Asymptotic Performance of Vector Quantizers with a Perceptual Distortion Measure

Jia Li Navin Chaddha Robert M. Gray* *

April 9, 1997

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, a class of distortion measure often used for perceptually meaningful distortion. The generalization involves a more rigorous derivation of a fixed rate result of Gardner and Rao and a new result for variable rate codes. 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 dB is shown to be linear in the relative entropy between the true and estimated probability densities.

1 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 ease of use. 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 empirically shown, however, that mean squared error does not correlate well with subjective (human) quality assessments [2, 3]. As a result, decreasing the mean squared error does not necessarily improve image quality. As standards for image quality become more demanding, code designers require distortion measures which are more consistent with human perception of images. As a result, perceptual distortion measures are receiving more attention. Algorithmic speed is less of an issue since the encoding is often off-line.

Since good perceptual distortion measures take the human eye's nonlinear perception of images into account, they cannot in general be modeled by simple difference distortion measures. For

*The authors are with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, CA 94305, U.S.A. This work was partially supported by the National Science Foundation under Grant No. NSF MIP-9016974. Email: jiali@isl.stanford.edu, navin@VXtreme.com, rmgray@stanford.edu

example, human eyes are far more sensitive to a particular range of intensity. When the pixels get too bright or too dark, the eyes will not notice large intensity variations. Consequently, 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 space frequency sensitivity and color sensitivity may play roles in the perceptual distortion measure as well. Considerable work has been done on developing objective image quality measures [4] consistent with human assessments and evaluating the efficiency of some popular quality measurements [5]. Nill [4] defined a quality measurement in the cosine transform domain incorporating a model of the human vision system which will be used as an example in our analysis.

A natural question is whether or not approximations and performance bounds for squared error distortion extend to more general perceptually motivated input-weighted quadratic distortion measures. Of particular interest here are the bounds resulting from asymptotic quantization or high-rate or high-resolution quantization approximations. 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. Gersho [6] developed approximations, conjectures, and bounds on the average distortion defined as powers of the Euclidean norm. By introducing the concept of inertial profile, Na and Neuhoff [7] proved a general formula similar to Bennett's integral for the average distortion of a high rate vector quantizer. Yamada et al. [1] generalized the lower bounds of Gersho [6] to difference distortion measures that are increasing functions of the norm of their argument. Gardner and Rao [8] extended the fixed rate coding results in [1] to a larger class of distortion measures $d(\mathbf{x}, \mathbf{y})$, where $d(\mathbf{x}, \mathbf{y})$ is a nonnegative function with continuous derivatives. Their distortion measure $d(\mathbf{x}, \mathbf{y})$ is used to model perceptual speech distortion. In this paper, we use a distortion measure $d(\mathbf{x}, \mathbf{y})$ similar to Gardner'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. [1] to our version of the distortion introduced by Gardner and Rao [8]. 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 2, we provide preliminaries in which basic notation and prerequisite results are introduced. In section 3 Gardner and Rao's [8] bounds on asymptotic average distortion for fixed rate codes are reviewed and a formal proof provided. The results are extended to variable rate coding in section 4. In section 5 the variable rate coding results are applied to two examples of perceptual distortion measures. The issue of source mismatch is addressed in section 6. The technique used in deriving the bounds in sections 3 and 4 is similar to that used by Yamada et al. [1]. Hence the notation parallels that of [1] 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.

2 Preliminaries

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

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

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

For a vector (i.e. a block space of pixels or a sub-image) 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 $L(\mathbf{x}, \mathbf{y})$, zero if and only if $\mathbf{x} = \mathbf{y}$, where $\mathbf{x}$ is the original vector and $\mathbf{y}$ is the quantized vector. For the derivation, we require regularity constraints on $L(\mathbf{x}, \mathbf{y})$, which are listed below.

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

$$B_{j,n}(\mathbf{y}) = \frac{1}{2} \frac{\partial^2 L(\mathbf{x}, \mathbf{y})}{\partial x_j \partial x_n} \bigg|_{\mathbf{x}=\mathbf{y}} \tag{1}$$

is positive definite almost everywhere.

Actually, combined with condition 1, $L(\mathbf{x}, \mathbf{y})$ being nonnegative and being zero if and only if $\mathbf{x} = \mathbf{y}$ implies that $B(\mathbf{y})$ is semi-positive definite. Gardner et al. [8] used a similar $d(\mathbf{x}, \mathbf{y})$ to model perceptual distortion measure for speech. The matrix $B(\mathbf{y})$ is the 'sensitivity matrix' in [8]. Here, we clarify the constraints on $L(\mathbf{x}, \mathbf{y})$. In [8], it is not explicitly stated that $B(\mathbf{y})$ must be positive definite and the order of continuous partial derivatives of $L(\mathbf{x}, \mathbf{y})$ must be three.

The requirement that $B(\mathbf{y})$ be positive definite may appear to be a very strong constraint because it means the distortion is dominated by the quadratic terms when $\mathbf{x}$ and $\mathbf{y}$ are very close. Nevertheless, these conditions are usually satisfied by perceptual distortion measures. Examples can be found in Eskicioglu and Fisher [5]. Also Nill's [4] definition of quality measure analyzed in detail later in this paper satisfies these conditions.

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

$$L(\mathbf{x}, \mathbf{y}_i) \approx (\mathbf{x} - \mathbf{y}_i)^t B(\mathbf{y}_i)(\mathbf{x} - \mathbf{y}_i) \quad . \tag{2}$$

For simplicity of notation, $B(\mathbf{y}_i)$ is denoted by B_i. The derivation for equation (2) is given in the Appendix. In the case when $L(\mathbf{x}, \mathbf{y}_i)$ exactly equals the right hand side of the above expression, all the results below hold.

