From the Proceedings of the International Workshop in Applied Probability 2008

Advances in nonlinear signal processing for heavy tailed noise

John P. Nolan[1]

Department of Mathematics and Statistics, American University,
4400 Massachusetts Avenue, NW
Washington, DC 20016-8050, USA
(e-mail: jpnolan@american.edu)

Abstract. Standard signal processing techniques perform poorly in the presence of impulsive, heavy tailed noise. We describe recent advances in nonlinear signal processing, which can perform significantly better than linear filters.

Keywords. Nonlinear signal processing, stable distributions, global optimization.

1 Introduction

Traditional signal processing uses linear algorithms to filter out noise. The theory of linear filters is well developed and optimal when the noise terms are normally distributed. However, linear filters will not work well when the noise is impulsive, heavy tailed. In the heavy tailed case, a linear filter is strongly influenced by large extreme values, causing the signal to be swamped by the noise. This occurs even when a large window is used.

The standard additive noise model for a signal is

$$x_t = s_t + n_t, \quad t = 1, 2, 3, \ldots \tag{1}$$

Here s_t is some signal we are interested in, and n_t are noise terms that corrupt the signal. This noise can be caused by natural events like lightening, sea clutter in naval radar systems, animal sounds (snapping shrimp) underwater, or man made sources like electro-mechanical noise in urban environments. The goal is to recover the unknown signal s_t as well as possible. This is done by computing an estimate $\widehat{s}_t$ using a window of the data, centered at t. (Padding is done at the ends of the array to handle end effects; this can be constant padding, symmetric padding or circular padding.)

We assume the noise terms n_t have i.i.d. symmetric α-stable distributions. This is a family of distributions with symmetric, bell-shaped densities with heavy tails that generalize the normal distribution. The Generalized Central

Limit Theorem shows that these distribution are the only possible limits of normalized sums of i.i.d. terms. There is no known formula for the pdf for general α, but the characteristic function has a simple form: we will say n is symmetric α-stable with index of stability $\alpha \in (0,2]$ and scale $\gamma > 0$ if

$$E \exp(iun) = \exp(-\gamma^\alpha |u|^\alpha).$$

More information on stable distributions can be found in Samorodnitsky and Taqqu (1994).

Two special cases are worth noting. When $\alpha = 2$, and $\gamma^2 = \sigma^2/2$, n is $\mathrm{N}(0, \sigma^2)$. (Note that in this case, $\gamma = \sigma/\sqrt{2}$, so the scale is not simply the standard deviation.) When $\alpha = 1$, the resulting distribution is (centered) Cauchy with scale γ. The Cauchy filter that arises below is called the myriad filter in the engineering literature, see Gonzales (1997) and Arce (2005). This paper generalizes those results to the cases where $\alpha \notin \{1, 2\}$.

2 Linear and stable filters

The linear filter uses a sliding window of width m to compute an estimate for each window: $\widehat{s}_t = \widehat{\theta}_{\mathrm{LINEAR}}(x_{t-k_1}, x_{t-k_1+1}, \ldots, x_{t+k_2})$, where $k_1 = \lfloor m/2 \rfloor$, $k_2 = m - k_1$ and

$$\widehat{\theta}_{\mathrm{LINEAR}}(x_1, \ldots, x_m) := (x_1 + x_2 + \cdots + x_m)/m. \tag{2}$$

The well established theory of linear filter shows this is optimal when $\alpha = 2$, but experience shows that the linear filter can be severely degraded when $\alpha < 2$. As is well known in the statistics literature, extreme values of the noise terms n_t can have a large effect on the sample mean.

