Robust Norms in Signal Processing and Compressive Sensing

Gonzalo R. Arce

Department Electrical & Computer Engineering
University of Delaware
arce@ee.udel.edu

August 9, 2016

Linear signal processing enjoys the rich theory of linear systems. Linear filters are also simple to implement.

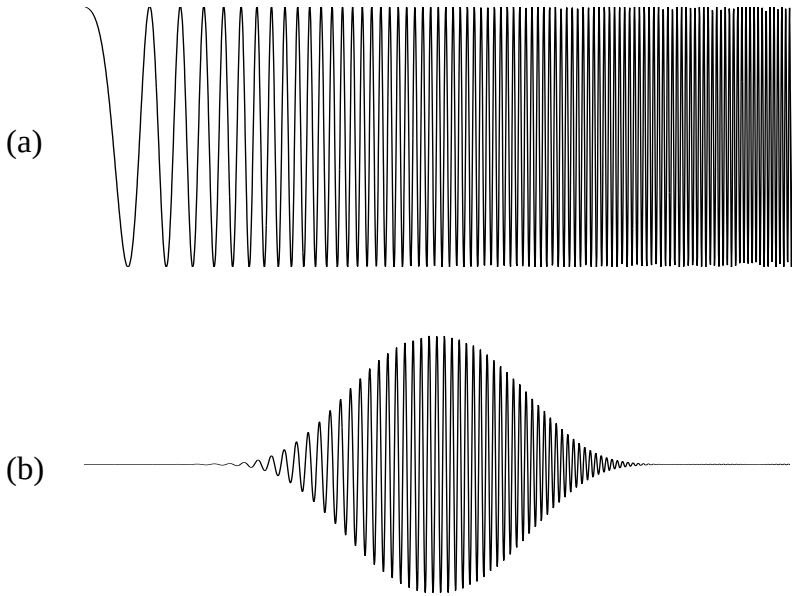


Figure: Frequency selective filtering: (a) chirp signal, (b) linear FIR filter output.

Consider again the bandpass filtering example using a chirp signal degraded by non-Gaussian noise. The linear FIR filter output is severely degraded.

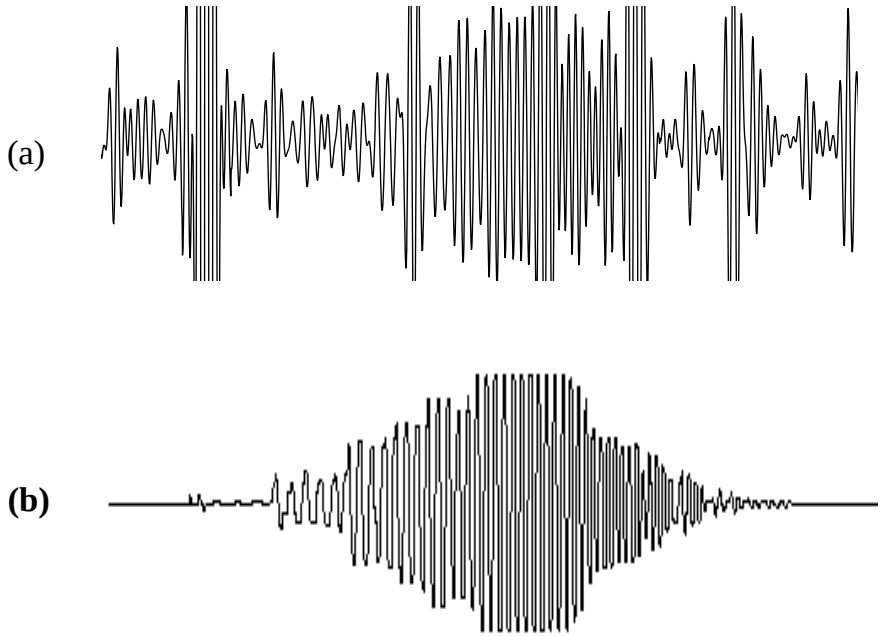


Figure: Frequency selective filtering in non-Gaussian noise: (a) linear FIR filter output, (b) nonlinear filter

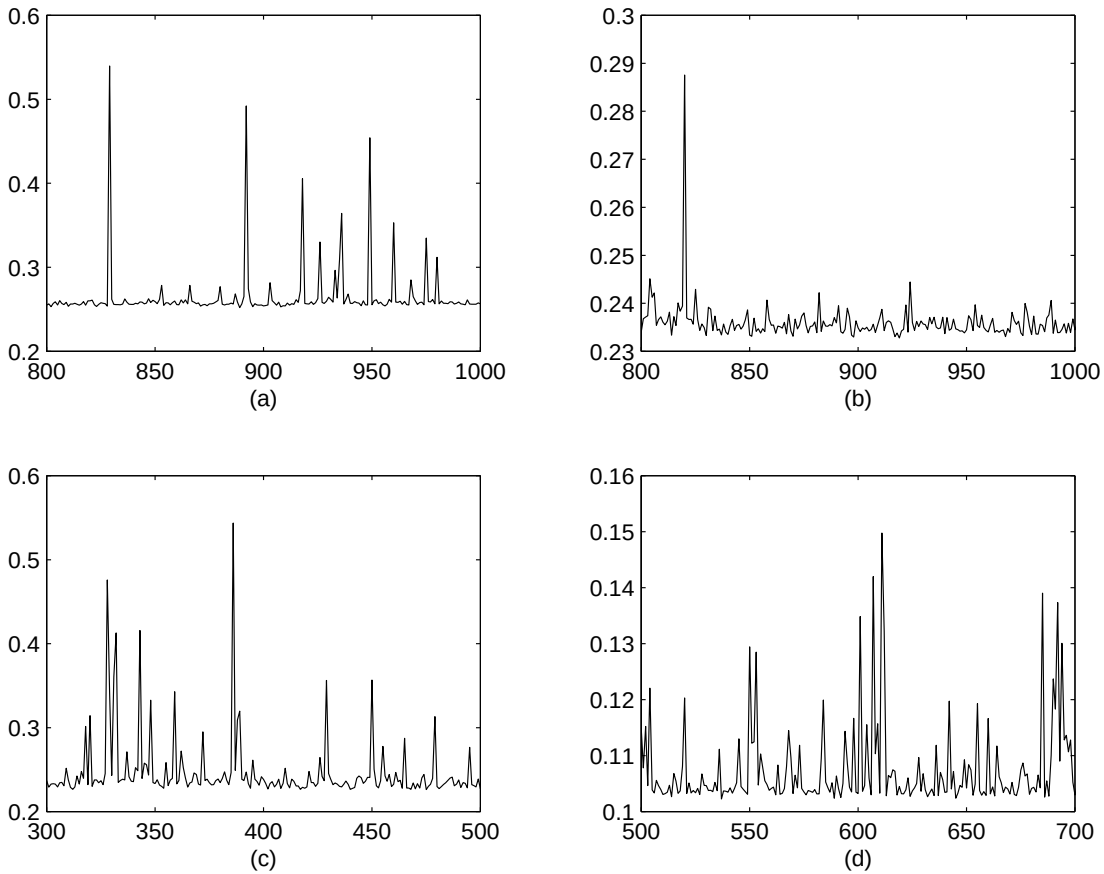


Figure: RTT time series measured in seconds between a host at the University of Delaware and hosts in (a) Australia (12:18 AM - 3:53 AM); (b) Sydney, Australia (12:30 AM - 4:03 AM); (c) Japan (2:52 PM - 6:33 PM); (d) London, UK (10:00 AM - 1:35 PM). All plots shown in 1 minute interval samples.

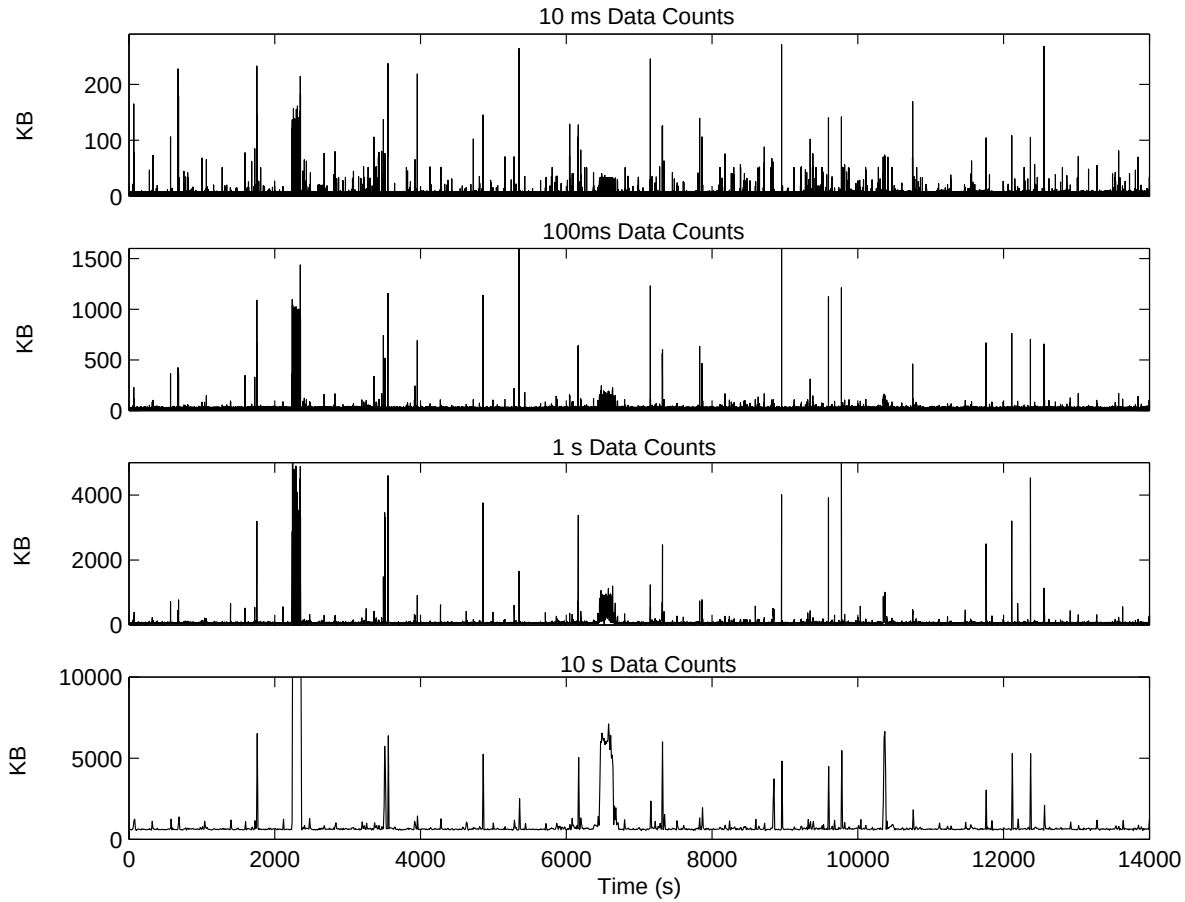


Figure: Byte counts measured over 14000 seconds in a web server of the ECE Department at the University of Delaware viewed through different aggregation intervals: from top to bottom, 10ms, 100ms 1s, 10s.

In this course we will considers two model families that encompass a large class of random processes with different tail characteristics:

- *generalized Gaussian* distribution
- *stable* distributions

The tail of a distribution can be measured by the mass of the tail (*order*), defined as $\mathrm{P_r}(X > x)$ as $x \to \infty$.

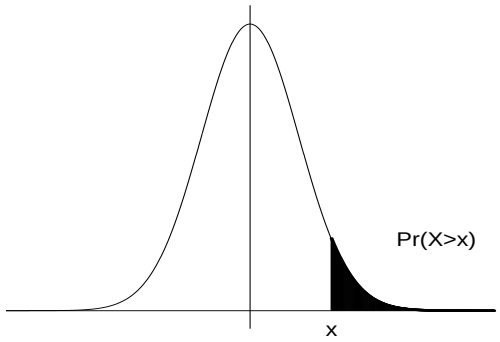


Figure: Mass of the tail of a Gaussian distribution

The justification for using stable distribution models lies in the generalized central limit theorem which includes the well known "traditional" CLT as a special case.

A random variable X is stable if it can be the limit of a normalized sum of i.i.d. random variables.

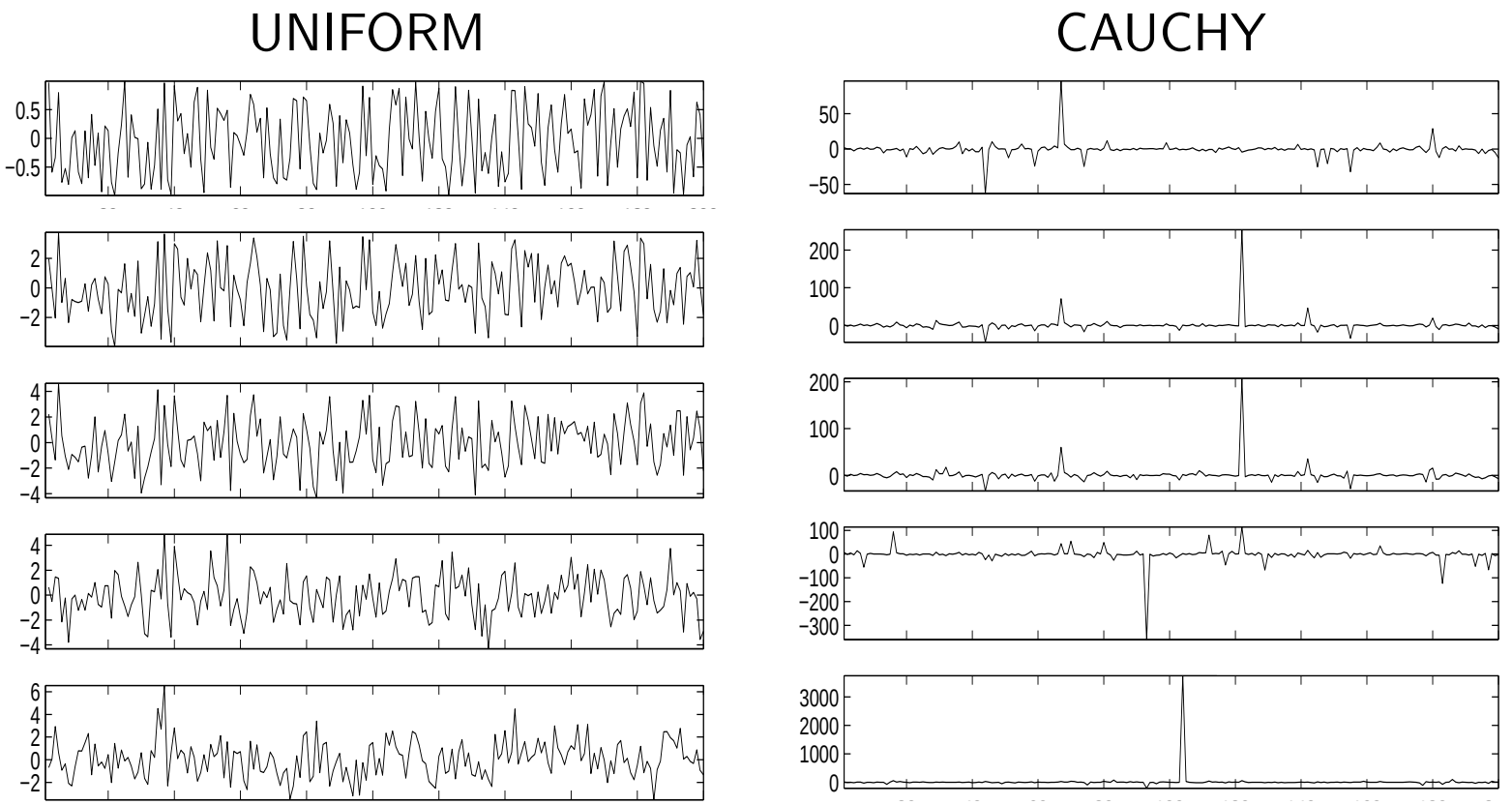


Figure: Traditional Vs. generalized CLT. The plots show the normalized sum of 1, 2, 3, 10 and 30 uniform(-1,1) or Cauchy(0,1) random variables

Generalized Gaussian Distributions

Definition (Generalized Gaussian Distribution)

The p.d.f. for the generalized Gaussian distribution is

$$f(x) = \frac{k}{2\sigma\Gamma(1/k)} \exp^{-(|x-\beta|/\sigma)^k}, \tag{2}$$

where $\Gamma(\cdot)$ is the Gamma function $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$.