A quantity we will frequently use later is the volume of the unit sphere in the $k-$dimensional space for the quadratic norm $\| \mathbf{x} \|^2 = \mathbf{x}^t B_i \mathbf{x}$, denoted as V_i.

$$V_i = V(\{\mathbf{u} : \sqrt{\mathbf{u}^t B_i \mathbf{u}} \leq 1\}) \quad .$$

From [1],

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

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

$$V(\{\mathbf{x} : \| \mathbf{x} - \mathbf{y}_i \| \leq a\}) = a^k V_i \quad . \tag{4}$$

For convenience, we generalize the Gish-Pierce function $M(v)$ of [1] to this case:

$$M_i(\nu) = \frac{1}{V_i} \int_{\mathbf{u}:\mathbf{u}^t B_i \mathbf{u} \leq 1} (\frac{\nu}{V_i^{1/k}} \mathbf{u})^t B_i (\frac{\nu}{V_i^{1/k}} \mathbf{u}) d\mathbf{u} \quad . \tag{5}$$

Substituting equation (3) into (5) yields

$$M_i(\nu) = \frac{[\det(B_i)]^{\frac{1}{2}}}{C_k} \int_{\mathbf{u}:\mathbf{u}^t B_i \mathbf{u} \leq 1} (B_i^{\frac{1}{2}} \frac{\nu}{V_i^{1/k}} \mathbf{u})^t (B_i^{\frac{1}{2}} \frac{\nu}{V_i^{1/k}} \mathbf{u}) d\mathbf{u} \quad ;$$

and changing variable $\mathbf{w} = (B(\mathbf{x}))^{\frac{1}{2}} \mathbf{u}$ results in

$$M_i(\nu) = \frac{1}{C_k} \int_{\mathbf{w}:\mathbf{w}^t \mathbf{w} \leq 1} (\frac{\nu}{V_i^{1/k}} \mathbf{w})^t (\frac{\nu}{V_i^{1/k}} \mathbf{w}) d\mathbf{w} \tag{6}$$

$M_i(\nu)$ is convex and monotonically increasing for $\nu \geq 0$.

3 Bounds for Asymptotically Optimal Performance with Fixed Rate Coding

In this section, we consider the case of fixed rate coding, that is, the rate is measured by $\log N$ where N is the total number of codewords. The distortion measure is the generalized perceptual distortion $L(\mathbf{x}, \mathbf{y})$ satisfying the two conditions. A lower bound on the asymptotic average distortion D and the corresponding optimal limiting density function $\lambda(\mathbf{x})$ are obtained.

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

$$\begin{aligned} D &= Ed(\mathbf{x}, Q(\mathbf{x})) = EL(\mathbf{x}, Q(\mathbf{x})) \\ &= \int p(\mathbf{x}) L(\mathbf{x}, Q(\mathbf{x})) d\mathbf{x} \\ &= \sum_{i=1}^{N} \int_{S_i} p(\mathbf{x}) L(\mathbf{x}, Q(\mathbf{x})) d\mathbf{x} \\ &= \sum_{i=1}^{N} \int_{S_i} p(\mathbf{x}) L(\mathbf{x}, \mathbf{y}_i) d\mathbf{x} \quad . \end{aligned}$$

Just as in all studies of asymptotic quantization, we assume the probability density $p(\mathbf{x})$ is sufficiently "smooth" to ensure that $P(\mathbf{x})$ is effectively constant over small bounded sets. If we define $P_i = \int_{S_i} p(\mathbf{x}) d\mathbf{x}$, then $p(\mathbf{x}) \approx P_i / V(S_i), \mathbf{x} \in S_i$ and the previous equation for D can be written as

$$D \approx \sum_{i=1}^{N} (P_i / V(S_i)) \int_{S_i} L(\mathbf{x}, \mathbf{y}_i) d\mathbf{x} \quad . \tag{7}$$

For high rate analysis, the approximation $L(\mathbf{x},\mathbf{y}_i) \approx (\mathbf{x}-\mathbf{y}_i)^t B_i(\mathbf{x}-\mathbf{y}_i)$ stated in equation (2) implies the further approximation of D as

$$D \approx \sum_{i=1}^{N}(P_i/V(S_i))\int_{S_i}(\mathbf{x}-\mathbf{y}_i)^t B_i(\mathbf{x}-\mathbf{y}_i)d\mathbf{x} \quad . \tag{8}$$

As in [1], we introduce two concepts: 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 $\| \mathbf{x} \|_i^2 = \mathbf{x}^t B_i \mathbf{x}$. By the scaling property stated in (4), it is easy to see that

$$R(S_i) = (V(S_i)/V_i)^{\frac{1}{k}} \quad . \tag{9}$$

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

$$\begin{aligned} T(S_i) &= \{\mathbf{x} : \sqrt{(\mathbf{x}-\mathbf{y}_i)^t B_i(\mathbf{x}-\mathbf{y}_i)} \leq R(S_i)\} \\ &= \{\mathbf{x} : \| \mathbf{x}-\mathbf{y}_i \| \leq R(S_i)\} \end{aligned} \tag{10}$$

where $\| \cdot \|$ is the quadratic norm $\| \mathbf{x} \| = \sqrt{\mathbf{x}^t B_i \mathbf{x}}$. $T(S_i)$ is essentially a sphere centered at $\mathbf{y}_i$ with radius $R(S_i)$. Obviously,

$$V(T(S_i)) = R(S_i)^k V_i = V(S_i) \quad . \tag{11}$$

A crucial inequality [1] we need to use is

$$\int_S d(\mathbf{x},\mathbf{y})d\mathbf{x} \geq \int_{T(S)} d(\mathbf{x},\mathbf{y})d\mathbf{x} \quad . \tag{12}$$

