Research Article

The High-Resolution Rate-Distortion Function under the Structural Similarity Index

Jan Østergaard,[1] Milan S. Derpich,[2] and Sumohana S. Channappayya[3]

[1] *Department of Electronic Systems, Aalborg University, 9220 Alborg, Denmark*
[2] *Department of Electronic Engineering, Federico Santa María Technical University, 2390123 Valparaíso, Chile*
[3] *PacketVideo Corporation, San Diego, CA 92121, USA*

Correspondence should be addressed to Jan Østergaard, jo@es.aau.dk

We show that the structural similarity (SSIM) index, which is used in image processing to assess the similarity between an image representation and an original reference image, can be formulated as a locally quadratic distortion measure. We, furthermore, show that recent results of Linder and Zamir on the rate-distortion function (RDF) under locally quadratic distortion measures are applicable to this SSIM distortion measure. We finally derive the high-resolution SSIM-RDF and provide a simple method to numerically compute an approximation of the SSIM-RDF of real images.

1. Introduction

A vast majority of the work on source coding with a fidelity criterion (i.e., rate-distortion theory) concentrates on the mean-squared error (MSE) fidelity criterion. The MSE fidelity criterion is used mainly due to its mathematical tractability. However, in applications involving a human observer it has been noted that distortion measures which include some aspects of human perception generally perform better than the MSE [1]. A great number of perceptual distortion measures are nondifference distortion measures and, unfortunately, even for simple sources, their corresponding rate-distortion functions (RDFs), that is, the minimum bitrate required to attain a distortion equal to or smaller than some given value, are not known. However, in certain cases it is possible to derive their RDFs. For example, for a Gaussian process with a weighted squared error criterion, where the weights are restricted to be linear time-invariant operators, the complete RDF was first found in [2] and later rederived by several others [3, 4]. Other examples include the special case of locally quadratic distortion measures for fixed rate vector quantizers and under high-resolution assumptions [5], results which are extended to variable-rate vector quantizers in [6, 7], and applied to perceptual audio coding in [8, 9].

In [10], Wang et al. proposed the structural similarity (SSIM) index as a perceptual measure of the similarity between an image representation and an original reference image. The SSIM index takes into account the cross-corelation between the image and its representation as well as the images first- and second-order moments. It has been shown that this index provides a more accurate estimate of the perceived quality than the MSE [1]. The SSIM index was used for image coding in [11] and was cast in the framework of ℓ_1-compression of images and image sequences in [12]. The relation between the coding rate of a fixed-rate uniform quantizer and the distortion measured by the SSIM index was first addressed in [13]. In particular, for several types of source distributions and under high-resolution assumptions, upper and lower bounds on the SSIM index were provided as a function of the operational coding rate of the quantizer [13].

In this paper, we present the high-resolution RDF for sources with finite differential entropy and under an SSIM index distortion measure. The SSIM-RDF is particularly important for researchers and practitioners within the image coding area, since it provides a lower bound on the number of bits that *any* coder, for example, JPEG, and so forth, will use when encoding an image into a representation,

which has an SSIM index not smaller than a prespecified level. Thus, it allows one to compare the performance of a coding architecture to the optimum performance theoretically attainable. The SSIM-RDF is nonconvex and does not appear to admit a simple closed-form expression. However, when the coding rate is high, that is, when each pixel of the image is represented by a high number of bits, say more than 0.5 bpp, then we are able to find a simple expression, which is asymptotically (as the bit rate increases) exact. For finite and small bit rates, our results provides an approximation of the true SSIM-RDF.

In order to find the SSIM-RDF, we first show that the SSIM index can be formulated as a locally quadratic distortion measure. We then show that recent results of Linder and Zamir [7] on the RDF under locally quadratic distortion measures are applicable, and finally obtain a closed form expression for the high-resolution SSIM-RDF. We end the paper by showing how to numerically approximate the high-resolution SSIM-RDF of real images.

2. Preliminaries

In this section, we present an important existing result on rate-distortion theory for locally quadratic distortion measures and also present the SSIM index. We will need these elements when proving our main results, that is, Theorems 2 and 3 in Section 3.

