

Rate Distortion Bounds via Threshold-based Classification

Claudio Weidmann
Communication Systems Department
Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland
Email: Claudio.Weidmann@epfl.ch

Abstract

We present upper bounds on the distortion rate function of memoryless sources with the mean squared error fidelity criterion. The bounds are particularly useful to characterize highly non-Gaussian sources with a peaked, heavy-tailed probability density function.

1 Introduction

We consider lossy encoding of a real-valued memoryless source using mean squared error as the distortion measure. The present paper will be restricted to finite variance sources with symmetric pdf, $f(x) = f(-x)$, but the results can be immediately generalized by observing that conditional variance is upper bounded by the second moment: $\mathrm{Var}(X|X \in \mathcal{S}) \leq \mathrm{E}[X^2|X \in \mathcal{S}]$.

Each source sample is classified by comparing its magnitude with a threshold T: the samples above threshold ($|x_i| \geq T$) are called *significant*, the others ($|x_i| < T$), *insignificant*. The significant samples are characterized by their ratio

$$\mu(T) = \Pr\{|X| \geq T\} = 2\int_T^\infty f(x)\,dx \tag{1}$$

and their unnormalized variance

$$A(T) = \mu(T)\,\mathrm{E}[X^2|\,|X| \geq T] = 2\int_T^\infty f(x)x^2\,\mathrm{d}x, \tag{2}$$

where $A(0) = \sigma^2$ is the variance of the source.

The classification decision is sent as side information to the decoder, using $h_b(\mu)$ nats[1] per sample. By upper bounding $D(R)$ of the significant samples and discarding the insignificant ones we obtain a low-rate bound, whereas also allocating rate to the insignificant samples will yield a high-rate bound. These bounds and some applications are presented in the remainder of the paper.

2 Low-Rate Bounds

The distortion rate function (drf) of the significant samples can be upper bounded with the drf of a Gaussian with the same variance. By adding the distortion from the discarded insignificant samples (which are quantized to zero, like in a deadzone quantizer) we can bound the drf of the original source, given a threshold $T \geq 0$ and rate $R \geq h_b(\mu(T))$:

$$D(R) \leq B(T,R) = A(T)\exp\left(-2\tfrac{R-h_b(\mu(T))}{\mu(T)}\right) + \sigma^2 - A(T). \tag{3}$$

We obtain a tighter parametric bound by temporarily fixing a threshold t and determining the rate $R^*(t)$ corresponding to the midpoint of the common tangent of two bounds $B(t,R)$ and $B(t+\Delta t, R)$, $\Delta t \to 0$. The following theorem was first presented in [4] (detailed proof in [3]):

Theorem 1 (Low-Rate Bound) *The distortion rate function of a memoryless source with symmetric pdf $f(x)$ and variance σ^2 is upper bounded by*

$$D(R^*(t)) \leq A(t)\left[\exp\left(-2\tfrac{R^*(t)-h_b(\mu(t))}{\mu(t)}\right) - 1\right] + \sigma^2, \quad \forall\, t \geq 0 : \exists R^*(t) \tag{4}$$

where the rate $R^*(t)$ is given by[2]

$$R^*(t) = h_b(\mu(t)) - \tfrac{1}{2}\mu(t)\left[2h_b'(\mu(t)) + \gamma(t) + W_{-1}\left(-\gamma(t)e^{-2h_b'(\mu(t))-\gamma(t)}\right)\right] \tag{5}$$

with the reciprocal normalized tail variance

$$\gamma(t) = \tfrac{\mu(t)}{A(t)}t^2 = \tfrac{t^2}{\mathrm{E}[X^2 |\, X \geq t]}. \tag{6}$$

To gain more insight into this bound we loosen it, in order to obtain simpler expressions. Since the bound (3) holds for all positive rates R and thresholds t, we can use an approximation to

[1] $h_b(\cdot)$ is the binary entropy function. Equations in this paper use natural logs, but the figures are labeled in bits.

[2] The expression for $R^*(t)$ involves Lambert's W function, which solves $W(x)e^{W(x)} = x$. The subscript -1 indicates the second real branch of W, taking values on $[-1, -\infty[$.