In our special case, $d(\mathbf{x},\mathbf{y_i}) = (\mathbf{x}-\mathbf{y}_i)^t B_i(\mathbf{x}-\mathbf{y}_i)$, and hence

$$\int_{S_i}(\mathbf{x}-\mathbf{y}_i)^t B_i(\mathbf{x}-\mathbf{y}_i)d\mathbf{x} \geq \int_{T(S_i)}(\mathbf{x}-\mathbf{y}_i)^t B_i(\mathbf{x}-\mathbf{y}_i)d\mathbf{x} \tag{13}$$

equality holds if and only if $T(S_i) = S_i$, that is, if S_i is a sphere under the norm. $\| \mathbf{x} \|_i = \sqrt{\mathbf{x}^t B_i \mathbf{x}}$.

Considering the right hand side of equation (13) and making change of variables, we get

$$\begin{aligned} \int_{T(S_i)}(\mathbf{x}-\mathbf{y}_i)^t B_i(\mathbf{x}-\mathbf{y}_i)d\mathbf{x} = \frac{V(S_i)}{V_i} \cdot \\ \int_{\mathbf{u}:\sqrt{\mathbf{u}^t B_i \mathbf{u}} \leq 1}\left(\frac{V(S_i)^{\frac{1}{k}}}{V_i^{\frac{1}{k}}}\mathbf{u}\right)^t \cdot B_i\left(\frac{V(S_i)^{\frac{1}{k}}}{V_i^{\frac{1}{k}}}\mathbf{u}\right)d\mathbf{u} \end{aligned}$$

where $\mathbf{u} = \frac{\mathbf{x}-\mathbf{y}_i}{R(S_i)}$. By substituting in equation (5), we can simplify the above equation to

$$\int_{T(S_i)}(\mathbf{x}-\mathbf{y}_i)^t B_i(\mathbf{x}-\mathbf{y}_i)d\mathbf{x} = V(S_i)M_i(V(S_i)^{\frac{1}{k}}) \quad . \tag{14}$$

Using equation (8), (13) and (14), we get

$$D \geq \sum_{i=1}^{N} P_i M_i(V(S_i)^{\frac{1}{k}}) \quad . \tag{15}$$

Following Gersho [6], we can define the $k-$dimensional reproduction vector density by

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

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

$$V(S_i) \approx (N\lambda(\mathbf{y}_i))^{-1} \tag{16}$$

for every bounded region S_i. Substituting the approximation to (15), we get

$$D \;\geq\; \sum_{i=1}^{N} P_i M_i((N\lambda(\mathbf{y}_i))^{-\frac{1}{k}}) \quad .$$

Consider the right hand side of the above inequality. Expanding M_i by equation (6) and using the integral approximation for the sum yields

$$D \;\geq\; \alpha \int p(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}} \lambda(\mathbf{x})^{-\frac{2}{k}} d\mathbf{x} \tag{17}$$

where

$$\begin{aligned} \alpha &= \frac{1}{C_k}[\int_{\mathbf{w}:\mathbf{w}^t\mathbf{w}\leq 1} (\frac{\mathbf{w}}{C_k^{\frac{1}{k}}})^t (\frac{\mathbf{w}}{C_k^{\frac{1}{k}}}) d\mathbf{w}] N^{-\frac{2}{k}} \\ &= \frac{k}{k+2} C_k^{-\frac{2}{k}} N^{-\frac{2}{k}} \quad . \end{aligned}$$

Since $\int \lambda(\mathbf{x})d\mathbf{x} = 1$, the above inequality for D is equivalent to

$$\begin{aligned} D &\geq \alpha[\int p(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}} \lambda(\mathbf{x})^{-\frac{2}{k}} d\mathbf{x}][\int \lambda(\mathbf{x})d\mathbf{x}]^{\frac{2}{k}} \\ &\geq \alpha[\int (p(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}})^{\frac{1}{1+2/k}} d\mathbf{x}]^{\frac{k+2}{k}} \quad . \end{aligned}$$

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

$$\lambda(\mathbf{x}) \propto (p(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}})^{\frac{k}{k+2}} \quad .$$

Subject to the unit integral constraint,

$$\lambda_{opt}(\mathbf{x}) = \frac{(p(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}})^{\frac{k}{k+2}}}{\int (p(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}})^{\frac{k}{k+2}} d\mathbf{x}} \quad . \tag{18}$$

We define $D_L(Q_{opt})$ as the lower bound of D with optimal $\lambda(\mathbf{x})$, which is

$$D_L(Q_{opt}) = \alpha \int (\lambda_{opt}(\mathbf{x}))^{-\frac{2}{k}} (\det(B(\mathbf{x})))^{\frac{1}{k}} p(\mathbf{x}) d\mathbf{x} \quad . \tag{19}$$

Hence, the lower bound for the asymptotic distortion is

$$\begin{aligned} D &\geq D_L(Q_{opt}) \\ &= \frac{k}{k+2} C_k^{-\frac{2}{k}} N^{-\frac{2}{k}} \cdot \\ &\quad \{\int [p(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}}]^{\frac{1}{1+2/k}} d\mathbf{x}\}^{\frac{k+2}{k}} \quad . \end{aligned} \tag{20}$$

This bound is originally due to Gardner and Rao [8] who provided an informal proof subject to weaker conditions. If we specialize to the MSE distortion, $\det(B(\mathbf{x})) = 1$. Hence $D_L(Q_{opt}) = \frac{k}{k+2} C_k^{-\frac{2}{k}} N^{-\frac{2}{k}} \parallel p \parallel_{\frac{k}{k+2}}$. This is just what has been proved in Gersho [6].