2.1. Rate-Distortion Theory for Locally Quadratic Distortion Measures. Let $x \in \mathbb{R}^n$ be a realization of a source vector process and let $y \in \mathbb{R}^n$ be the corresponding reproduction vector. A distortion measure $d(x, y)$ is said to be locally quadratic if it admits a Taylor series (i.e., it possesses derivatives of all orders in a neighborhood around the points of interest) and furthermore, if the second-order terms of its Taylor series dominate the distortion asymptotically as $y \rightarrow x$ (corresponding to the high-resolution regime). In other words, if $d(x, y)$ is locally quadratic, then it can be written as $d(x, y) = (x - y)^T B(x)(x - y) + \mathcal{O}(\|x - y\|^3)$, where $B(x)$ is an input-dependent positive-definite matrix and where for y close to x, the quadratic term (i.e., $(x - y)^T B(x)(x - y)$) is dominating [7]. We use upper case X when referring to the stochastic process generating a realization x and use $h(X)$ to denote the differential entropy of X, provided it exists. The determinant of a matrix B is denoted $\det(B)$ and $\mathbb{E}$ denotes the expectation operator.

The RDF for locally quadratic distortion measures and smooth sources was found by Linder and Zamir [7] and is given by the following theorem.

Theorem 1 (see [7]). *Suppose $d(x, y)$ and X satisfy some mild technical conditions (see conditions (a)–(g) in Section II.A in [7]), then*

$$\lim_{D \to 0} \left[R(D) + \frac{n}{2} \log_2 \left(\frac{2\pi e D}{n} \right) \right]$$

$$= h(X) + \frac{1}{2} \mathbb{E} \left[\log_2 (\det(B(X))) \right], \quad (1)$$

where $R(D)$ is the RDF of X (in bits per block) under distortion $d(x, y)$, and $h(X)$ denotes the differential entropy of X. (The distribution of image coefficients and transformed image coefficients of natural images can in general be approximated sufficiently well by smooth models [14, 15]. Thus, the regularity conditions of Theorem 1 are satisfied for many naturally ocurring images.)

2.2. The Structural Similarity Index. Let $x, y \in \mathbb{R}^n$ where $n \geq 2$. We define the following empirical quantities: the sample mean $\mu_x \triangleq (1/n) \sum_{i=0}^{n-1} x_i$, the sample variance $\sigma_x^2 \triangleq (1/(n-1))(x - \mu_x)^T (x - \mu_x) = (x^T x/(n-1)) - (n\mu_x^2/(n-1))$, and the sample cross-variance $\sigma_{xy} = \sigma_{yx} \triangleq (1/(n-1))(x - \mu_x)^T (y - \mu_y) = (x^T y/(n-1)) - (n\mu_x\mu_y/(n-1))$. We define μ_y and σ_y^2 similarly.

The SSIM index studied in [10] is defined as.

$$\text{SSIM}(x, y) \triangleq \frac{\left(2\mu_x\mu_y + C_1 \right)\left(2\sigma_{xy} + C_2 \right)}{\left(\mu_x^2 + \mu_y^2 + C_1 \right)\left(\sigma_x^2 + \sigma_y^2 + C_2 \right)}, \quad (2)$$

where $C_i > 0$, $i = 1, 2$. The SSIM index ranges between -1 and 1, where positive values close to 1 indicate a small perceptual distortion. We can define a distortion "measure" as one minus the SSIM index, that is,

$$d(x, y) \triangleq 1 - \frac{\left(2\mu_x\mu_y + C_1 \right)\left(2\sigma_{xy} + C_2 \right)}{\left(\mu_x^2 + \mu_y^2 + C_1 \right)\left(\sigma_x^2 + \sigma_y^2 + C_2 \right)}, \quad (3)$$

which ranges between 0 and 2 and where a value close to 0 indicates a small distortion. The SSIM index is locally applied to $N \times N$ blocks of the image. Then, all block indexes are averaged to yield the SSIM index of the entire image. We treat each block as an n-dimensional vector where $n = N^2$.

3. Results

In this section, we present the main theoretical contributions of this paper. We will first show that $d(x, y)$ is locally quadratic and then use Theorem 1 to obtain the high-resolution RDF for the SSIM index.