$R^*(t)$ at low rates without giving up the bounding property. From [1] we have the following series expansion for the Lambert W function:

$$W_{-1}(z) = \ln(-z) - \ln(-\ln(-z)) - \sum_{k\geq 0} \ln(1 + p_k/v_k), \tag{7}$$

where $v_{n+1} = v_n + p_n$ and $p_{n+1} = -ln(1 + p_n/v_n)$, with starting values $v_0 = \ln(-z)$ and $p_0 = \ln(-ln(-z))$, respectively. By retaining only some terms in this series we approximate (5) and obtain the following bounds, in order of increasing weakness:

1.

$$R^*(t) \approx R_1(t) = h_b(\mu(t)) + \frac{\mu(t)}{2}\left\{-\ln\gamma(t) + \ln\left[-\ln\gamma(t) + 2h_b'(\mu(t)) + \gamma(t)\right]\right\}, \tag{8}$$

$$D(R_1(t)) \leq \frac{\mu(t)\,t^2}{-\ln\gamma(t) - 2\ln\mu(t) + 2\ln(1-\mu(t)) + \gamma(t)} + \sigma^2 - A(t). \tag{9}$$

2.

$$R^*(t) \approx R_2(t) = h_b(\mu(t)) + \frac{\mu(t)}{2}\left[-\ln\gamma(t) + \ln\left(-2\ln\mu(t)\right)\right], \tag{10}$$

$$D(R_2(t)) \leq \frac{A(t)\gamma(t)}{-2\ln\mu(t)} + \sigma^2 - A(t) = \frac{\mu(t)\,t^2}{-2\ln\mu(t)} + \sigma^2 - A(t). \tag{11}$$

3.

$$R^*(t) \approx R_3(t) = h_b(\mu(t)) - \frac{\mu(t)}{2}\ln\gamma(t), \tag{12}$$

$$D(R_3(t)) \leq A(t)\gamma(t) + \sigma^2 - A(t) = \mu(t)\,t^2 + \sigma^2 - A(t). \tag{13}$$

These expressions display the interaction between $\mu(t)$, $A(t)$ and the tail parameter $\gamma(t)$ in bounding distortion rate. The weakest bound (13) can be used to upper bound the slope of $D(R)$ at $R = 0$.

Theorem 2 *Let $f(x)$ be a symmetric, finite variance pdf that satisfies the following conditions: (i) $\lim_{t\to\infty} f(t) = 0$ and $f'(t)$ exists for $t \to \infty$, and (ii) $\mu(t) > 0$ (and $A(t) > 0$) for any finite $t \geq 0$. Then the slope of $D(R)$ at $R = 0$ satisfies*

$$D'(0) \leq \lambda_0 = -2\left(\lim_{t\to\infty}\frac{f(t)}{f'(t)}\right)^2. \tag{14}$$

Proof: The main idea is that at $R = 0$ the slope of $D(R)$ must be more negative than the slope of the upper bound (13). Let $B(t) = A(t)\gamma(t) + \sigma^2 - A(t)$, the right-hand side of (13). The theorem is proved by evaluating the following limit:

$$D'(0) \le \lim_{t\to\infty} \frac{B'(t)}{R_3'(t)}.$$

□

Example Consider a source with a generalized Gaussian density $f(t) = \frac{\beta}{2\alpha\Gamma(\beta^{-1})} \exp\left(-\frac{|t|^\beta}{\alpha^\beta}\right)$ with two parameters: $\alpha > 0$ and the *shape parameter* $\beta > 0$. We differentiate f to get

$$\frac{f(t)}{f'(t)} = -\frac{\alpha^\beta}{\beta} t^{1-\beta}.$$

Taking the limit for $t \to \infty$ we can distinguish three cases:

(a) $\beta > 1$: This case includes the Gaussian density ($\beta = 2$); the slope bound is $D'(0) \le \lambda_0 = 0$. This is trivially true for any distortion rate function and hence not very useful, except for confirming the weakness of bound (13).

(b) $\beta = 1$ (Laplacian density): $D'(0) \le \lambda_0 = -2\alpha^2 = -\sigma^2$.

