RATE-DISTORTION RESULTS FOR GENERALIZED GAUSSIAN DISTRIBUTIONS

Aurélia Fraysse[1], *Béatrice Pesquet-Popescu*[1] *and Jean-Christophe Pesquet*[2]

(1) ENST, Signal and Image Processing Department
46, rue Barrault, 75634 Paris Cédex 13, FRANCE
(2) IGM and UMR CNRS 8049, Université de Paris-Est,
5, Bd Descartes, 77454 Marne la Vallée Cédex 2, France.
Email: {fraysse,pesquet}@tsi.enst.fr,pesquet@univ-mlv.fr

ABSTRACT

In this paper, we provide operational rate-distortion results for memoryless Generalized Gaussian sources. Close approximations of the entropy are provided for these sources, after a uniform scalar quantization at low/high resolution. Asymptotic expressions of the distortion for an arbitrary p-th order error measure are also given. The resulting approximations at low/high bitrate of the operational rate-distortion function are thus compared with the Shannon optimal bound showing the overall good performance of uniform quantization rules.

Index Terms— rate-distortion theory, entropy, quantization, Gaussian distributions, asymptotic performances

1. INTRODUCTION

The Generalized Gaussian (GG) distribution (also known as the Exponential Power distribution) provides an ubiquitous probabilistic model for a great variety of symmetrically distributed random phenomena. In particular, it has been successfully exploited to model sparse data such as the wavelet coefficients of natural images [1]. It was then commonly used in wavelet based signal/image processing tasks both for compression and restoration applications. One of the main advantages of GG distributions is to offer intermediate models between the two most popular power distributions, which are the Gaussian and Laplace laws.

In a transform coding setting, under appropriate assumptions (e.g. the orthogonality of the transform), the rate-distortion performance can be easily related to the individual characteristics of the quantized coefficients. It appears therefore useful to investigate the operational rate-distortion properties of these coefficients, by adopting a GG model for the original (unquantized) ones. However, even in the case of a uniform scalar quantization, no simple explicit formula for the rate-distortion function is available. So, one must resort to asymptotic expansions to obtain insightful expressions.

The objective of this paper is to study asymptotic rate-distortion properties of a GG memoryless source X after a symmetric scalar uniform quantization. In this direction, several developments can be found in the literature. First of all, the well-known results in [2] show the efficiency of uniform quantization at high bitrate. Furthermore, in [3], an approach for designing entropy scalar constrained quantizers for exponential and Laplace distribution was proposed and comparisons were performed with uniform quantizers. More recently in [4], the asymptotic behaviour of a uniform quantizer with centroid reconstruction levels and an offset parameter was characterized at low resolution for a memoryless Gaussian source and a squared error distortion measure. Our work will extend the latter results on two fronts, first by considering the more flexible class of GG distributions and, secondly by adopting more general distortion measures.

In Section 2, we provide close approximations of the entropy of the quantized source $\overline{X}$, from which we deduce low and high resolution expressions. In Section 3, we study the asymptotic behaviour of the distortion obtained with the p-th order moment error measure. We are particularly interested in the low resolution characteristics which allow us in Section 4 to determine the slope factors of the corresponding operational rate-distortion function. We then study the dependence of this quantity with respect to the GG distribution exponent and the parameter p and examine the relations existing with the Shannon rate-distortion bound. This leads us to discuss the cases when the scalar uniform quantization attains the optimal slope factor at low bitrate.

2. ENTROPY OF QUANTIZED GG SOURCES

We first recall some basic facts about GG random variables. The probability density function of a GG random variable is

$$\forall \xi \in \mathbb{R}, \qquad f(\xi) = \frac{\beta \omega^{1/\beta}}{2\Gamma(1/\beta)} e^{-\omega|\xi|^\beta} \tag{1}$$

where $\beta > 0$ is the exponent parameter, $\omega > 0$ is the scaling factor and Γ is the gamma function. In the following, we will restrict to "heavy tail" log-concave distributions within this class by choosing $\beta \in [1, 2]$. We can remark that, for $\beta =$

1, this density corresponds to the Laplace distribution and, for $\beta = 2$, to the Gaussian one. In addition, the differential entropy of this distribution is given in [5] by:[1]

$$h_\beta(\omega) = -\int_{-\infty}^{\infty} f(\xi)\ln f(\xi)\,d\xi = \ln\Big(\frac{2\Gamma(1/\beta)}{\beta\omega^{1/\beta}}\Big)+\frac{1}{\beta}. \quad (2)$$