Theorem 2. *$d(x, y)$, as defined in (3), is locally quadratic.*

Proof. See the appendix. □

Theorem 3. *The high-resolution RDF $R(D)$ for the source X under the distortion measure $d(x, y)$, defined in (3) and where $h(X) < \infty$ and $0 < \mathbb{E}\|X\|^2 < \infty$, is given by*

$$\lim_{D \to 0} \left[R(D) + \frac{n}{2} \log_2 (2\pi e D) \right]$$

$$= h(X) + \frac{1}{2} \mathbb{E} \left[(n-1)\log_2 (a(X)) + \log_2 (a(X) + b(X)n) \right]$$

$$+ \frac{n}{2} \log_2 (n), \quad (4)$$

where $a(X)$ and $b(X)$ are given by

$$a(X) = \frac{1}{n-1} \cdot \frac{1}{2\sigma_x^2 + C_2}, \tag{5}$$

$$b(X) = \frac{1}{n^2} \cdot \frac{1}{2\mu_x^2 + C_1} - \frac{1}{n(n-1)} \cdot \frac{1}{2\sigma_x^2 + C_2}. \tag{6}$$

Proof. Recall from Theorem 2 that $d(x, y)$ is locally quadratic. Moreover, the weighting matrix $B(X)$ in (1), which is also known as a *sensitivity* matrix [5], is given by (A.8), see the appendix. In the appendix, it is also shown that $B(x)$ is positive definite since $a(x) > 0$, $a(x) + b(x)n > 0$, for all x, where $a(x)$ and $b(x)$ are given by (5) and (6), respectively. From (A.9), it follows that

$$\mathbb{E}\left[\log_2(\det(B(X)))\right]$$
$$= \mathbb{E}\left[(n-1)\log_2(a(X)) + \log_2(a(X) + b(X)n)\right]. \tag{7}$$

At this point, we note that the main technical conditions required for Theorem 1 to be applicable is boundedness in the following sense [7]: $h(X) < \infty$, $0 < \mathbb{E}\|X\|^2 < \infty$, $\mathbb{E}[\log_2(\det(B(X)))] < \infty$, and $\mathbb{E}(\mathrm{trace}\{B^{-1}(X)\})^{3/2} < \infty$ and furthermore uniformly bounded third-order partial derivatives of $d(X, Y)$. The first two conditions are satisfied by the assumptions of the Theorem. The next two conditions follow since all elements of $B(x)$ are bounded for all x (see the proof of Theorem 2). Moreover, due to the positive stabilization constants C_1 and C_2, $\mathrm{trace}\{B(x)\}^{-1}$ is clearly bounded. Finally, it was established in the proof of Theorem 2 that the third-order derivatives of $d(X, Y)$ are uniformly bounded. Thus, the proof now follows simply by using (7) in (1). □

3.1. Evaluating the SSIM Rate-Distortion Function. In this section we propose a simple method for estimating the SSIM-RDF in practice based on real images. Conveniently, we do *not* need to encode the images in order to find their corresponding high-resolution RDF. Thus, the results in this section (as well as the results in the previous sections) are *independent* of any specific coding architecture.

In practice, the source statistics are often not available and must therefore be found empirically from the image data. Towards that end, one may assume that the individual vectors $\{x(i)\}_{i=1}^{M}$ (where $x(i)$ denotes the ith $N \times N$ subblock of the image and M denotes the total number of subblocks in the image) of the image constitute approximately independent realizations of a vector process. In this case, we can approximate the expectation by the empirical arithmetic mean, that is,

$$\mathbb{E}\left[\log_2(\det(B(X)))\right] \approx \frac{1}{M}\sum_{i=1}^{M}(n-1)\log_2(a(x(i)))$$
$$+ \log_2(a(x(i)) + b(x(i))n), \tag{8}$$

where $a(x(i))$ and $b(x(i))$ indicates that the functions a and b defined in (5) and (6) are used on the ith subblock

Table 1: Estimated $(1/2n)\mathbb{E}[\log_2(\det(B(X)))] + \log_2(N)$ values for some 512×512 8-bit grey images and block sizes $n = N^2$, $N = 4, 8$, and 16.

Image	$N = 4$	$N = 8$	$N = 16$
Baboon	−4.57	−4.77	−5.00
Pepper	−3.16	−3.51	−4.12
Boat	−3.66	−3.99	−4.45
Lena	−3.13	−3.49	−4.08
F16	−2.83	−3.14	−3.65

Table 2: Estimated $(1/n)h(x)$ (in bits/dim or equivalently bits per pixel (bpp)) for different 512×512 8-bit grey images and block sizes $n = N^2$, $N = 4, 8$ and 16.

Image	$N = 4$	$N = 8$	$N = 16$
Baboon	6.18	6.06	6.03
Pepper	4.75	4.55	4.49
Boat	5.10	4.92	4.88
Lena	4.63	4.41	4.38
F16	4.32	4.14	4.13

$x(i)$. Several estimates of $(1/2n)\mathbb{E}[\log_2(\det(B(X)))] + \log_2(N)$ using (8) are shown in Table 1, for various images commonly considered in the image processing literature.

