

Low-Resolution Scalar Quantization for Gaussian Sources and Absolute Error

Daniel Marco, *Member, IEEE*, and David L. Neuhoff, *Fellow, IEEE*

Abstract—This correspondence considers low-resolution scalar quantization for a memoryless Gaussian source with respect to absolute error distortion. It shows that slope of the operational rate-distortion function of scalar quantization is infinite at the point D_{\max} where the rate becomes zero. Thus, unlike the situation for squared error distortion, or for Laplacian and exponential sources with squared or absolute error distortion, for a Gaussian source and absolute error, scalar quantization at low rates is far from the Shannon rate-distortion function, i.e., far from the performance of the best lossy coding technique.

Index Terms—Absolute error, entropy constrained quantization, Gaussian, low rate, low resolution, scalar quantization.

I. INTRODUCTION

This correspondence considers the asymptotic low-resolution, i.e., low-rate, performance of scalar quantization for a Gaussian source and absolute error distortion measure. One motivation is to increase understanding of the performance of scalar quantization. While there are well-known asymptotically accurate formulas for the rate-distortion performance of scalar quantizers for the high-resolution region, only a few results are known for the low-resolution region, see [1] and the references therein. Another motivation comes from transform coding where a sizable fraction of coefficients might need to be coded at low rates, and one would like to know how well scalar quantization performs relative to ideal lossy source coding.

The correspondence follows a somewhat similar path to that taken in [1], where a Gaussian source and squarederror distortion measure is considered. Specifically, we find the slope of the operational rate-distortion function of scalar quantization, $R(D)$, at $D = D_{\max}$, where D_{\max} is the minimum distortion attainable with zero rate. This slope determines the speed with which $R(D) \rightarrow 0$ as $D \rightarrow D_{\max}$.

Following [1], we write

$$R(D) = s \left(1 - \frac{D}{D_{\max}} \right) [1 + o_{D \rightarrow D_{\max}}]$$

where $o_{D \rightarrow D_{\max}}$ is a quantity that tends to zero as D goes to D_{\max} , and s is the magnitude of the slope with respect to normalized distortion.

The values of s in the case of exponential and Laplacian sources with both absolute and squared error distortion measures have been given in [2]; the values in the case of a uniform source and both distortion measures can be deduced from [3]; the value of s for Gaussian source and squared error was provided in [1]; finally, the value of s for a Gaussian source and absolute error is given in this correspondence. Table I summarizes these values.

Manuscript received December 30, 2005; revised November 27, 2006. The work of D. Marco was supported by the Center for the Mathematics of Information at the California Institute of Technology. The work of D. L. Neuhoff was supported by the National Science Foundation under Grant CCR-0329715. The material in this correspondence was presented at the IEEE International Symposium on Information Theory, Seattle, WA, July 2006.

D. Marco was with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: idaniel@ist.caltech.edu).

D. L. Neuhoff is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: neuhoff@eecs.umich.edu).

Communicated by W. Szpankowski, Associate Editor for Source Coding. Digital Object Identifier 10.1109/TIT.2006.890714

TABLE I
MAGNITUDE OF THE SLOPE OF THE OPERATIONAL RATE-DISTORTION FUNCTION $R(D)$ AT $D = D_{\max}$

	exponential	Laplacian	uniform	Gaussian
squared error	0	0	∞	$\frac{\log_2 e}{2}$
absolute error	1	$\log_2 e$	∞	∞

We observe from Table I that since the slopes of $R(D)$ at $D = D_{\max}$ equal 0 for exponential and Laplacian sources with squared error, they must equal the slopes of the corresponding Shannon rate-distortion functions (because the latter's magnitudes could be no larger). Further, for a Gaussian source with squared error, and for Laplacian and exponential sources with absolute error, the Shannon rate-distortion functions are known [4], [5],¹ and their slopes match the corresponding slopes of $R(D)$. Thus, in low resolution, scalar quantization for these sources and distortion measures is asymptotically optimal, i.e., as good as any quantization technique—vector or otherwise. For a uniform source with both distortion measures, and for a Gaussian source with absolute error, the slopes of $R(D)$ at $D = D_{\max}$ are negatively infinite, whereas the slopes of the corresponding Shannon rate-distortion functions cannot be negatively infinite, i.e., they must be finite (because these functions are convex). Thus, for these sources and distortion measures, low-resolution scalar quantization is far from optimal. Furthermore, the negatively infinite slopes of $R(D)$ at $D = D_{\max}$ for the mentioned sources and distortion measures imply that the corresponding operational rate-distortion functions are not convex (a fact that is already known in the case of uniform sources [3]).