The scale is determined by $\sigma > 0$; impulsiveness related to $k > 0$.

- The standard Gaussian distribution is a special case for $k = 2$.
- For $k = 1$, the Laplacian, distribution is

$$f(x) = \frac{1}{2\sigma} e^{-|x-\beta|/\sigma}. \tag{3}$$

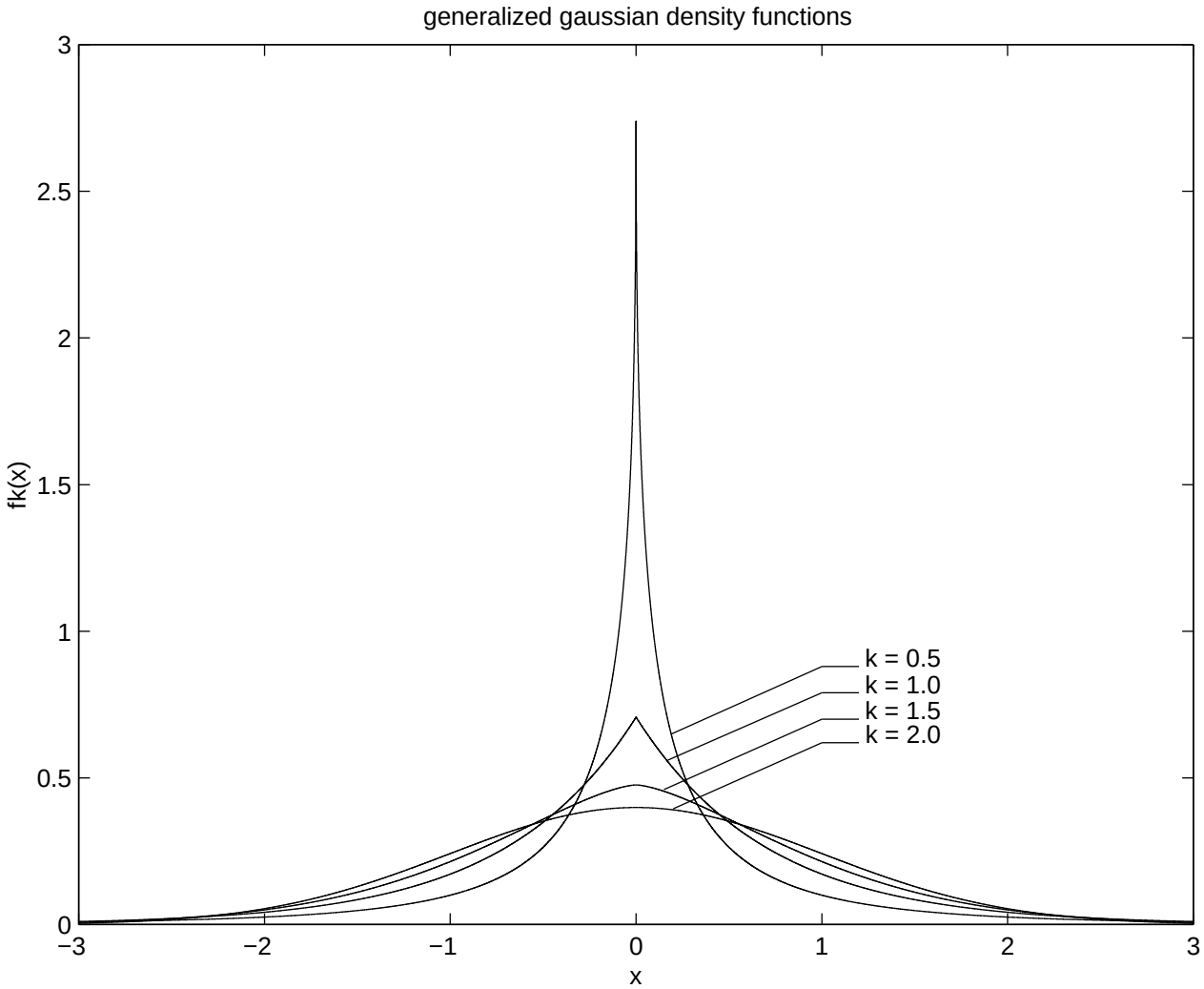


Figure: Generalized Gaussian density functions for different values of k.

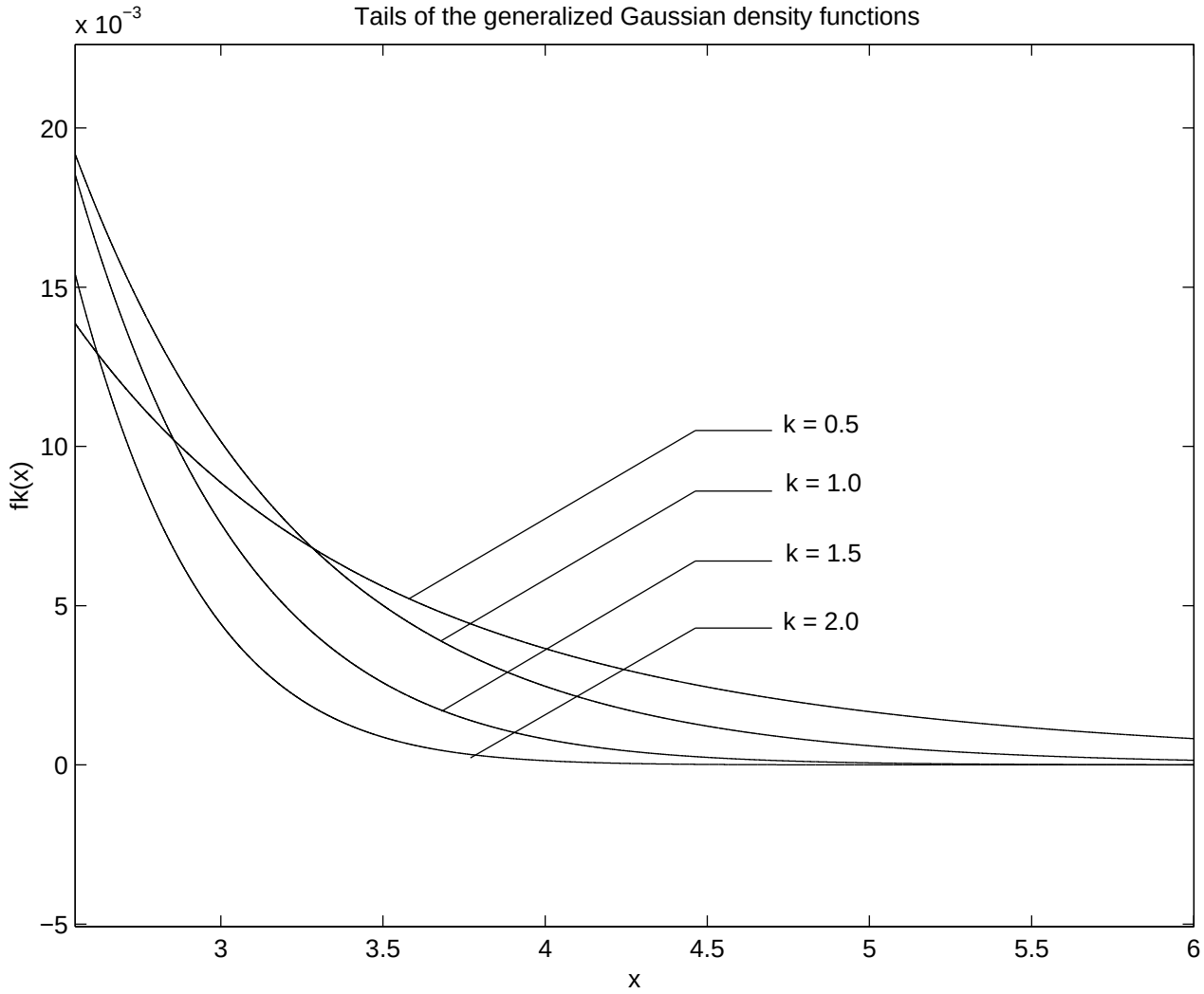


FIGURE: Tails of the Generalized Gaussian density functions for different k.

Definition (Stable Random Variables)

A random variable X is *stable* if for X_1 and X_2 independent copies of X and for arbitrary positive constants a and b, there are constants c and d such that

$$aX_1 + bX_2 \stackrel{d}{=} cX + d. \tag{4}$$

Shape of X is preserved under addition up to scale and shift.

For Gaussian random variables, $c^2 = a^2 + b^2$ and $d = (a + b - c)\mu$ where μ is the mean of the parent Gaussian distribution.

Other stable distributions are the Cauchy and Lévy distributions. The density function, for $X \sim \text{Cauchy}(\gamma, \beta)$ has the form

$$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x-\beta)^2}, \quad -\infty < x < \infty. \tag{5}$$

The Lévy density function is totally skewed concentrating on $(0, \infty)$. The density function for $X \sim \text{Lévy}(\gamma, \delta)$ has the form

$$f(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{1}{(x-\delta)^{3/2}} \exp\left(-\frac{\gamma}{2(x-\delta)}\right), \quad -\delta < x < \infty. \tag{6}$$

Symmetric Stable Distributions

Symmetric α-stable or $S\alpha S$ distributions are defined when the skewness parameter δ is set to zero. These can be characterized by the characteristic function

$$\phi(\omega) = E\exp(j\omega X) = \int_{-\infty}^{\infty} \exp(j\omega x) f(x) dx \tag{7}$$

Definition (Characteristic Function of SαS Distributions)

A *symmetrically stable* random variable is characterized by

$$\phi(\omega) = e^{-\gamma|\omega|^{\alpha}}. \tag{8}$$

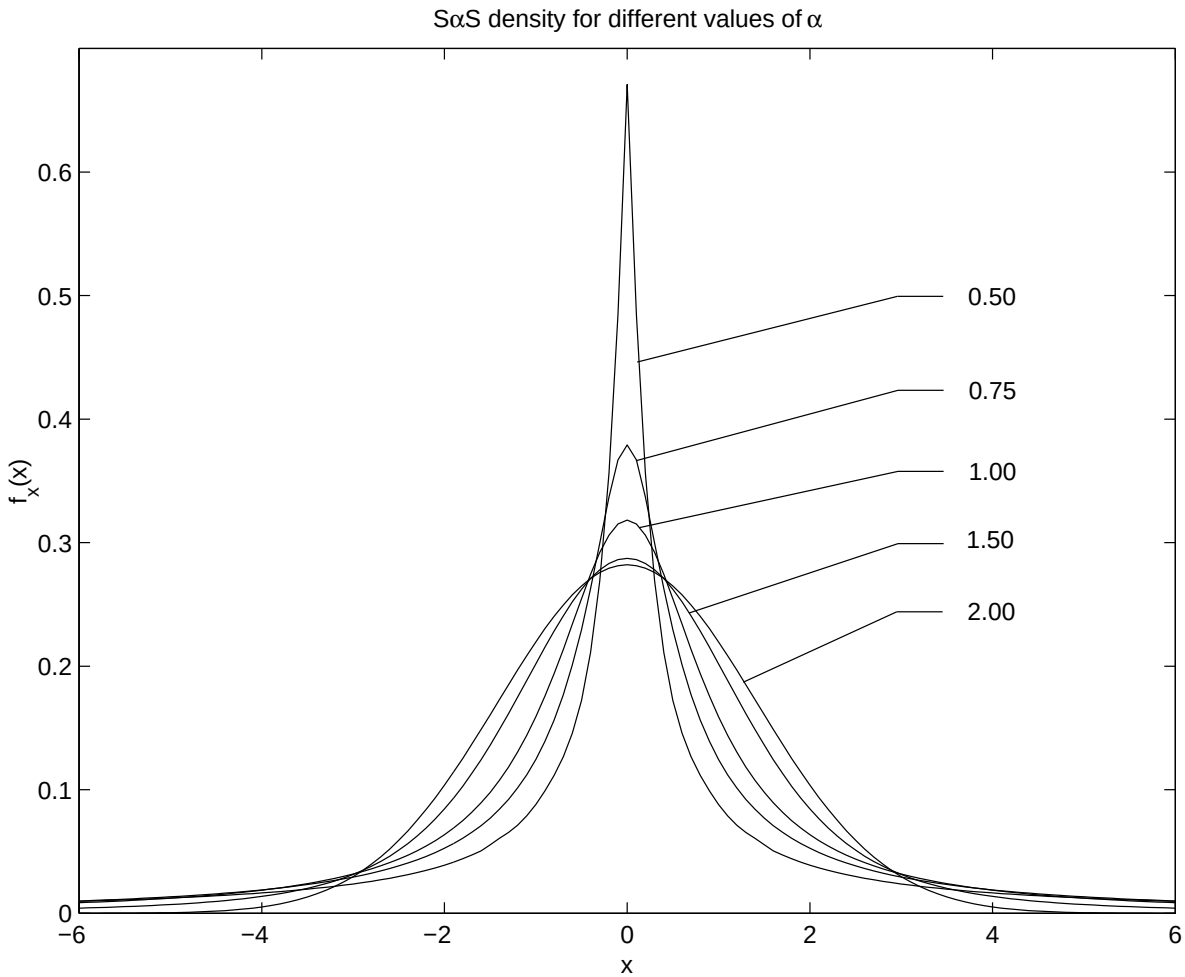


Figure: Density functions of Symmetric stable distributions for different values of the tail constant α.