In order to obtain the high-resolution RDF of the image, according to Theorem 3, we also need the differential entropy $h(X)$ of the image, which is usually not known a priori in practice. Thus, we need to numerically estimate $h(X)$, for example, by using the average empirical differential entropy over all blocks of the image. In order to do this, we apply the two-dimensional KLT on each of the subblocks of the image in order to reduce the correlation within the subblocks (since the KLT is an orthogonal transform, this operation will not affect the differential entropy.) Then we use a nearest-neighbor entropy-estimation approach to approximate the marginal differential entropies of the elements within a subblock [16]. Finally, we approximate $h(X)$ by the sum of the marginal differential entropies, which yields the values presented in Table 2.

4. Simulations

In this section, we use the JPEG codec on the images and measure the corresponding SSIM values of the reconstructed images. In particular, we use the baseline JPEG coder implementation available via the *imwrite* function in Matlab. Then, we compare these operational results to the information theoretic estimated high-resolution SSIM RDF obtained as described in the previous section. We are interested in the high-resolution region, which corresponds to small $d(x, y)$ values (i.e., values close to zero) or equivalently large SSIM values (i.e., values close to one). Figure 1 shows the high-resolution SSIM-RDF for $d(x, y)$ values below 0.27, corresponding to SSIM values above 0.73. Notice that the rate becomes negative at large distortions (i.e., small rates), which happens because the high-resolution assumption is clearly not satisfied and the approximations are therefore
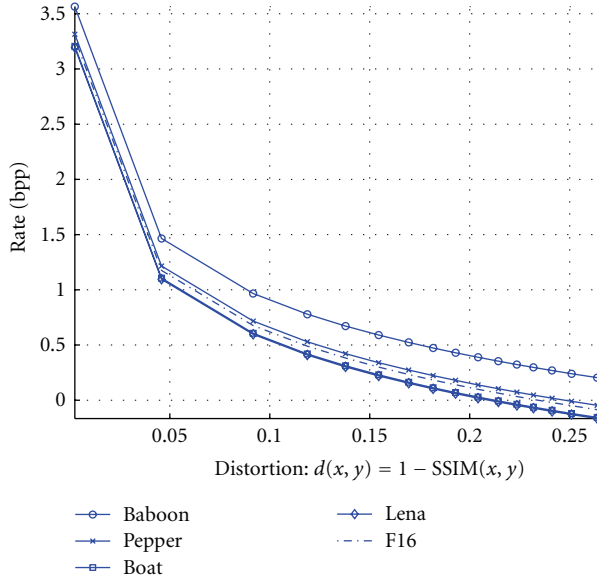
Figure 1: High-resolution RDF under the similarity measure $d(x, y) = 1 - \text{SSIM}(x, y)$ for different images and using an 8×8 block size.

Figure 2: Operational RDF using the JPEG coder on the Lena image under the similarity measure $d(x, y) = 1 - \text{SSIM}(x, y)$ for block size 8×8. For comparison we have also shown the high-resolution SSIM-RDF (thin line).

not accurate. Thus, it does not make sense to evaluate the asymptotic SSIM-RDF of Theorem 3 at large distortions.

5. Discussion

The information-theoretic high-resolution RDF characterized by Theorem 3 constitutes a lower bound on the operationally achievable minimum rate for a given SSIM distortion value. As discussed in [17], achieving the high-resolution RDF could require the use of optimal compounding, which may not be feasible in some cases. Thus, the questions of whether the RDF obtained in Theorem 3 is achievable and how to achieve it, remain open. Nevertheless, we can obtain a loose estimate of how close a practical coding scheme could get to the high-resolution SSIM-RDF by evaluating the operational performance of, for example, the baseline JPEG. Figure 2 shows the operational RDF for the JPEG coder used on the Lena image and using block sizes of 8×8. For comparison, we have also shown the SSIM-RDF. It may be noticed that the operational curve is up to 2 bpp above the corresponding SSIM-RDF (a similar behavior is observed for the other four images in the test set).