(c) $\beta < 1$: We have $D'(0) \le \lambda_0 = -\infty$, i.e. $D(R)$ decays very rapidly at low rates.

Thus Theorem 2 suffices to establish that for generalized Gaussians with $\beta < 1$, i.e. those which are more peaked than a Laplacian, $D(R)$ is tangent to the D axis at $(R, D) = (0, \sigma^2)$. The abrupt change in the slope bound for β around 1 is related to the tail decay of the source density, which is super-exponential for $\beta > 1$ and sub-exponential for $\beta < 1$.

The different bounds are compared in Figure 1 for two densities of the generalized Gaussian family. As the shape parameter gets smaller, the density is more peaked and the low rate decay of the bounds becomes steeper.

3 High-Rate Bound

Now we also allocate rate to the insignificant samples, i.e. the classification side information is used to switch between high-variance ($|x_i| \ge T$) and low-variance ($|x_i| < T$) codebooks.

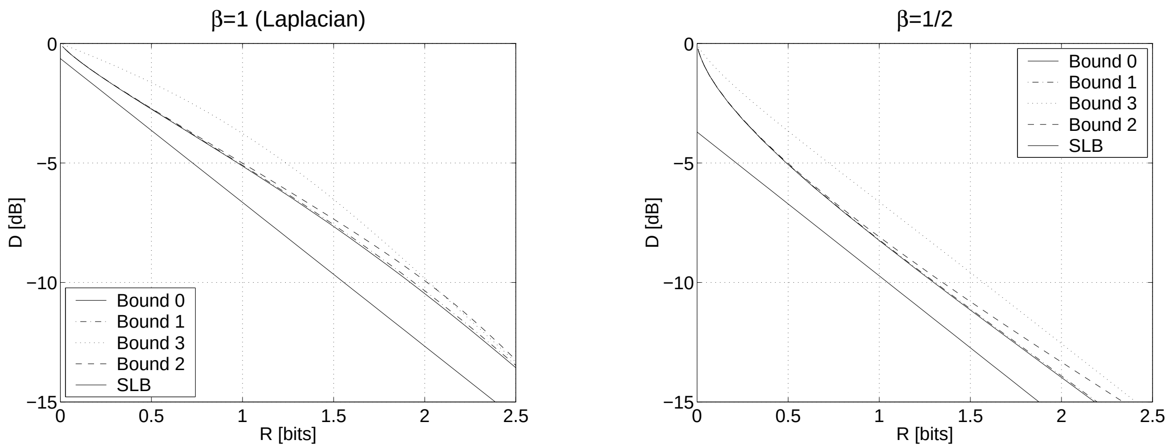

Figure 1: Comparison of upper bounds for generalized Gaussian $D(R)$. Legend: "Bound 0" is the original bound (4), "Bound 1" is Eq. (9), "Bound 2" is Eq. (11) and "Bound 3" is Eq. (13); "SLB" is the Shannon lower bound.

Theorem 3 (High-Rate Bound) *Let the variances of the insignificant and the significant samples be* $\sigma_0^2(t) = \mathrm{E}[X^2 | \, |X| < t] = \frac{\sigma^2 - A(t)}{1-\mu(t)}$ *and* $\sigma_1^2(t) = \mathrm{E}[X^2 | \, |X| \geq t] = \frac{A(t)}{\mu(t)}$*, respectively. Then for all* $R \geq R_{min}(t) = h_b(\mu(t)) + \frac{1}{2} \ln \frac{\sigma_1^2(t)}{\sigma_0^2(t)}$ *distortion rate of a memoryless source is upper bounded by*

$$D(R) \leq B_{hr}(t, R) = c(t) e^{-2R}, \tag{15}$$

where $c(t) = \exp\left[3h_b(\mu(t)) + (1-\mu(t)) \ln(\sigma^2 - A(t)) + \mu(t) \ln A(t)\right]$*. The best asymptotic upper bound for* $R \to \infty$ *is obtained by numerically searching the* $t_0 \in [0, \infty)$ *that minimizes* $c(t)$*. Since* $\lim_{t \to +0} c(t) = \sigma^2$*, the Gaussian upper bound is always a member of this family.*

The low-rate and high-rate bounds coincide in the minimum of the latter, i.e. as expected there is a smooth transition between the two bounds. The quantities t, $\mu(t)$ and $A(t)$ needed to compute the bounds are easily estimated from a sample, yielding empirical bounds as shown in Figure 2.

