# Low Rate Scalar Quantization for Gaussian Sources and Absolute Error

Daniel Marco
Electrical Engineering Department
California Institute of Technology
Pasadena CA 91125, USA
Email: idaniel@caltech.edu

David L. Neuhoff
Electrical Engineering and Computer Science Department
University of Michigan
Ann Arbor MI 48109, USA
Email: neuhoff@umich.edu

***Abstract*— This paper considers low resolution scalar quantization for a memoryless Gaussian source with respect to absolute error distortion. It shows that slope of the 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.**

## I. Introduction

This paper considers the asymptotic low resolution, i.e. low rate, performance of scalar quantization for a Gaussian source and absolute error distortion measure. It follows a path somewhat similar to that taken in [1], where a Gaussian source and squared error 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 $R(D) \to 0$ as $D \to D_{\max}$.

Following [1], we write

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

where $o_{D\to D_{\max}}$ 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 paper. Table I below summarizes these values.

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

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

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). Furthermore, 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][3], 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.

The remainder of this paper 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 paper is that the source to be quantized is stationary, memoryless and Gaussian with zero mean and variance $\sigma^2$. 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

[1]The work of D. Marco was supported by the Center of Mathematics of Information at California Institute of Technology. The work of D. L. Neuhoff was supported by NSF Grant ANI-0112801.

[2]This work was submitted to the IEEE transactions on Information Theory.

[3]Reference [5] makes a small 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 $\mathcal{R}(D) = -\ln\left(2(1 - e^{-\alpha D})\right)$, rather than the formula given in (24) of [5].

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 paper 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

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

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

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

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 $\sigma^2$ we have $D_{\max} = \sqrt{\frac{2}{\pi}}\sigma$.

Following the notation in [1], let the entropy function be defined as $\mathcal{H}(\ldots, z_{-1}, z_0, z_1, \ldots) = -\sum_{k=-\infty}^{\infty} z_k \log z_k$, where $0 < z_k \le 1$ 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 $x \to \infty$ and $y \to \infty$. If this quantity depends on parameters other than $x$ and $y$, 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 below.

*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

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

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

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

where $(a)$ is due to the fact that $r_o$ is the median of $(-a, b)$, and $(b)$ follows from having $\int_x^{\infty} t\, G(t)\, dt = G(x)$. Next, we observe that $D_{\max} = \int_{-\infty}^{\infty} |x| f(x)\, dx = 2\sigma\, G(0)$. Therefore,

$$D_{\max} - D_o = \sigma\Big(G(\frac{a}{\sigma}) + G(\frac{b}{\sigma}) + 2\big[G(0) - G(\frac{r_o}{\sigma})\big]\Big) \ . \tag{2}$$

Finally, we have

$$\begin{aligned} 2\sigma\big[G(0) - G(\frac{r_0}{\sigma})\big] &= 2\sigma \int_0^{\frac{r_0}{\sigma}} x G(x)\, dx \\ &\le 2\sigma G(0)\big(\frac{r_0}{\sigma}\big)^2 \\ &= \sigma\Big(G(\frac{a}{\sigma}) + G(\frac{b}{\sigma})\Big) o_{a,b} \ , \end{aligned} \tag{3}$$

where the last equality is due to having $\left(\frac{r_0}{\sigma}\right)^2 = \left(G\left(\frac{a}{\sigma}\right) + G\left(\frac{b}{\sigma}\right)\right) o_{a,b}$, as shown by Lemma A1 of the appendix. The lemma now follows from (2) and (3) □

The following theorem is the principle result of this paper.

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

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

*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

$$H(q_D) \ \le \ R_{\sigma^2}(D) + \varepsilon(D) \quad \text{and} \quad d(q_D) \le D \ , \tag{4}$$

where $\varepsilon(D)$ is some function of $D$ such that $\varepsilon(D) > 0$ and $\lim_{D \to D_{\max}} \frac{\varepsilon(D)}{D_{\max} - D} = 0$. (The choices of $q_D$ and $\varepsilon(D)$ 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 $D \to D_{\max}$, $A_D, B_D \to \infty$. 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

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

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

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

Finally, we have that

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

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, $\lim_{D\to D_{\max}} \frac{R_{\sigma^2}(D)}{D_{\max}-D} = \liminf_{D\to D_{\max}} \frac{R_{\sigma^2}(D)}{D_{\max}-D} = \infty$, as needed to show. □