The gap between the SSIM-RDF and the operational RDF based on JPEG encoding as can be observed in Figure 2 can be explained by the following observations. First, the JPEG coder aims at minimizing a frequency-weighted MSE rather than maximizing the SSIM index. Second, JPEG is a practical algorithm with reduced complexity and is therefore not rate-distortion optimal even for the weighted MSE. Third, the differential entropy as well as the expectation of the log of the determinant of the sensitivity matrix are empirically found—based on a finite amount of image data. Thus, they are only *estimates* of the true values. Finally, the SSIM-RDF becomes exact in the asymptotic limit where the coding rate
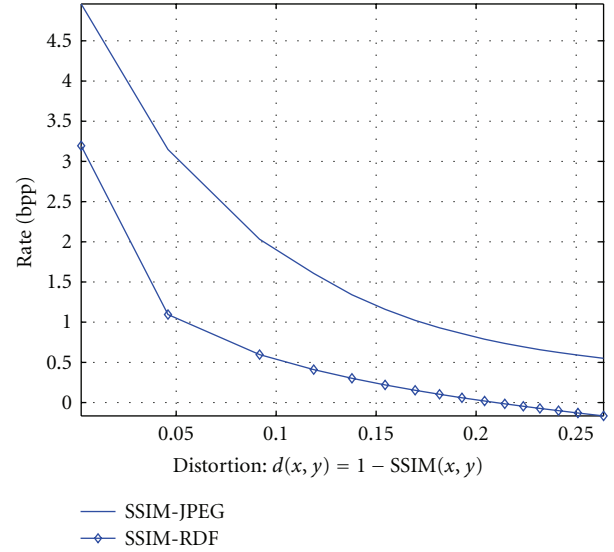
diverges towards infinity (i.e., for small distortions). At finite coding rates, it is an approximation. Nevertheless, within these limitations, the numerical evaluation of the SSIM-RDF presented here suggests that significant compression gains could be obtained by an SSIM-optimal image coder, at least at high-rate regimes. To obtain further insight into this question, the corresponding RDF under MSE distortion (MSE-RDF) for the Lena image is shown in Figure 3. We can see that the excess rate of JPEG with respect to the MSE-RDF at high rates is not greater than 1.4 bpp. This suggests that a JPEG-like algorithm aimed at minimizing SSIM distortion could reduce at least a fraction of the bit rate gap seen in Figure 2.

It is interesting to note that, in the MSE case, we have $B(x) = I$, which implies that $\log_2(|\det(B(x))|) = 0$. Thus, the difference between the SSIM-RDF and the MSE-RDF, under high-resolution assumptions, is constant (e.g., independent of the bit-rate). In fact, if the MSE is measured per dimension, then the rate difference is given by the values in Table 1, that is, $(1/2n)\mathbb{E}[\log_2(\det(B(X)))] + \log_2(N)$. It follows that the SSIM-RDF is simply a shifted version of the MSE-RDF at high resolutions. Moreover, the gap between the curves illustrates the fact that, in general, a representation of an image which is MSE optimal is not necessarily also SSIM optimal.

6. Conclusions

We have shown that, under high-resolution assumptions, the RDF for a range of natural images under the commonly used SSIM index has a simple form. In fact, the RDF only depends upon the differential entropy of the source image as well as the expected value of a function of the *sensitivity* matrix of the image. Thus, it is independent of any specific
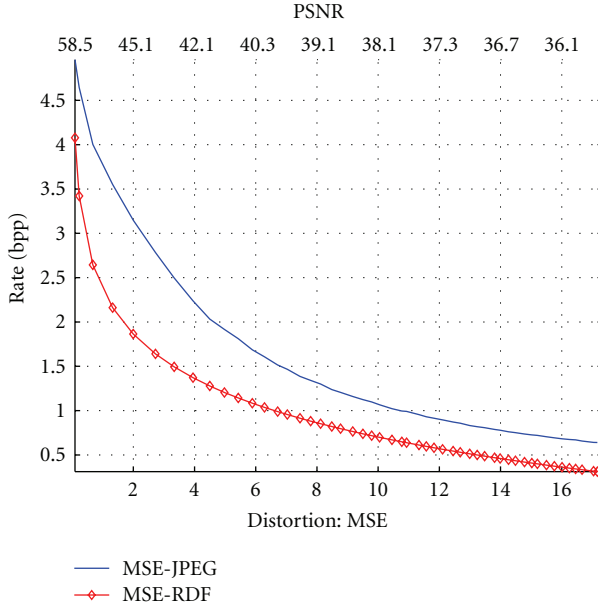
FIGURE 3: Operational RDF using the JPEG coder on the Lena image under the MSE distortion measure. For comparison we have also shown the high-resolution MSE-RDF (thin line). The horizontal axes on the top and the bottom show the PSNR and MSE, respectively.

coding architecture. Moreover, we also provided a simple method to estimate the SSIM-RDF in practice for a given image. Finally, we compared the operational performance of the baseline JPEG image coder to the SSIM-RDF and showed by approximate numerical evaluations that potentially significant perceptual rate-distortion improvements could be obtained by using SSIM-optimal encoding techniques.

