Scalar quantisation of heavy-tailed signals

P.Tsakalides, P.Reveliotis and C.L.Nikias

Abstract: Efficient stochastic data processing presupposes proper modelling of the statistics of the data source. The authors address the issues that arise when the data to be processed exhibits statistical properties which depart significantly from those implied under the Gaussianity assumption. First, they present a study on the modelling of coefficient data obtained when applying the wavelet transform (WT) to images. They show that WT coefficients are heavy-tailed and can be modelled with alpha-stable distributions. Then, they introduce an alternative to the common mean-square error (MSE) quantiser for the efficient, scalar quantisation of heavy-tailed data by means of distortion minimisation. The proposed quantiser is based on a particular member of the family of alpha-stable distributions, namely the Cauchy distribution, and it employs a distortion measure based on the mean square root absolute value of the quantisation error. Results of the performance of this quantiser when applied to simulated as well as real data are also presented.

1 Introduction

The processing of data in digital form is intrinsically related to the problem of quantisation. The precise formulation of this problem was first addressed in the literature by Max in [1] for the case of the MSE criterion. Similar results were also obtained by Lloyd in [2]. These papers give the necessary conditions for the design of the optimal quantiser, by means of the minimisation of the MSE. They also present the thresholds for the optimal quantiser in the case of data following the Gaussian distribution. Sufficient conditions for the optimality of a MSE quantiser were first investigated by Fleischer [3] and further examined by Trushkin [4] for a broader class of optimisation criteria. In general, two approaches have been proposed in the literature for the optimal design of a quantiser. The first is based on an iterative method for the solution of a system of equations, determining the stationary points of the distortion measure, which results in a locally optimum quantiser [1, 2]. The second method is based on a computationally demanding search, using dynamic programming [5]. It is also worth noting that nonsymmetric quantisation schemes have been found to be optimal for certain symmetric distributions and for symmetric error weighting functions [5, 6].

Classical work on optimal quantiser design has assumed that sources may be fitted to a certain fixed statistical model. Recent applications of variable bit rate transmission of multimedia data over communication networks require adaptive quantisers that estimate the model statistics on the fly and without *a priori* information. Ortega and Vetterli [7] have proposed one such scalar quantiser that adapts its parameters such as bin size, reconstruction level, codeword length, and dynamic range, to the changing statistics of the source signal.

Although the optimal quantisation problem has been completely solved for data following certain distributions including the Gaussian, Laplacian or Rayleigh, applications exist where the data does not follow any of these distributions, following instead a so-called heavy-tailed marginal distribution. This type of data has been observed in such diverse fields as telecommunications, finance and economics, radar and sonar, and speech and image compression. Examples include file lengths, CPU time to complete a job, interarrival times between packets in network communications [8, 9], stock returns and interest rate movements in economics [10], clutter returns in radar [11], and as demonstrated in this paper, coefficients in state-of-the-art image coders based on wavelets.

In the past, optimum entropy constrained quantisers have been designed for a class of non-Gaussian memoryless sources [12]. In this paper, we claim that alpha-stable distributions are sufficiently flexible and rich to appropriately model wavelet coefficients in image coding applications. Our modelling results give rise to new questions related to rate distortion theory problems within the alpha-stable framework.

2 Data modelling with alpha-stable distributions

In this Section, we introduce the statistical model that will be used to describe sources of a heavy-tailed nature. The model is based on the class of symmetric α-stable ($S\alpha S$) distributions and is well-suited for characterising distributions that exhibit heavy tails. A review of the state of the art of stable processes from a statistical point of view is provided by a collection of papers edited by Cambanis *et*

IEE Proceedings online no. 20000470

DOI: 10.1049/ip-vis:20000470

Paper first received 16th June 1999 and in final revised form 21st March 2000

P. Tsakalides is with the VLSI Design Laboratory, Department of Electrical and Computer Engineering, University of Patras, 261 10 Rio, Greece

P. Reveliotis is with Philips Research, Software Systems and Architecture, Briarcliff Manor, NY 10510, USA

C.L. Nikias is with the Integrated Media Systems Center, Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2561, USA

al. [13], while textbooks in the area have been written by Samorodnitsky and Taqqu [14], and by Nikias and Shao [15].

2.1 Stable distribution and its properties

The appeal of $S\alpha S$ distribution as a statistical model for signals derives from some important theoretical and empirical reasons. First, stable random variables satisfy the stability property which states that linear combinations of jointly stable variables are indeed stable. Secondly, stable processes arise as limiting processes of sums of independent, identically-distributed (i.i.d.) random variables via the generalised central limit theorem. Actually, the only possible nontrivial limit of normalized sums of i.i.d. terms is stable. However, strong empirical evidence suggests that many data sets in several physical and economic systems exhibit heavy-tail features that justify the use of stable models [16].

The symmetric α-stable ($S\alpha S$) distribution is best defined by its characteristic function

$$\varphi(\omega) = \exp(j\delta\omega - \gamma|\omega|^{\alpha}) \quad (1)$$

where α is the characteristic exponent. Taking values $0 < \alpha \leq 2$, $\delta(-\infty < \delta < \infty)$ is the location parameter and γ $(\gamma > 0)$ is the dispersion of the distribution. For values of α in the interval (1,2], the location parameter δ corresponds to the mean of the $S\alpha S$ distribution, while for $0 < \alpha \leq 1$, δ corresponds to its median. The dispersion parameter γ determines the spread of the distribution around its location parameter δ, similar to the variance of the Gaussian distribution.

The characteristic exponent α is the most important parameter of the $S\alpha S$ distribution and it determines the shape of the distribution. The smaller the characteristic exponent α is, the heavier are the tails of the $S\alpha S$ density. This implies that random variables following $S\alpha S$ distributions with small characteristic exponents are highly impulsive. Gaussian processes are stable processes with $\alpha = 2$ while Cauchy processes result when $\alpha = 1$. In fact, no closed-form expressions for the general $S\alpha S$ probability density function (pdf) are known except for the Gaussian and the Cauchy members.

Although the $S\alpha S$ density behaves approximately like a Gaussian density near the origin, its tails decay at a lower rate than the Gaussian density tails [14]. Indeed, let X be a non-Gaussian $S\alpha S$ random variable. Then, as $x \rightarrow \infty$

$$P(X > x) \sim c_{\alpha} x^{-\alpha} \quad (2)$$

where $c_{\alpha} = \Gamma(\alpha)\sin(\pi\alpha/2)/\pi, \Gamma(\cdot)$ is the Gamma function, and the statement $h(x) \sim g(x)$ as $x \rightarrow \infty$ means that $\lim_{x\rightarrow\infty} h(x)/g(x) = 1$. Hence, the tail probabilities are asymptotically power laws. In other words, while the Gaussian density has exponential tails, the stable densities have algebraic tails. Fig. 1 shows the tail behaviour of several $S\alpha S$ densities including the Cauchy and the Gaussian. We should note that because eqn. 2 gives exactly the tail probability of the Pareto distribution, the term stable Paretian laws is used to distinguish between the fast decay of the Gaussian law and the Pareto-like tail behaviour when $\alpha < 2$.