Let us now define the quantized random variable of interest, which is obtained by a uniform scalar quantization rule. For a given quantization step size $q > 0$, and for X distributed according to a GG law, the quantized random variable $\overline{X}$ is given by:

$$\overline{X} = r_0 = 0, \qquad \text{if } |X| < \tfrac{q}{2}, \quad (3)$$

and, for all $i \in \mathbb{Z}$ such that $i \neq 0$,

$$\overline{X} = r_i, \qquad \text{if } (|i| - \tfrac{1}{2})q \leq |X| < (|i| + \tfrac{1}{2})q \quad (4)$$

where the quantization levels are given by

$$\forall i \geq 1, \qquad r_i = -r_{-i} = (i+\zeta)q \quad (5)$$

with $-1/2 \leq \zeta \leq 1/2$. Note that we will not consider any saturation effect. The parameter ζ serves to adjust the values of the quantizer reconstruction levels. The most commonly used quantization rule corresponds to reconstruction levels located at the midpoints of the decision intervals, that is $\zeta = 0$. This rule is usually used in wavelet based image compression techniques and constitutes the basic ingredient of many encoding strategies.

Let us now turn our attention to the entropy of $\overline{X}$. This one is given by:

$$H_f(\overline{q}) = -\sum_{i=-\infty}^{\infty} \mathsf{P}(\overline{X} = r_i)\ln \mathsf{P}(\overline{X} = r_i) \quad (6)$$

where $\overline{q} = \omega^{1/\beta} q$ is the normalized quantization step-size. Except in the Laplace case, where $\beta = 1$, this entropy cannot be expressed in a simple manner. However it can be written in terms of the incomplete Gamma function, for which asymptotic approximations are known [6]. Let us define the normalized incomplete Gamma function Q_a, $a > 0$, as follows:

$$\forall \xi \in \mathbb{R}, \qquad Q_a(\xi) = \frac{1}{\Gamma(a)}\int_0^{\xi} \theta^{a-1}e^{-\theta}d\theta. \quad (7)$$

For all $i \geq 1$, let $\overline{q}_i = (i - 1/2)\overline{q}$. It can be noticed that the probability of the r_i reconstruction level is expressed as

$$\forall i \geq 1, \quad \mathsf{P}(\overline{X} = r_i) = \mathsf{p}_i = \frac{1}{2}\Big(Q_{1/\beta}(\overline{q}_{i+1}^{\beta}) - Q_{1/\beta}(\overline{q}_i^{\beta})\Big), \quad (8)$$

whereas the zero reconstruction level occurs with probability $\mathsf{P}(\overline{X} = r_0) = \mathsf{p}_0 = Q_{1/\beta}(\overline{q}_1^{\beta})$. From the fact that the density function f is decreasing over $\mathbb{R}_+$ and the positivity of the Kullback-Leibler divergence, we then deduce the following approximation of H_f:

$$H_f(\overline{q}) = H_f^{(n)}(\overline{q}) + \Delta_n \quad (9)$$

where

$$H_f^{(n)}(\overline{q}) = -\mathsf{p}_0 \ln \mathsf{p}_0 - 2\sum_{i=1}^{n-1} \mathsf{p}_i \ln \mathsf{p}_i + \big(h_\beta(1) - \ln\overline{q}\big)\big(1 - Q_{1/\beta}(\overline{q}_n^{\beta})\big) + \frac{\overline{q}_n}{\Gamma(1/\beta)}\, e^{-\overline{q}_n^{\beta}} \quad (10)$$

and

$$0 \leq \Delta_n \leq \overline{\Delta}_n = \frac{\beta\overline{q}}{\Gamma(1/\beta)}\Big(\frac{2n}{2n-1}\Big)^{\beta-1} e^{-\overline{q}_n^{\beta}}. \quad (11)$$

As we can see in Fig. 1, $H_f^{(2)}$ provides a tight lower approximation of H_f, whereas $H_f^{(3)} + \Delta_3$ gives a close upper bound.

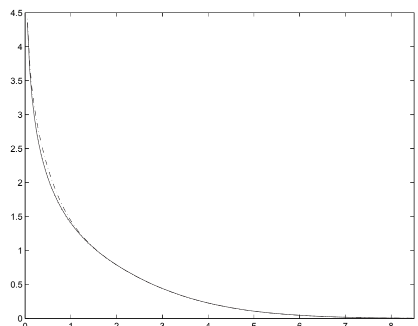

Fig. 1. Entropy of a uniformly quantized GG source (solid line) versus normalized quantization step for $\beta = 4/3$, and its lower (almost superimposed) and upper (dashdot line) approximations.

