Rate Distortion Function for Alpha-Stable Sources

Ercan E. Kuruoglu and Jia Wang

Abstract

In this paper, we develop a numerical approximation based on the Blahut-Arimoto algorithm to the rate distortion function of sources with alpha-stable distribution both for the symmetric and the skewed cases and provide bounds for its lossy coding performance.

Index Terms

Rate distortion function, alpha-stable distribution, Blahut-Arimoto algorithm.

I. INTRODUCTION

Heavy-tailed or impulsive phenomena abound in many real life signal processing applications. Many man made signals such as web teletraffic or web transmission times [1], email based communications, and timing of individual human actions [2], SAR images of urban areas [3], or natural signals such as astronomical images [4] demonstrate impulsive characteristics.

In particular, it has been observed by various researchers that the wavelet coefficients of audio [6], and various types of images show heavy tailed characteristics and suggested Laplace distribution [5], generalized Gaussian distributions [7], Cauchy distribution [8] and α-stable distributions [10], [11].

E.E. Kuruoglu is with Istituto di Scienza e Technologie dell'Informazione, Consiglio Nazionale delle Ricerche, I-56124 Pisa, Italy, ercan.kuruoglu@isti.cnr.it.

Jia Wang is with the Institute of Image Communication and Information Processing, Electronics Engineering Department, Shanghai Jiao Tong University, China 200240, jiawang@sjtu.edu.cn.

Manuscript received May XX, 2015; revised XXX XX, 2015.

These observations seem to be contadicting the central limit theorem which indicates the Gaussian distribution for processes formed as the summation of a large number of variables. However, there exists a generalised version of the central limit theorem which states that the sum of a large number of random variables with power-law tail distributions decreasing as $1/|x|^{\alpha+1}$ where $0 < \alpha < 2$ (and therefore having infinite variance) will tend to a stable distribution $S_\alpha(x)$ as the number of variables approaches infinity [9].

This theoretical justification has provided the alpha-stable distributions with wide acceptance as impulsive data models which is also supported experimentally by works such as, [10] where Achim et al. show its superiority to Laplace distribution in modelling the wavelet coefficients of biomedical ultrasound images and in [11] for SAR images. Despite the potential of α-stable distributions in modelling various types of impulsive and skewed data, unfortunately, to the best of our knowledge, there is no work on the source coding properties of α-stable distribution and its rate-distortion function other than in [8] where Tsakalides and Nikias indicated a number of open problems for the study of source coding systems for heavy-tailed distributions.

Recently, with the increasing popularity of compressed sensing and the need for efficient modelling of sparse data, α-stable distributions have been used in the framework of compressed sensing in works such as [12], [13], [14].

It is, therefore, of fundamental importance to develop the rate distortion function of this distribution family which would provide us with bounds in lossy and lossless source coding and insight into the limit of success of compressed sensing schemes with alpha-stable data models. In particular, the rate distortion function can help us with the design of more realistic source coders and compressed sensing schemes. In this paper, we develop a numerical approximation to the rate distortion function for alpha-stable distributions.

II. Alpha-stable distribution

The alpha-stable distributions have received great interest in the last decade due to their success in modeling data which are too impulsive to be accommodated by the Gaussian distribution. The alpha-stable distribution family was known since early 20th century and it became a popular model for various type of signals.

The alpha-stable distribution is a generalization of the Gaussian distribution that accommodates for impulsive and skewed characteristics as well. The distributions are defined by their characteristic functions:

$$\phi(t) = \exp[i\delta t - |\gamma t|^{\alpha} B_{t,\alpha}] \tag{1}$$

$$B_{t,\alpha} = \begin{cases} 1 - i\beta \mathrm{sgn}(t) \tan(\frac{\pi\alpha}{2}) & if \quad \alpha \neq 1 \\ 1 + i\beta \mathrm{sgn}(t) \frac{2}{\pi} \log|t| & if \quad \alpha = 1 \end{cases} \tag{2}$$