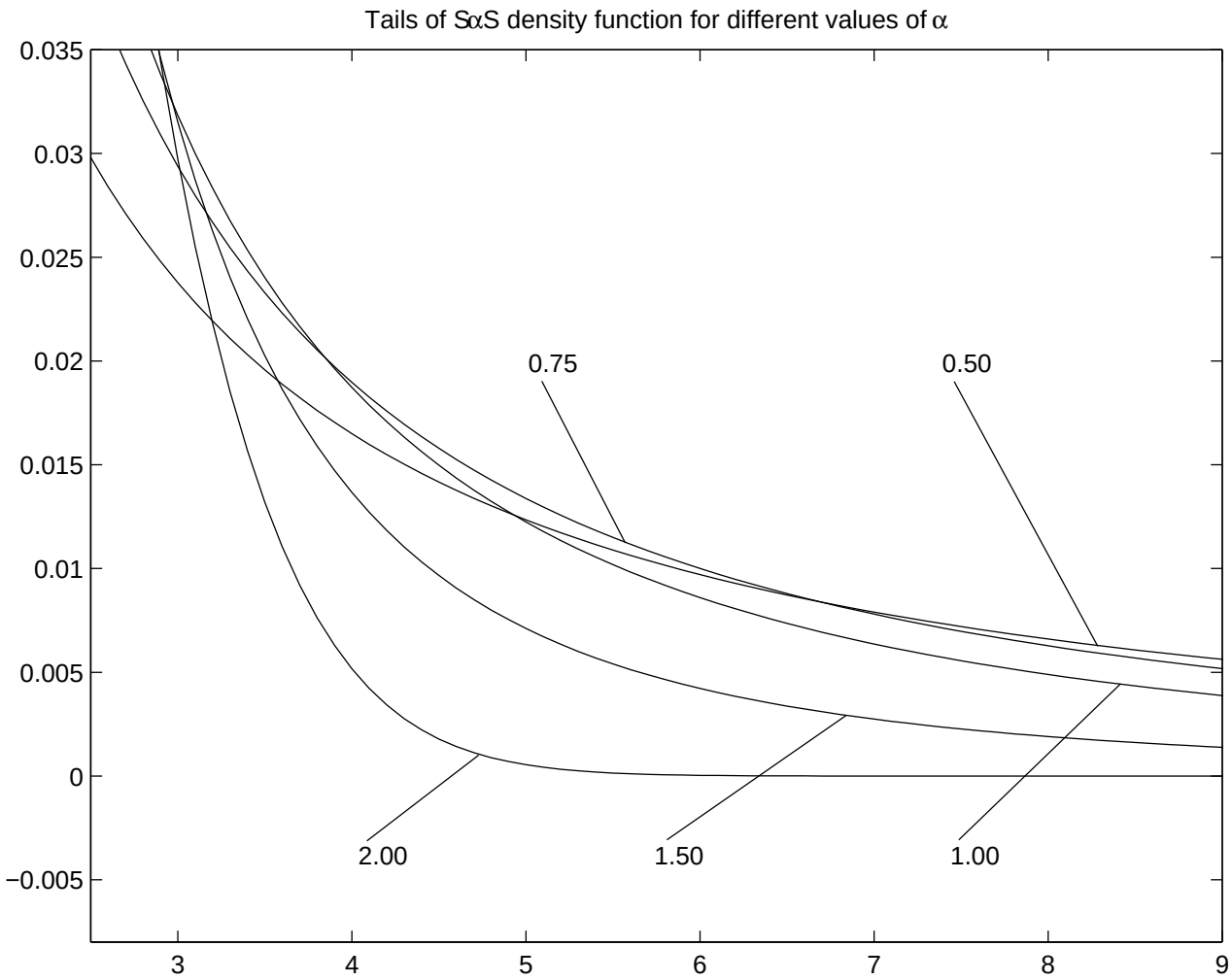


FIGURE: Tails of symmetric stable distributions for different values of the tail constant α.

Generalized Central Limit Theorem

Theorem (Generalized Central Limit Theorem)

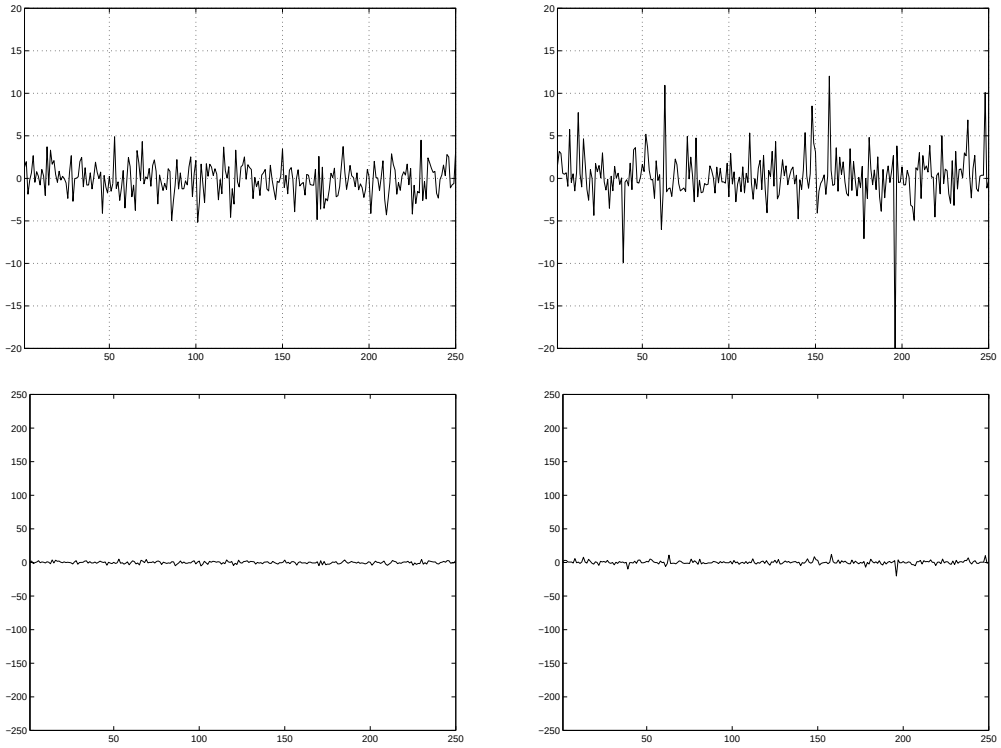
Let X_1, $X_2, \cdots$ be an independent, identically distributed sequence of (possibly shift corrected) random variables. There exist constants a_n such that as $n \to \infty$ the sum

$$a_n(X_1 + X_2 + \cdots) \xrightarrow{d} Z \tag{10}$$

if and only if Z is a stable random variable with some $0 < \alpha \leq 2$.

The generalized CLT constitutes a strong argument compelling the use of stable models in practice.

Figures 7 and 8 illustrate the impulsive behavior of symmetric stable processes as the characteristic exponent α is varied.



$\alpha = 2$ $\alpha = 1.7$

FIGURE: Impulsive behavior of i.i.d. α-stable signals as the tail constant α is varied. Signals are plotted twice under two different scales.

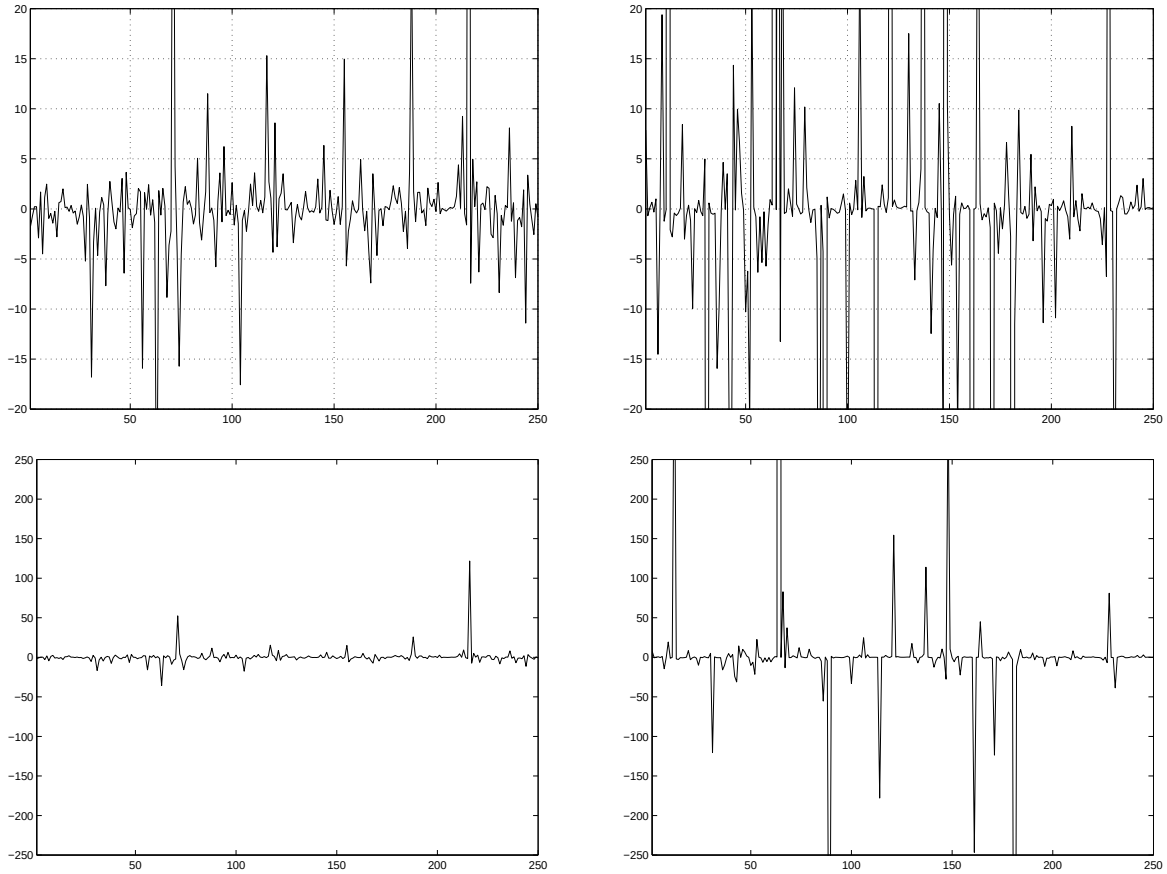


Figure: Impulsive behavior of i.i.d. α-stable signals as the tail constant α is varied. Signals are plotted twice under two different scales.

The location estimate:

Suppose that a constant signal β is transmitted through a channel which adds Gaussian noise Z_i. Several independent observations X_i are measured giving

$$X_i = \beta + Z_i \qquad i = 1, 2, \cdots, N.$$

Given $X_1,\ X_2, \cdots,\ X_N$, the goal is to derive a "good" estimate of β.
Estimates of this kind are known as *location estimates*, a *key* in the formulation of the optimal filtering problem.

Several methods of estimating β are possible.

- The sample mean:

$$\bar{\beta}_N = \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

- The sample median $\tilde{\beta}_N = \tilde{X}$.
- The *trimmed-mean* (the largest and smallest samples are first discarded and the remaining $N-2$ samples are averaged.)

Which one of these estimators, if any, is correct will depend on the criterion which is selected.

Location Estimation in Gaussian Noise

Assume that $X_1, X_2, \cdots, X_N$, are i.i.d. Gaussian with a constant but unknown mean β. The Maximum Likelihood estimate of location is the value $\hat{\beta}$ which maximizes the likelihood function

$$\begin{aligned} f(X_1, X_2, \cdots, X_N; \beta) &= \prod_{i=1}^{N} f(X_i - \beta) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \, e^{-(X_i-\beta)^2/2\sigma^2} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} e^{-\sum_{i=1}^{N}(X_i-\beta)^2/2\sigma^2}. \end{aligned} \tag{5}$$

The ML estimate of location is the value $\hat{\beta}$ which minimizes the least squares sum

$$\hat{\beta}_{ML} = \arg\min_{\beta} \sum_{i=1}^{N} (X_i - \beta)^2. \tag{6}$$

The value that minimizes the sum, results in the sample mean

$$\hat{\beta}_{ML} = \frac{1}{N} \sum_{i=1}^{N} X_i. \tag{7}$$

Note that the sample mean is unbiased since $E\{\hat{\beta}_{ML}\} = (1/N) \sum_{i=1}^{N} E\{X_i\} = \beta$. As a ML estimate, it is efficient having its variance, in (1), reach the Cramér-Rao bound.

Location Estimation in Generalized Gaussian Noise

In the generalized Gaussian distribution case, the Maximum Likelihood estimate of location is

$$\begin{aligned} f(X_1, X_2, \cdots, X_N; \beta) &= \prod_{i=1}^{N} f_\gamma(X_i - \beta) \\ &= \prod_{i=1}^{N} C\ e^{-|X_i-\beta|^\gamma/\sigma} \\ &= C^N e^{-\sum_{i=1}^{N}|X_i-\beta|^\gamma/\sigma}, \end{aligned} \tag{8}$$

where C is a normalizing constant, and γ is the dispersion parameter. Maximizing the likelihood function is equivalent to

$$\tilde{\beta}_{ML} = \arg\min_{\beta} \sum_{i=1}^{N} |X_i - \beta|^\gamma. \tag{9}$$

$$\tilde{\beta}_{ML} = \arg\min_{\beta} \sum_{i=1}^{N} |X_i - \beta|^{\gamma}.$$

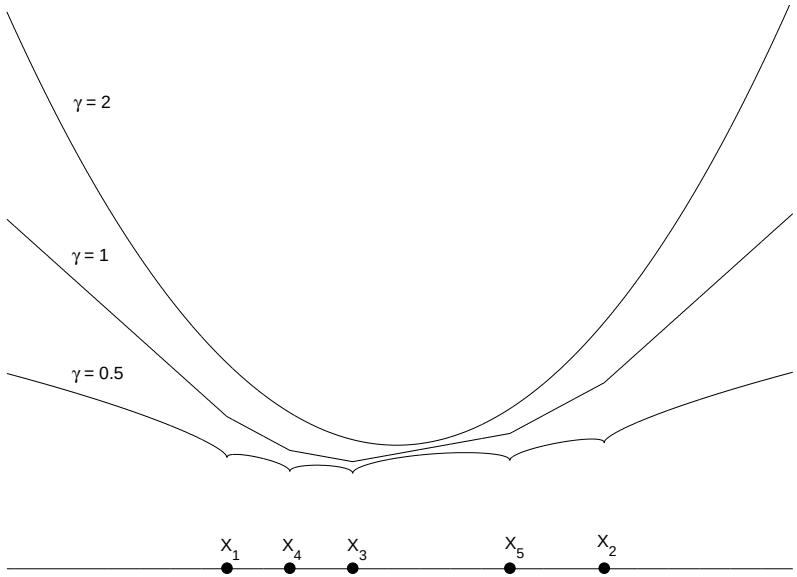


Figure: Cost functions for the observation samples $X_1 = -3, X_2 = 10, X_3 = 1, X_4 - 1, X_5 = 6$ for $\gamma = 0.5, 1,$ and 2.

When the dispersion parameter is 1, the model is Laplacian and the optimal estimator minimizes

$$\tilde{\beta}_{ML} = \arg\min_{\beta} \sum_{i=1}^{N} |X_i - \beta|. \tag{10}$$

The solution to the above is the sample median as it is shown next. Define the cost function in (10) as $L_1(\beta)$. For values of β in the interval $-\infty < \beta \leq X_{(1)}$, $L_1(\beta)$ is simplified to

$$L_1(\beta) = \sum_{i=1}^{N} (X_{(i)} - \beta) = \sum_{i=1}^{N} X_{(i)} - N\beta. \tag{11}$$

This, as a direct consequence that in this interval, $X_{(1)} \geq \beta$.