The alpha-stable tail power law provided one of the earliest approaches in estimating the stability index α of real measurements [14]. The empirical distribution of the data, plotted on a log–log scale, should approach a straight line with slope $-\alpha$ if the data is stable. Another approach is based on quantiles [17]. Maximum likelihood (ML) methods developed by DuMouchel [18] and by Brorsen and Yang [19] are asymptotically efficient but were considered difficult to compute. Recently, Nolan [20] showed that ML estimation of stable parameters is feasible by designing an efficient program.

One consequence of heavy tails is that only moments of order less than α exist for the non-Gaussian alpha-stable family members, i.e.

$$E|X|^{p} < \infty \quad \text{for} \quad p < \alpha \quad (3)$$

As a result, stable Paretian laws have infinite variance. In the past, the infinite variance property of the $S\alpha S$ family has caused skeptics to dismiss the stable model. With the same reasoning, one could argue that the routinely used Gaussian distribution, which has infinite support, should also be dismissed as a model of bounded measurements. In practice, one should remember that it is important to capture the shape of the distribution and that the variance is only one measure of the spread of a density [20].

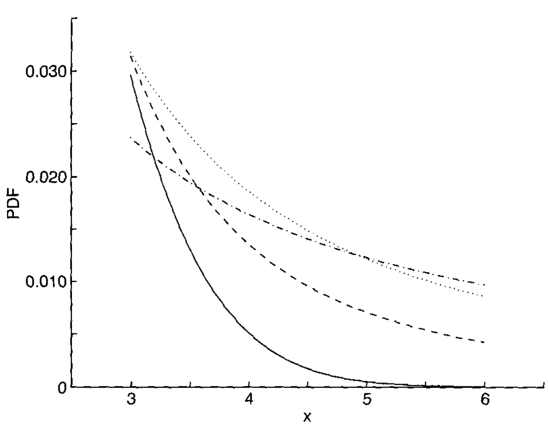


Fig. 1 *Tail behaviour of symmetric alpha-stable densities for $\alpha = 0.5$, 1.0 (Cauchy), 1.5, and 2.0 (Gaussian)*

- · - · $\alpha = 0.5$
· · · · · $\alpha = 1.0$
– – – $\alpha = 1.5$
—— $\alpha = 2.0$

2.2 Alpha-stable modelling of wavelet coefficient data

In this Section, we present results on the modelling of coefficient data obtained when applying the two-dimensional (2-D) wavelet transform in a set of test images. In image compression applications, the 2-D wavelet transform is applied along both the vertical and horizontal directions decomposing the image into four regions referred as 'image subbands.' The LL subband contains the low-pass information and is essentially a low-resolution version of the original image. The HL (LH) subband contains high- (low-) pass information horizontally and low- (high-) pass information vertically. Finally, the HH subband contains the high-pass information in both directions.

The first step in compressing wavelet data is to apply a quantiser to the coefficients. After quantisation, an entropy coder compresses the quantisation indices. Nearly all popular coders partition the subband data into different classes taking into account the different statistics in each class [21, 22]. Wavelet data in different subbands contribute to the quality of the reconstructed image in different degrees. Hence, an optimal bit allocation is needed among the subbands. Instead of directly controlling the exact number of bits assigned to a subband, one can instead adjust the quantisation step size to each subband. Choosing the optimal quantiser based on the data statistics of each subband may result in better overall compression and coding results.

Having in mind these observations, our goal is to study whether the stable family of distributions provides a flexible modelling tool for wavelet data. We proceed in two steps: First, we assess whether the data deviate from the normal distribution and if they have heavy tails. To determine that, we make use of the normal probability plots. If the normality assumption is violated, we test whether the data has infinite variance by employing the so-called 'converging variance' test [23]. Then, we check if the data is in the stable domain of attraction by estimating the characteristic exponent α directly from the data and by providing the related confidence intervals. Several methods have been proposed for estimating stable parameters. Here, we use the maximum likelihood method described by Nolan in [24], which gives reliable estimates and provides the most tight confidence intervals. As further stability diagnostics, we employ the so-called 'stabilised' p–p plots defined by Michael [25] as well as probability density plots that give a good indication of whether the $S\alpha S$ fit matches the data near the mode and on the tails of the distribution.

Fig. 2–4 illustrate the above described two-step process and show comparative results on the modelling of the statistics of the wavelet coefficients of the 'Lena' image. The normal probability plot of Fig. 2 provides strong evidence that the underlying distribution is not normal. The plus signs in the plot show the empirical probability against the data value for each point in the sample. The plus signs are in a curve that does not follow the straight Gaussian line and thus, the normality assumption is violated for this data. Fig. 3 depicts one of the oldest tests for determining whether the data has finite variance [23]. It plots the sample variance $S_n^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$, based on the first n observations, as a function of n for two representative subrecords of the wavelet coefficients. If the data came from a population with finite variance, then by the law of large numbers, S_n^2 should converge to the population variance $\sigma^2 = E(X-\mu)^2$ as n increases. Otherwise, as is the case here, S_n^2 oscillates no matter how large the sample is and the graph typically shows large jumps.

While non-Gaussian stable densities are heavy-tailed, not all heavy-tailed distributions are stable. Hence, in Fig. 4 we assess the stability of the data. First, the characteristic exponent is estimated and the data sample is fitted with the

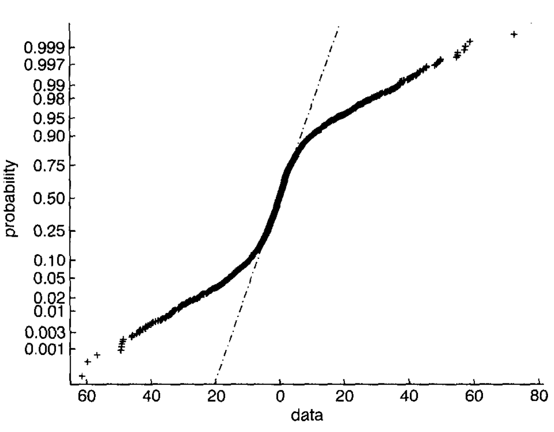


Fig. 2 *Two-step modelling process for wavelet coefficients of 'Lena' image: characterisation of data non-Gaussianity and heavy tails - normal probability plot*