The remainder of this correspondence is organized as follows. Section II provides background and introduces notation. The main result is given in Section III. Section IV offers concluding remarks. Finally, one lemma proof is left to the Appendix.

II. BACKGROUND

The assumption throughout the correspondence is that the source to be quantized is stationary, memoryless, and Gaussian with zero mean and variance σ^2 . We denote this source by $\mathcal{N}(0, \sigma^2)$.

A scalar quantizer q is a partition of the real line into cells S_k , each of which contains a reconstruction level r_k such that when the input lies in S_k , the output of the quantizer is r_k . The number of cells may be finite or infinite. P_k is the probability of the input lying in S_k . The (output) entropy of quantizer q is given by $H(q) = -\sum_k P_k \log P_k$, where all logarithms in this correspondence have base 2. The mean absolute error induced by the quantizer is

$$d(q) = \int_{-\infty}^{\infty} |x - q(x)| f(x) dx = \sum_k \int_{S_k} |x - r_k| f(x) dx$$

where f is the Gaussian density of the source. It is well known that reconstruction levels at cell medians minimize mean absolute error, for a given partition. Specifically, for a contiguous cell S_k , i.e., $S_k = [a_k, b_k)$, where it matters not if the interval is open or closed on either side, the median r_k satisfies

$$\int_{a_k}^{r_k} f(x) dx = \int_{r_k}^{b_k} f(x) dx. \quad (1)$$

¹Reference [5] makes an error in applying its Theorem 2 to compute the rate-distortion function, with respect to absolute error, of an exponential source. Specifically, for $f(x) = \alpha e^{-\alpha x}$, $\alpha > 0$, a correct application of this theorem yields $\mathcal{R}(D) = -\ln(2(1 - e^{-\alpha D}))$, rather than the formula given in [5, eq. (24)].

The operational rate-distortion function of scalar quantization for a Gaussian source with variance σ^2 and absolute error distortion measure is defined as follows:

$$R_{\sigma^2}(D) = \inf_{d(q) \leq D} H(q)$$

which specifies the least entropy of any scalar quantizer with distortion D or less.

Let D_{\max} denote the minimum distortion attainable when the rate is zero. Specifically, for a Gaussian source with variance σ^2 we have $D_{\max} = \sqrt{\frac{2}{\pi}}\sigma$.

Following the notation in [1], let the entropy function be defined as

$$\mathcal{H}(\dots, z_{-1}, z_0, z_1, \dots) = - \sum_{k=-\infty}^{\infty} z_k \log z_k$$

where $0 < z_k \leq 1$ for all k , are a finite or countably infinite set of numbers that need not sum to one. Let $o_{x,y}$ denote a quantity that converges to zero when both $x \rightarrow \infty$ and $y \rightarrow \infty$. If this quantity depends on parameters other than x and y , then its convergence to zero is uniform in such parameters. Finally, $G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ denotes the Gaussian density with zero mean and unit variance.

III. MAIN RESULT

The following lemma is used to show Theorem 2 later.

Lemma 1: Consider a scalar quantizer applied to a $\mathcal{N}(0, \sigma^2)$ source. If the cell containing the origin has boundaries $-a$ and b , has reconstruction level at the median, and contributes D_o to the mean absolute error of the quantizer, then

$$D_{\max} - D_o = \sigma \left(G\left(\frac{a}{\sigma}\right) + G\left(\frac{b}{\sigma}\right) \right) [1 + o_{a,b}].$$