From (9) and using the approximations of Q_a in [6], we can derive the following asymptotic formulas for $H_f(\overline{q})$. When $\overline{q} \to 0$,

$$H_f(\overline{q}) = h_\beta(1) - \ln\overline{q} + O(\overline{q}). \quad (12)$$

The two first terms in the right-hand side actually correspond to Bennett's formula [2]. More interestingly, we have at low resolution when $\overline{q} \to \infty$,

$$H_f(\overline{q}) = -\mathsf{p}_0 \ln \mathsf{p}_0 - (1-\mathsf{p}_0)\ln\Big(\frac{1-\mathsf{p}_0}{2}\Big) + O(\overline{q}e^{-(3\overline{q}/2)^{\beta}}) \quad (13)$$

$$= \frac{\overline{q}e^{-\overline{q}^{\beta}/2^{\beta}}}{2\Gamma(1/\beta)}\Big(1 + O\Big(\frac{\ln\overline{q}}{\overline{q}^{\beta}}\Big)\Big). \quad (14)$$

Eq. (14) provides a more accurate approximation than (13). However, (13) shows that the low resolution behaviour of the entropy is similar to that of a 3-state discrete source. The quality of these approximations can be observed in Fig. 2.

[1] For simplicity, the entropies will be expressed in Nats.

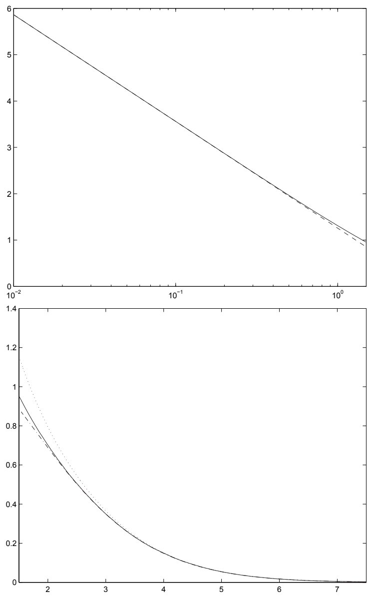

Fig. 2. High resolution (top) and low resolution (bottom) approximations of the entropy versus $\overline{q}$ for $\beta = 3/2$. The approximation in (12) is plotted in dashed line, those in (13) and (14) are plotted in dashdot line and dotted line, respectively.

3. ASYMPTOTIC DISTORTION RESULTS

We now focus on the distortion expressed in terms of the p-th order moment of the quantization error. We subsequently define the normalized distortion as:

$$\overline{d}_{p,\zeta}(\overline{q}) = \omega^{p/\beta}\mathsf{E}\{|X-\overline{X}|^p\} \quad (15)$$
$$= 2\omega^{p/\beta}\left(\int_0^{\frac{q}{2}} \xi^p f(\xi)d\xi + \sum_{i=1}^{\infty}\int_{(i-\frac{1}{2})q}^{(i+\frac{1}{2})q} |\xi - r_i|^p f(\xi)d\xi\right) \quad (16)$$

where p is any real exponent greater than or equal to 1. Notice that $p = 2$ corresponds to the mean square error criterion and $p = 1$ to the mean absolute one. We will see in the next section that considering other values of p may be of interest.

In the following, we derive asymptotic expressions of the distortion at low and high resolution. As f is a decreasing function over $\mathbb{R}_+$, and using once again approximations of the incomplete Gamma function, we find, for $\overline{q} \to 0$,

$$\overline{d}_{p,\zeta}(\overline{q}) = \frac{\nu\overline{q}^p}{p+1}\big(1 + O(\overline{q})\big), \quad (17)$$

where $\nu = \left(\frac{1}{2}+\zeta\right)^{p+1} + \left(\frac{1}{2}-\zeta\right)^{p+1}$. When $\zeta = 0$ and $p = 2$, the classical formula for the mean square quantization error can be recognized.