Using the inertial profile of Na and Neuhoff [7], we can rewrite inequality (13) as an equality

$$\int_{S_i} (\mathbf{x} - \mathbf{y}_i)^t B_i (\mathbf{x} - \mathbf{y}_i) d\mathbf{x} = m(S_i) \cdot \int_{T(S_i)} (\mathbf{x} - \mathbf{y}_i)^t B_i (\mathbf{x} - \mathbf{y}_i) d\mathbf{x}$$

where $m(S_i)$ is the normalized moment of inertia of the cell S_i about the point $\mathbf{y}_i$. Suppose there is a function $m(\mathbf{x})$ (usually smooth) such that for any $\mathbf{x}$, the cells in the vicinity of $\mathbf{x}$ have normalized moment of inertia (about their respective quantization points) approximately equal to $m(\mathbf{x})$, the so called the inertial profile. Then, the inequality (17) can be modified to the equality below

$$D = \alpha \int p(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}} m(\mathbf{x}) \lambda(\mathbf{x})^{-\frac{2}{k}} d\mathbf{x} \quad .$$

As $\det(B(\mathbf{x}))$ varies with $\mathbf{x}$, it is not obvious what the optimal $m(\mathbf{x})$ is even for low vector dimension. If we suppose a suboptimal quantizer is used, for instance, $m(\mathbf{x})$ is taken as a constant, the optimal limiting density $\lambda(\mathbf{x})$ is given by expression (18).

Gardner and Rao [8] argued that in high rate quantization, although expression (12) takes a form of a lower bound, it can also be viewed as an approximation, that is, $\int_S d(\mathbf{x}, \mathbf{y}) d\mathbf{x} \approx \int_{T(S)} d(\mathbf{x}, \mathbf{y}) d\mathbf{x}$. 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 occurred in this approximation was investigated in [9] for spaces up to dimension 10. Consequently, the inequalities in (15) and (17) become approximations. And the lower bound $D_L(Q_{opt})$ is approximately $D(Q_{opt})$, so

$$D(Q_{opt}) \approx \frac{k}{k+2} C_k^{-\frac{2}{k}} N^{-\frac{2}{k}} \cdot \{\int [p(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}}]^{\frac{1}{1+2/k}} d\mathbf{x}\}^{\frac{k+2}{k}} \quad . \tag{21}$$

4 Bounds for Asymptotically Optimal Performance with Variable Rate Coding

In this section, we consider variable rate coding. In this case, the rate is the entropy of the encoded source. As in the previous section, we derive a lower bound on the asymptotic average distortion D and the corresponding optimal limiting density function $\lambda(\mathbf{x})$.

To start the derivation, express $D_L(Q_{opt})$ as a function of $\lambda_{opt}(\mathbf{x})$. From (19), we obtain

$$D_L(Q_{opt}) = \frac{kC_k^{-\frac{2}{k}}}{k+2} \int [N\lambda_{opt}(\mathbf{x})]^{-\frac{2}{k}} (\det(B(\mathbf{x})))^{\frac{1}{k}} p(\mathbf{x}) d\mathbf{x} \quad . \tag{22}$$

For the convenience of derivation, we define

$$f(\mathbf{x}) = \frac{(\det(B(\mathbf{x})))^{\frac{1}{k}} p(\mathbf{x})}{\int (\det(B(\mathbf{x})))^{\frac{1}{k}} p(\mathbf{x}) d\mathbf{x}} \quad .$$

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

Substituting $f(\mathbf{x})$ into (22) simplifies the equation to

$$\begin{aligned} D_L(Q_{opt}) = \frac{kC_k^{-\frac{2}{k}}}{k+2} [\int (\det(B(\mathbf{x})))^{\frac{1}{k}} p(\mathbf{x}) d\mathbf{x}] \cdot \\ \int [N\lambda_{opt}(\mathbf{x})]^{-\frac{2}{k}} f(\mathbf{x}) d\mathbf{x} \quad . \end{aligned}$$

From equation (28) of [1], the entropy of the encoded source is approximately

$$H_Q \cong h(p) - E\{\log \frac{1}{N\lambda(\mathbf{x})}\} \quad .$$

We denote $g(\mathbf{x}) = N\lambda(\mathbf{x})$.

Now, finding Q_{opt} with fixed entropy H_Q becomes a condition constrained optimization problem stated as follows:

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

As $f(\mathbf{x})$ and $p(\mathbf{x})$ are fixed functions in the minimization, the constraint $\int \log(g(\mathbf{x}))p(\mathbf{x})d\mathbf{x} \leq C$ is equivalent to $\int \log(g(\mathbf{x})^{-\frac{2}{k}} \frac{f(\mathbf{x})}{p(\mathbf{x})}) p(\mathbf{x}) d\mathbf{x} \geq C'$ where C' is another constant. By Jensen's inequality

$$\log \int (g(\mathbf{x})^{-\frac{2}{k}} \frac{f(\mathbf{x})}{p(\mathbf{x})}) p(\mathbf{x}) d\mathbf{x} \geq \int \log(g(\mathbf{x})^{-\frac{2}{k}} \frac{f(\mathbf{x})}{p(\mathbf{x})}) p(\mathbf{x}) d\mathbf{x} \quad .$$