where $\alpha \in (0, 2], \beta \in [-1, 1], \gamma \in (0, \infty), \delta \in (-\infty, \infty) \in -\infty < \delta < +\infty$. $\alpha(0 < \alpha \leq 2)$ is the characteristic exponent and sets the degree of impulsiveness of the distribution (Fig. 2). The smaller the value of α, the greater the frequency and the size of extreme events. For $\alpha = 2$, the distribution corresponds to the Gaussian distribution; $\alpha = 1$ and $\beta = 0$ corresponds to the Cauchy distribution. β is the symmetry parameter and determines the skewness of the distribution (Fig. 3). $\beta = 0$ implies that the distribution is symmetric. γ is the scale parameter, which measures of the spread of the samples from a distribution around the mean (Fig. 4) similar to the role of variance for the Gaussian distribution. δ is the location parameter and basically corresponds to a shift in the x-axis of the probability density function.

III. Rate distortion function for alpha-stable distribution

The fundamental results of rate-distortion theory are due to Shannon's coding theorems [15] which provide an achievable bound on the performance of source coding methods. This bound is often expressed as the rate-distortion function $R(D)$ for a given source and separates the regions that can or cannot be attained by any coding system. The rate distortion function is formally defined as:

$$\min_{Q_{Y|X}(y|x)} I_Q(Y;X) \text{ subject to } D_Q \leq D^* \tag{3}$$

where D_Q and D^* are the distortions between X and Y for a given $Q_{Y|X}(y\,|x\,)$ and the prescribed maximum distortion, respectively. When the mean squared error is used as distortion measure, for continuous amplitude signals, we have:

$$D_Q = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_X(x)Q_{Y|X}(y\,|x\,)(x-y)^2 dxdy \tag{4}$$

An analytical solution to this minimization problem is often difficult to obtain.

For calculating the rate-distortion function of discrete-alphabet sources, there exists an elegant numerical algorithm developed by Blahut, known as the Blahut-Arimoto (BA) algorithm [16]. For a continuous-alphabet source, the source can be approximated by a discrete source-alphabet using a high-rate scalar quantizer and the Blahut-Arimoto algorithm may be used on this discretised source to approximate the $R(D)$ curve.

It is known that the BA algorithm converges to the rate distortion function for discrete sources [16], [18]. For a continuous source, since we first discretize/quantize it to a discrete source before applying the BA algorithm, one may wonder whether this two-step procedure converges to the rate distortion function of the original continuous source. This is indeed true and we briefly sketch the proof as follows:

Assume $\tilde{X}$ to be the discretized version of the continuous source X and $\hat{X}$ is the optimal description achieving $R_{\tilde{X}}(D)$ which is the rate-distortion function of the discrete source $\tilde{X}$. Then suppose $\mathbb{E}d(X,\hat{X}) = D+\epsilon$, where ϵ is a positive real number. Due to the fact that $X - \tilde{X} - \hat{X}$ form a Markov chain and $\tilde{X}$ is a function of X, $I(X;\hat{X}) = I(\tilde{X};\hat{X})$. Thus

$$I(\tilde{X};\hat{X}) \geq R(D+\epsilon), \tag{5}$$

where $R(D)$ is the rate-distortion function of X. Moreover, for given positive real number ϵ', one can make the quantization step size sufficiently small such that for any $\hat{X}'$ with $\mathbb{E}d(X,\hat{X}') \leq D-\epsilon'$, we have $\mathbb{E}d(\tilde{X},\hat{X}') \leq D$. Thus

$$I(\tilde{X};\hat{X}) \leq R(D-\epsilon'). \tag{6}$$

TABLE I

APPROXIMATION ERROR FOR VARIOUS NUMBER OF SAMPLES. (R_T: THEORETICAL $R(D)$ FUNCTION OF GAUSSIAN DISTRIBUTION, R_E: EXPERIMENTAL $R(D)$ FUNCTION USING THE BA ALGORITHM.

# of samples	$S = \sum(\|R_T - R_E\| \times \Delta D)$
8	0.2928
16	0.0621
32	0.0148
64	0.0027
128	0.0015

Then it can be seen that by letting $\epsilon, \epsilon' \to 0$, $R_{\tilde{X}}(D)$ converges to $R(D)$ due to the continuity of the rate distortion function.