For N odd there is an integer k, such that the slopes over the intervals $(X_{(k-1)}, X_{(k)}]$ and $(X_{(k)}, X_{(k+1)}]$, are negative and positive, respectively. From (14), these two conditions are satisfied if both

$$k < \frac{N}{2} \qquad \text{and} \qquad k > \frac{N}{2} - 1$$

hold. Both constraints are met when $k = \frac{N+1}{2}$

$$\begin{aligned} \hat{\beta}_{ML} &= \arg\min_{\beta} \sum_{i=1}^{N} |X_i - \beta| \\ &= \begin{cases} X_{(\frac{N+1}{2})} & N \text{ odd} \\ \left(X_{(\frac{N}{2})}, X_{(\frac{N}{2})}\right] & N \text{ even} \end{cases} \\ &= \text{MEDIAN}(X_1, X_2, \cdots, X_N). \end{aligned} \tag{15}$$

Location Estimation in Stable Noise

Maximum likelihood estimation requires the knowledge of the density function. Among the class of symmetric stable densities, only the Gaussian ($\alpha = 2$) and Cauchy ($\alpha = 1$) distributions have closed-form expressions.

- The only non-Gaussian distribution for which we have a closed form expression is the Cauchy distribution.
- ML estimates under the Cauchy model can be made tunnable acquiring remarkable efficiency over the entire spectrum of stable distributions.

Given a set of i.i.d. samples $X_1, X_2, \cdots, X_N$ obeying the Cauchy distribution with scaling factor k,

$$f(x-\beta) = \frac{k}{\pi}\ \frac{1}{k^2+(x-\beta)^2}, \tag{16}$$

the location parameter β is to be estimated from the data samples as the value $\hat{\beta}_k$ which maximizes the likelihood function

$$\hat{\beta}_k = \arg\max_{\beta} \prod_{i=1}^{N} f(X_i-\beta) = \arg\max_{\beta} \left(\frac{k}{\pi}\right)^N \prod_{i=1}^{N} \frac{1}{k^2+(X_i-\beta)^2} \tag{17}$$

This is equivalent to minimizing

$$G_k(\beta) = \prod_{i=1}^{N} [k^2 + (X_i - \beta)^2]. \tag{18}$$

Given $k > 0$, the ML location estimate is known as the sample *myriad* and is given by

$$\begin{aligned} \hat{\beta}_k &= \arg\min_{\beta} \prod_{i=1}^{N} \left(k^2 + (X_i - \beta)^2\right) \\ &= \text{MYRIAD}\{k; X_1, X_2, \cdots, X_N\}. \end{aligned} \tag{19}$$

The sample myriad involves the free parameter k (refered to as the *linearity parameter*). The behavior of the myriad is markedly dependent on the value of k.

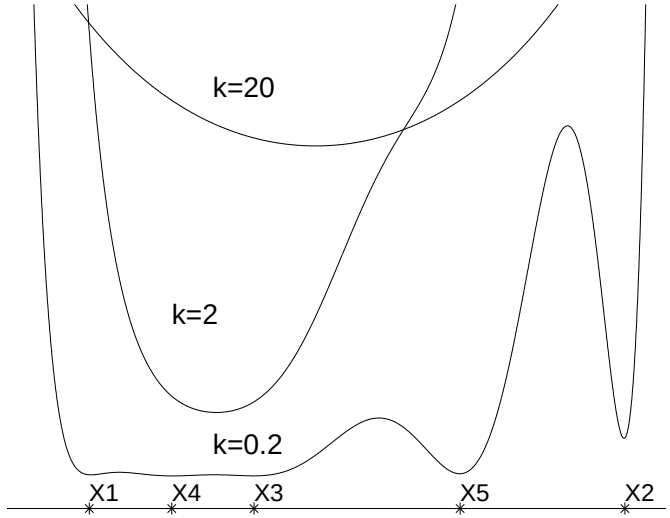


Figure: Myriad cost functions for the observation samples $X_1 = -3, X_2 = 10, X_3 = 1, X_4 - 1, X_5 = 6$ for $k = 20, 2, 0.2$.

Least Logarithmic Deviation

The sample myriad minimizes

$$G_k(\beta) = \prod_{i=1}^{N} [k^2 + (X_i - \beta)^2].$$

Since the logarithm is a strictly monotonic function, then the sample myriad will also minimize $\log G_k(\beta)$.

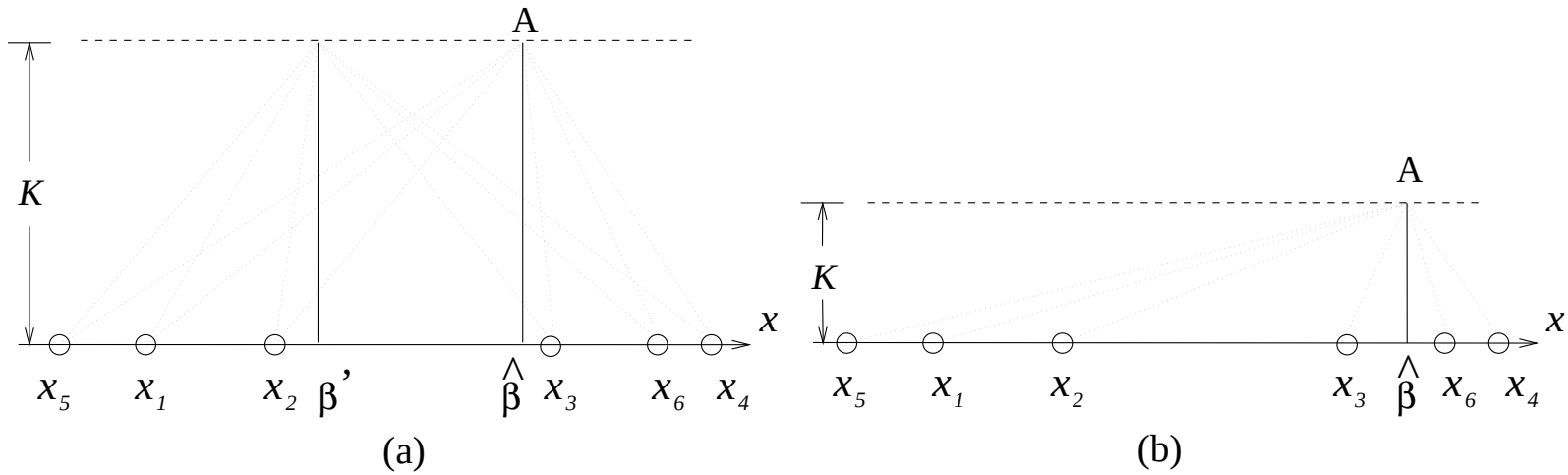
$$\text{MYRIAD}\{k; X_1, \cdots, X_N\} = \arg\min_{\beta} \sum_{i=1}^{N} \log\left[k^2 + (X_i - \beta)^2\right]. \qquad (20)$$

Geometrical Interpretation

The observations $X_1,\ X_2, \cdots,\ X_N$ are placed along the real line. Next, a vertical bar that runs horizontally through the real line is added. The length of the vertical bar is equal to k. Each of the terms

$$\left(k^2 + (X_i - \beta)^2\right) \tag{21}$$

in (20), represents the distance from point A, at the end of the vertical bar, to the sample point X_i.



The sample myriad, $\hat{\beta}_k$, indicates the position of the bar for which the product of distances from point A to the samples $X_1,\ X_2, \cdots,\ X_N$ is minimum. Any other value, such as $x = \beta'$, produces a higher product of distances.

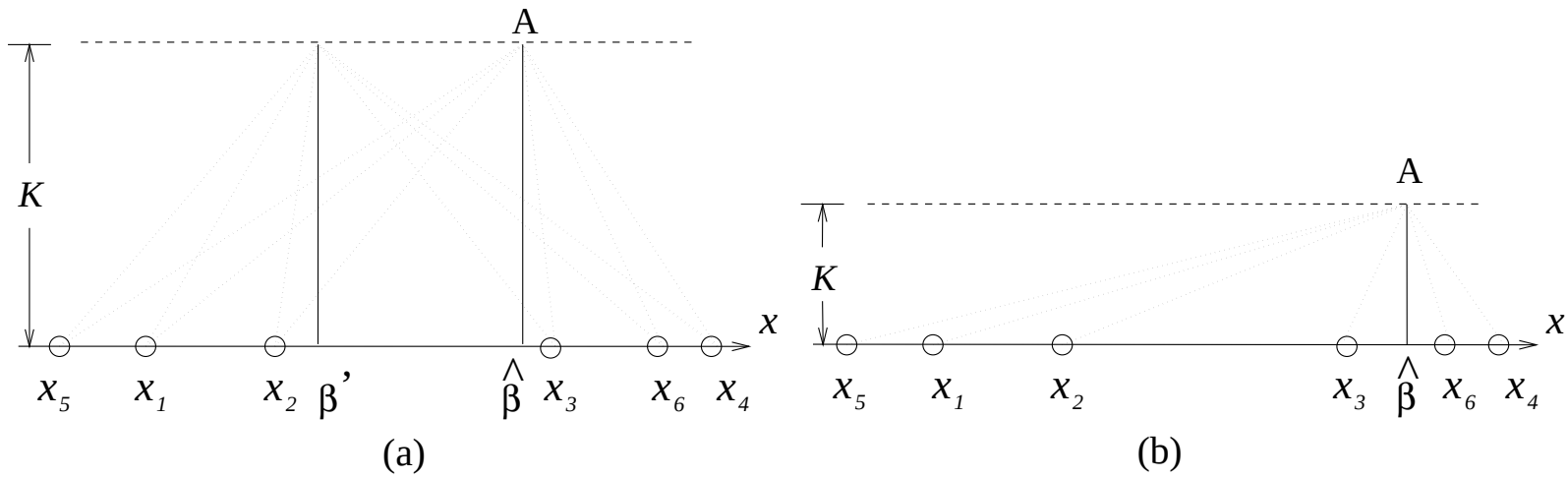


Figure: (a) The sample myriad, $\hat{\beta}$, minimizes the product of distances from point A to all samples. (b) the myriad as k is reduced.

Weighted Median Smoothers

- Running medians are temporally blind.
- All observation samples are treated equally regardless of their location within the observation window.
- Better smoothers are obtained if weighting is allowed.

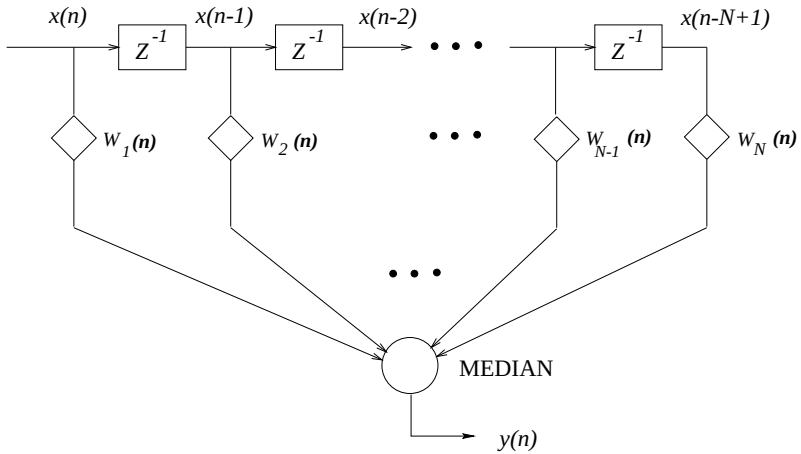


Figure: The weighted median smoothing operation .

Statistical Foundations

- Although time–series, in general, exhibit temporal correlation, the independent but not identically distributed model can be used to synthesize the mutual correlation.
- The estimate $Y(n)$ can rely more on the sample $X(n)$ than on the other samples. $X(n)$ is more reliable than $X(n-1)$ or $X(n+1)$, which in turn are more reliable than $X(n-2)$ or $X(n+2)$, and so on.
- By assigning different variances (reliabilities) to the independent but not identically distributed location estimation model, the temporal correlation used in time-series smoothing is synthesized.

For the generalized Gaussian distribution, where the samples have a common location parameter β, but different scale parameter σ_i. The ML estimate of location is

$$G_p(\beta) = \sum_{i=1}^{N} \frac{1}{\sigma_i^p} |X_i - \beta|^p. \tag{6}$$

In the special case of the Gaussian distribution ($p = 2$), the ML estimate reduces to the normalized weighted average

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (X_i - \beta)^2 = \frac{\sum_{i=1}^{N} W_i \cdot X_i}{\sum_{i=1}^{N} W_i} \tag{7}$$

where $W_i = 1/\sigma_i^2 > 0$.

In the case of a Laplacian distribution ($p = 1$), the ML estimate minimizes

$$G_1(\beta) = \sum_{i=1}^{N} \frac{1}{\sigma_i} |X_i - \beta|. \tag{8}$$

where $W_i \stackrel{\triangle}{=} 1/\sigma_i > 0$. $G_1(\beta)$ is piecewise linear and convex.
The value β minimizing (8) is one of the samples $X_1,\ X_2, \ldots,\ X_N$. The weighted median output is defined as

$$\begin{aligned} Y(n) &= \arg\min_{\beta} \sum_{i=1}^{N} W_i |X_i - \beta| \\ &= \text{MEDIAN}[W_1 \diamond X_1(n),\ W_2 \diamond X_2(n), \cdots,\ W_N \diamond X_N(n)] \end{aligned}$$

where $W_i > 0$ and $\diamond$ is the replication operator defined as

$$W_i \diamond X_i = \overbrace{X_i,\ X_i, \cdots,\ X_i}^{W_i \text{ times}}.$$

Weighted Median Smoothing Computation

Given the weight vector $\mathbf{W} = \langle 1,\ 2,\ 3,\ 2,\ 1 \rangle$. For $\mathbf{X}(n) = [12,\ 6,\ 4,\ 1,\ 9]$, the weighted median smoother output is

$$\begin{aligned} Y(n) &= \text{MEDIAN}\,[\ 1\diamond 12,\ 2\diamond 6,\ 3\diamond 4,\ 2\diamond 1,\ 1\diamond 9\ ] \\ &= \text{MEDIAN}\,[\ 12,\ 6,\ 6,\ 4,\ 4,\ 4,\ 1,\ 1,\ 9\ ] \\ &= \text{MEDIAN}\,[\ 1,\ 1,\ 4,\ 4,\ \underline{4},\ 6,\ 6,\ 9,\ 12\ ] \\ &= 4 \end{aligned} \tag{9}$$

The standard median output for the given input is $Y(n) = 6$.

The Center Weighted Median Smoother

WM smoothers can emphasize or deemphasize specific input samples. The sample most correlated with the desired estimate is the center observation sample. This observation leads to the center weighted median (CWM) smoother:

$$Y(n) = \text{MEDIAN}[X_1, \ldots, X_{c-1}, W_c \diamond X_c, X_{c+1}, \ldots, X_N],$$

where W_c is an odd positive integer and $c = (N+1)/2 = N_1 + 1$ is the index of the center sample.
When $W_c = 1$, we get a median smoother, and for $W_c \geq N$, the CWM reduces to an identity operation.