Appendix

Proof of Theorem 2

We need to show that the second-order terms of the Taylor series of $d(x, y)$ are dominating in the high-resolution limit where $y \to x$. In order to do this, we show that the Taylor series coefficients of the zero- and first-order terms vanish whereas the coefficients of the second- and third-order terms are nonzero. Then, we upper bound the remainder due to approximating $d(x, y)$ by its second-order Taylor series. This upper bound is established via the third-order partial derivatives of $d(x, y)$. We finally show that the second-order terms decay more slowly towards zero than the remainder as y tends to x.

Let us define $f \triangleq ((2\mu_x\mu_y + C_1)/(\mu_x^2 + \mu_y^2 + C_1))$ and $g \triangleq ((2\sigma_{xy} + C_2)/(\sigma_x^2 + \sigma_y^2 + C_2))$ and let $h = fg$. It follows that $d(x, y) = 1 - h$ and we note that the second-order partial derivatives with respect to y_i and y_j for any i, j, are given by

$$\frac{\partial^2 h}{\partial y_i \partial y_j} = g\frac{\partial^2 f}{\partial y_i \partial y_j} + f\frac{\partial^2 g}{\partial y_i \partial y_j} + \frac{\partial f}{\partial y_i}\frac{\partial g}{\partial y_j} + \frac{\partial f}{\partial y_j}\frac{\partial g}{\partial y_i}. \quad (A.1)$$

Clearly $f|_{y=x} = g|_{y=x} = 1$, where $(\cdot)|_{y=x}$ indicates that the expression $(\cdot)$ is evaluated at the point $y = x$. Since $\partial\mu_y/\partial y_i = 1/n, \partial\sigma_y^2/\partial y_i = (2/(n-1))(y_i - \mu_y)$, and $\partial\sigma_{yx}/\partial y_i = (1/(n-1))(x_i - \mu_x)$, it is easy to show that $\partial f/\partial y_i|_{y=x} = \partial g/\partial y_i|_{y=x} = 0$, for all i. Thus, the coefficients of the zero- and first-order terms of the Taylor series of $d(x, y)$ are zero. Moreover, it follows from (A.1) that $\partial^2 h/\partial y_i\partial y_j|_{y=x} = \partial^2 f/\partial y_i\partial y_j|_{y=x} + \partial^2 g/\partial y_i\partial y_j|_{y=x}$, for all i, j. With this, and after some algebra, it can be shown that

$$\frac{\partial^2 h}{\partial y_i \partial y_j}\bigg|_{y=x}$$
$$= \begin{cases} -\dfrac{2}{n^2}\dfrac{1}{2\mu_x^2 + C_1} + \dfrac{2}{n(n-1)}\dfrac{1}{2\sigma_x^2 + C_2} & \text{if } i \neq j, \\[2ex] -\dfrac{2}{n^2}\dfrac{1}{2\mu_x^2 + C_1} - \dfrac{2}{n}\dfrac{1}{2\sigma_x^2 + C_2} & \text{if } i = j. \end{cases}$$
$$(A.2)$$

We now let $h^{(m)}$ denote the mth partial derivative of h with respect to some m variables and note that from Leibniz generalized product rule [18] it follows that $h^{(3)} = gf^{(3)} + 3g^{(1)}f^{(2)} + 3g^{(2)}f^{(1)} + g^{(3)}f$. When evaluated at $y = x$, this reduces to $h^{(3)}|_{y=x} = f^{(3)}|_{y=x} + g^{(3)}|_{y=x}$ since $f^{(1)}|_{y=x}$ and $g^{(1)}|_{y=x}$ are both zero. For the third-order derivatives of f, we have, for all i, j, k,

$$\frac{\partial^3 f}{\partial y_i\partial y_j\partial y_k}\bigg|_{y=x} = \frac{12}{n^3}\frac{\mu_x}{(2\mu_x^2 + C_1)^2}. \quad (A.3)$$