Assuming natural logarithms, we can rewrite the above inequality as

$$\begin{aligned} \int (g(\mathbf{x})^{-\frac{2}{k}} \frac{f(\mathbf{x})}{p(\mathbf{x})}) p(\mathbf{x}) d\mathbf{x} &= \int g(\mathbf{x})^{-\frac{2}{k}} f(\mathbf{x}) d\mathbf{x} \\ &\geq e^{\int \log(g(\mathbf{x})^{-\frac{2}{k}} \frac{f(\mathbf{x})}{p(\mathbf{x})}) p(\mathbf{x}) d\mathbf{x}} \\ &\geq e^{C'} \quad , \end{aligned} \tag{23}$$

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

$$\lambda_{opt}(\mathbf{x}) = \frac{(\det(B(\mathbf{x})))^{\frac{1}{2}}}{\int_{\mathbf{x}\in G}(\det(B(\mathbf{x})))^{\frac{1}{2}}d\mathbf{x}} \quad . \tag{24}$$

From the result, we see that the quantization density is higher where $\det(B(\mathbf{x}))$ is larger. This accords with intuition because the area with larger $\det(B(\mathbf{x}))$ 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, $\lambda_{opt}(\mathbf{x})$ is shown to be constant in [1], which follows here by setting $\det(B(\mathbf{x}))$ to a constant.

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

$$\begin{aligned} H_Q = h(p) + \log N + \frac{1}{2}\int \log(\det(B(\mathbf{x})))p(\mathbf{x})d\mathbf{x} \\ - \log(\int (\det(B(\mathbf{x})))^{\frac{1}{2}}d\mathbf{x}) \end{aligned} \tag{25}$$

and

$$D_L(Q_{opt}) = \frac{kC_k^{-\frac{2}{k}}}{k+2}N^{-\frac{2}{k}}[\int (\det(B(\mathbf{x})))^{\frac{1}{2}}d\mathbf{x}]^{\frac{2}{k}} \quad . \tag{26}$$

Combining equations (25) and (26), we get the lower bound for the asymptotic distortion

$$\begin{aligned} D &\geq D_L(Q_{opt}) \\ &= \frac{kC_k^{-\frac{2}{k}}}{k+2}[\int (\det(B(\mathbf{x})))^{\frac{1}{2}}d\mathbf{x}]^{\frac{2}{k}} \cdot e^{A(\mathbf{x})} \end{aligned} \tag{27}$$

where

$$\begin{aligned} A(\mathbf{x}) = -\frac{2}{k}(H_Q - h(p) + \log(\int (\det(B(\mathbf{x})))^{\frac{1}{2}}d\mathbf{x}) \\ -\frac{1}{2}\int \log(\det(B(\mathbf{x})))p(\mathbf{x})d\mathbf{x}) \quad . \end{aligned}$$

Further calculation simplifies (27) to

$$\begin{aligned} D &\geq D_L(Q_{opt}) \\ &= \frac{kC_k^{-\frac{2}{k}}}{k+2} \cdot e^{-\frac{2}{k}(H_Q - h(p) - \frac{1}{2}\int \log(\det(B(\mathbf{x})))p(\mathbf{x})d\mathbf{x})} \quad . \end{aligned}$$

The above inequality shows how a non-constant $\det(B(\mathbf{x}))$ affects the lower bound of asymptotic distortion. In the case of MSE distortion measure, $\det(B(\mathbf{x}))$ is a constant and the expression (26) can be simplified to the same form as in [6].

5 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 $\lambda(\mathbf{x})$ for the perceptual distortion developed by Nill [4] and the input weighted squared error measure [10] are derived.

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

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

where

$$\begin{aligned}
H(r) &= \text{rotationally symmetric spatial frequency} \\
&\quad \text{response of human visual system,} \\
&\quad r = \sqrt{u^2 + v^2} \\
F, \hat{F} &= \text{transform coefficients of original and} \\
&\quad \text{quantized sub-image, respectively} \\
u, v &= \text{coordinates in the transform domain.} \\
M, N &= \text{number of coefficients in orthogonal } u, v \\
&\quad \text{directions} \\
\mathbf{x} &= [F(0,0), ..., F(0, N-1), F(1,0), ..., F(1, N-1), \\
&\quad ..., ..., F(M-1, N-1)] \\
\hat{\mathbf{x}} &= [\hat{F}(0,0), ..., \hat{F}(0, N-1), \hat{F}(1,0), ..., \hat{F}(1, N-1), \\
&\quad ..., ..., \hat{F}(M-1, N-1)] \\
W(\mathbf{x}) &= \text{weighting factor proportional to sub-image's} \\
&\quad \text{intensity level variance.} \\
&\quad \text{It is a quadratic function of } x_i\text{'s.}
\end{aligned}$$

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

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

$$\begin{aligned}
\text{when} \quad j \neq n, \quad B_{j,n}(\mathbf{x}) &= 0 \\
\text{when} \quad j = n, \text{suppose} \quad F(u,v) &= x_n \\
B_{j,n}(\mathbf{x}) &= H^2(r) W(\mathbf{x}) \\
\text{where} \quad r &= \sqrt{u^2 + v^2}
\end{aligned}$$

Hence

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

As $\prod_{u=0}^{M-1}\prod_{v=0}^{N-1} H^2(\sqrt{u^2+v^2})$ is a constant with respect to $\mathbf{x}$,

$$\lambda(\mathbf{x}) \propto W(\mathbf{x})^{\frac{M\cdot N}{2}} \quad .$$

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 $\frac{M\cdot N}{2}$th power of the intensity variance.