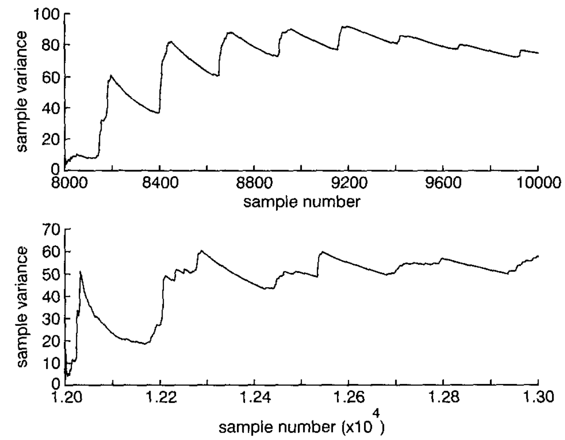


Fig. 3 *Two-step modelling process for wavelet coefficients of 'Lena' image: characterisation of data non-Gaussianity and heavy tails - converging variance test for two different data regions*

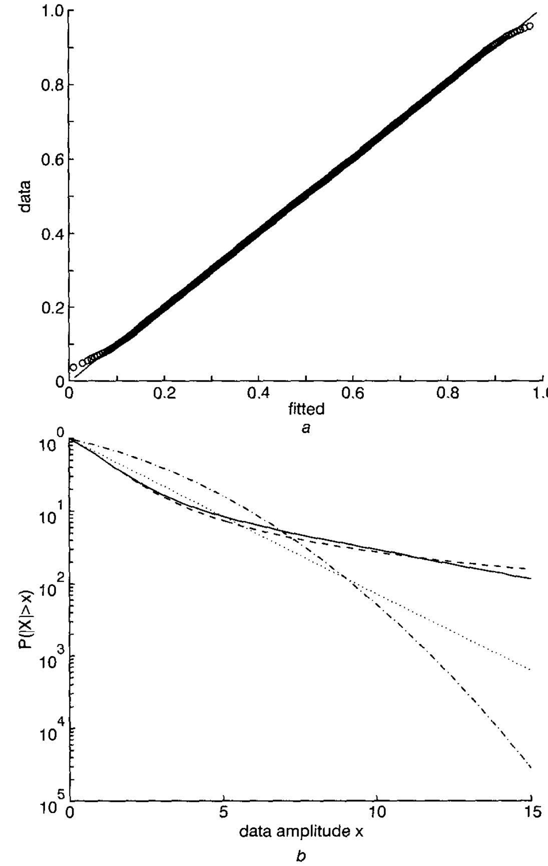


Fig. 4 *Two-step modeling process for the wavelet coefficients of the 'Lena' image: diagnostics for assessing stability*

a Stabilised p–p plot for $S\alpha S$ fit
b Amplitude probability density (APD) comparison. Gaussian has standard deviation $\sigma = 3.578$, $S\alpha S$ has characteristic exponent $\alpha = 1.362$ and dispersion $\gamma = 1.710$, Laplacian has parameter $\lambda = 0.492$
Note that the $S\alpha S$ distribution models both the mode and the tails of the empirical density more accurately than the Gaussian or the Laplace distributions
——— empirical
– – – – – $S\alpha S$
. . . . Laplacian
– · – · – Gaussian

corresponding stable distribution. For the particular case shown here, the characteristic exponent of the $S\alpha S$ distribution which best fits the data was estimated to be $\hat{\alpha} = 1.362$. Naturally, the real question is whether the stable fit accurately describes the data. The p–p and density plots in Figs. 4*a* and *b*, respectively, show a close stable fit. In particular, Fig. 4*b* shows that the $S\alpha S$ distribution is superior to the Gaussian and Laplacian distributions because it provides a better fit to both the mode and the tails of the empirical density of the actual data.

We also applied the 2-D Haar wavelet transform to a set of test images, which can be found in Matlab's Wavelet Toolbox. Because of limited space, we cannot present the complete modelling analysis as we did for the 'Lena' image. Instead, we iterated a separable four-band decomposition three times (octave-band decomposition) and we observed the statistics of each subband, i.e. we modelled the wavelet coefficients of each subband by using the alpha-stable family. As noted in the bibliography [26], the residual low-pass band exhibits Gaussian statistics, while all the other wavelet subbands have diverse and highly varying statistical properties because the image data are subjected to multiple combinations of high-pass and low-pass filtering. Hence, coding algorithms that exploit local variations in statistics have been successfully applied to the field of lossless and lossy image coding.

Our findings in terms of wavelet coefficient characterisation for the test images are summarised in Table 1, which shows the ML estimates of the stability index α and the associated 95% confidence intervals. At each decomposition level (scale), first we model one subband at a time and then we group the subbands together considering the data a single statistical entity. The three-scale WT decomposition and the subband terminology used in Table 1 is shown in Fig 5. Naturally, for a particular image size, the confidence intervals for the ML parameter estimates become larger at each iteration since the subband size is divided by four. The Table demonstrates that the coeffi-

Table 1: Characterisation of actual wavelet coefficient data using alpha-stable distributions. Maximum likelihood parameter estimates and 95% confidence intervals for the $S\alpha S$ characteristic exponent, α

