{-2} B(x)\big)^{1/(k+2)} = 1 + \ln \big(\hat{p}(x)^{-2} B(x)\big) \frac{1}{k+2} + \frac{\ln^2 \big(\hat{p}(x)^{-2} B(x)\big) \big(\hat{p}(x)^{-2} B(x)\big)^{z(x)}}{2} \left(\frac{1}{k+2}\right)^2$$

where $z(x) \in (0, 1/(k+2))$.

Hence

$$\int \hat{p}(x) \big(\hat{p}(x)^{-2} B(x)\big)^{1/(k+2)} \, dx = 1 + \frac{1}{k+2} E_{\hat{p}} \ln \big(\hat{p}(x)^{-2} B(x)\big) + \frac{1}{(k+2)^2} E_{\hat{p}} \ln^2 \big(\hat{p}(x)^{-2} B(x)\big) \big(\hat{p}(x)^{-2} B(x)\big)^{z(x)}.$$

Given the condition $E_{\hat{p}} \ln^2(\hat{p}(x)^{-2} B(x)) < \infty$, we know $E_{\hat{p}} \ln(\hat{p}(x)^{-2} B(x)) < \infty$. With the condition $E_{\hat{p}}(\hat{p}(x)^{-2} B(x))^{\epsilon} < \infty$ combined, we can prove

$$E_{\hat{p}} \ln^2(\hat{p}(x)^{-2} B(x))(\hat{p}(x)^{-2} B(x))^{z(x)} < \infty.$$

The proof is slightly more involved.

As $z(x) \in (0, 1/(k+2))$, we can make k big enough so that $z(x) < \epsilon/2$. Since there exists constant C so that

$$\text{for } \hat{p}(x)^{-2} B(x) > C \qquad \ln^2 \big(\hat{p}(x)^{-2} B(x)\big) < \big(\hat{p}(x)^{-2} B(x)\big)^{\epsilon/2}$$

as a result

$$\begin{aligned} &E_{\hat{p}} \ln^2 \big(\hat{p}(x)^{-2} B(x)\big) \big(\hat{p}(x)^{-2} B(x)\big)^{z(x)} \\ &\quad < E_{\hat{p}} \ln^2 \big(\hat{p}(x)^{-2} B(x)\big) \cdot \mathrm{Max} \Big(1, \big(\hat{p}(x)^{-2} B(x)\big)^{z(x)}\Big) \\ &\quad < E_{\hat{p}} \ln^2 \big(\hat{p}(x)^{-2} B(x)\big) + E_{\hat{p}} \ln^2 \big(\hat{p}(x)^{-2} B(x)\big) \\ &\qquad \cdot \big(\hat{p}(x)^{-2} B(x)\big)^{\epsilon/2} \\ &\quad < \Big(1 + C^{\epsilon/2}\Big) E_{\hat{p}} \ln^2 \big(\hat{p}(x)^{-2} B(x)\big) + E_{\hat{p}} \big(\hat{p}(x)^{-2} B(x)\big)^{\epsilon} \\ &\quad < \infty. \end{aligned}$$

Then it is not hard to see that $\int \hat{p}(x)(\hat{p}(x)^{-2} B(x))^{1/(k+2)} \, dx$ converges to 1 as $k \to \infty$. Hence, its logarithm, the first term in (19), converges to 0.

The method used to treat the second and third terms is very similar. With the conditions

$$E_p \big(\ln p(x)^{-2} B(x)\big)^2 < \infty$$
$$E_p \big(\ln \hat{p}(x)^{-2} B(x)\big)^2 < \infty$$
$$E_p \big(p(x)^{-2} B(x)\big)^{\epsilon} < \infty$$

and

$$E_p \big(\hat{p}(x)^{-2} B(x)\big)^{\epsilon} < \infty$$

added in, we can show with details omitted

$$\begin{aligned} &-(k+2) \ln \int p(x) \big(p(x)^{-2} B(x)\big)^{1/(k+2)} \, dx \\ &\quad = -(k+2) \ln \left(1 + \frac{1}{k+2} \int p(x) \ln \big(p(x)^{-2} B(x)\big) dx + o\left(\frac{1}{k+2}\right)\right) \\ &\quad = -(k+2) \left(\frac{1}{k+2} \int p(x) \ln \big(p(x)^{-2} B(x)\big) dx + o\left(\frac{1}{k+2}\right)\right) \\ &\quad = -\int p(x) \ln \big(p(x)^{-2} B(x)\big) dx \end{aligned}$$