The second example is the input weighted squared error measure [10], which is defined as

$$L(\mathbf{y},\mathbf{x}) = (\mathbf{y}-\mathbf{x})^t W(\mathbf{y})(\mathbf{y}-\mathbf{x}) \quad .$$

For our analysis here, we suppose $W(\mathbf{y}) = \| \mathbf{y} \|^{-2} \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. We can rewrite the distortion as

$$\begin{aligned} L(\mathbf{y},\mathbf{x}) &= \frac{\| \mathbf{y}-\mathbf{x} \|^2}{\| \mathbf{y} \|^2} \\ &= \frac{\sum_{i=0}^{k-1}(y_i - x_i)^2}{\sum_{i=0}^{k-1} y_i^2} \quad . \end{aligned}$$

Using the previously given procedure,

$$\begin{aligned} \text{when} \quad j \neq n \quad B_{j,n}(\mathbf{x}) &= 0 \\ \text{when} \quad j = n \quad B_{j,n}(\mathbf{x}) &= \frac{1}{\| \mathbf{x} \|^2} \end{aligned}$$

Hence

$$\begin{aligned} \det(B(\mathbf{x})) &= \frac{1}{\| \mathbf{x} \|^{2k}} \\ \text{and} \quad \lambda(\mathbf{x}) &\propto \frac{1}{\| \mathbf{x} \|^{k}} \quad . \end{aligned}$$

So we see for this input weighted squared error measure, the limiting density of codewords is inversely proportional to the kth power of the norm of $\mathbf{x}$.

6 Source Mismatch

The previous analysis was based on the assumption that the probability density function $p(\mathbf{x})$ of the source is known. However, this is usually not the situation in practice. In real life, $p(\mathbf{x})$ must be estimated. For high rate variable rate coding, $p(\mathbf{x})$ is not required since $\lambda_{opt}(\mathbf{x})$ is unrelated to $p(\mathbf{x})$. For fixed rate coding, $\lambda_{opt}(\mathbf{x})$ does depend on $p(\mathbf{x})$ 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 $\mathbf{x} = (x_1, ..., x_k)$, x_i's are i.i.d. random variables. We analyze the asymptotic case when $k \to \infty$. The result can be easily generalized to $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k)$ where $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ir})$, $\mathbf{x}_i$'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 $\det(B(\mathbf{x}))$ changes with k to get a reasonable result. The assumption we make here is that $\det(B(\mathbf{x})) = \prod_{i=1}^{k} \det(B(\mathbf{x}_i))$, where, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k)$, $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ir})$. This assumption is essentially saying that all the sub-vectors $\mathbf{x}_i$ play an equal and independent role in determining the distortion. In the following derivation, $\mathbf{x}_i$ is a scalar, hence $\det(B(x_i)) = B(x_i)$. Consequently, the assumption becomes $\det(B(\mathbf{x})) = \prod_{i=1}^{k} B(x_i)$.

Recall the optimal $\lambda(\mathbf{x})$ of (18). If we estimate $p(\mathbf{x})$ by $\hat{p}(\mathbf{x})$, the optimal $\lambda(\mathbf{x})$ is

$$\hat{\lambda}_{opt}(\mathbf{x}) = \frac{[\hat{p}(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}}]^{\frac{k}{k+2}}}{\int [\hat{p}(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}}]^{\frac{k}{k+2}} d\mathbf{x}} \quad .$$

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

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

where

$$\hat{\lambda}_{opt}(x_i) = \frac{[\hat{p}(x_i)(\det(B(x_i)))^{\frac{1}{k}}]^{\frac{k}{k+2}}}{\int [\hat{p}(x_i)(\det(B(x_i)))^{\frac{1}{k}}]^{\frac{k}{k+2}} dx_i} \quad .$$

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

The asymptotic distortion becomes

$$D(\hat{Q}_{opt}) \approx \hat{D}_L = \alpha \int p(\mathbf{x})(\det(B(\mathbf{x})))^{\frac{1}{k}} \hat{\lambda}_{opt}(\mathbf{x})^{-\frac{2}{k}} d\mathbf{x} \quad .$$

As the true optimal distortion $D(Q_{opt})$ is given by (21), the increase of distortion in dB is $10 \log \frac{D(\hat{Q}_{opt})}{D(Q_{opt})}$. We choose dB as a measure of performance loss because in practice, SNR or PSNR is common evaluation of performance, and the decrease of SNR or PSNR is equal to the increase of distortion in dB. With $D(\hat{Q}_{opt})$ and $D(Q_{opt})$ substituted in and some algebra, we get

$$\begin{aligned} \ln \frac{D(\hat{Q}_{opt})}{D(Q_{opt})} &\approx \ln \frac{\hat{D}_L}{D_L(Q_{opt})} \\ &= \ln\left(\frac{\int p(x)B(x)^{\frac{1}{k}} \hat{\lambda}_{opt}(x)^{-\frac{2}{k}} dx}{\int p(x)B(x)^{\frac{1}{k}} \lambda_{opt}(x)^{-\frac{2}{k}} dx}\right)^k \\ &= k \ln \int p(x)\hat{p}(x)^{-\frac{2}{k+2}} B(x)^{\frac{1}{k+2}} dx \\ &\quad +2 \ln \int \hat{p}(x)^{\frac{k}{k+2}} B(x)^{\frac{1}{k+2}} dx \\ &\quad -(k+2) \ln \int [p(x)B(x)^{\frac{1}{k}}]^{\frac{k}{k+2}} dx \quad . \end{aligned} \tag{28}$$

The following limits are proved in the Appendix under assumptions (30) $\sim$ (33).