Here we define a robust technique that is optimal for the case when the noise terms are stable. Let $\rho(x) = -\log f(x)$ be the negative of the log density, and define a cost function

$$C(\theta; x_1, \ldots, x_m) = \sum_{i=1}^{m} \rho(x_i - \theta) \tag{3}$$

We define the stable filter to be the value of θ that minimizes the cost: $\widehat{s}_t = \widehat{\theta}_{\mathrm{STABLE}}(x_{t-k_1}, x_{t-k_1+1}, \ldots, x_{t+k_2})$, where

$$\widehat{\theta}_{\mathrm{STABLE}}(x_1, \ldots, x_m) = \arg\min_\theta C(\theta; x_1, x_2, \ldots, x_m). \tag{4}$$

Since $\rho(x) = -\log f(x)$, the minimum of the cost function is exactly the maximum likelihood estimate of the location parameter θ. Note that in the Gaussian $\alpha = 2$ case, $\rho(x) = -x^2/2$, and the minimum can be found explicitly; it is simply $\widehat{\theta}_{\mathrm{LINEAR}}$. For $0 < \alpha < 2$, the filters are nonlinear with no closed formula, and the minimum in (4) must be found numerically. This is discussed in the following section.

There are several generalizations of the filter that are given by replacing the cost function (3). The simplest extension is to allow non-negative weights:

$$C_{\text{weighted}}(\theta; x_1, \ldots, x_m) = \sum_{i=1}^{m} \rho(w_i(x_i - \theta)) \tag{5}$$

In detection problems, one looks for a known pattern $s_1, \ldots, s_n$ in the signal, which is contaminated by stable noise:

$$x_t = \theta s_t + n_t, \quad t = 1, 2, 3, \ldots$$

The null case corresponds to $\theta = 0$ (no signal), the case where a signal is present corresponds to $\theta \neq 0$. There are two ways to implement this. The first is to allow signed weights as in the myriad filter in Arce (2005). This is a way to approximate a matched filter, using cost function

$$C_{\text{signed}}(\theta; x_1, \ldots, x_m) = \sum_{i=1}^{m} \rho(|s_i|((\operatorname{sign} s_i)x_i - \theta)) = \sum_{i=1}^{m} \rho(s_i x_i - |s_i|\theta).$$

The second way is a true stable matched filter, with cost function

$$C_{\text{matched}}(\theta; x_1, \ldots, x_m) = \sum_{i=1}^{m} \rho(x_i - \theta s_i). \tag{6}$$

The $\widehat{\theta}$ obtained by minimizing (6) gives an estimate of the strength of the original pattern in the signal.

Finally, it is possible to use non-symmetric stable distributions for the noise terms. This may be useful in cases like Kuruoglu and Zerubia (2003). In the skewed case, one should use a parameterization that centers the distribution at the mode.

3 Numerical issues and examples

While there is no known closed formula for stable densities (Cauchy and Gaussian cases excluded), there are now reliable and accurate routines for computing general stable densities and cumulative distribution functions numerically. The basic algorithms used to do this are described in Nolan (1997). These routines are incorporated into a software package called STABLE, which is used to compute $\rho(x) = -\log f(x)$, shown at the top of Figure 1.

One way to minimize the cost function is to search for a zero of the derivative of the cost function. The evaluation of $C'(\theta)$ requires the score function $\rho'(x) = -f'(x)/f(x)$. This function can be computed numerically, see the middle of Figure 1. In the $\alpha = 2$ Gaussian case, $\rho'(x) = x$ is linear, but when $\alpha < 2$, the score function is not linear. In the engineering literature,