In addition, at low resolution, different approximations of the distortion are obtained, depending on the values of p and ζ. Three cases can be distinguished:

- if $\zeta \neq -1/2$ or $p \geq 2$, then

$$\overline{d}_{p,\zeta}(\overline{q}) = \mu_p - \frac{\overline{q}^{p+1}e^{-\overline{q}^\beta/2^\beta}}{2^{p+1}\Gamma(1/\beta)\tilde{q}^\beta} \left(1 - (1+2\zeta)^p + \frac{p}{\beta\tilde{q}^\beta}\big(1 + (1+2\zeta)^{p-1}\big) + O\Big(\frac{1}{\overline{q}^{2\beta}}\Big)\right); \quad (18)$$

- if $\zeta = -1/2$ and $p = 1$, then

$$\overline{d}_{p,\zeta}(\overline{q}) = \mu_1 - \frac{\overline{q}^2 e^{-\overline{q}^\beta/2^\beta}}{4\Gamma(1/\beta)\tilde{q}^\beta}\left(1 + O\Big(\frac{1}{\overline{q}^{2\beta}}\Big)\right); \quad (19)$$

- if $\zeta = -1/2$ and $1 < p < 2$, then

$$\overline{d}_{p,\zeta}(\overline{q}) = \mu_p - \frac{\overline{q}^{p+1}e^{-\overline{q}^\beta/2^\beta}}{2^{p+1}\Gamma(1/\beta)\tilde{q}^\beta}\left(1 + \frac{p}{\beta\tilde{q}^\beta} + O\Big(\frac{1}{\overline{q}^{p\beta}}\Big)\right), \quad (20)$$

where

$$\mu_p = \omega^{p/\beta}\frac{\mathsf{E}[|X|^p]}{\epsilon} = \frac{\Gamma\big((p+1)/\beta\big)}{\Gamma(1/\beta)}. \quad (21)$$

The proof of these results is quite technical. We refer to [7] for more details. It can also be remarked that, for $\zeta = 0$, the zeroth order term in (18) vanishes, thus yielding the simpler formula

$$\overline{d}_{p,\zeta}(\overline{q}) = \mu_p - \frac{p\overline{q}^{p+1}e^{-\overline{q}^\beta/2^\beta}}{2^p\beta\Gamma(1/\beta)\tilde{q}^{2\beta}}\left(1 + O\Big(\frac{1}{\overline{q}^\beta}\Big)\right). \quad (22)$$

In contrast, when $\zeta \neq 0$, the zeroth order term is prevalent. In particular, when $\zeta > 0$, we have, for all $p \geq 1$, $\overline{d}_{p,\zeta}(\overline{q}) > \mu_p$ when $\overline{q}$ is large enough. This shows that, as expected, choosing $\zeta > 0$ is a poor reconstruction strategy. An illustration of these results is provided in Fig. 3.

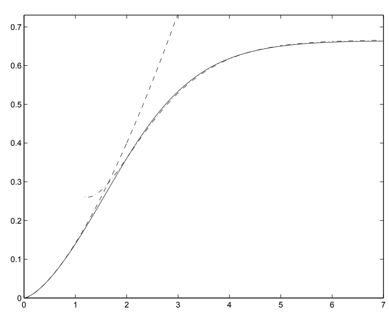

Fig. 3. Distortion (solid line) versus $\overline{q}$ and its low resolution (dashdot line) and high resolution (dash line) approximations, when $p = \beta = 3/2$.

4. EFFICIENCY OF SCALAR UNIFORM QUANTIZATION

Using the approximations of the entropy and distortion derived in two previous sections, we are now able to state the main results of this paper by calculating the operational rate-distortion function defined by

$$\forall \overline{D} > 0, \qquad R_p(\overline{D}) = \inf_{\zeta \in [-1/2, 1/2]} R_{p,\zeta}(\overline{D}). \tag{23}$$

where

$$\forall \zeta \in [-1/2, 1/2], \qquad R_{p,\zeta}(\overline{D}) = \inf_{\{\overline{q}>0 \mid \overline{d}_{p,\zeta}(\overline{q}) \leq \overline{D}\}} H_f(\overline{q}). \tag{24}$$

Obviously, according to (17), when $\overline{D}$ tends to zero, so does $\overline{q}$. Therefore, combining (17) with (12) yields, when $\overline{D} \to 0$,

$$R_p(\overline{D}) = h_\beta(1) - \ln 2 - \frac{1}{p} \ln((p+1)\overline{D}) + O(\overline{D}^{1/p}). \tag{25}$$

In addition, for small distortion, it can be verified that the optimum rate-distortion performance is obtained by taking $\zeta = 0$, that is when the reconstruction levels are the midpoints of the decision intervals.