## IV. Conclusions

This paper 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 has been 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,

$$(\frac{r_0}{\sigma})^2 \;=\; \Big(G(\frac{a}{\sigma}) + G(\frac{b}{\sigma})\Big) o_{a,b} \;.$$

*Proof:* From (1) we obtain that

$$Q(\frac{r_0}{\sigma}) \;=\; \frac{Q\big(\frac{-a}{\sigma}\big)+Q\big(\frac{b}{\sigma}\big)}{2} \;=\; \frac{1-Q\big(\frac{a}{\sigma}\big)+Q\big(\frac{b}{\sigma}\big)}{2} \;. \tag{A1}$$

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

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

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

$$\begin{aligned} (\frac{r_0}{\sigma})^2 \;&\leq\; \Big(\frac{\frac{1}{2}-Q(\frac{r_0}{\sigma})}{G(\frac{r_0}{\sigma})}\Big)^2 \;\overset{(a)}{=}\; \Big(\frac{\frac{1}{2}-\frac{1}{2}+\frac{Q(\frac{a}{\sigma})}{2}-\frac{Q(\frac{b}{\sigma})}{2}}{G(\frac{r_0}{\sigma})}\Big)^2 \\ &\overset{(b)}{=}\; \Big(\frac{Q(\frac{a}{\sigma})-Q(\frac{b}{\sigma})}{\frac{2}{\sqrt{2\pi}}[1+o_{a,b}]}\Big)^2 \;\overset{(c)}{=}\; \Big(Q(\frac{a}{\sigma})-Q(\frac{b}{\sigma})\Big)\, o_{a,b} \\ &\overset{(d)}{=}\; \Big(\frac{G(\frac{a}{\sigma})}{a/\sigma}[1+o_a] - \frac{G(\frac{b}{\sigma})}{b/\sigma}[1+o_b]\Big)\, o_{a,b} \\ &=\; \Big(G(\frac{a}{\sigma})[1+o_a]\,o_a - G(\frac{b}{\sigma})[1+o_b]\,o_b\Big)\, o_{a,b} \\ &=\; \Big(G(\frac{a}{\sigma})+G(\frac{b}{\sigma})\Big) o_{a,b} \;, \end{aligned}$$

where $(a)$ follows from (A1), $(b)$ is obtained from the fact that $\frac{r_0}{\sigma}\to 0$ as both $a$ and $b$ tend to infinity, $(c)$ follows from having $Q(\frac{a}{\sigma})-Q(\frac{b}{\sigma})\to 0$ as both $a$ and $b$ tend to infinity, and $(d)$ derives from having $Q(x) = \frac{1}{x}\,G(x)\,\big[1+o_x\big]$ for any $x>0$, which is obtained from the fact that for any $x>0$, $\frac{1}{x}(1-\frac{1}{x^2})\,G(x) < Q(x) < \frac{1}{x}\,G(x)$, as shown in [7, pp. 82-83]. □

## References


[1] D. Marco and D. L. Neuhoff, "Low resolution scalar quantization for Gaussian sources and squared error," *IEEE Trans. Info. Theory*, vol. 52, pp. 1689–1697, Apr. 2006.

[2] G. J. Sullivan, "Efficient scalar quantization of exponential and Laplacian random variables," *IEEE Trans. Info. Theory*, vol. 42, pp. 1365–1374, Sep. 1996.

[3] A. Gyorgy and T. Linder, "Optimal entropy-constrained scalar quantization of a uniform source," *IEEE Trans. Info. Theory*, vol. 46, pp. 2704–2711, Nov. 2000.

[4] T. Berger, *Rate Distortion Theory*, Prentice-Hall, Englewood Cliffs, 1971.

[5] K. Yao and H. H. Tan, "Absolute error rate-distortion functions for sources with constrained magnitudes," *IEEE Trans. Info. Theory*, vol. 24, pp. 499–503, July 1978.

[6] A. Gyorgy and T. Linder, "On the Structure of optimal entropy-constrained scalar quantizers," *IEEE Trans. Info. Theory*, vol. 48, pp. 416–427, Feb. 2002.

[7] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, John-Wiley & Sons Inc., New York, 1967.