$$
\begin{aligned}
&\lim_{k\to\infty} \ln \int \hat{p}(x)^{\frac{k}{k+2}} B(x)^{\frac{1}{k+2}} dx = 0 \\
&\lim_{k\to\infty} k \ln \int p(x)\hat{p}(x)^{-\frac{2}{k+2}} B(x)^{\frac{1}{k+2}} dx \\
&\qquad = \int p(x) \ln(\hat{p}(x)^{-2} B(x)) dx \\
&\lim_{k\to\infty} -(k+2) \ln \int [p(x) B(x)^{\frac{1}{k}}]^{\frac{k}{k+2}} dx \\
&\qquad = -\int p(x) \ln(p(x)^{-2} B(x)) dx \quad .
\end{aligned} \tag{29}
$$

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

1.

$$E \ln^2(\hat{p}(x)^{-2} B(x)) < \infty \tag{30}$$

with respect to both $\hat{p}$ and p

2.

$$E_p \ln^2(p(x)^{-2} B(x)) < \infty \tag{31}$$

3.

$$\exists \epsilon > 0 \quad st. \quad E(\hat{p}(x)^{-2} B(x))^{\epsilon} < \infty \tag{32}$$

with respect to both $\hat{p}$ and p

4.

$$E_p (p(x)^{-2} B(x))^{\epsilon} < \infty \quad . \tag{33}$$

In many practical situations, the support of x is a bounded closed set. If this is the case and if $\hat{p}(x)$ and $B(x)$ are continuous functions, it is easy to see that the above four conditions hold.

Back to the limit results, by substituting the three terms to (28), we get

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

where $D(p(x) \parallel \hat{p}(x))$ with bit as dimension is the relative entropy of distributions $p(x)$ and $\hat{p}(x)$. Changing the base of logarithm, we finally obtain the loss in dB when $k \to \infty$

$$10 \log \frac{\hat{D}_L}{D_L(Q_{opt})} = 6D(p(x) \parallel \hat{p}(x)) \quad . \tag{34}$$

It is interesting to notice the limit loss is independent of $B(x)$, which means the effect of $B(x)$ is washed out when $k \to \infty$. Equation (34) also shows the linear dependence of the performance loss on the relative entropy of the true distributions and the estimated one. Since relative entropy is a formal measure of the closeness of two pdf functions, the result shows the straightforward way for the probability density estimation to affect the average distortion of the quantizer.

Acknowledgments

The authors gratefully acknowledge the helpful comments of William Gardner and Ken Zeger.

A Appendix

We prove in the appendix the approximating relation stated in equation (2) and the three limits in (29).

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

$$L(\mathbf{x}, \mathbf{y}_i) \approx (\mathbf{x} - \mathbf{y}_i)^t B(\mathbf{y}_i)(\mathbf{x} - \mathbf{y}_i)$$

where $\mathbf{y}_i$ in our consideration here is a constant vector. Readers need to keep this in mind to avoid confusion. Since $L(\mathbf{x}, \mathbf{y}_i)$ has continuous partials of third order, Taylor's theorem [11] implies that

$$L(\mathbf{x}, \mathbf{y}_i) = \nabla L(\mathbf{x}, \mathbf{y}_i) \mid_{\mathbf{x}=\mathbf{y}_i} \cdot (\mathbf{x} - \mathbf{y}_i) + (\mathbf{x} - \mathbf{y}_i)^t B(\mathbf{y}_i)(\mathbf{x} - \mathbf{y}_i) + R_2(\mathbf{x} - \mathbf{y}_i, \mathbf{y}_i) \quad .$$

We discuss the three terms on the right hand side of the above equation one by one. First, $\nabla L(\mathbf{x}, \mathbf{y}_i) \mid_{\mathbf{x}=\mathbf{y}_i}$ is the gradient of $L(\mathbf{x}, \mathbf{y}_i)$ calculated at y_i. It is equivalent to

$$\nabla L(\mathbf{x}, \mathbf{y}_i) \mid_{\mathbf{x}=\mathbf{y}_i} = \left(\frac{\partial L}{\partial x_1}, \frac{\partial L}{\partial x_2}, \ldots, \frac{\partial L}{\partial x_k} \right) \mid_{\mathbf{x}=\mathbf{y}_i} \quad .$$

As $L(\mathbf{x}, \mathbf{y}_i) \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{y}_i$, $\mathbf{x} = \mathbf{y}_i$ is a local extremum. From the local extremum theorem [11], $\nabla L(\mathbf{x}, \mathbf{y}_i) \mid_{\mathbf{x}=\mathbf{y}_i} = 0$.

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

$$B_{j,n}(\mathbf{y}_i) = \frac{1}{2} \frac{\partial^2 L(\mathbf{x}, \mathbf{y}_i)}{\partial x_j \partial x_n} \mid_{\mathbf{x}=\mathbf{y}_i} \quad .$$

We assume that $B(\mathbf{y}_i)$ is positive definite. As a result, $(\mathbf{x} - \mathbf{y}_i)^t B(\mathbf{y}_i)(\mathbf{x} - \mathbf{y}_i) \geq \mu_{min} \parallel \mathbf{x} - \mathbf{y}_i \parallel^2$, where μ_{min} is the minimum eigenvalue of $B(\mathbf{y}_i)$.

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

In conclusion, we get the approximation

$$L(\mathbf{x},\mathbf{y}_i) \approx (\mathbf{x}-\mathbf{y}_i)^t B(\mathbf{y}_i)(\mathbf{x}-\mathbf{y}_i) \quad . \tag{35}$$

Now we prove the three limits in (29), 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 equation (28)

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

By Taylor's formula with the remainder [12] for exponential function expanded at origin

$$\begin{aligned}
(\hat{p}(x)^{-2}B(x))^{\frac{1}{k+2}} &= 1 + \ln(\hat{p}(x)^{-2}B(x))\frac{1}{k+2} + \\
&\frac{\ln^2(\hat{p}(x)^{-2}B(x))(\hat{p}(x)^{-2}B(x))^{z(x)}}{2}\left(\frac{1}{k+2}\right)^2
\end{aligned}$$

where $z(x) \in (0, \frac{1}{k+2})$.