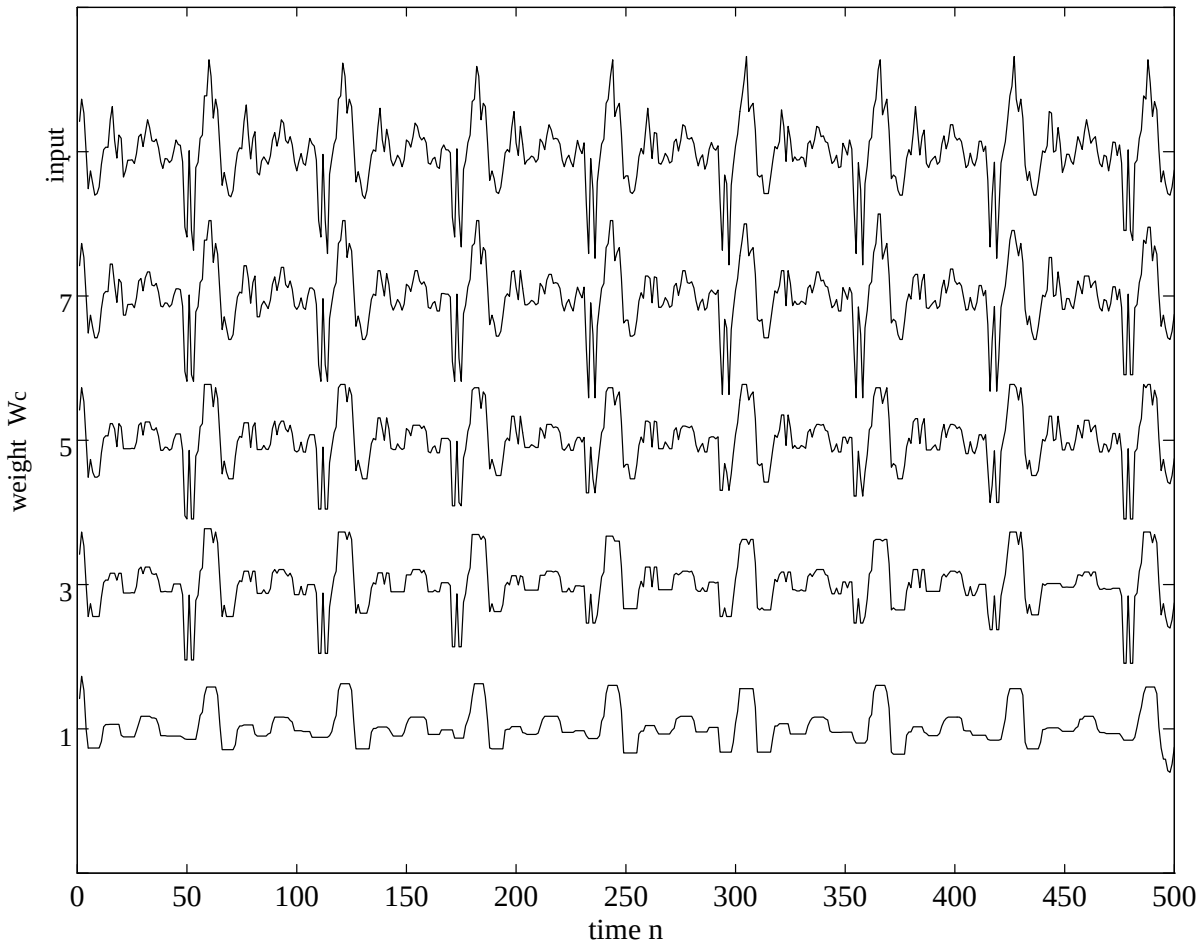


FIGURE: Effects of increasing the center weight of a CWM smoother of size $N = 9$ operating on the voiced speech "a".

The output of a CWM smoother is equivalent to computing

$$Y(n) = \text{MEDIAN}\left[X_{(k)},\ X_c,\ X_{(N+1-k)}\right], \tag{14}$$

where $k = (N + 2 - W_c)/2$ for $1 \leq W_c \leq N$, and $k = 1$ for $W_c > N$.
Since $X_c = X(n)$, the output of the smoother is identical to the input as long as $X(n)$ lies in the interval $\left[X_{(k)}, X_{(N+1-k)}\right]$.
If $X_c > X_{(N+1-k)}$ the smoother outputs $X_{(N+1-k)}$, guarding against a possible aberrant data point being taken as the output. Similarly, the smoother's output is $X_{(k)}$ if the sample $X(n)$ is smaller than this order statistic.

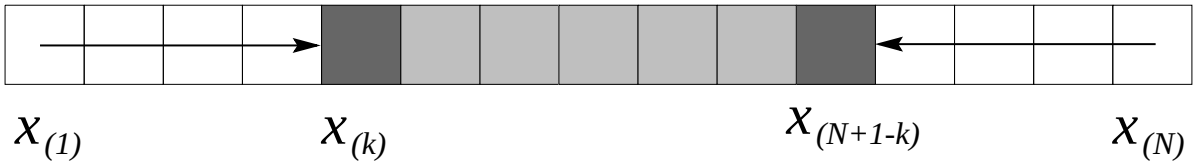


FIGURE: The center weighted median smoothing operation.

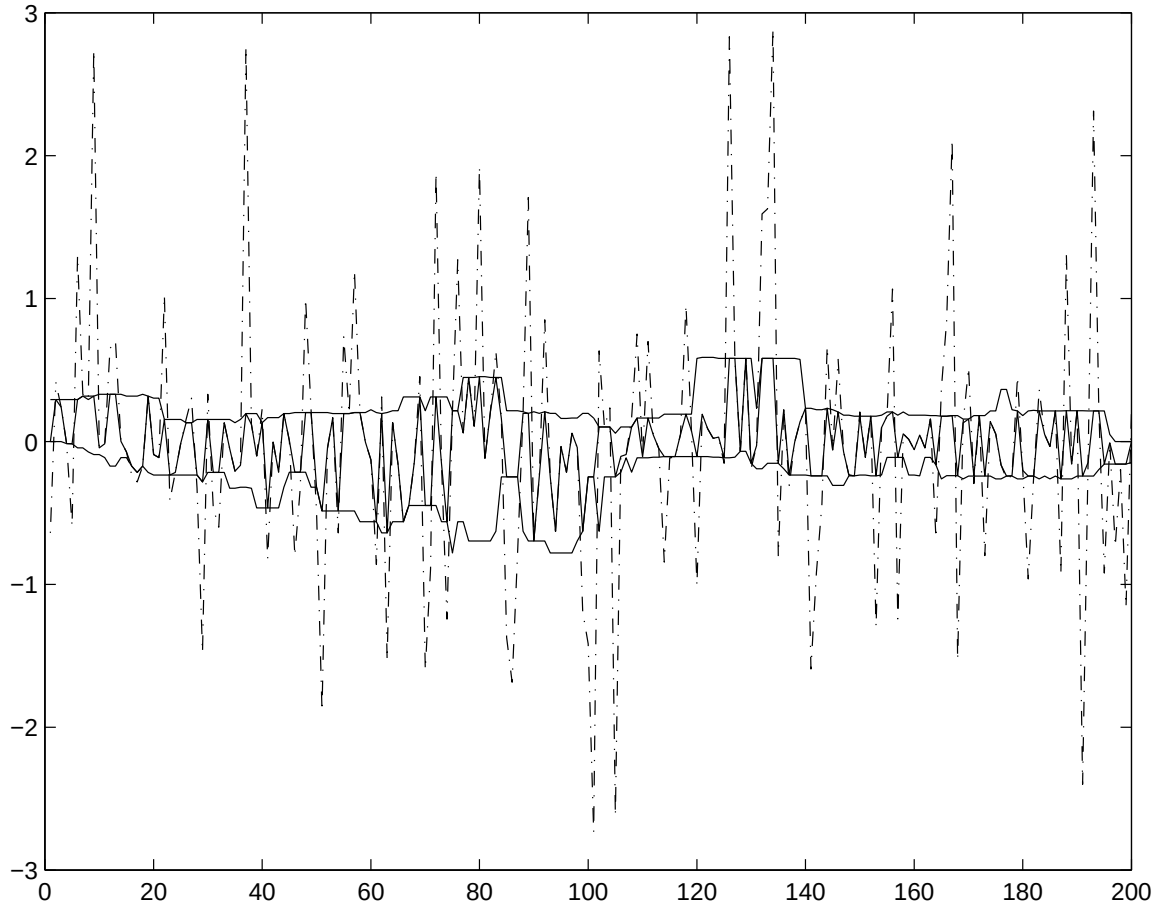


FIGURE: An example of the CWM smoother operating on a Laplacian distributed sequence with unit variance. Shown are the input and output sequences as well as the trimming statistics $X_{(k)}$ and $X_{(N+1-k)}$. $N = 25$ and $k = 7$.

Application of CWM Smoother To Image Cleaning

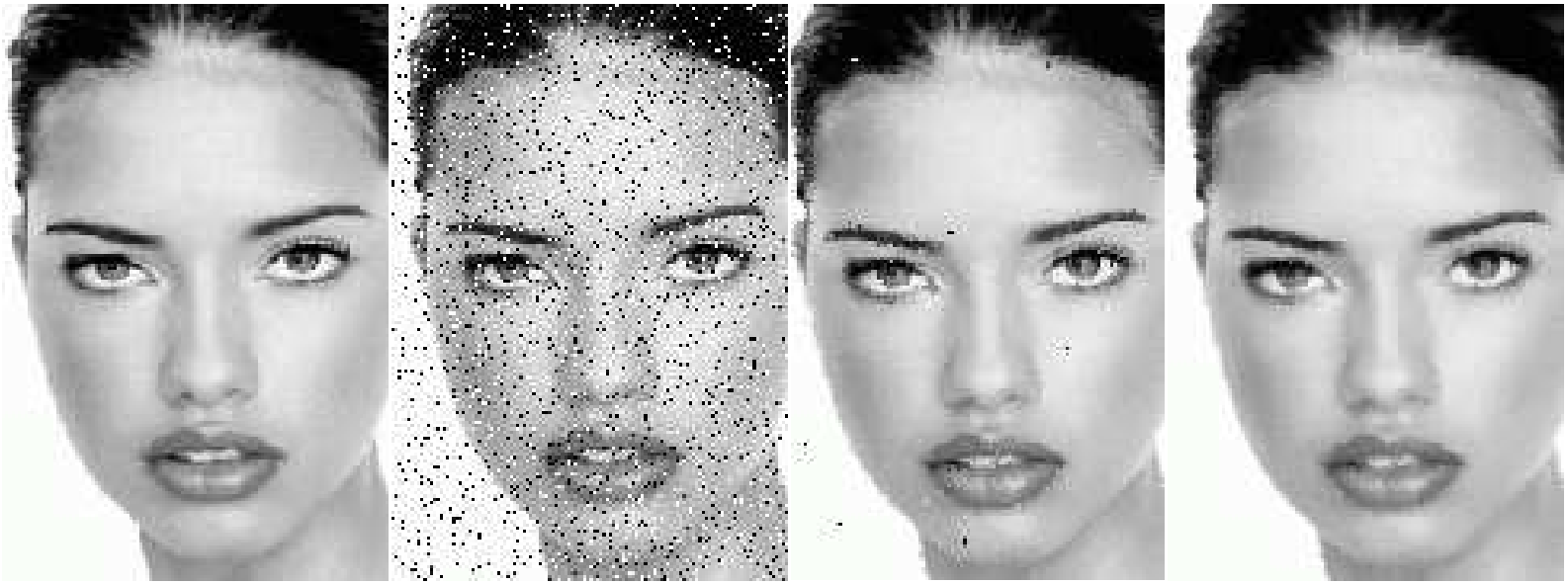


FIGURE: Impulse noise cleaning with a 5×5 CWM smoother: (a) original, (b) image with salt and pepper noise, (c) CWM smoother with $W_c = 15$, (d) CWM smoother with $W_c = 5$.

Figure: (Enlarged) Noise-free image (left), 5×5 median smoother (center), and 5×5 mean smoother (right).

Running Myriad Smoothers

Given an observation vector $\mathbf{X}(n) = [X_1(n), X_2(n), \ldots, X_N(n)]$ and a fixed positive (tunable) value of K, the running myriad smoother output at time n is computed as

$$\begin{aligned} Y_K(n) &= \text{MYRIAD}[K; X_1(n), X_2(n), \ldots, X_N(n)] \\ &= \arg\min_{\beta} \prod_{i=1}^{N} \left[K^2 + (X_i(n) - \beta)^2\right]. \qquad (4) \\ &= \arg\min_{\beta} \sum_{i=1}^{N} \log\left[K^2 + (X_i(n) - \beta)^2\right]. \qquad (5) \end{aligned}$$

The myriad $Y_K(n)$ is thus the value of β that minimizes the above cost function.

The definition of the sample myriad involves the free-tunable parameter K. This parameter will be shown to play a critical role in characterizing the behavior of the myriad.

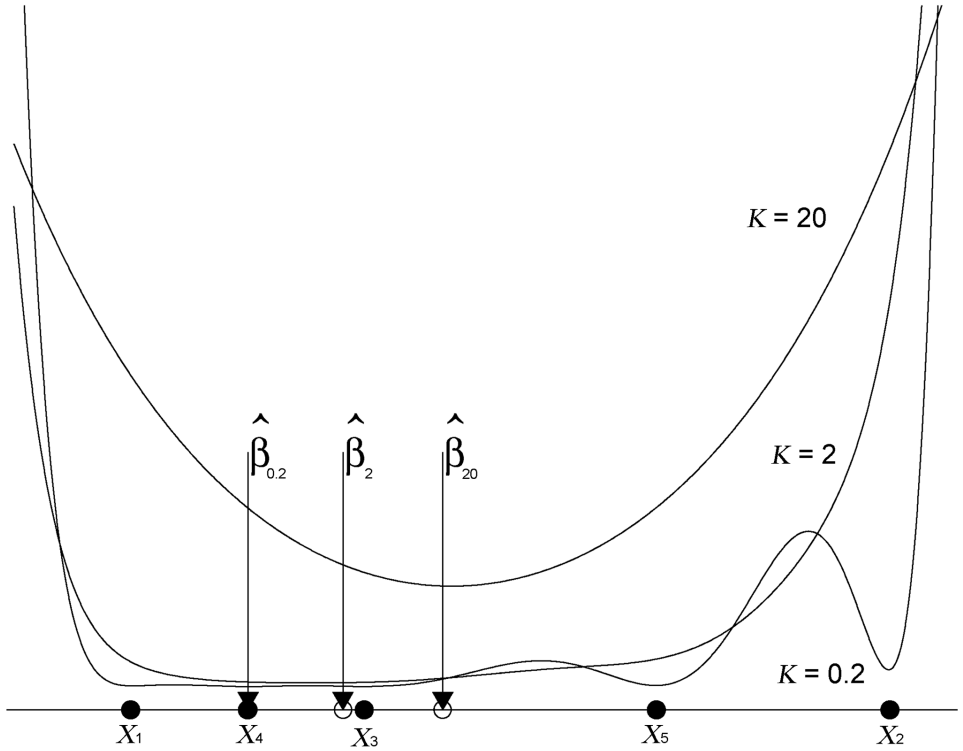


Figure: Myriad cost functions for different values of k

Property (Linear Property)

Given a set of samples, $X_1, X_2, \ldots, X_N$, the sample myriad $\hat{\beta}_K$ converges to the sample average as $K \to \infty$. This is,

$$\begin{aligned} \lim_{K\to\infty} \hat{\beta}_K &= \lim_{K\to\infty} \text{MYRIAD}(K; X_1, \ldots, X_N) \\ &= \frac{1}{N}\sum_{i=1}^{N} X_i. \end{aligned} \tag{6}$$

Definition (Sample mode-myriad)

Given a set of samples $X_1,\ X_2, \ldots,\ X_N$, the mode-myriad estimator, $\hat{\beta}_0$, is defined as

$$\hat{\beta}_0 = \lim_{K \to 0} \hat{\beta}_K, \tag{10}$$

where $\hat{\beta}_K = \text{MYRIAD}(K; X_1, X_2, \ldots, X_N)$.