IV. EXPERIMENTAL RESULTS

A. *Effect of the number of samples on the approximation of $R(D)$*

For $\alpha = 2$, the α-stable distribution corresponds to the Gaussian distribution with mean δ and the variance is $2\gamma^2$. The rate distortion function for a $N(0, \sigma^2)$ source with squared error distortion measure (in the rest of this paper we use this distortion measure) can be analytically calculated to be [17]:

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D} & 0 \le D \le \sigma^2 \\ 0 & D > \sigma^2 \end{cases} \tag{7}$$

When the number of samples increases, the experimental results become better approximations to the theoretical results (Fig. 1a). As can be seen clearly in Fig. 1.b and Table I, 128 samples are enough for a very close approximation.

For smaller α, for example $\alpha = 1$, as the number of samples increases, we also can see that the gap between the two experimental curves become smaller (Fig. 1.c). In this case, even 64 samples are enough.

B. *$R(D)$ curves for various α values*

α determines the weight in the tails. At a particular distortion, the minimum rate description required increases as α decreases (Fig. 2). That is, for more impulsive data, we need higher description length to have the same distortion. Therefore, source coding schemes based on the Gaussian assumption can lead to significantly more distortion if the source is α-stable with small α.

C. $R(D)$ curves for various β values

As α decreases, the effect of β on the pdf (and R(D)) becomes more pronounced: the left tail gets lighter and lighter for $\beta > 0$. In Fig. 3b, we can see that, at a particular distortion value, the minimum rate description required decreases as the parameter β increases, that is, the more skewed the distribution is, higher compression rate can be achieved; however, the effect of the symmetry parameter β is less pronounced in the $R(D)$ function when compared to α.

D. $R(D)$ curves for various γ values

At a particular distortion, the minimum rate description required increases as the dispersion, γ, increases (Fig. 4), that is, the more dispersed the data is, the higher rate of compression can be achieved for the same distortion. As in α, the effect of γ is very pronounced on the $R(D)$ curve.

E. $R(D)$ curves for various δ values

When δ varies, the curves of rate-distortion function are the same (Fig. 5). This can be explained by the fact that the entropy and hence mutual information) are independent of the location parameters and from Equation 3 also is $R(D)$.

V. Conclusions

For α-stable sources, we have demonstrated experimentally that the BA algorithm can be used to approximate the $R(D)$ curve of the alpha-stable distribution. The dependence of the $R(D)$ curve on the parameter values is demonstrated. The $R(D)$ function is very sensitive to the value of the shape parameter α and the scale parameter γ, while it is not affected by the location parameter and is affected only slightly by the symmetry parameter, β. It has been observed that for impulsive sources, i.e. for small α, for a given distortion, required rate description is larger. With the curves presented in this paper, we provide bounds for the achievable rates for impulsive sources modeled with alpha stable distributions which we believe can be useful in the design of realistic source coding systems.

ACKNOWLEDGMENT

E.E. Kuruoglu gratefully acknowledges the partial support of 111 project of Shanghai Jiao Tong University Image Communications and Information Processing Institute and partial support from the short term mobility program of CNR, Italy. J. Wang acknowledges J. Chen of Mc-Master University for discussions on convergence.

REFERENCES

[1] X.Y. Zhu and J. Yu and J. Doyle :'Heavy tails, generalized coding, and optimal Web layout', INFOCOM, pp. 1617-1626, 2001.

[2] A. Vazquez, J.G. Oliveira, Z. Dezs, K.-I. Goh, I. Kondor and A.-L. Barabsi, 'Modeling bursts and heavy tails in human dynamics,' Physical Review E 73, 036127, 2006.

[3] E.E. Kuruoglu,and J. Zerubia, 'Modeling SAR Images with a generalization of the Rayleigh distribution', IEEE Trans. Image Process., 13, pp. 527-533, Apr. 2004.

[4] D. Herranz, E. E. Kuruoglu and L. Toffolatti, "An alpha-stable approach to model point source P(D) distributions in CMB sky maps," Astronomy and Astrophysics, vol. 424, no. 3, 1081-1096, August 2004.

[5] S.G. Chang, B. Yu and M. Vetterli, "Wavelet thresholding for multiple noisy image copies," IEEE Trans. Image Processing, vol. 9, no. 9, pp. 1631-1635, 2000.