First iteration				
	Image subbands			
Image	LH	HL	HH	Combined
Boat (512 × 512)	1.215± 0.011	1.285± 0.012	1.715± 0.012	1.341± 0.007
Mandrill (512 × 512)	1.233± 0.012	1.468± 0.013	1.594± 0.013	1.377± 0.007
Cameraman (256 × 256)	0.853± 0.018	0.859± 0.018	0.994± 0.020	0.887± 0.011
Woman2 (128 × 128)	1.391± 0.048	1.208± 0.045	1.263± 0.046	1.226± 0.026
Detfingr (301 × 301)	1.543± 0.020	1.411± 0.020	1.738± 0.019	1.341± 0.012
Wbarb (256 × 256)	1.293± 0.023	1.129± 0.021	1.268± 0.023	1.207± 0.013
Detail (359 × 371)	1.668± 0.016	1.558± 0.017	1.563± 0.013	1.593± 0.009
Tire (205 × 232)	0.861± 0.021	0.879± 0.022	1.136± 0.025	0.900± 0.013
Second iteration				
	Image subbands			
Image	LL,LH	LH,LL	LH,LH	Combined
Boat	1.006± 0.020	0.996± 0.020	1.198± 0.022	1.017± 0.012
Mandrill	1.265± 0.024	1.449± 0.025	1.401± 0.025	1.336± 0.014
Cameraman	0.721± 0.032	0.754± 0.033	0.798± 0.035	0.752± 0.019
Woman2	1.369± 0.095	1.308± 0.093	1.356± 0.095	1.281± 0.053
Detfingr	1.523± 0.041	1.569± 0.041	1.727± 0.038	1.406± 0.023
Wbarb	1.178± 0.044	1.018± 0.041	1.183± 0.044	1.077± 0.024
Detail	1.568± 0.034	1.525± 0.034	1.866± 0.026	1.539± 0.020
Tire	0.778± 0.039	0.748± 0.038	0.924± 0.044	0.779± 0.023
Third iteration				
	Image subbands			
Image	LLL,LLH	LLH,LLL	LLH,LLH	Combined
Boat	1.066± 0.042	0.960± 0.039	1.038± 0.041	1.132± 0.035
Mandrill	1.558± 0.050	1.523± 0.050	1.532± 0.051	1.493± 0.029
Cameraman	0.703± 0.063	0.709± 0.063	0.672± 0.061	0.699± 0.036
Woman2	1.830± 0.157	1.454± 0.194	1.457± 0.193	1.305± 0.107
Detfingr	1.467± 0.081	1.856± 0.062	1.872± 0.062	1.532± 0.047
Wbarb	1.219± 0.090	1.116± 0.086	1.106± 0.085	1.054± 0.048
Detail	1.613± 0.067	1.531± 0.067	1.861± 0.052	1.552± 0.039
Tire	0.694± 0.073	0.747± 0.077	0.820± 0.082	0.727± 0.043

The tabulated key parameter α defines the degree of non-Gaussianity as deviations from the value $\alpha = 2$, which corresponds to the Gaussian condition. We iterate a separable four-band decomposition three times (octave-band decomposition). The size of the test images is shown in parentheses.

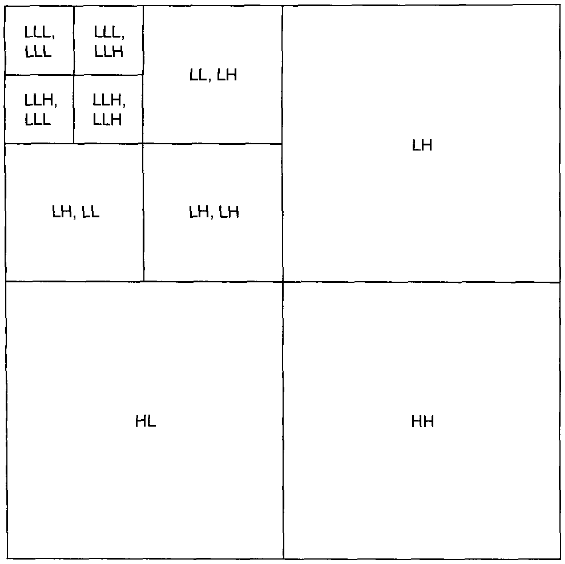


Fig. 5 *Three-scale, octave-band decomposition of an image into unequal subbands*

The spectral decomposition, the ordering, and the names of the subbands are shown

cients of different subbands and decomposition levels exhibit various degrees of non-Gaussianity. The important observation is that all subbands exhibit distinctly non-Gaussian characteristics, with values of α varying between 0.7–1.8, away from the Gaussian point of $\alpha = 2$. Our modelling results clearly point to the need for the use of quantisers that take into consideration the non-Gaussian heavy-tailed character of the data to achieve close to optimal quantisation performance.

3 Quantisation of a Cauchy source

In its general form, the problem of optimum scalar quantisation can be considered as the task of defining the decision levels $d_0 < d_1 < \cdots < d_M$ and the reconstruction levels $r_1 < r_2 < \cdots < r_M$ to form the following partitioning of the data dynamic range $R = [L, U)$:

$$[L, U) = \bigcup_{k=0}^{M-1} [d_k, d_{k+1}) \tag{4}$$

and represent all the data values x lying within the subrange $[d_k, d_{k+1})$ with the reconstruction level r_k so that a distortion measure $D(e)$ is minimised, where e is the quantisation error, defined by

$$e = x - r_k \tag{5}$$

In other words, e is the difference of the reconstruction level from the data value, which it represents. For stochastic data, the distortion measure is defined as the expected value of an error weighting function:

$$D(e) = E[f(e)] = \int_L^U f(e) p(x)\, \mathrm{d}x \tag{6}$$

where $p(x)$ is the PDF of the data distribution and $f(e)$ is the error weighting function. For the specific case under consideration of data following a $S\alpha S$ distribution with $1 \le \alpha < 2$, to define completely the quantisation problem one has to determine $p(x)$ as well as $f(e)$.

Given that the PDF of a general non-Gaussian $S\alpha S$ distribution cannot be defined in closed form, except for $\alpha = 1$, the only available choice for $p(x)$ is the Cauchy probability density function

$$p(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x - \mu)^2} \tag{7}$$

where μ is the location parameter and γ is the dispersion. Note that if a Cauchy random variable (r.v.) X follows the distribution described by eqn. 7, then $(X - \mu)/\gamma$ is also a Cauchy r.v with location parameter equal to zero and dispersion equal to one.

However, the choice of $f(e)$ is constrained by the fact that for Cauchy random variables, only moments of order less than one can be defined. Furthermore, $f(e)$ should be a symmetric and monotonically increasing function of e. For our analysis we have set

$$f(e) = \sqrt{|e|} \tag{8}$$

which is a choice that satisfies the above-mentioned conditions. This choice of error weighting function results in a distortion measure called the 'mean square-root absolute error' (MSRAE). An additional condition that the quantiser should be symmetric has been set, so that the mean of the quantisation error for $S\alpha S$ random variables with $\alpha > 1$ is always zero. Note that for a Cauchy r.v ($\alpha = 1$) the mean is not defined. For the symmetric quantiser due to the corresponding symmetry of $p(x)$ and $f(e)$ one decision level has to be set at zero and moreover the problem can be reduced to defining the quantiser for only positive values of data. The complete quantiser, having $N = 2M$ reconstruction levels, can be obtained by mirroring the defined thresholds (d_k and r_k) with respect to the y-axis.

Taking the above considerations into account, the quantisation problem for $S\alpha S$ random variables can be formulated as follows: For a given number of levels M, determine the decision levels $d_0 < d_1 < \cdots < d_M$ and the reconstruction levels $r_1 < r_2 < \cdots < r_M$ so that

$$D = \sum_{k=1}^{M} \int_{d_{k-1}}^{d_k} \sqrt{|x - r_k|} \frac{1}{\pi} \frac{1}{1 + x^2}\, \mathrm{d}x \tag{9}$$

where $d_0 = 0$, $d_M = \infty$, is minimised.