An Upper Bound on Differential Entropy We exploit the trivial fact that an upper bound to $D(R)$ is also an upper bound to the Shannon lower bound.

Corollary 4 *Let* $\mu_0 = \mu(t_0)$ *and* $A_0 = A(t_0)$ *be the quantities yielding the best asymptotic upper*

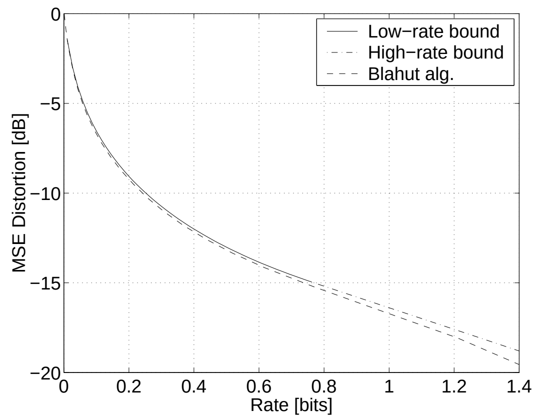

Figure 2: Empirical upper bounds and $D(R)$ for the coefficients of a wavelet image transform.

bound in Theorem 3. Define the probability mass functions

$$\boldsymbol{\mu}_0 = [\mu_0, 1-\mu_0], \quad \boldsymbol{a}_0 = [\tfrac{A_0}{\sigma^2}, 1-\tfrac{A_0}{\sigma^2}].$$

If the underlying source pdf $f(x)$ is absolutely continuous on $\mathbb{R}$, its differential entropy $h(X)$ is upper bounded by

$$h(X) \leq \tfrac{1}{2}\ln(2\pi e\sigma^2) + h_b(\mu_0) - \tfrac{1}{2}D(\boldsymbol{\mu}_0 \| \boldsymbol{a}_0). \tag{16}$$

For $t_0 = 0$, that is $\mu_0 = 1$, the bound (16) reduces to the well known Gaussian upper bound on differential entropy. We are most interested in highly compressible sources with a peaked, heavy-tailed pdf, which have a much smaller entropy than a Gaussian with the same variance. In that case the divergence term will be very large, and the side information term $h_b(\mu_0)$ becomes negligible. This entropy bound generalizes and quantifies the concept that the more confined a distribution is, the smaller its entropy [2, Sec. 20].

4 Conclusion

The presented bounds are useful tools to study the $D(R)$ behavior of the highly non-Gaussian sources (with peaked, heavy-tailed pdf) that often appear in practical compression applications. They can be estimated from a sample and thus complement Blahut's algorithm, especially at very low, resp. very high rates, where it is hard to get precise results with the latter algorithm.

Acknowledgment

The author would like to thank his thesis advisor, Prof. Martin Vetterli, for his encouragement and advice, which were crucial for this research.

References

[1] R. M. Corless, D. J. Jeffrey, and D. E. Knuth. A sequence of series for the Lambert W function. In *Proc. ISSAC '97*, pages 197–204, Maui, 1997. Also at http://www.apmaths.uwo.ca/~rcorless/papers/LambertW/.

[2] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:379–423, 623–656, July and October 1948. Also in *Claude Elwood Shannon: Collected Papers*, N.J.A Sloane and A.D. Wyner, Eds., IEEE Press 1993, pp. 5-83.

[3] C. Weidmann. *Oligoquantization in Low-Rate Lossy Source Coding*. PhD thesis, Swiss Federal Institute of Technology, Lausanne, July 2000.

[4] C. Weidmann and M. Vetterli. Rate distortion behavior of threshold-based nonlinear approximations. In *Proc. Data Compression Conference*, pages 333–342, Snowbird, Utah, March 2000.