[6] N. Ruiz, M. Rosa, F. Lopez, D. Martinez and R. Mata, "New algorithm for searching minimum bit rate wavelet representations with application to multiresolution-based perceptual audio coding," Proceedings 15th International Conference on Pattern Recognition (ICPR-2000), pp. 286-289, Sep. 3-7, 2000 Barcelona, Spain.

[7] S. Kasaei, M. Deriche and B. Boashash, "An efficient quantization technique for wavelet coefficients of fingerprint images," Signal Processing, vol. 62, no. 3, pp. 361-366, 1997.

[8] P. Tsakalides and P. Reveliotis and C.L. Nikias, 'Scalar Quantization of Heavy-Tailed Signals', IEE Proceedings. Vision, Image and Signal Processing, 147(5), Oct. 2000, pp.475-484

[9] G. Samorodnitsky and M. Taqqu, Non-Gaussian Stable Processes: Stochastic Models with Infinite Variance, Chapman and Hall, London, 1994.

[10] A.M. Achim, A. Bezerianos, P. Tsakalides, 'Novel Bayesian multiscale method for speckle removal in medical ultrasound images,' IEEE Trans. Medical Imaging, 20 (8), pp. 772-783, 2001.

[11] A.M. Achim, P. Tsakalides, A. Bezerianos, 'SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling,' IEEE Trans. Geoscience and Remote Sensing, 40 (8), pp. 1773-1784, 2003.

[12] A. Achim, B. Buxton, G. Tzagkarakis, P. Tsakalides, Compressive sensing for ultrasound RF echoes using ?-stable distributions (2010) 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10, art. no. 5626210, pp. 4304-4307.

[13] G. Tzagkarakis, P. Tsakalides, Greedy sparse reconstruction of non-negative signals using symmetric alpha-stable distributions (2010) European Signal Processing Conference, pp. 417-421.

[14] D. Otero, G.R. Arce, Generalized restricted isometry property for alpha-stable random projections (2011) ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, art. no. 5947148, pp. 3676-3679.

[15] C.E. Shannon: 'Coding theorems for a discrete source with a fidelity criterion,' IRE Nat. Conv. Rec., 7, 1959, pp. 142-163.

[16] R.E. Blahut, Computation of channel capacity and rate-distortion function, IEEE Trans. Inform. Theory, IT-18(4), July, 1972, pp. 460-473.

[17] T.M. Cover and J.A. Thomas, Elements of Information Theory, Wiley-Interscience press, 1991.

[18] I. Csiszar, On the computation of rate-distortion function, IEEE Trans. Inform. Theory, IT-20(2), 1974, pp.122-124.

Fig. 1. (a) R(D)s calculated using BA algorithm for the Gaussian distribution ($\alpha = 2, \beta = 0, \gamma = 1, \delta = 0$), with varying number of samples. (b) Zooming in to show the detail of (a). (c) R(D) functions calculated using BA algorithm for the Cauchy distribution ($\alpha = 1.0, \beta = 0, \gamma = 1, \delta = 0$), with varying sample numbers.