From low resolution results, and by invoking arguments similar to those in [4], we can quantify the slope of the tangent line to the rate-distortion curves at the extreme point corresponding to a zero rate and a μ_p distortion. We get

$$\lim_{\overline{D} \to (\mu_p)^-} \frac{R_p(\overline{D})}{\mu_p - \overline{D}} = \begin{cases} \infty & \text{if } p < \beta \\ 1 & \text{if } p = \beta \\ 0 & \text{if } p > \beta. \end{cases} \tag{26}$$

At this point, it is interesting to compare the above results with the Shannon optimal bound. We recall the expression of the Shannon rate-distortion function [8]:

$$\mathcal{R}_p(\overline{D}) = \inf_{\{\widehat{X} \mid \mathsf{E}[|X-\widehat{X}|^p] \leq \omega^{-p/\beta}\overline{D}\}} \mathcal{I}(X; \widehat{X}) \tag{27}$$

where $\mathcal{I}(X; \widehat{X})$ is the mutual information between the GG random variable X of interest and some arbitrary real-valued random variable $\widehat{X}$ defined on the same probability space. When $\overline{D} \to 0$, the asymptotic expression of $\mathcal{R}_2$ is well-known [8]. This result can be easily extended to any order p:

$$\mathcal{R}_p(\overline{D}) = h_\beta(1) - h_p(1) - \frac{1}{p} \ln(p\overline{D}) + o(1). \tag{28}$$

This shows that, at high bitrate, the performance loss related to the uniform scalar quantization is limited by $\ln \Gamma(1 + 1/p) + \frac{1}{p} + \frac{1}{p} \ln(\frac{p}{1+p}) + o(1)$, which for $p = 1$ corresponds to a maximum difference of about 0.4427 bit.

At low bitrate, we find that for $p \geq \beta$, the slope factor of R_p is equal to that of $\mathcal{R}_p$. Thus, in this case, uniform quantization is an asymptotically optimal coding method for GG sources. On the other hand, when $p < \beta$, the quantization procedure is not optimal anymore. Furthermore, for $p = \beta$, it can be proved that the optimal asymptotic performance is obtained when $\zeta = -1/2$, that is for positive quantization levels which are equal to the lower bound of the associated decision intervals. This emphasizes the suboptimality of midpoint reconstruction levels (i.e. $\zeta = 0$) in spite of their frequent use in practice. This behaviour can be compared with the result in [4] for centroid reconstruction levels, in the Gaussian case using a mean square error criterion. In the latter case, the centroids of the positive decision intervals are given, when $\overline{q} \to \infty$, by

$$r_i^* = (i - 1/2)q\big(1 + O(q^{-\beta})\big). \tag{29}$$

So, at low bitrate, the centroids indeed converge to the lower bounds of the decision intervals.

5. CONCLUSION

In this paper, we have derived accurate approximations of the operational rate-distortion function of a uniformly quantized GG source at high resolution level. However, this remains somehow suboptimal with respect to the Shannon optimal bound. Furthermore, at low bitrate, asymptotic formula for both the entropy and the distortion have been given. These have allowed us to determine the slope factor of the operational rate-distortion function. This result generalizes those in [4] as we show that the slope factor of the rate-distortion function R_p is the optimal one, provided that the order p of the distortion measure is greater than or equal to the exponent β and the positive (resp. negative) quantizer reconstruction levels are chosen equal to the lower (resp. upper) bounds of the decision intervals.

6. REFERENCES

[1] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, 1989.

[2] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. on Information Theory*, vol. 14, no. 5, pp. 676–683, 1968.

[3] G. Sullivan, "Efficient scalar quantization of exponential and Laplacian random variables," *IEEE Trans. on Information Theory*, vol. 42, no. 5, pp. 1365–1374, 1996.

[4] D. Marco and D. Neuhoff, "Low-resolution scalar quantization for Gaussian sources and squared error," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1689–1697, 2006.

[5] W. Szepanski, "Δ-entropy and rate distortion bounds for generalized Gaussian information sources and their application to image signals," *Electronics Letters*, vol. 16, no. 3, pp. 109–111, 1980.

[6] W. Gautschi, "The incomplete gamma functions since Tricomi," in *Tricomi's ideas and contemporary applied mathematics, Atti dei Convegni Lincei*, Accademia Nazionale dei Lincei, Roma, 1998, number 147, pp. 203–237.

[7] A. Fraysse, B. Pesquet Popescu, and J.-C. Pesquet, "On the uniform quantization of a class of sparse sources," submitted, 2007.

[8] T. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.