The quantisation problem, as defined in eqn. 9, is a highly nonlinear optimisation problem. The stationary points of the cost function $D(e)$ are given by the conditions

$$\frac{\partial D}{\partial d_k} = 0, \ k = 1, \ldots, M - 1 \tag{10}$$

$$\frac{\partial D}{\partial r_k} = 0, \ k = 1, \ldots, M \tag{11}$$

It can be easily seen, that these conditions result, respectively, in the following relationships for d_k and r_k:

$$d_k = \frac{r_k + r_{k+1}}{2}, \ k = 1, \ldots, M - 1 \tag{12}$$

$$\int_{d_{k-1}}^{r_k} \frac{1}{\sqrt{r_k - x}} \frac{1}{1 + x^2}\, \mathrm{d}x - \int_{r_k}^{d_k} \frac{1}{\sqrt{x - r_k}} \frac{1}{1 + x^2}\, \mathrm{d}x = 0 \tag{13}$$

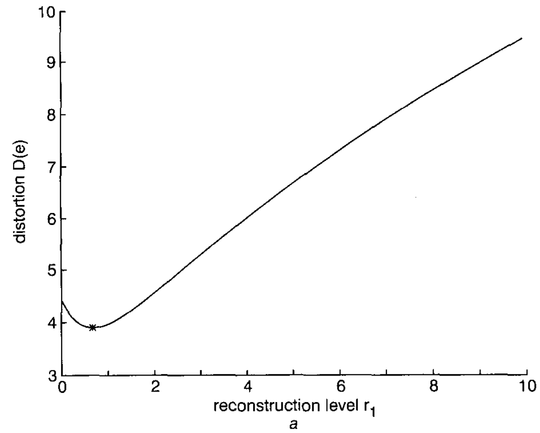
for $k = 1, \ldots, M$.

Eqn. 13 is a nonlinear integral equation for r_k. Hence, numerical iterative methods based on root finding by bisection and on polynomial interpolation and extrapolation [27, 28] have been applied for the solution of the highly nonlinear system of eqns. 12 and 13. The resulting values of the thresholds (d_k and r_k), for various values of the total number of reconstruction levels $N = 2M$, are

Table 2: Placement of decision (d_k) and reconstruction (r_k) levels for the Cauchy quantiser

	N=2		N=4	
k	d_k	r_k	d_k	r_k
1	∞	0.6735	1.4270	0.4719
2			∞	2.3821
D	3.9037		3.3077	
H	1.0000		1.9640	
	N=8		N=16	
k	d_k	r_k	d_k	r_k
1	0.7036	0.3058	0.3806	0.1818
2	1.9588	1.1015	0.8380	0.5794
3	6.0641	2.8161	1.4913	1.0966
4	∞	9.3122	2.5885	1.8860
5			4.7687	3.2909
6			10.1505	6.2465
7			29.7345	14.0544
8			∞	45.4145
D	2.6746		2.0652	
H	2.8541		3.7201	

shown in Table 2. This Table shows also the values of the distortion D, as well as the entropy H achieved by the presented quantisation schemes. The values of H given in the Table denote the minimum average number of binary digits required to code the quantiser output. It can be seen from the Table that this number is a large fraction (around 0.9) of $\log_2 N$. When $N=2^n$, with n being an integer, a simple n binary digit code for the outputs of the quantiser makes near optimum use of the transmission capacity of the system. Naturally, as the number of quantisation levels N increases the entropy increases.



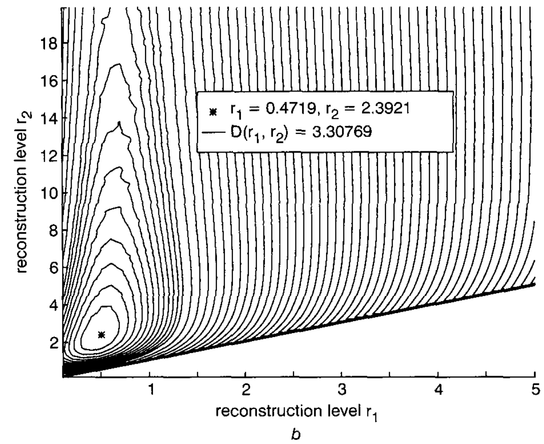


Fig. 6 *Plots of distortion D*

a Distortion D as a function of reconstruction level r_1 for $N=2$ (compare with Table 2)
* $r_1=0.6735$, $D(r_1)=3.9036$
b Distortion D as a function of reconstruction levels r_1 and r_2 for $N=4$ (compare with Table 2)

Table 3: Placement of decision (d_k) and reconstruction (r_k) levels for the optimum MSE quantiser of a Gaussian (Laplacian) source

	N=2		N=4	
k	d_k	r_k	d_k	r_k
1	∞	0.7979 (0.7071)	0.9816 (1.1269)	0.4528 (0.4198)
2			∞	1.5104 (1.8340)
MSE	0.3634 (0.5000)		0.1175 (0.1762)	
H	1.0000 (1.0000)		1.9111 (1.7282)	
	N=8		N=16	
k	d_k	r_k	d_k	r_k
1	0.5006 (0.5332)	0.2451 (0.2334)	0.2582 (0.2644)	0.1284 (0.1240)
2	1.0500 (1.2527)	0.7560 (0.8330)	0.5224 (0.5667)	0.3880 (0.4048)
3	1.7479 (2.3796)	1.3439 (1.6725)	0.7995 (0.9198)	0.6568 (0.7287)
4	∞	2.1519 (3.0867)	1.0993 (1.3444)	0.9423 (1.1110)
5			1.4371 (1.8776)	1.2562 (1.5778)
6			1.8435 (2.5971)	1.6180 (2.1773)
7			2.4008 (3.7240)	2.0690 (3.0169)
8			∞	2.7326 (4.4311)
MSE	0.0345 (0.0545)		0.0095 (0.0154)	
H	2.8248 (2.5654)		3.7652 (3.4747)	

For easy reference, we include in Table 3 the optimum MSE quantiser parameters for Gaussian and Laplacian sources ([29] p. 153) Comparing Tables 2 and 3, it should be noted that the values of both the decision and reconstruction levels of the Lloyd–Max quantiser for the Cauchy distribution are significantly greater in absolute value than the corresponding levels of the Lloyd–Max MSE quantisers for the Gaussian and Laplacian densities, as they account for the much heavier tails of the Cauchy distribution.