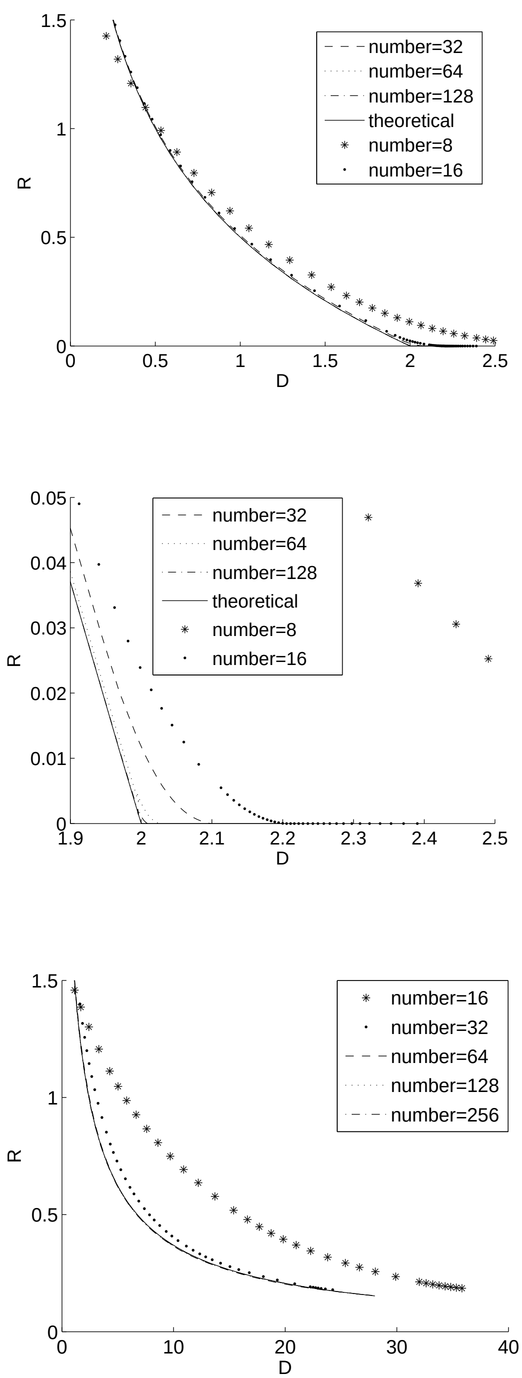
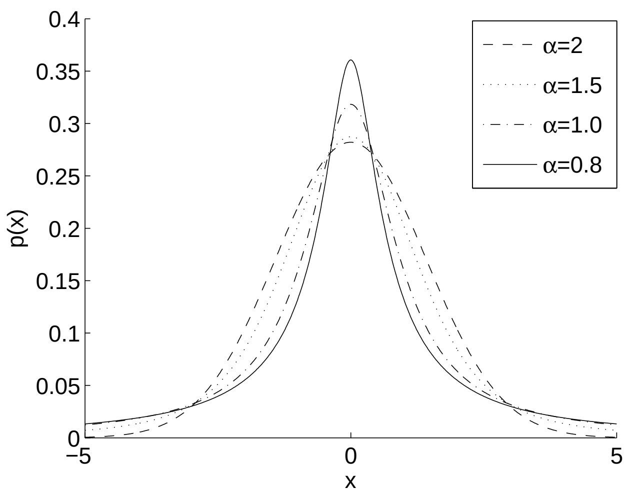


Fig. 2. (a) Probability density function for the α-stable distribution ($\beta = 0, \gamma = 1, \delta = 0$), with varying α. (b) R(D) functions calculated using BA algorithm for these α-stable distributions



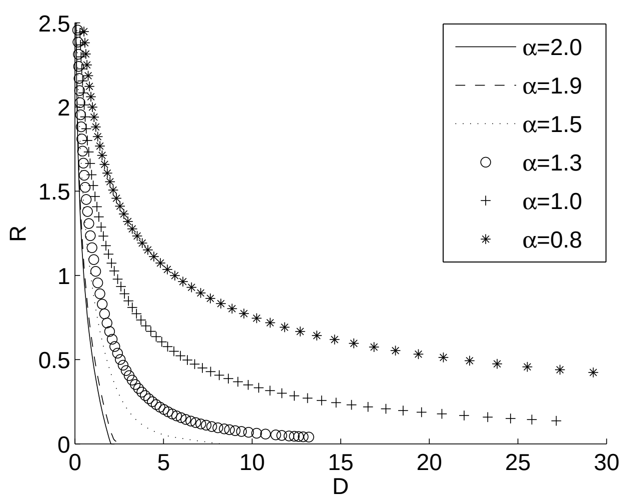


Fig. 3. (a) Probability density function of the α-stable distribution ($\alpha = 0.8, \gamma = 1, \delta = 0$), with varying β. (b) R(D) functions calculated using BA algorithm