Hence

$$\begin{aligned}
\int \hat{p}(x)(\hat{p}(x)^{-2}B(x))^{\frac{1}{k+2}} dx &= 1 + \frac{1}{k+2}E_{\hat{p}}\ln(\hat{p}(x)^{-2}B(x)) + \\
&\frac{1}{(k+2)^2}E_{\hat{p}}\ln^2(\hat{p}(x)^{-2}B(x))(\hat{p}(x)^{-2}B(x))^{z(x)} \quad .
\end{aligned}$$

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 $E_{\hat{p}}\ln^2(\hat{p}(x)^{-2}B(x))(\hat{p}(x)^{-2}B(x))^{z(x)} < \infty$. It is slightly more mathematics involved.

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

$$\begin{aligned}
\text{for}\quad & \hat{p}(x)^{-2}B(x) > C \\
& \ln^2(\hat{p}(x)^{-2}B(x)) \;<\; (\hat{p}(x)^{-2}B(x))^{\frac{\epsilon}{2}}
\end{aligned}$$

as a result

$$
\begin{aligned}
&E_{\hat{p}} \ln^2(\hat{p}(x)^{-2}B(x))(\hat{p}(x)^{-2}B(x))^{z(x)} \\
&\quad < E_{\hat{p}} \ln^2(\hat{p}(x)^{-2}B(x))(\hat{p}(x)^{-2}B(x))^{\frac{\epsilon}{2}} \\
&\quad < E_{\hat{p}} \ln^2(\hat{p}(x)^{-2}B(x))C^{\frac{\epsilon}{2}} + E_{\hat{p}}(\hat{p}(x)^{-2}B(x))^{\epsilon} \\
&\quad < \infty \qquad .
\end{aligned}
$$

Then it is not hard to see that $\int \hat{p}(x)(\hat{p}(x)^{-2}B(x))^{\frac{1}{k+2}}dx$ converges to 1 as $k \to \infty$. Hence, the logarithm of it, which is the first term in (28), converges to 0.

The method to treat the second and third terms is very similar. With the conditions $E_p(\ln p(x)^{-2}B(x))^2 < \infty$, $E_p(\ln \hat{p}(x)^{-2}B(x))^2 < \infty$, $E_p(p(x)^{-2}B(x))^{\epsilon} < \infty$ and $E_p(\hat{p}(x)^{-2}B(x))^{\epsilon} < \infty$ added in, we can show with details omitted

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

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

To sum up, we proved the following three limits

$$
\begin{aligned}
&\lim_{k\to\infty} \ln \int \hat{p}(x)^{\frac{k}{k+2}}B(x)^{\frac{1}{k+2}}dx = 0 \\
&\lim_{k\to\infty} k\ln \int p(x)\hat{p}(x)^{-\frac{2}{k+2}}B(x)^{\frac{1}{k+2}}dx \\
&\qquad\qquad = \int p(x)\ln(\hat{p}(x)^{-2}B(x))dx \\
&\lim_{k\to\infty} -(k+2)\ln \int [p(x)B(x)^{\frac{1}{k}}]^{\frac{k}{k+2}}dx \\
&\qquad\qquad = -\int p(x)\ln(p(x)^{-2}B(x))dx \quad .
\end{aligned}
$$

References

[1] Y. Yamada, S. Tazaki and R. M. Gray, "Asymptotic Performance of Block Quantizers with Difference Distortion Measures," *IEEE Trans. Inform. Theory* , vol. IT-26, no. 4, pp. 6–14, Jan. 1980.

[2] 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.

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

[4] N. B. Nill, "A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment," *IEEE Transactions on Communications* , vol. COM-33, no. 6, pp. 551–557, June 1985.

[5] A. M. Eskicioglu and P. S. Fisher, "Image Quality Measures and Their Performance," *IEEE Transactions on Communications* , vol. 43, no. 12, pp. 2959–2965, Dec. 1995.

[6] A. Gersho, "Asymptotically Optimum Block Quantization," *IEEE Trans. Inform. Theory* , vol. IT-25, no. 4, pp. 373–380, July 1979.

[7] S. Na and D. L. Neuhoff, "Bennett's Integral for Vector Quantizers," *IEEE Trans. Inform. Theory* , vol. 41, no. 4, pp. 886–900, July 1995.

[8] W. R. Gardner and B. D. Rao, "Theoretical Analysis of the High-Rate Vector Quantization of LPC Parameters," *IEEE Transactions on Speech and Audio Processing* , vol. 3, no. 5, pp. 367–381, Sep. 1995.

[9] J. Conway and N. Sloane, "Voronoi Regions of Lattices, Second Moments of Polytopes, and Quantization," *IEEE Trans. on Inform. Theory* , vol. 28, pp. 211–226, March 1982.

[10] R. M. Gray and E.D. Karnin, "Multiple Local Optima in Vector Quantizers," *IEEE Trans. on Inform. Theory*, vol. 28, pp. 256–261, March 1982.

[11] J. E. Marsden and A. J. Tromba, *Vector Calculus*, pp. 201–206, W.H.Freeman and Company, 1981.

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

[13] T. Linder and R. Zamir, "High-resolution source coding for non-difference distortion measures: the rate distortion function," submitted to *IEEE Trans. on Inform. Theory.*"

[14] T. Linder, R. Zamir, and K. Zeger, "High-resolution source coding for non-difference distortion measures: multidimensional companding," submitted to *IEEE Trans. on Inform. Theory.*"