Property (Mode Property)

The mode-myriad $\hat{\beta}_0$ is always equal to one of the most repeated values in the sample. Furthermore,

$$\hat{\beta}_0 = \arg\min_{X_j \in \mathcal{M}} \prod_{i=1, X_i \neq X_j}^{N} |X_i - X_j|, \tag{11}$$

where $\mathcal{M}$ is the set of most repeated values.

Example

Denoising of a very impulsive signal.

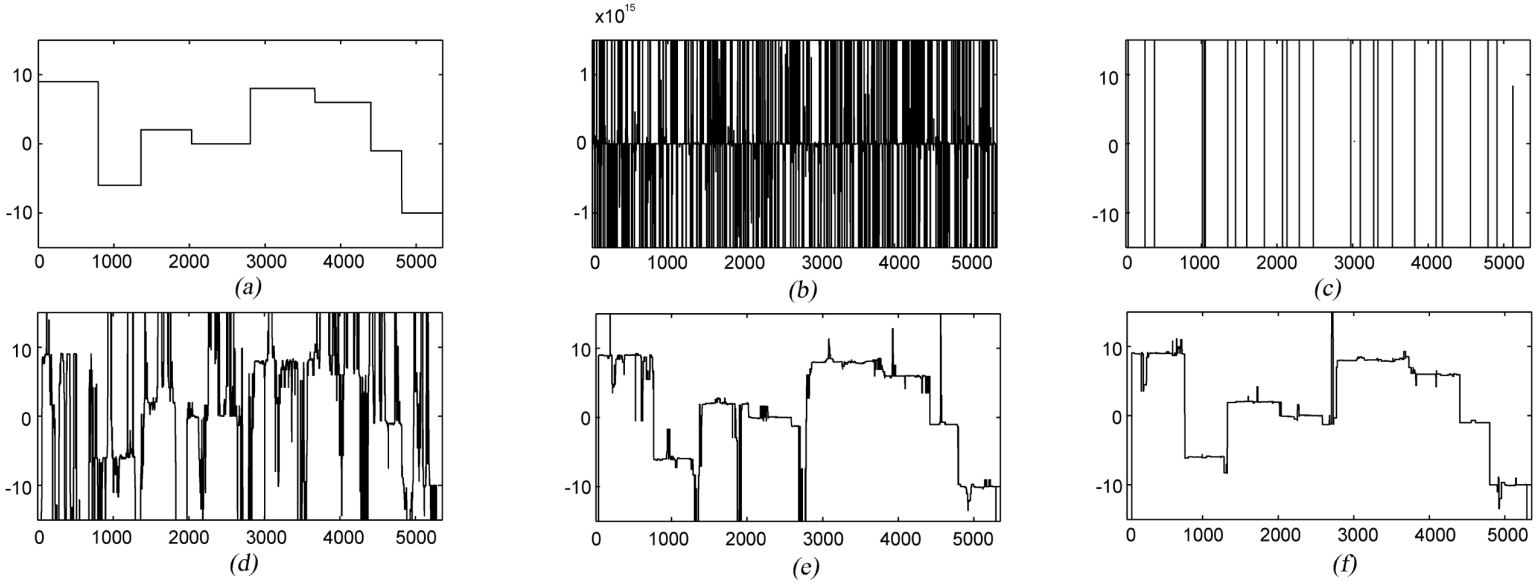


Figure: Running smoothers in stable noise ($\alpha = 0.2$). All smoothers of size 121; (a) original blocks signal, (b) corrupted signal with stable noise, (c) the output of the running mean, (d) the running median, (e) the running FLOM smoother, and (f) the running mode-myriad smoother.

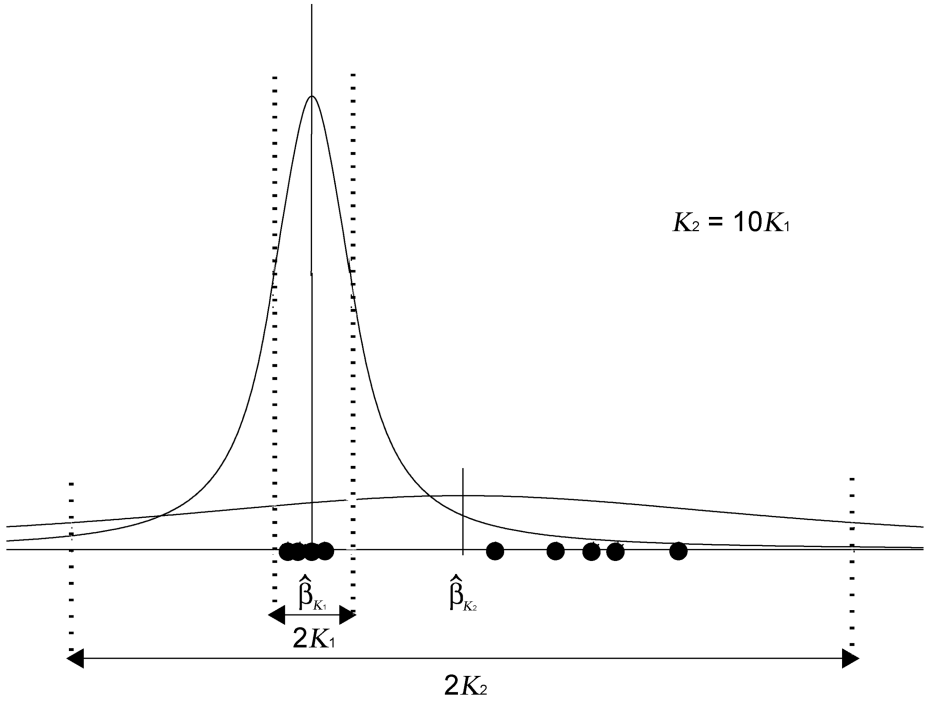


Figure: The role of the linearity parameter when the myriad is looked as a maximum likelihood estimator. When K is large, the generating density function is spread and the data are visualized as well-behaved (the optimal estimator is the sample average). For small values of K, the generating density becomes highly localized, and the data are visualized as very impulsive (the optimal estimator is a cluster locator).

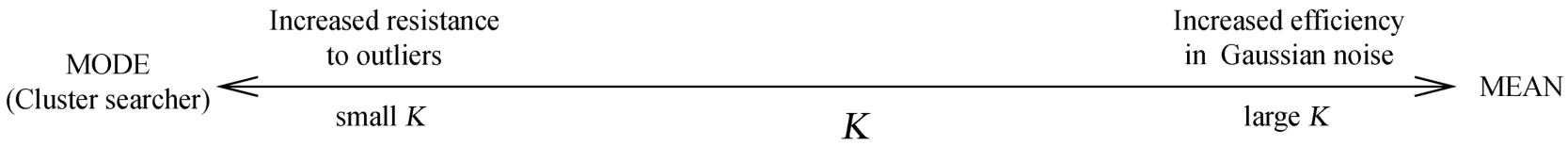


FIGURE: Functionality of the myriad as K is varied. Tuning the linearity parameter K adapts the behavior of the myriad from impulse-resistant mode-type estimators (small K) to the Gaussian-efficient sample mean (large K).

Empirical selection of K:

- Linear type $K \approx X_{(N)} - X_{(1)}$
- Mode type $K \approx \min_{i,j} |X_i - X_j|$

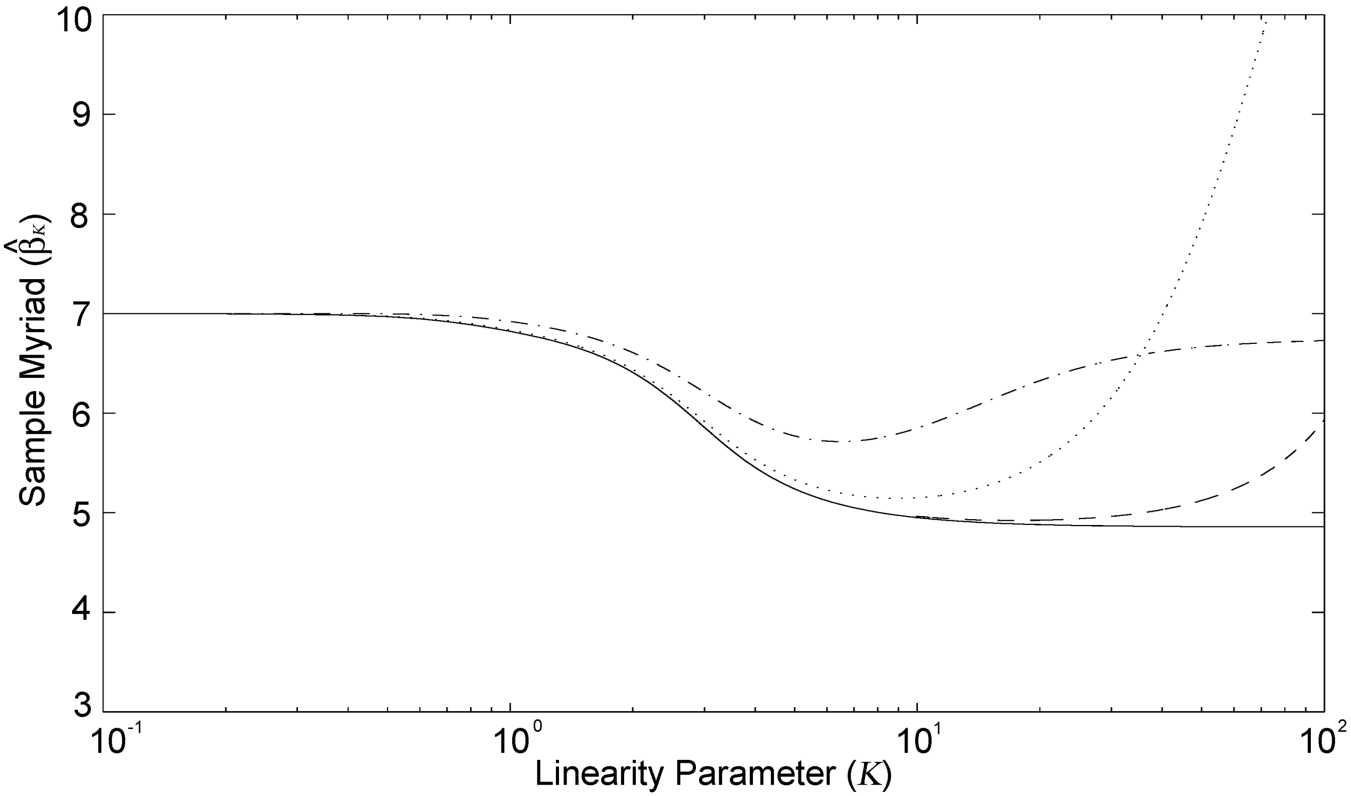


Figure: Values of the myriad as a function of K for the following data sets: (solid) original data set $= 0, 1, 3, 6, 7, 8, 9$; (dash-dot) original set plus an additional observation at 20; (dotted) additional observation at 100; (dashed) additional observations at 800, -500, and 700.

Optimality of the Sample Myriad

Optimality In The α-Stable Model

Proposition

Let $T_{\alpha,\gamma}(X_1, X_2, \ldots, X_N)$ denote the maximum likelihood location estimator derived from a symmetric α-stable distribution with characteristic exponent α and dispersion γ. Then,

$$\lim_{\alpha \to 0} T_{\alpha,\gamma}(X_1, X_2, \ldots, X_N) = \text{MYRIAD}\{0; X_1, X_2, \ldots, X_N\}. \tag{14}$$

The α-stable triplet of optimality points satisfied by the myriad:

- $\alpha = 2 \leftrightarrow K = \infty$
- $\alpha = 1 \leftrightarrow K = \gamma$
- $\alpha = 0 \leftrightarrow K = 0$

Proposition

Let α and γ denote the characteristic exponent and dispersion parameter of a symmetric α-stable distribution. Let $K_o(\alpha, \gamma)$ denote the optimal tuning value of K in the sense that $\hat{\beta}_{K_o}$ minimizes a given performance criterion (usually the variance) among the class of sample myriads with non negative linearity parameter. Then,

$$K_o(\alpha, \gamma) = K_o(\alpha, 1)\gamma. \tag{15}$$

A simple empirical formula is

$$K(\alpha) = \sqrt{\frac{\alpha}{2-\alpha}}, \tag{16}$$

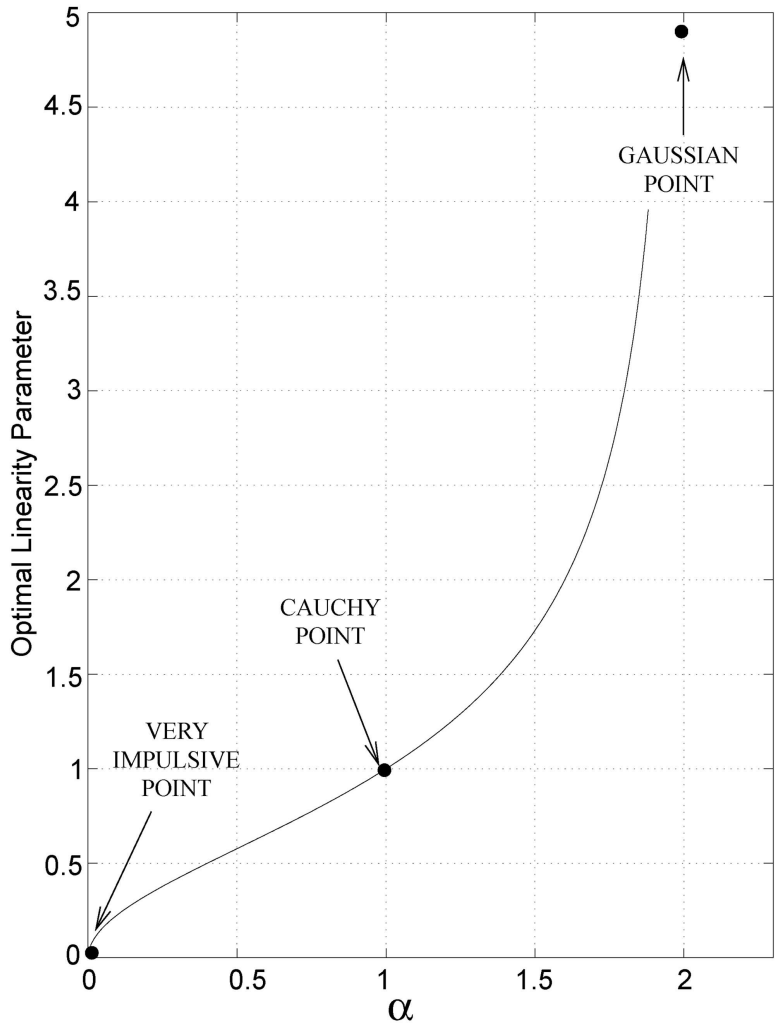


FIGURE: Empirical α-K curve for α-stable distributions. The curve values at $\alpha = 0, 1,$ and 2 constitute the optimality points of the α-stable triplet.