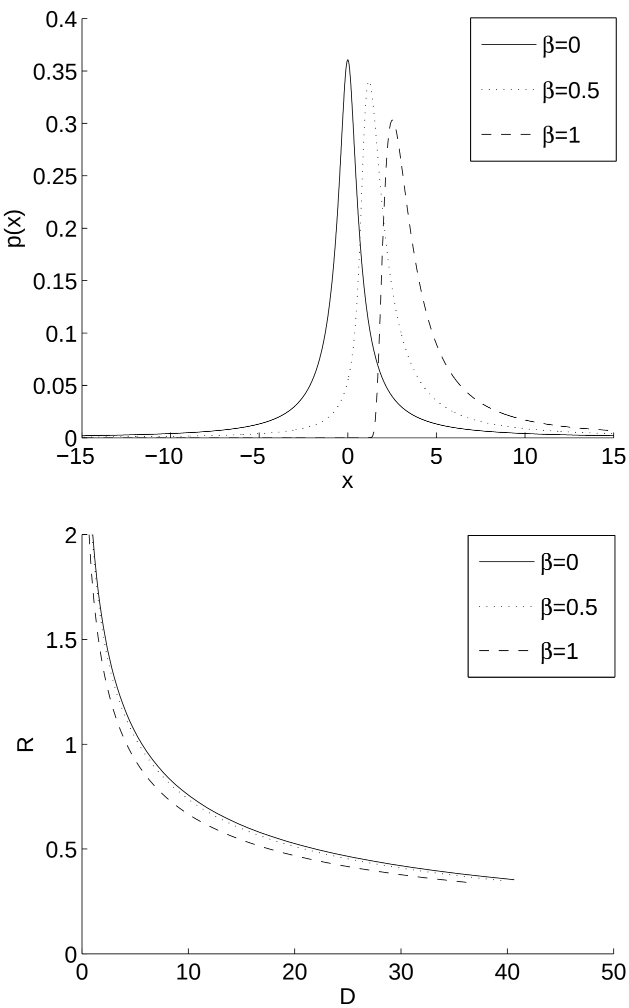


Fig. 4. (a) Probability density function of the α-stable distribution ($\alpha = 1.5, \beta = 0, \delta = 0$), with varying γ (b) R(D) functions calculated using BA algorithm

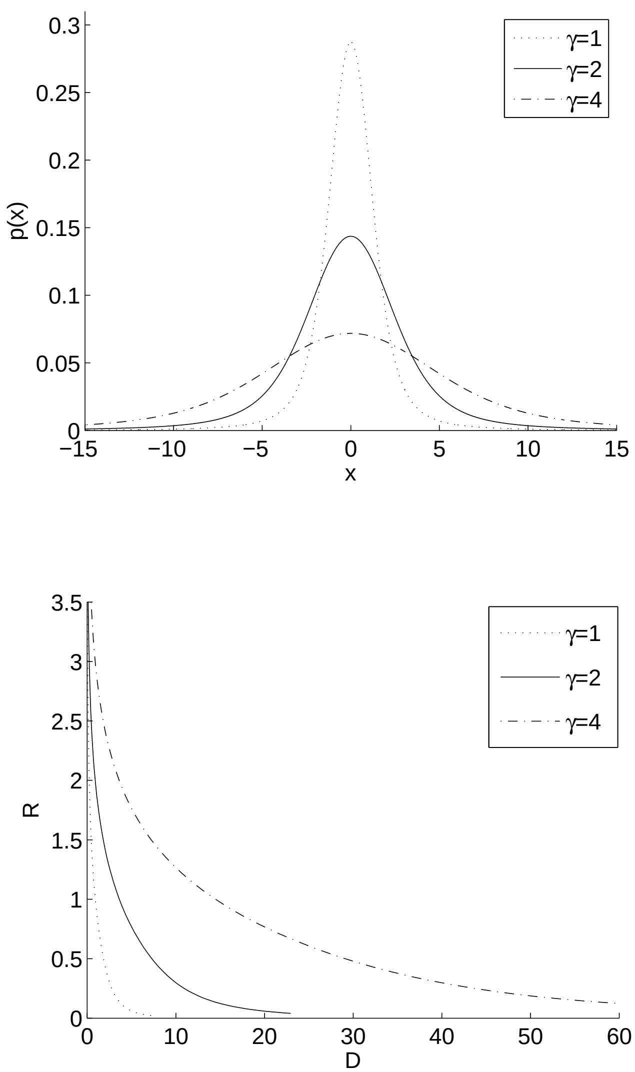


Fig. 5. R(D)functions calculated using BA algorithm for the α-stable distribution ($\alpha = 1.5, \beta = 0, \gamma = 1$), with varying δ.