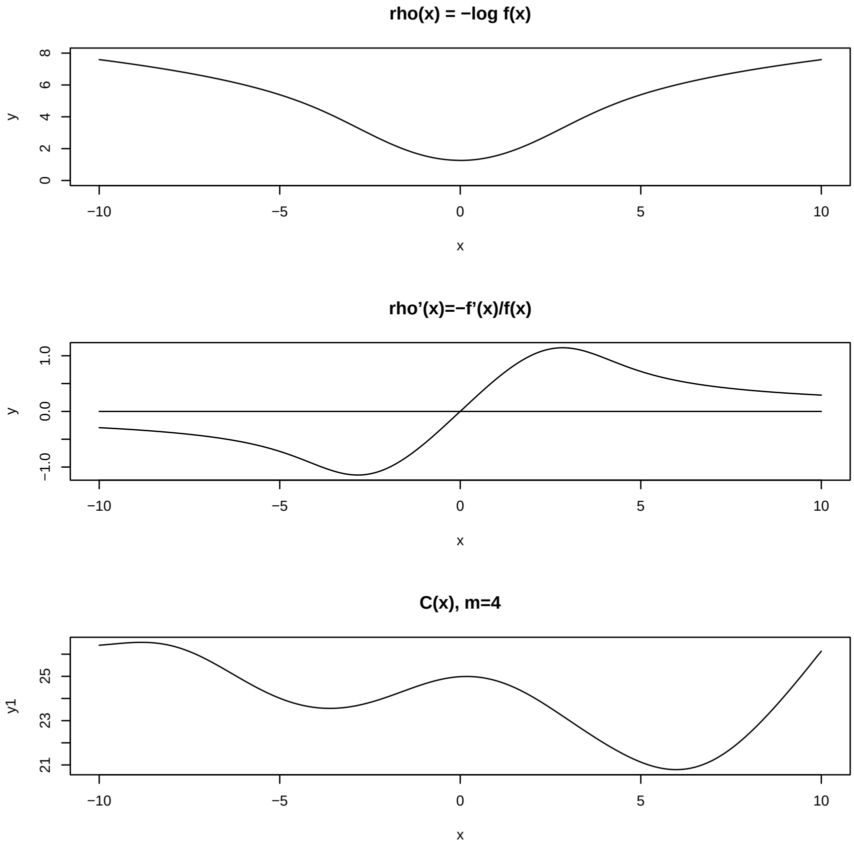

Fig. 1. $\rho(x)$, $\rho'(x)$ and an example of a non-convex cost function for a standardized symmetric stable distribution with $\alpha = 1.7$.

$\rho'(x)$ is called the nonlinear function. We note that unlike the Gaussian case, when $\alpha < 2$, the extreme values are downplayed. In fact, the most important values are neither at the mode nor on the tails, but rather in the "shoulders" or intermediate part of the range.

Direct computation of $f(x)$ uses numerical evaluation of an integral and can be slow, especially when sampling rates are high. To speed up the running of the stable filter, we developed an approximation to the density that is approximately 6,000 times as fast as the method based on numerical quadrature; a fast approximation to $\rho'(x)$ is also included in the software. These routines make it feasible to develop and test stable filters. It is feasible to do real time processing of data streams with low sampling rates, but not yet feasible with high sampling rates.

There is a hidden problem in minimizing the cost function: $\rho(x)$ is non-convex, which means the cost function is non-convex in general. The bottom graph in Figure 1 shows an example where the cost function has multiple local minimums. Minimizing the cost function is numerically challenging, because a local minimization routine will not guarantee the global minimum. For this reason, we developed a global minimization routine based on branch-and-bound to guarantee that the global minimum is found, see Núñez, et. al. (2008). That method bounds the cost function on an interval using only function evaluations. We are investigating methods using $\rho'(\cdot)$ and $\rho''(\cdot)$, which would allow faster optimization.

We end with some examples showing the improved performance possible with a stable filter. Figure 2 below compares the linear and stable filter in three different simulated cases. The first case is the case where the signal is identically zero, so the input to the filter is just i.i.d. stable noise and the perfect output would be a zero. The second example is then there is a level change in the signal. The last example shows a sinusoidal signal corrupted by stable noise. In all three cases, the stable filter significantly outperforms a linear filter, where the spikes in the input data propagate through the filter.

References

Arce, G. R. (2005) *Nonlinear Signal Processing*, Wiley, NY.

Gonzalez, J. (1997): Robust Techniques for Wireless Communications in NonGaussian Environments, PhD Dissertation, University of Delaware, USA.

Kuruoglu, E. and Zerubia, J (2003): Skewed alpha-stable distributions for modelling textures, *Pattern Recognition Letters*, *24*, 339 - 348.

Nikias, C. L. and Shao, M (1995): *Signal Processing with Alpha-Stable Distributions and Applications*, Wiley, NY.