Weighted Myriad Smoothers

Given N observations $\{X_i\}_{i=1}^N$ and nonnegative weights $\{W_i \geq 0\}_{i=1}^N$, let the input and weight vectors be defined as $\mathbf{X} \stackrel{\triangle}{=} [X_1, X_2, \ldots, X_N]^T$ and $\mathbf{W} \stackrel{\triangle}{=} [W_1, W_2, \ldots, W_N]^T$, respectively. For a given *nominal* scale factor K, the underlying random variables are assumed to be independent and Cauchy distributed with a common location parameter β, but varying scale factors $\{S_i\}_{i=1}^N$: $X_i \sim \text{Cauchy}(\beta, S_i)$:

$$f_{X_i}(X_i; \beta, S_i) = \frac{1}{\pi} \frac{S_i}{S_i^2 + (X_i - \beta)^2}, \quad -\infty < X_i < \infty, \tag{17}$$

and where

$$S_i \stackrel{\triangle}{=} \frac{K}{\sqrt{W_i}} > 0, \; i = 1, 2, \ldots, N. \tag{18}$$

Definition (Weighted myriad)

Let $\mathbf{W} = [W_1, W_2, \ldots, W_N]$ be a vector of nonnegative weights. Given $K > 0$, the weighted myriad of order K for the data $X_1, X_2, \ldots, X_N$ is defined as

$$\begin{aligned} \hat{\beta}_K &= \text{MYRIAD}\{K; W_1 \circ X_1, \ldots, W_N \circ X_N\} \\ &= \arg\min_{\beta} \sum_{i=1}^{N} \log\left[K^2 + W_i(X_i - \beta)^2\right], \end{aligned} \tag{22}$$

where $W_i \circ X_i$ represents the weighting operation in (22). In some situations, the following equivalent expression can be computationally more convenient

$$\hat{\beta}_K = \arg\min_{\beta} \prod_{i=1}^{N} \left[K^2 + W_i(X_i - \beta)^2\right]. \tag{23}$$

Figure: (*a*) Original image, (*b*) Image with 5% salt-and-pepper noise (PSNR=17.75dB)

Figure: ((*c*) smoothed with 5×5 center weighted median with $W_c = 15$(PSNR=37.48dB), (*d*) smoothed with 5×5 center weighted myriad with $W_c = 10,000$ and $K = (X_{(21)} + X_{(5)})/2$ (PSNR=39.98dB)

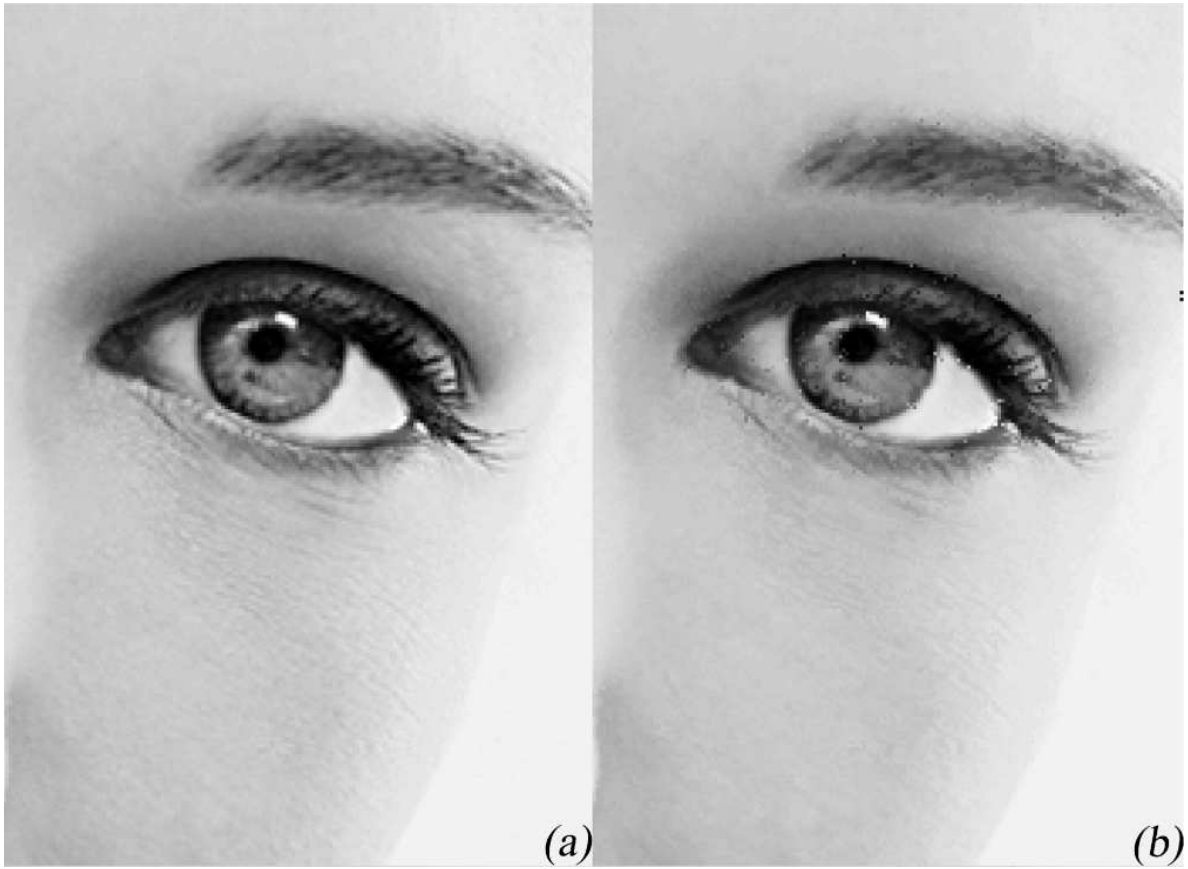


Figure: Comparison of different filtering schemes (Enlarged). (*a*) Original Image, (*b*) Image smoothed with a center weighted median (PSNR=37.48dB)

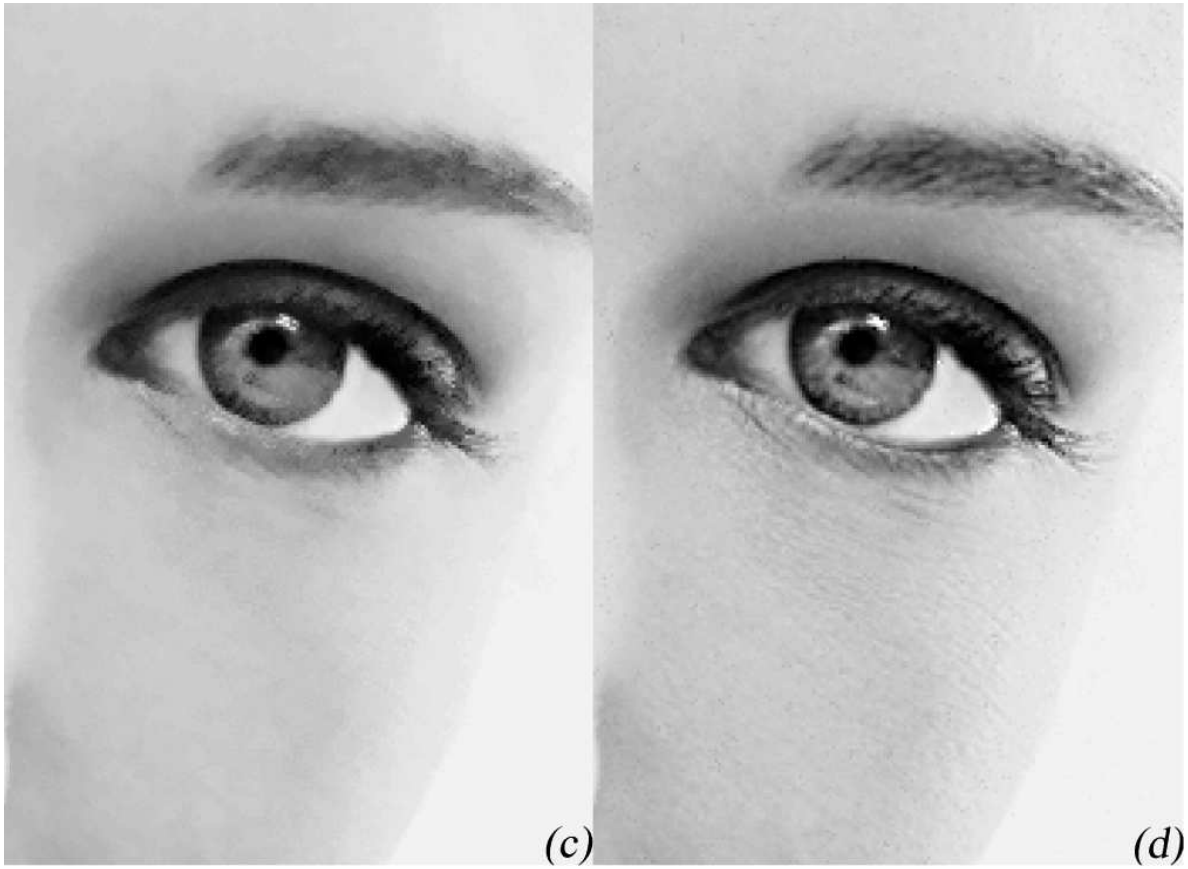


Figure: Comparison of different filtering schemes (Enlarged). (c) Image smoothed with a 5×5 permutation weighted median (PSNR=35.55dB), (d) Image smoothed with the center weighted myriad (PSNR=39.98dB).

Figure: Output of the Center weighted myriad smoother for different values of the center weight W_c (a) Original image, (b) 100 (PSNR=36.74dB)

Figure: Output of the Center weighted myriad smoother for different values of the center weight W_c (*c*) 10,000 (PSNR=39.98dB), (*d*) 1,000,000 (PSNR=38.15dB).

Myriadization

- First, design a constrained linear smoother for Gaussian or noiseless environments using FIR filter (smoother) design techniques.
- Then, plug in these smoother coefficients into weighted myriad structure.
- Choose the suitable K according to the impulsiveness of the environment.
- Note that the smoother coefficients W_i must be non-negative and satisfy the normalization condition $\sum_{i=1}^{N} W_i = 1$

Example

Robust Low Pass Filter Design

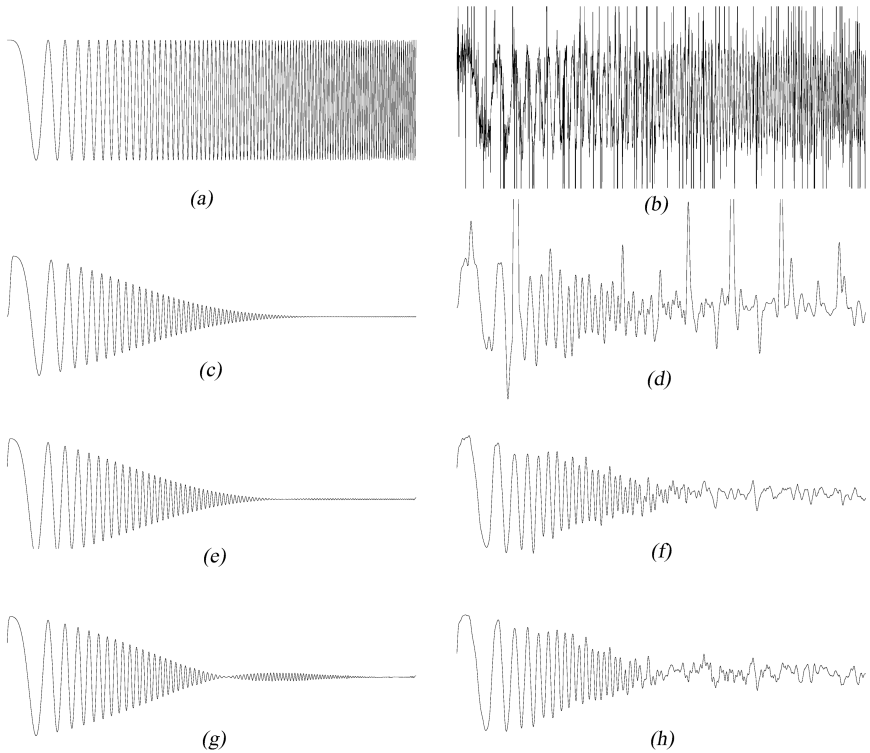


Figure: Myriadizing a linear low-pass smoother in an impulsive environment: (a) chirp signal, (b) chirp in additive impulsive noise, (c) ideal (no noise) myriad smoother output with $K = \infty$, (e) $K = 0.5$, and (g) $K = 0.2$; Myriad smoother output in the presence of noise with (d) $K = \infty$, (f) $K = 0.5$, and (h) $K = 0.2$.

Example: Robust Band-Pass Filter Design

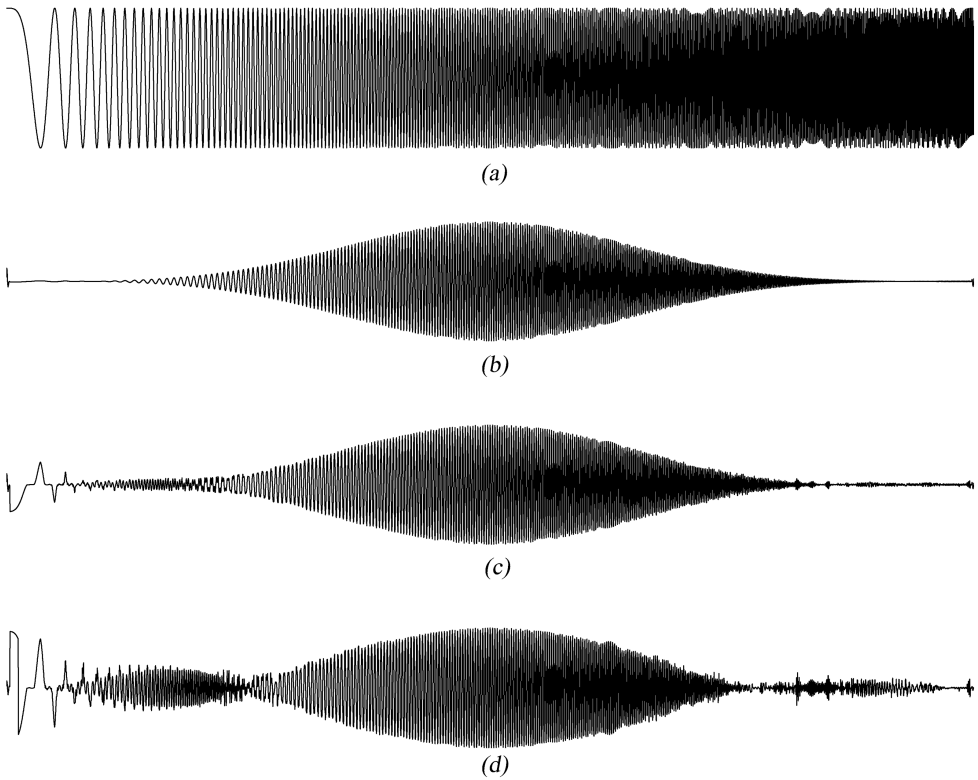


Figure: Myriadizing a linear band-pass filter in an impulsive environment: (a) chirp signal, (b) ideal (no noise) myriad smoother output with $K = \infty$, (c) $K = 0.5$, and (d) $K = 0.2$.