The values of thresholds in Table 2 have been found to provide a locally optimal quantiser. This is indicated in Fig. 6. When $N=4$, the distortion D is a function of three variables, namely the two reconstruction levels r_1 and r_2 as well as the decision level d_1. Setting $d_1=(r_1+r_2)/2$ yields according to eqn. 12 all the candidate triplets (r_1,r_2,d_1) for being the optimal points. By these means, distortion D can be considered as a function of just the two variables r_1 and r_2. Fig. 6*b* depicts the contour plot of this function $D(r_1,r_2)$ indicating the local minimum for the values of r_1 and r_2 given in Table 2.

4 Experimental results

The performance of the above presented quantiser was tested in comparison with the performance of both the optimal mean square error (MSE) Gaussian and Laplacian quantisers, when applied to the same data.

Given that the Cauchy-based quantiser (here denoted as Cauchy quantiser for simplicity) has been constructed with reference to data following $S\alpha S$ distributions, at a first stage, simulated data were generated for $\alpha=1,1.1,\ldots,2$. Note that this range of values for α covers the majority of the image subband cases shown in Table 1. In each trial, the statistical parameters of the generated data, namely the mean and the standard deviation, as well as the location parameter and the dispersion, were estimated and the thresholds of the three quantisers for the standard distributions were scaled and translated to fit the data distribution. The data was then quantised according to each of the three quantisation schemes, using quantisers with $N=16$ levels. Based on the quantised data and the original data the following measures of distortion were computed:

- Mean square error (MSE), defined as

$$MSE=\frac{1}{K}\sum_{i=1}^{K}(x_i-\hat{x}_i)^2 \quad (14)$$

- Mean absolute error (MAE), defined as

$$MAE=\frac{1}{K}\sum_{i=1}^{K}|x_i-\hat{x}_i| \quad (15)$$

- Mean square root absolute error (MSRAE), defined as

$$MSRAE=\frac{1}{K}\sum_{i=1}^{K}\sqrt{|x_i-\hat{x}_i|} \quad (16)$$

where $\hat{x}_i$ is the quantised value of the data value x_i and K is the number of generated data. The above three error measures are special cases of the general L_P norm defined as

$$L_p=\sum_{i=1}^{K}|x_i-\hat{x}_i|^p/K$$

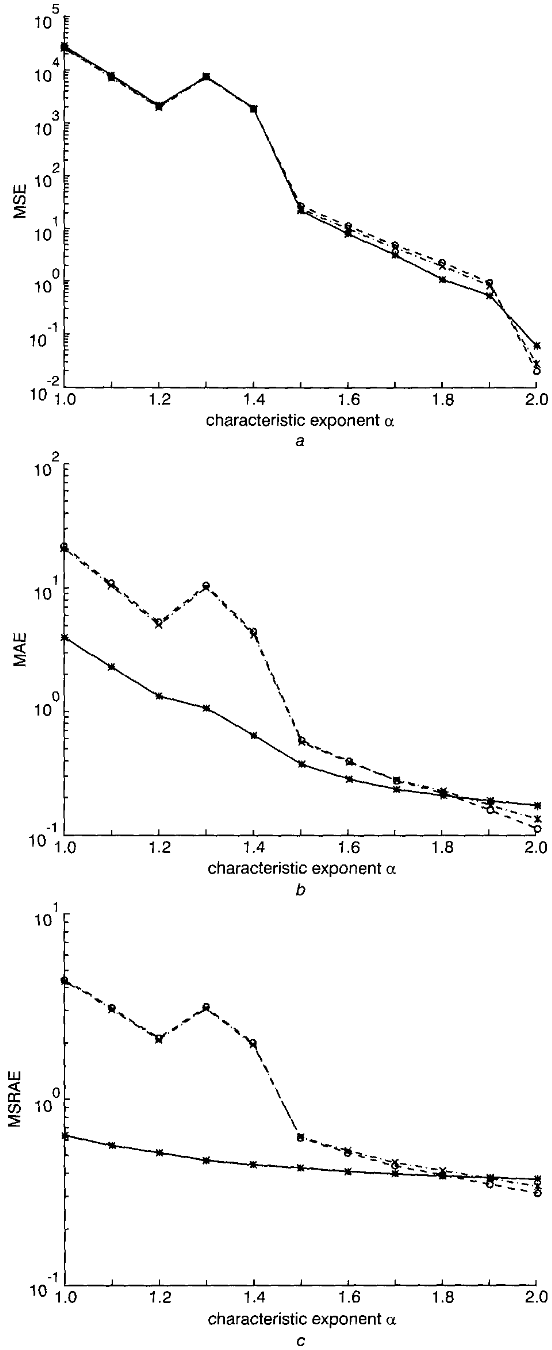


Fig. 7 *Metrics of quantisation error for simulated data*
The metrics are plotted as a function of the characteristic exponent that describes the statistics of the alpha-stable source
* Cauchy
O Gauss
× Laplace
a MSE
b MAE
c MSRAE

for $p=2$, 1, and 0.5, respectively. The results for each quantisation scheme, for each distortion measure, and for each value of α for $K=65000$ are shown in Fig. 7.

Fig. 7*a* is mainly determined by the fact that $S\alpha S$ random variables with $\alpha<2$ have in theory infinite variance. This results in the great order of magnitude for MSE, especially for $\alpha\leq 1.5$ for the three quantisation schemes. This Figure implies that MSE is not a suitable measure of distortion for

$S\alpha S$ random variables with $\alpha < 2$. However, note that the optimality of the Gaussian MSE quantiser is evident for $\alpha = 2$.

Fig. 7*b* however, shows that the Cauchy quantiser achieves better performance, with respect to MAE, for $S\alpha S$ distributions which significantly depart from the Gaussian case (i.e. for values of $\alpha \leq 1.8$). Note that MAE is well defined for $\alpha > 1$, since $S\alpha S$ random variables with $\alpha > 1$ have finite first moments. Moreover, MAE seems to be a more objective measure of distortion than MSE and MSRAE, since it weights all errors with the same factor. On the contrary MSE emphasises errors with absolute value greater than one and de-emphasises errors with absolute value less than one, while MSRAE performs in exactly the opposite way. Fig. 8*c* shows the superior performance of the Cauchy quantiser with respect to MSRAE for the same range of values of α. One should note that for $\alpha = 2$, i.e. for a Gaussian source, the Gaussian quantiser achieves the best performance according to all three measures. Similarly, when $\alpha = 1$, i.e. for a Cauchy source, the Cauchy quantiser is the best. This fact seems to imply that performance is mostly determined by the distribution used in the design of the quantiser rather than by the error weighting function. Most importantly, for sources other than Cauchy and Gaussian that correspond to values $1 \leq \alpha \leq 1.8$, the Cauchy quantiser exhibits less error than the Gaussian and the Laplacian.