Nolan, J. P. (1997): Numerical calculation of stable densities and distribution functions, *Communications in Statistics -Stochastic Models*, *13*, 759-774.

Núñez, R. C., Gonzalez, J. G., Arce, G. R. and Nolan, J. P. (2008): Fast and accurate computation of the myriad filter via branch-and-bound search. *IEEE Trans. Signal Processing*, to appear.

Samorodnitsky, G. and Taqqu, M. (1994): *Stable Non-Gaussian Random Processes*, Chapman and Hall, N. Y.

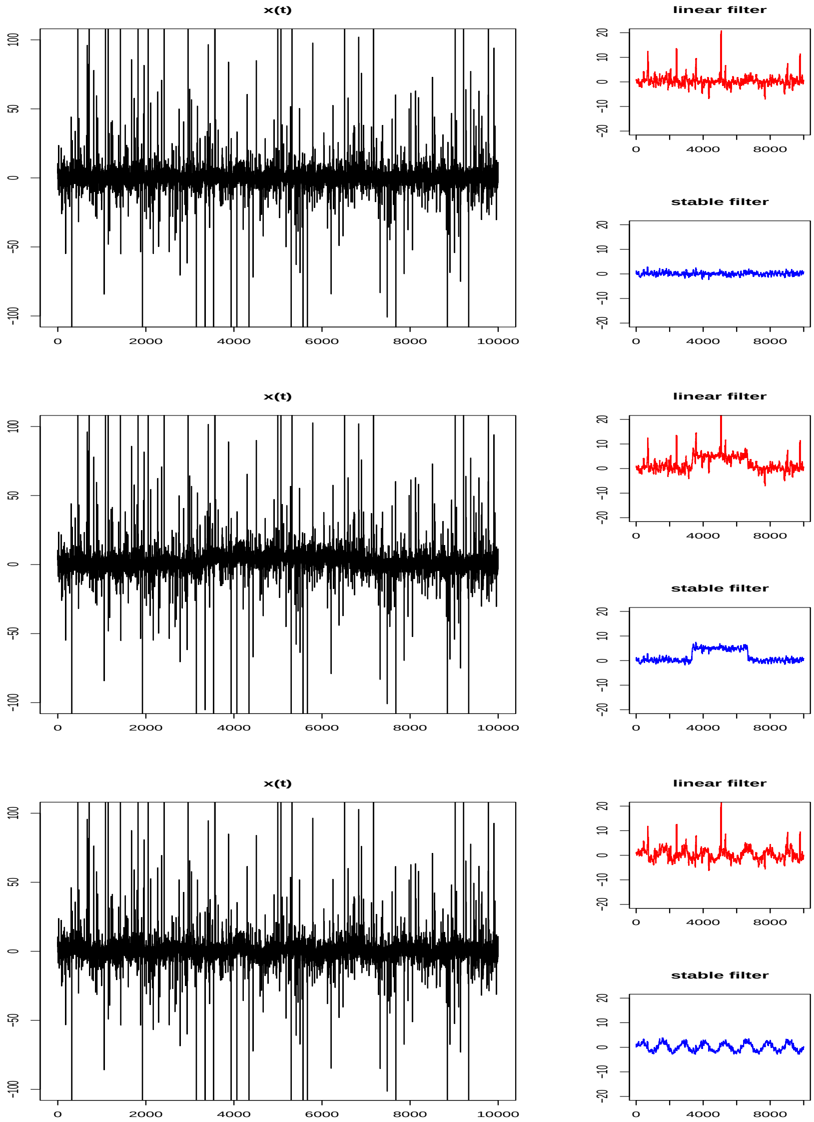

Fig. 2. Signal filtering with stable noise, $\alpha = 1.3$, $\gamma = 2$, $n = 10000$, and window width $m = 50$. The top plot shows the null case with no signal, the middle shows a level change, the bottom shows a sinusoidal signal. The large graph shows the signal with noise, the smaller plots show the output of the two filters. Note the difference in the scales for the input and the output.