Moreover, if $i \neq j \neq k$ and $i \neq k$, we obtain

$$\frac{\partial^3 g}{\partial y_i\partial y_j\partial y_k}\bigg|_{y=x} = -\frac{4}{n(n-1)^2}\frac{1}{(2\sigma_x^2 + C_2)^2}$$
$$\times \Big[(x_i - \mu_x) + (x_j - \mu_x) + (x_k - \mu_x)\Big], \quad (A.4)$$

whereas if any two indices are equal, for example, $i \neq j = k$, we obtain

$$\frac{\partial^3 g}{\partial y_i\partial y_j\partial y_j}\bigg|_{y=x} = -\frac{8}{n(n-1)^2}\frac{x_j - \mu_x}{(2\sigma_x^2 + C_2)^2}$$
$$+ \frac{4}{(n-1)^2}\frac{(x_i - \mu_x)(1 - 1/n)}{(2\sigma_x^2 + C_2)^2}. \quad (A.5)$$

Finally, if $i = j = k$, we obtain

$$\frac{\partial^3 g}{\partial y_i\partial y_i\partial y_i}\bigg|_{y=x} = \frac{12}{(n-1)^2}\frac{(x_i - \mu_x)(1 - 1/n)}{(2\sigma_x^2 + C_2)^2}. \quad (A.6)$$

Let $\mathcal{B}$ be an n-dimensional ball of radius ϵ centered at x, let $\xi = y - x$, and let $\mathcal{T}_2(\xi)$ be the second-order Taylor series of $d(x, x + \xi)$ centered at x (i.e., at $\xi = 0$). It follows that

$$\mathcal{T}_2(\xi) \triangleq -\frac{1}{2}\sum_{i,j}\frac{\partial^2 h(x, y)}{\partial y_i\partial y_j}\bigg|_{y=x}\xi_i\xi_j = \xi^T B(x)\xi, \quad (A.7)$$

where $B(x)$ is given by half the second-order partial derivatives of $d(x, y)$, that is (see (A.2)),

$$
B(x) = \frac{1}{n^2} \frac{1}{2\mu_x^2 + C_1} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}
$$

$$
- \frac{1}{n} \frac{1}{2\sigma_x^2 + C_2} \begin{bmatrix} -1 & \dfrac{1}{n-1} & \cdots & \dfrac{1}{n-1} \\ \dfrac{1}{n-1} & -1 & \cdots & \dfrac{1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{1}{n-1} & \dfrac{1}{n-1} & \cdots & -1 \end{bmatrix},
$$

$$(A.8)$$

which has full rank and is well defined for $1 < n < \infty$. This can be rewritten as

$$
B(x) = a(x)I + b(x)J, \tag{A.9}
$$

where I is the identity matrix, J is the all-ones matrix,

$$
a(x) = \frac{1}{n-1} \frac{1}{2\sigma_x^2 + C_2}, \tag{A.10}
$$

$$
b(x) = \frac{1}{n^2} \frac{1}{2\mu_x^2 + C_1} - \frac{1}{n(n-1)} \frac{1}{2\sigma_x^2 + C_2}. \tag{A.11}
$$

Thus, $B(x)$ has eigenvalues $\lambda_0 = a(x) + b(x)n$ and $\lambda_i = a(x)$, $i = 1, \ldots, n-1$. Since $B(x)$ is symmetric, the quadratic form $\xi^T B(x)\xi$ is lower bounded by

$$
\xi^T B(x)\xi \geq \lambda_{\min}\|\xi\|^2, \tag{A.12}
$$

where $\lambda_{\min} = \min\{\lambda_i\}_{i=0}^{n-1} = \min\{a(x) + nb(x), a(x)\} > 0$, which implies that $B(x)$ is positive definite.

On the other hand, it is known from Taylor's theorem that for any $y \in \mathfrak{B}$, the remainder $\mathcal{R}_2(\xi)$, where

$$
\mathcal{R}_2(\xi) \triangleq d(x, x + \xi) - \mathcal{T}_2(\xi), \tag{A.13}
$$

is upper bounded by

$$
|\mathcal{R}_2(\xi)| < \phi \sum_{i,j,k} |\xi_i \xi_j \xi_k|, \tag{A.14}
$$

where

$$
\phi \leq \sup_{y \in \mathfrak{B}} \left| \frac{\partial^3 h}{\partial y_i \partial y_j \partial y_k} \right|, \tag{A.15}
$$

that is, ϕ is upper bounded by the supremum over the set of third-order coefficients of the Taylor series of h. Since for real images, the pixel values are finite, and since $C_i > 0$, $i = 1, 2$, it follows from (A.3)–(A.6) that the third-order derivatives are *uniformly* bounded and ϕ is therefore finite. Moreover, for