At a second stage, instead of using simulated data, the 'Lena' wavelet coefficient data were quantised using the Cauchy, Gaussian and Laplacian quantisers with $N = 16$ reconstruction levels. Once again, the above-mentioned three distortion measures were computed in each case. The results are shown in Table 4. The relatively modest MSE values imply that our data lack the presence of extreme outliers. However, the non-Gaussian nature of the data is evident by the fact that the Gaussian quantiser has the worst MSE performance. In addition, the MAE values in Table 4 show that despite the lack of extreme outliers, the Cauchy quantiser achieves a good MAE performance. Finally, the superiority of the Cauchy quantiser in terms of MSRAE is one more indication of the non-Gaussian nature of the data.

The problem with the above three error measures, or with any other L_p metric, is determining their connection to visual interpretation in image coding. A human observer, in viewing images, does not compute any of the above measures. Hence, to visually study the merit of the proposed $S\alpha S$ WT coefficient modeleing and the Cauchy quantiser, we chose the 512×512, 8-bit 'boat' test image, shown in Fig. 8*a* to compute its WT, quantise its coefficients and evaluate the reconstructed image both visually and by means of the extensively used peak signal-to-noise ratio (*PSNR*) defined as

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \qquad (17)$$

Table 4: Quantisation error metrics for the Gaussian, Laplacian and Cauchy quantisers employed on the wavelet coefficient data shown in Fig. 2

	Cauchy	Gaussian	Laplacian
MSE	3.52	6.23	2.85
MAE	0.71	0.84	0.73
MSRAE	0.69	0.77	0.75

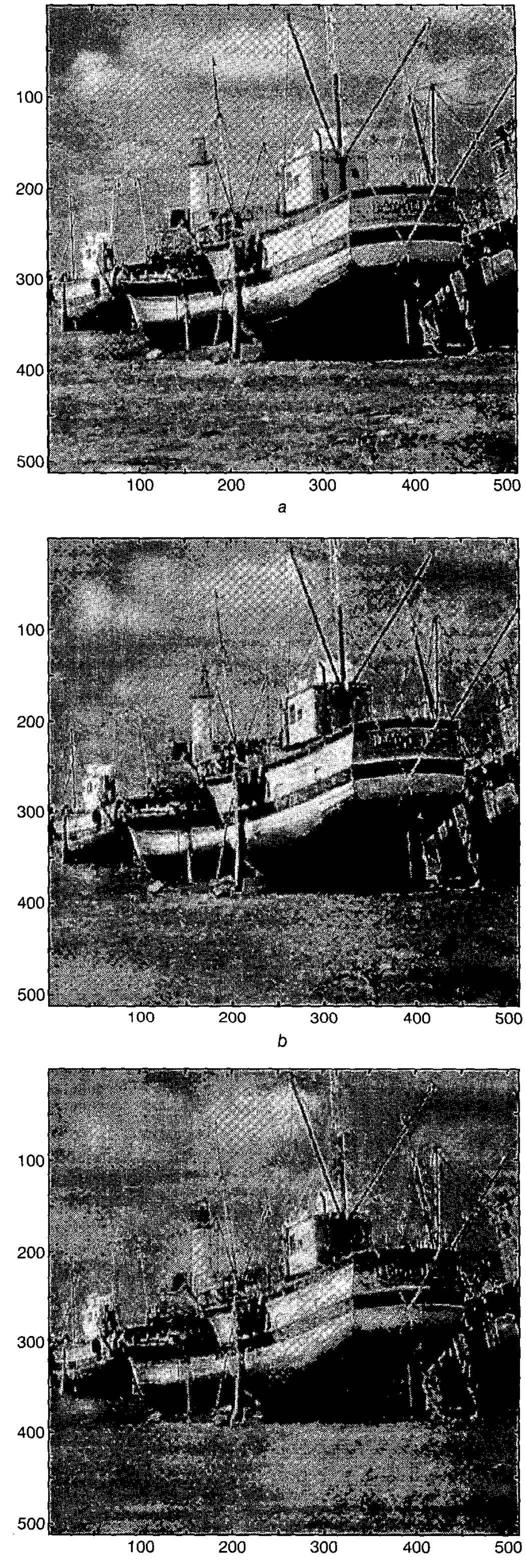


Fig. 8 *Performance of Cauchy and Laplace quantisation schemes used in wavelet coding of the 'boat' image*
a Original 512×512 'boat' image at 8 bits/pixel (bpp)
b Cauchy quantiser results, 0.7 bpp, PSNR-Cauchy = 24.67 dB
c Laplace quantiser results, 0.7 bpp, PSNR-Laplace = 22.90 dB.

Table 5: Coding results for the 'boat', 'cameraman' and 'mandrill' test images showing peak-signal-to-noise ratios (PSNR) for the Cauchy and Laplace quantisation schemes of the wavelet coefficients and various compression ratios

Bits/pixel	Compression	PSNR-Cauchy (dB)	PSNR-Laplacian (dB)
Boat (512 × 512)			
0.7	11.4:1	24.67	22.90
0.5	16:1	23.16	21.85
0.25	32:1	20.86	20.74
0.125	64:1	20.28	20.26
Cameraman (256 × 256)			
0.7	11.4:1	20.65	19.73
0.5	16:1	18.96	18.81
0.25	32:1	17.67	17.32
0.125	64:1	16.76	16.50
Mandrill (512 × 512)			
0.7	11.4:1	23.39	23.15
0.5	16:1	22.66	22.17
0.25	32:1	21.12	21.09
0.125	64:1	20.83	20.76

where *MSE* denotes the L_2 error metric between the original and reconstructed images. We quantised the low-pass band based on its Gaussian statistics, while for all other bands we used the Cauchy against the Laplace quantisers.

Fig. 8 show a comparison among the original 'boat' image (Fig. 8*a*), the Cauchy encoded images (Fig. 8*b*), and the Laplace encoded image (Fig. 8*c*) at bit rate equal to 0.7 bits/pixels (bpp). The subjective quality of the Cauchy encoded image is superior to the quality of the Laplace encoded one, a fact that is also translated to a higher (by 1.77 dB) PSNR value for the Cauchy coder. Table 5 presents the achieved PSNR values for the Cauchy and Laplace coders, for various test images and compression ratios. From the Table it can be seen that the PSNR performance improvement of the Cauchy over the Laplacian coder depends on the test image and the compression ratio. It seems that the achieved improvement is more significant at lower compression ratios.