Proof: Let r_o denote the median of the cell $(-a, b)$, and let f denote the Gaussian density with zero mean and variance σ^2 . We evaluate D_o as follows:

$$\begin{aligned} D_o &= \int_{-a}^{r_o} (r_o - x)f(x)dx + \int_{r_o}^b (x - r_o)f(x)dx \\ &= r_o \left[\int_{-a}^{r_o} f(x) dx - \int_{r_o}^b f(x) dx \right] \\ &\quad - \int_{-a}^{r_o} xf(x) dx + \int_{r_o}^b xf(x) dx \\ &\stackrel{(a)}{=} \int_{r_o}^b xf(x) dx - \int_{-a}^{r_o} xf(x) dx \\ &\stackrel{(b)}{=} \sigma \left(2G\left(\frac{r_o}{\sigma}\right) - G\left(\frac{a}{\sigma}\right) - G\left(\frac{b}{\sigma}\right) \right), \end{aligned}$$

where (a) is due to the fact that r_o is the median of $(-a, b)$, and (b) follows since $\int_x^\infty tG(t)dt = G(x)$. Next, we observe that

$$D_{\max} = \int_{-\infty}^{\infty} |x|f(x)dx = 2\sigma G(0).$$

Therefore,

$$D_{\max} - D_o = \sigma \left(G\left(\frac{a}{\sigma}\right) + G\left(\frac{b}{\sigma}\right) + 2 \left[G(0) - G\left(\frac{r_o}{\sigma}\right) \right] \right). \quad (2)$$

Finally, we have

$$\begin{aligned} 2\sigma \left[G(0) - G\left(\frac{r_o}{\sigma}\right) \right] &= 2\sigma \int_0^{\frac{r_o}{\sigma}} xG(x)dx \\ &\leq 2\sigma G(0) \left(\frac{r_o}{\sigma}\right)^2 \\ &= \sigma \left(G\left(\frac{a}{\sigma}\right) + G\left(\frac{b}{\sigma}\right) \right) o_{a,b} \quad (3) \end{aligned}$$

where the last equality is due to having

$$\left(\frac{r_o}{\sigma}\right)^2 = \left(G\left(\frac{a}{\sigma}\right) + G\left(\frac{b}{\sigma}\right) \right) o_{a,b}$$

as shown by Lemma A1 of the Appendix. The lemma now follows from (2) and (3). \square

The following theorem is the principal result of this correspondence.

Theorem 2: For a $\mathcal{N}(0, \sigma^2)$ source and absolute error distortion, the operational rate-distortion function of scalar quantization satisfies

$$\lim_{D \rightarrow D_{\max}} \frac{R_{\sigma^2}(D)}{D_{\max} - D} = \infty.$$

Proof: It suffices to consider only scalar quantizers with contiguous cells, as follows from [6]. By definition of $R_{\sigma^2}(D)$, for any $D \in (0, D_{\max})$ there exists a quantizer q_D such that

$$H(q_D) \leq R_{\sigma^2}(D) + \epsilon(D) \quad \text{and} \quad d(q_D) \leq D \quad (4)$$

where $\epsilon(D)$ is some function of D such that $\epsilon(D) > 0$ and $\lim_{D \rightarrow D_{\max}} \frac{\epsilon(D)}{D_{\max} - D} = 0$. (The choices of q_D and $\epsilon(D)$ are not unique, but any fixed choices will do.) Let $S_{o,D} = (-A_D, B_D)$ denote the cell of q_D containing the origin (it is immaterial if the cell is open or closed on either side). As $D \rightarrow D_{\max}$, $A_D, B_D \rightarrow \infty$. Note that either A_D or B_D (but not both simultaneously) might be infinite. Let $D_{o,D}$ be the contribution to distortion of cell $S_{o,D}$, where the reconstruction level of $S_{o,D}$ lies at the median of the cell. It follows from Lemma 1 that

$$D_{\max} - D_{o,D} = \sigma \left(G\left(\frac{A_D}{\sigma}\right) + G\left(\frac{B_D}{\sigma}\right) \right) [1 + o_{A_D, B_D}]. \quad (5)$$

Next, applying Lemma 5 from [1], which shows that $\mathcal{H}(Q(x)) = \frac{\log e}{2} xG(x) [1 + o_x]$, where $Q(x) = \int_x^\infty G(t)dt$ is the usual “ Q function,” we obtain

$$\begin{aligned} &\mathcal{H}\left(Q\left(\frac{A_D}{\sigma}\right)\right) + \mathcal{H}\left(Q\left(\frac{B_D}{\sigma}\right)\right) \\ &= \frac{\log e}{2} \left(\frac{A_D}{\sigma} G\left(\frac{A_D}{\sigma}\right) + \frac{B_D}{\sigma} G\left(\frac{B_D}{\sigma}\right) \right) [1 + o_{A_D, B_D}]. \quad (6) \end{aligned}$$