all ξ such that $\|\xi\|^2 \leq \varepsilon$, it follows using (A.7), (A.12), and (A.14) that

$$
\lim_{\|\xi\| \to 0} \frac{|\mathcal{R}_2(\xi)|}{|\mathcal{T}_2(\xi)|} \leq \lim_{\|\xi\| \to 0} \frac{\left(\max_{i \in \{1,\ldots,n\}} |\xi_i|^3\right) n^3 \phi}{\lambda_{\min}\|\xi\|^2} \tag{A.16}
$$

$$
\leq \lim_{\|\xi\| \to 0} \frac{n^3 \phi}{\lambda_{\min}} \frac{\|\xi\|^3}{\|\xi\|^2} \tag{A.17}
$$

$$
= \lim_{\|\xi\| \to 0} \frac{n^3 \phi}{\lambda_{\min}} \|\xi\| = 0, \tag{A.18}
$$

where (A.16) follows since $|\xi_i \xi_j \xi_k| \leq \max_{i \in \{1,\ldots,n\}} |\xi_i|^3$, and the sum in (A.14) runs over all possible combinations of third-order partial derivatives of a vector of length n, that is, $\sum_{i,j,k} 1 = n^3$. Furthermore, (A.17) follows by use of (A.12) and the fact that $|\xi_i|^3 < \|\xi\|^3$. Finally, (A.18) follows from the fact that ϕ is bounded by (A.15). Since the limit of (A.18) exists and is zero, we deduce that the second-order terms of the Taylor series of $d(x, y)$ are asymptotically dominating as y tends to x. This completes the proof.

Acknowledgments

References

[1] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*, Morgan Claypool Publishers, 2006.

[2] R. L. Dobrushin and B. S. Tsybakov, "Information transmission with additional noise," *IRETransactions on Information Theory*, vol. 8, pp. 293–304, 1962.

[3] R. A. McDonald and P. M. Schultheiss, "Information rates of Gaussian signals under criteria constraining the error spectrum," *Proceedings of the IEEE*, vol. 52, no. 4, pp. 415–416, 1964.

[4] D. J. Sakrison, ""The rate distortion function of a Gaussian process with a weighted square error criterion," *IEEE Transactions on Information Theory*, 1968.

[5] 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, 1995.

[6] J. Li, N. Chaddha, and R. M. Gray, "Asymptotic performance of vector quantizers with a perceptual distortion measure," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1082–1091, 1999.

[7] T. Linder and R. Zamir, "High-resolution source coding for non-difference distortion measures: the rate-distortion function," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 533–547, 1999.

[8] J. Østergaard, R. Heusdens, and J. Jensen, "On the rate loss in perceptual audio coding," in *Proceedings of the IEEE Benelux/DSP Valley Signal Processing Symposium*, pp. 27–30, Antwerpen, Belgium, March 2006.

[9] R. Heusdens, W. B. Kleijn, and A. Ozerov, "Entropy-constrained high-resolution lattice vector quantization using a perceptually relevant distortion measure," in *Proceedings of the IEEE Asilomar Conference on Signals, Systems, and Computers (Asilomar CSSC '07)*, pp. 2075–2079, Pacific Grove, Calif, USA, November 2007.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[11] Z. Wang, Q. Li, and X. Shang, "Perceptual image coding based on a maximum of minimal structural similarity criterion," in *Proceedings of the International Conference on Image Processing (ICIP '07)*, vol. 2, pp. 121–124, September 2007.

[12] J. Dahl, J. ∅stergaard, T. L. Jensen, and S. H. Jensen, "$\ell 1$ compression of image sequences using the structural similarity index measure," in *Proceedings of the Data Compression Conference (DCC '09)*, pp. 133–142, Snowbird, Utah, USA, March 2009.

[13] S. S. Channappayya, A. C. Bovik, and R. W. Heath Jr., "Rate bounds on SSIM index of quantized images," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1624–1639, 2008.

[14] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1661–1666, 2000.

[15] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural scenes," *Advances in Neural Information Processing Systems*, vol. 12, pp. 855–861, 2000.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, NY, USA, 2nd edition, 2001.

[17] T. Linder, R. Zamir, and K. Zeger, "High-resolution source coding for non-difference distortion measures: multidimensional companding," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 548–561, 1999.

[18] T. M. Apostol, *Mathematical Analysis*, Addison-Wesley, New York, NY, USA, 2nd edition, 1974.