$$
\begin{aligned}
&k \ln \int p(x)\left(\hat{p}(x)^{-2}B(x)\right)^{1/(k+2)}\,dx \\
&\quad = k \ln\left(1 + \frac{1}{k+2}\int p(x)\,\ln\left(\hat{p}(x)^{-2}B(x)\right)dx \right. \\
&\qquad\qquad \left. + o\left(\frac{1}{k+2}\right)\right) \\
&\quad = k\cdot\left(\frac{1}{k+2}\int p(x)\,\ln\left(\hat{p}(x)^{-2}B(x)\right)dx \right. \\
&\qquad\qquad \left. + o\left(\frac{1}{k+2}\right)\right) \\
&\quad = \int p(x)\,\ln\left(\hat{p}(x)^{-2}B(x)\right)dx.
\end{aligned}
$$

To sum up, we proved the following three limits:

$$
\begin{aligned}
&\lim_{k\to\infty} \ln \int \hat{p}(x)^{k/(k+2)}B(x)^{1/(k+2)}\,dx = 0 \\
&\lim_{k\to\infty} k \ln \int p(x)\hat{p}(x)^{-(2/(k+2))}B(x)^{1/(k+2)}\,dx \\
&\qquad = \int p(x)\,\ln\left(\hat{p}(x)^{-2}B(x)\right)dx \\
&\lim_{k\to\infty} -(k+2)\ln \int p(x)^{k/(k+2)}B(x)^{1/(k+2)}\,dx \\
&\qquad = -\int p(x)\,\ln\left(p(x)^{-2}B(x)\right)dx.
\end{aligned}
$$

ACKNOWLEDGMENT

The authors gratefully acknowledge the helpful comments of W. Gardner and K. Zeger. They also wish to thank one of the reviewers for very useful suggestions.

REFERENCES

[1] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 525–536, July 1974.

[2] F. X. J. Lukas and Z. L. Budrikis, "Picture quality prediction based on a visual model," *IEEE Trans. Commun.*, vol. COM-30, pp. 1679–1692, July 1982.

[3] D. J. Sakrison, "On the role of the observer and a distortion measure in image transmission," *IEEE Trans. Commun.*, vol. COM-25, pp. 1251–1267, Nov. 1977.

[4] N. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Commun.*, vol. COM-33, pp. 551–557, June 1985.

[5] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.

[6] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, July 1948.

[7] P. Zador, "Asymptotic quantization of continuous random variables," Bell Labs., 1966, unpublished memorandum.

[8] P. Elias, "Bounds and asymptotes for the performance of multivariate quantizers," *Ann. Math. Stat.*, vol. 41, pp. 1249–1259, 1970.

[9] A. Gersho, "Asymptotically optimum block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.

[10] S. Na and D. L. Neuhoff, "Bennett's integral for vector quantizers," *IEEE Trans. Inform. Theory*, vol. 41, pp. 886–900, July 1995.

[11] Y. Yamada, S. Tazaki, and R. M. Gray, "Asymptotic performance of block quantizers with difference distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 6–14, Jan. 1980.

[12] W. R. Gardner and B. D. Rao, "Theoretical analysis of the high-rate vector quantization of LPC parameters," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 367–381, Sept. 1995.

[13] T. Linder and R. Zamir, "High-resolution source coding for nondifference distortion measures: The rate distortion function," *IEEE Trans. Inform. Theory*, vol. 45, pp. 533–547, Mar. 1999.

[14] T. Linder, R. Zamir, and K. Zeger, "High-resolution source coding for nondifference distortion measures: Multidimensional companding," *IEEE Trans. Inform. Theory*, vol. 45, pp. 548–561, Mar. 1999.

[15] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, pp. 2959–2965, Dec. 1995.

[16] J. Conway and N. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 211–226, Mar. 1982.

[17] L. Fejes Toth, *Lagerungen in der Ebene auf der Kugel und im Raum.* Berlin, Germanyn: Springer-Verlag, 1953. p. 81.

[18] R. M. Gray and E. D. Karnin, "Multiple local optima in vector quantizers," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 256–261, Mar. 1982.

[19] F. Itakura and S. Saito, "Analysis synthesis telephony based upon maximum likelihood method," in *Repts. 6th Int. Congr. Acoustics* Y. Kohasi, Ed. (Tokyo, Japan, 1968), nos. C-5-5 and C17-20.

[20] F. Itakura, "Maximum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 67–72, Feb. 1975.

[21] J. E. Marsden and A. J. Tromba, *Vector Calculus.* San Francisco, CA: Freeman, 1981, pp. 201–206.

[22] E. W. Swokowski, *Calculus with Analytic Geometry.* PWS Publishers, 1983, p. 483.