5 Conclusions and future work

We introduced a new representation of wavelet coefficients for image processing applications, based on symmetric α-stable distributions to better model the heavy-tailed nature of the wavelet data. The proposed representation addresses the inability of the Gaussian or other distributions with exponential tails to describe data with a power law tail behavior. The significance of these results was demonstrated by considering the source quantisation problem. As shown through experiments with simulated as well as real data, the introduced Cauchy quantiser appears to be a useful tool for discretising data which follow $S\alpha S$ distributions with characteristic exponent α significantly less than two.

There are certainly some further issues that should be addressed in future work. Theoretical issues concerning the uniqueness of the solutions of the highly nonlinear system of eqns. 12–13 and the absolute optimality of the quantiser are still open. Furthermore, one should study the effects of using other error weighting functions for the Cauchy quantiser, having the general form $f(e) = |e|^p, 0 < p < 1$ (we have examined the case for $p = \frac{1}{2}$), especially for values of p close to one. Most importantly, the issue of optimal quantisation of the general alpha-stable distribution is still open and is currently being addressed by the authors.

6 Acknowledgements

The work in this paper was supported by the US Office of Naval Research under Contract N00014-92-J-1034 and by the Greek General Secretariat for Research and Technology under Program EIIET II, Code 97EΛ - 152.

7 References

1 MAX, J.: 'Quantizing for minimum distortion', *IRE Trans. Inf. Theory*, 1960, **6**, pp. 7–12

2 LLOYD, S.P.: 'Least squares quantization in PCM', *IEEE Trans. Inf. Theory*, 1982, **IT-28**, pp. 129–137

3 FLEISCHER, P.E.: 'Sufficient conditions for achieving minimum distortion in a quantizer'. IEEE International Conv. Rec., 1964, pp. 104–111

4 TRUSHKIN, A.V.: 'Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions', *IEEE Trans. Inf. Theory*, 1982, **IT-28**, pp. 187–198

5 SHARMA, D.K.: 'Design of absolutely optimal quantizers for a wide class of distortion measures', *IEEE Trans. Inf. Theory*, 1978, **IT-24**, pp. 693–702

6 KABAL, P.: 'Quantizers for the gamma distribution and other symmetrical distributions', *IEEE Trans. Acoust. Speech Signal Process.*, 1984, **ASSP-32**, pp. 836–841

7 ORTEGA, A., and VETTERLI, M.: 'Adaptive scalar quantization without side information', *IEEE Trans. Image Process.*, 1997, **6**, pp. 665–676

8 WILLINGER, W., TAQQU, M.S., LELAND, W.E., and WILSON, D.V.: 'Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements', *Stat. Sci.*, 1995, **10**, pp. 67–85

9 RESNICK, S.: 'Why non-linearities can ruin the heavy-tailed modelers' day, *in* ADLER, R., FELDMAN, R., and TAQQU, M.S. (Eds.): 'A practical guide to heavy tails: Statistical techniques and applications' (Birkhauser, Boston, 1998), pp. 219–239

10 McCULLOCH, J.H.: 'Financial applications of stable distributions', *in* MADDALA, G.S., and RAO, C.R. (Eds.): 'Statistical methods in finance (Handbook of statistics 14)' (Elsevier Science, Amsterdam, 1996), 393–425

11 TSAKALIDES, P., RASPANTI, R., and NIKIAS, C.L.: 'Angle/doppler estimation in heavy-tailed clutter backgrounds', *IEEE Trans. Aerosp. Electron. Syst.*, 1999, **AES-35**, pp. 419–436

12 FARVARDIN, N., and MODESTINO, J.W.: 'Optimum quantizer performance for a class of non-Gaussian memoryless sources', *IEEE Trans. Inf. Theory*, 1984, **IT-30**, pp. 485–497

13 CAMBANIS, S., SAMORODNITSKY, G., TAQQU, M.S.: 'Stable processes and related topics' (Birkhauser, Boston, 1991)

14 SAMORODNITSKY, G., and TAQQU, M.S.: 'Stable non-Gaussian random processes: Stochastic models with infinite variance' (Chapman and Hall, New York, 1994)

15 NIKIAS, C.L., and SHAO, M.: 'Signal processing with alphastable distributions and applications' (John Wiley and Sons, New York, 1995)

16 ADLER, R., FELDMAN, R., and TAQQU, M.S.: 'A guide to heavy tails: Statistical techniques and applications' (Birkhauser, Boston, 1998)

17 FAMA, E.F., and ROLL, R.: 'Some properties of symmetric stable distributions', *J. Am. Stat. Assoc.*, 1968, **63**, pp. 817–836

18 DUMOUCHEL, W.H.: 'Stable distributions in statistical inference'. PhD thesis, Dept. of Statistics, Yale University, 1971

19 BRORSEN, B.W., and YANG, S.R.: 'Maximum likelihood estimates of symmetric stable distribution parameters', *Commun. Stat.-Simul. Comput.*, 1990, **19**, pp. 1459–1464

20 NOLAN, J.P.: 'Maximum likelihood estimation and diagnostics for stable distributions. Tech. Rep., Dept. of Math. and Stat., American University, 1999

21 SAID, A., and PEARLMAN, W.: 'A new fast and efficient image coder based on set partitioning on hierarchical trees', *IEEE Trans. Circuits Syst. Video Technol.*, 1996, **6**, pp. 243–250

22 TAUBMAN, D., and ZAKHOR, A.: 'Multirate 3-D subband coding of video', *IEEE Trans. Image Process.*, 1994, **3**, pp. 572–588

23 ADLER, R.J., FELDMAN, R.E., and GALLAGHER, C.: 'Analysing stable time series', *in* ADLER, R., FELDMAN, R., and TAQQU, M.S. (Eds.): 'A practical guide to heavy tails: Statistical techniques and applications' (Birkhauser, Boston, 1998, pp. 133–158)
24 NOLAN, J.P.: 'Numerical calculation of stable densities and distribution functions', *Commun. Stat.-Stochastic Models*, 1997, **13**, pp. 759–774
25 MICHAEL, J.R.: 'The stabilized probability plot', *Biometrika*, 1983, **70**, pp. 11–17
26 VILLASENOR, J.D., and WEN, J.: 'Low-complexity compression of run length coded image sub-bands', *in* TOPIWALA, P.N. (Ed.): 'Wavelet image and video compression' (Kluwer Academic, Boston, 1998) pp. 221–235
27 GILL, P.E., MURRAY, W., and WRIGHT, M.H.: 'Practical optimization' (Academic Press, London, 1981)
28 FLETCHER, R.: 'Practical methods of optimization. Volume 1: Unconstrained optimization' (John Wiley and Sons, New York, 1980)
29 PRATT, W.K.: 'Digital image processing' (John Wiley and Sons, New York, 1991)