Finally, we have that

$$\begin{aligned} \liminf_{D \rightarrow D_{\max}} \frac{R_{\sigma^2}(D)}{D_{\max} - D} &\stackrel{(a)}{\geq} \liminf_{D \rightarrow D_{\max}} \frac{H(q_D) - \epsilon(D)}{D_{\max} - d(q_D)} \\ &\stackrel{(b)}{\geq} \liminf_{D \rightarrow D_{\max}} \frac{\mathcal{H}\left(Q\left(\frac{A_D}{\sigma}\right)\right) + \mathcal{H}\left(Q\left(\frac{B_D}{\sigma}\right)\right)}{D_{\max} - D_{o,D}} \\ &\stackrel{(c)}{=} \infty \end{aligned}$$

where (a) follows from (4), (b) is due to an elementary property of entropy and from having $D_{o,D} \leq d(q_D)$, and (c) derives from (5) and (6). Thus,

$$\lim_{D \rightarrow D_{\max}} \frac{R_{\sigma^2}(D)}{D_{\max} - D} = \liminf_{D \rightarrow D_{\max}} \frac{R_{\sigma^2}(D)}{D_{\max} - D} = \infty$$

as needed to show. \square

IV. CONCLUSION

This correspondence considered the asymptotic low-resolution performance of scalar quantizers for a Gaussian source with absolute error distortion measure. This performance is determined by the slope of the operational rate-distortion function of such quantizers at $D = D_{\max}$. It was shown that the slope of the operational rate-distortion function of scalar quantization is infinite, and hence does not match the slope of the Shannon rate-distortion function, which is finite. Consequently, scalar quantization is not an optimal coding technique, in asymptotically low rate, for the given source and distortion measure. This is somewhat surprising since, as noted earlier, scalar quantization is optimal for a Gaussian source and squared error distortion measure, and for Laplacian source and both squared and absolute error distortion measures.

APPENDIX

Lemma A1: Let $-a$ and b be the boundaries of the cell containing the origin for a quantizer applied to a $\mathcal{N}(0, \sigma^2)$ source. Let r_o be the median of $(-a, b)$. Then

$$\left(\frac{r_o}{\sigma}\right)^2 = \left(G\left(\frac{a}{\sigma}\right) + G\left(\frac{b}{\sigma}\right)\right)_{o_{a,b}}$$

Proof: From (1) we obtain that

$$Q\left(\frac{r_o}{\sigma}\right) = \frac{Q\left(\frac{-a}{\sigma}\right) + Q\left(\frac{b}{\sigma}\right)}{2} = \frac{1 - Q\left(\frac{a}{\sigma}\right) + Q\left(\frac{b}{\sigma}\right)}{2}. \quad (A1)$$

Next, let $v \in \mathbb{R}$ be arbitrary. If $v \geq 0$, then

$$Q(v) = \int_v^\infty G(x)dx = \frac{1}{2} - \int_0^v G(x)dx \leq \frac{1}{2} - vG(v)$$

from which it follows that $0 \leq v \leq \frac{\frac{1}{2} - Q(v)}{G(v)}$. Similarly, if $v < 0$, then

$$\begin{aligned} Q(v) &= 1 - Q(|v|) = 1 - \left(\frac{1}{2} - \int_0^{|v|} G(x)dx\right) \\ &= \frac{1}{2} + \int_0^{|v|} G(x)dx \geq \frac{1}{2} + |v|G(v) = \frac{1}{2} - vG(v) \end{aligned}$$

from which it follows that $\frac{\frac{1}{2} - Q(v)}{G(v)} \leq v < 0$. These two bounds to v imply that $v^2 \leq \left(\frac{\frac{1}{2} - Q(v)}{G(v)}\right)^2$. This is now used as follows:

$$\begin{aligned} \left(\frac{r_o}{\sigma}\right)^2 &\leq \left(\frac{\frac{1}{2} - Q\left(\frac{r_o}{\sigma}\right)}{G\left(\frac{r_o}{\sigma}\right)}\right)^2 \stackrel{(a)}{=} \left(\frac{\frac{1}{2} - \frac{1}{2} + \frac{Q\left(\frac{a}{\sigma}\right)}{2} - \frac{Q\left(\frac{b}{\sigma}\right)}{2}}{G\left(\frac{r_o}{\sigma}\right)}\right)^2 \\ &\stackrel{(b)}{=} \left(\frac{Q\left(\frac{a}{\sigma}\right) - Q\left(\frac{b}{\sigma}\right)}{\frac{2}{\sqrt{2\pi}}[1 + o_{a,b}]}\right)^2 \stackrel{(c)}{=} \left(Q\left(\frac{a}{\sigma}\right) - Q\left(\frac{b}{\sigma}\right)\right)_{o_{a,b}} \\ &\stackrel{(d)}{=} \left(\frac{G\left(\frac{a}{\sigma}\right)}{a/\sigma}[1 + o_a] - \frac{G\left(\frac{b}{\sigma}\right)}{b/\sigma}[1 + o_b]\right)_{o_{a,b}} \\ &= \left(G\left(\frac{a}{\sigma}\right)[1 + o_a]_{o_a} - G\left(\frac{b}{\sigma}\right)[1 + o_b]_{o_b}\right)_{o_{a,b}} \\ &= \left(G\left(\frac{a}{\sigma}\right) + G\left(\frac{b}{\sigma}\right)\right)_{o_{a,b}}, \end{aligned}$$

where (a) follows from (A1), (b) is obtained from the fact that $\frac{r_o}{\sigma} \rightarrow 0$ as both a and b tend to infinity, (c) follows since $Q\left(\frac{a}{\sigma}\right) - Q\left(\frac{b}{\sigma}\right) \rightarrow 0$ as both a and b tend to infinity, and (d) derives from having $Q(x) = \frac{1}{x}G(x)[1 + o_x]$ for $x > 0$, which is obtained from the fact that for any $x > 0$,

$$\frac{1}{x} \left(1 - \frac{1}{x^2}\right) G(x) < Q(x) < \frac{1}{x} G(x)$$

as shown in [7, pp. 82–83]. \square

REFERENCES

- [1] D. Marco and D. L. Neuhoff, "Low resolution scalar quantization for Gaussian sources and squared error," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1689–1697, Apr. 2006.
- [2] G. J. Sullivan, "Efficient scalar quantization of exponential and Laplacian random variables," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1365–1374, Sep. 1996.
- [3] A. György and T. Linder, "Optimal entropy-constrained scalar quantization of a uniform source," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2704–2711, Nov. 2000.
- [4] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [5] K. Yao and H. H. Tan, "Absolute error rate-distortion functions for sources with constrained magnitudes," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 4, pp. 499–503, Jul. 1978.
- [6] A. György and T. Linder, "On the structure of optimal entropy-constrained scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 416–427, Feb. 2002.
- [7] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1967.

On Divergence-Power Inequalities

Jacob Binia, *Member, IEEE*

Abstract—Expressions for (entropy-power inequality (EPI) Shannon type) divergence-power inequalities (DPIs) in two cases (time-discrete and time-continuous) of stationary random processes are given. The new expressions connect the divergence rate of the sum of independent processes, the individual divergence rate of each process, and their power spectral densities. All divergences are between a process and a Gaussian process with same second-order statistics, and are assumed to be finite. A new proof of the Shannon EPI, based on the relationship between divergence and causal minimum mean-square error (CMMSE) in Gaussian channels with large signal-to-noise ratio, is also shown.

Index Terms—Causal minimum mean-square error (CMMSE), divergence rate, divergence-power inequality, entropy-power inequality.

I. INTRODUCTION

The Shannon entropy-power inequality (EPI) is expressed with differential entropies of random variables or vectors. The aim of this correspondence is twofold. First, we give results for the (EPI Shannon type) divergence-power inequality (DPI) in two cases: time-discrete and time-continuous stationary random processes. Simple expressions connect the divergence rate of the sum of independent processes, the individual divergence rates of each process, and their power spectral densities. All divergences are between a process and a Gaussian process with same second-order statistics, and are assumed to be finite.

Second, we show in the Appendix a new proof of the EPI based on the relationship between divergence and causal minimum mean-square error (CMMSE) in Gaussian channels with large signal-to-noise ratio. The proof is similar to the new and simple one that was based on the relationship between mutual information and minimum mean-square error (MMSE) in Gaussian channels ([1], [2]).

Manuscript received September 7, 2006.

The author is with the New Elective-Engineering Services Ltd, 34970 Haifa, Israel (e-mail: biniaja@netvision.net.il).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2006.890715