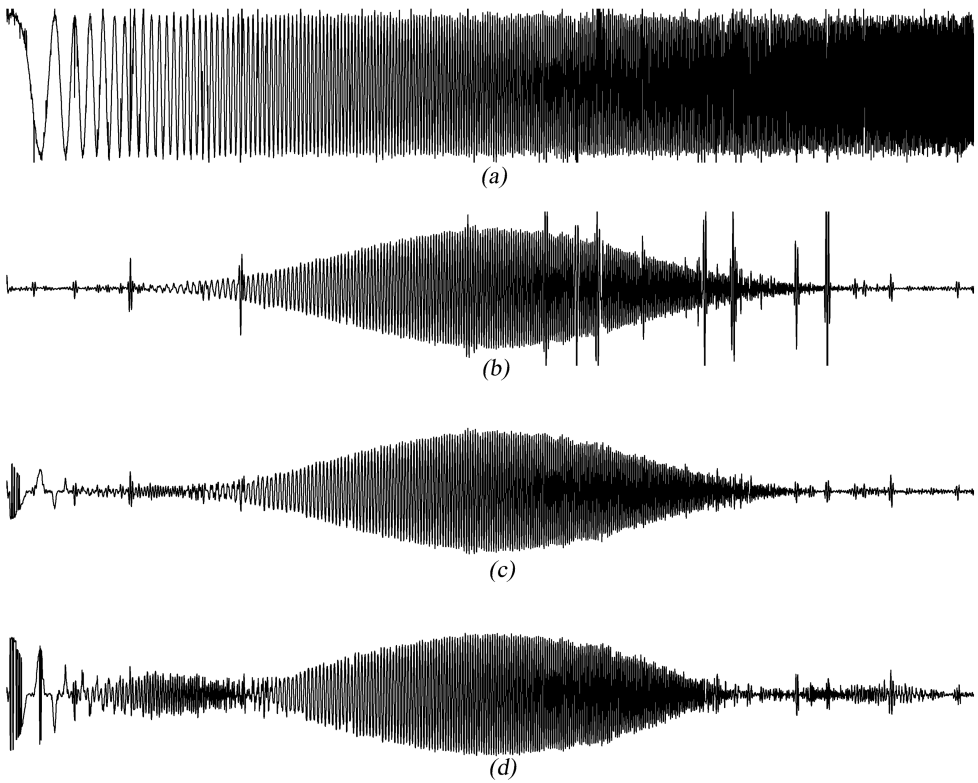


Figure: Myriadizing a linear band-pass filter in an impulsive environment (continued): (a) chirp in additive impulsive noise. Myriad filter output in the presence of noise with (b) $K = \infty$, (c) $K = 0.5$, and (d) $K = 0.2$

Measurements corrupted by noise

Measurements corrupted by noise can be modeled as:

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{z},$$

where $\mathbf{z}$ is a zero-mean additive white noise.

Under some characteristics of the noise:

- Notably finite second order statistics or
- bounded noise in the ℓ_2 sense,
- Having a measurement matrix $\mathbf{\Phi}$ that satisfies the RIP condition.
- Then, there are algorithms to stably recover the sparse signals from the noisy measurements.

Algorithms to recover the sparse signal

- **Basis Pursuit Denoising (BPD)**

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y}-\mathbf{\Phi x}\|_2 \leq \epsilon,$$

for some small $\epsilon > 0$, then $\|\mathbf{x}-\hat{\mathbf{x}}\|_2 \leq C\epsilon$, for some constant C.

- **ℓ_1-regularized least squares (ℓ_1-LS)**

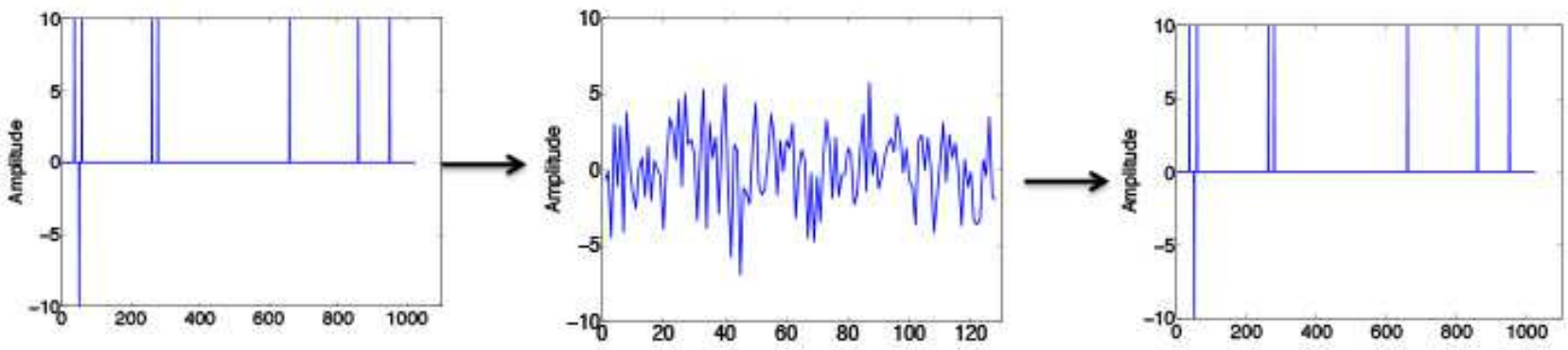
$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{y}-\mathbf{\Phi x}\|_2^2 + \lambda\|\mathbf{x}\|_1,$$

where λ is a regularization parameter that balances the weight between the data fidelity term and the ℓ_1 regularization term.

- **Greedy algorithms** as
  - Orthogonal matching pursuit (OMP) and Regularized OMP
  - Iterative hard thresholding (IHT)

Impulsive Noise Effects

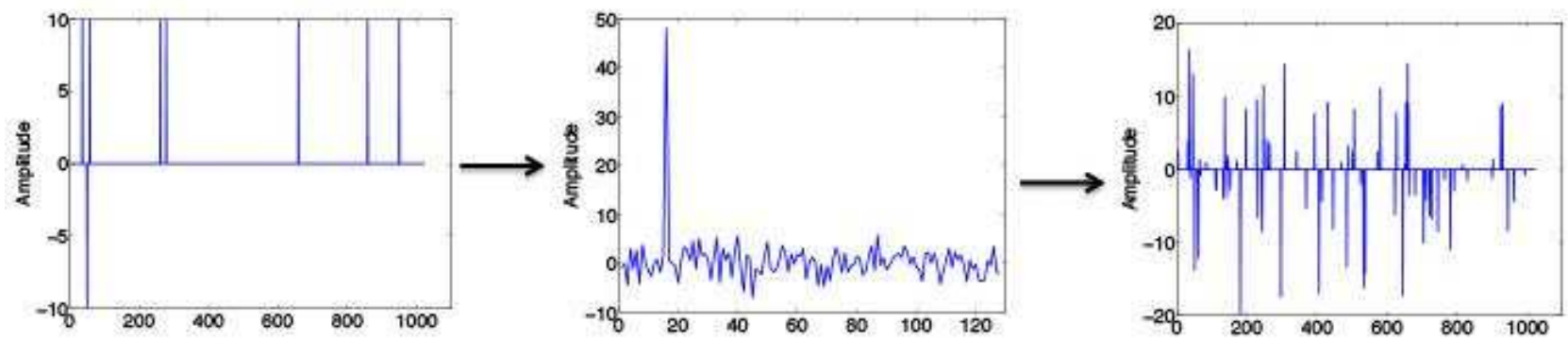
$$\mathbf{x} \quad \rightarrow \quad \mathbf{y} = \mathbf{\Phi x} \quad \rightarrow \quad \hat{\mathbf{x}}$$



$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_0 \text{ subject to } \|\mathbf{y} - \mathbf{\Phi x}\|_2 \leq \epsilon.$$

Impulsive Noise Effects

$$\mathbf{x} \quad \rightarrow \quad \mathbf{y} = \mathbf{\Phi x} + \mathbf{w} \quad \rightarrow \quad \hat{\mathbf{x}}$$



Traditional noise aware CS systems consider finite variance noise models

ℓ_2 norm as *data-fitting*

- Perform adequately under the assumption that the contaminating noise has finite second order statistics.
- Tends to be very sensitive to outliers or gross error present in the measurements

Problem

How to address the CS reconstruction problem when the measurements are corrupted by sparse or impulsive noise?

- Impulsive noise has infinite or very large variance breaking the assumptions of traditional LS-based recovery algorithms.

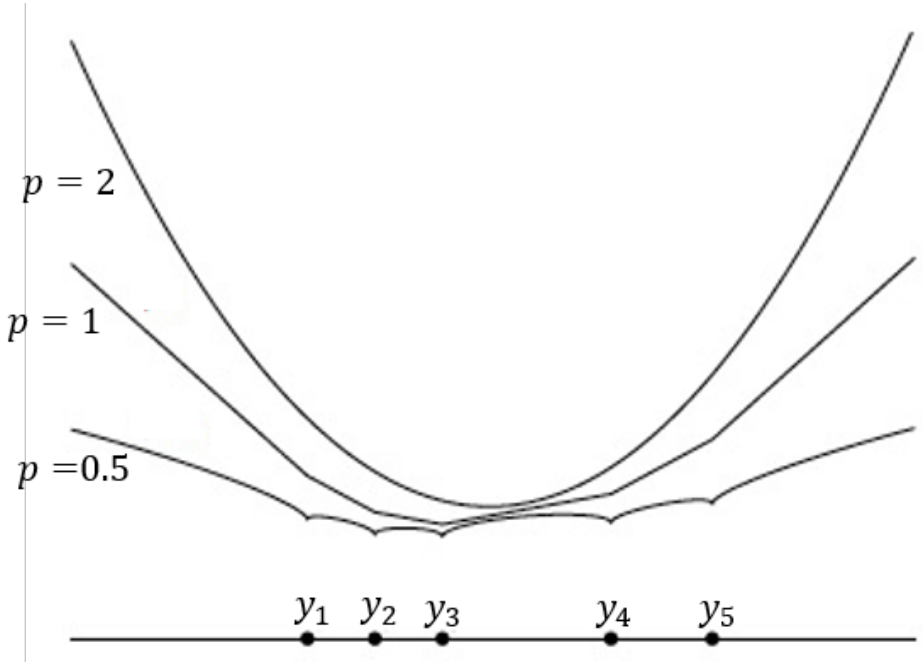
Strategy

Replace the commonly used ℓ_2 norm by M-estimators as data fidelity functions

ℓ_p estimator

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n} \frac{1}{\sigma_i^p} |y_i - \boldsymbol{\alpha}|^p.$$

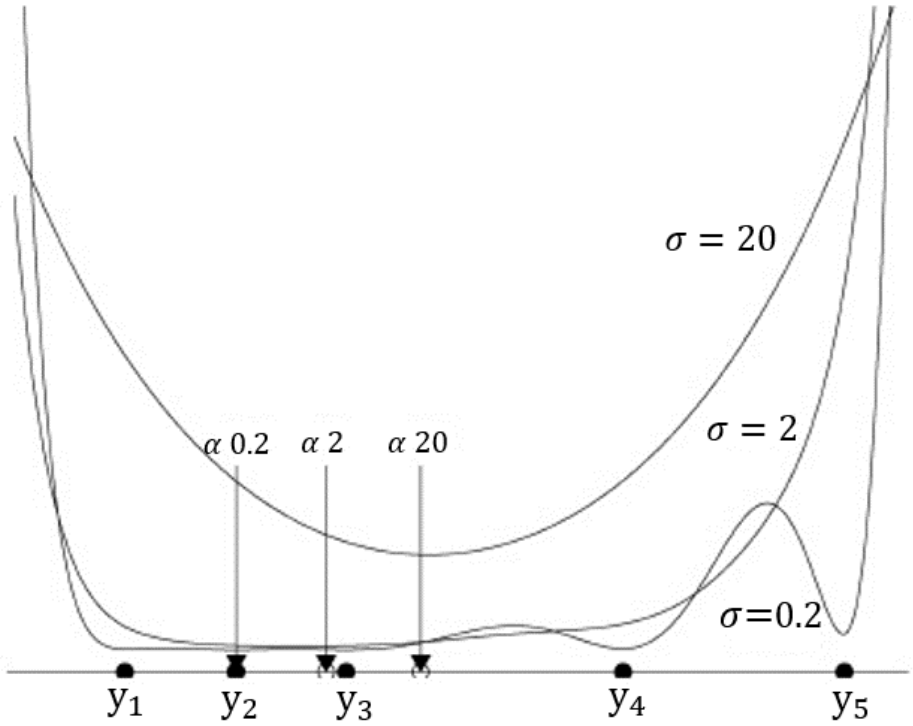
where y_i are the observation samples.



Myriad estimator

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n} \log[\sigma^2 + (y_i - \boldsymbol{\alpha})^2]$$

where y_i are the observation samples.



Lorentzian norm (LL_2 norm)

The cost function of the myriad estimator can be extended to define a robust metric, the **Lorentzian norm** for vectors in $\mathbb{R}^m$

$$\|\mathbf{u}\|_{LL_2,\gamma} = \sum_{i=1}^{m} \log\left(1 + \frac{u_i^2}{\gamma^2}\right), \quad \gamma > 0, u \in \mathbb{R}^m.$$

- Does not penalize large deviations, leading to more robust error metrics when outliers are present.
- The robustness depends on the scale parameter γ

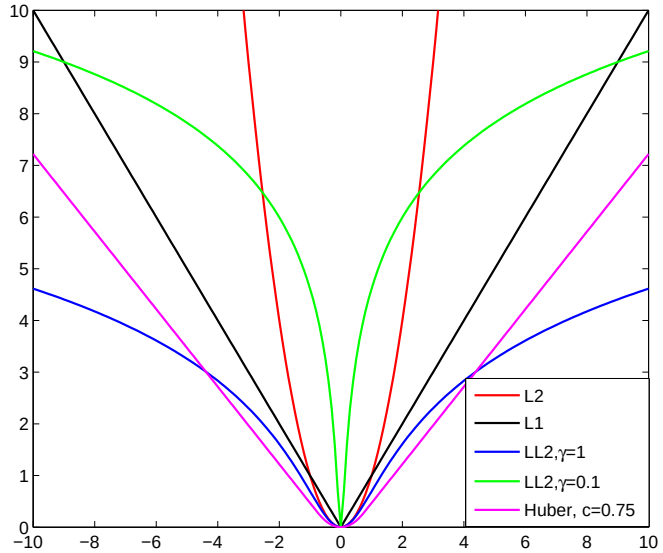
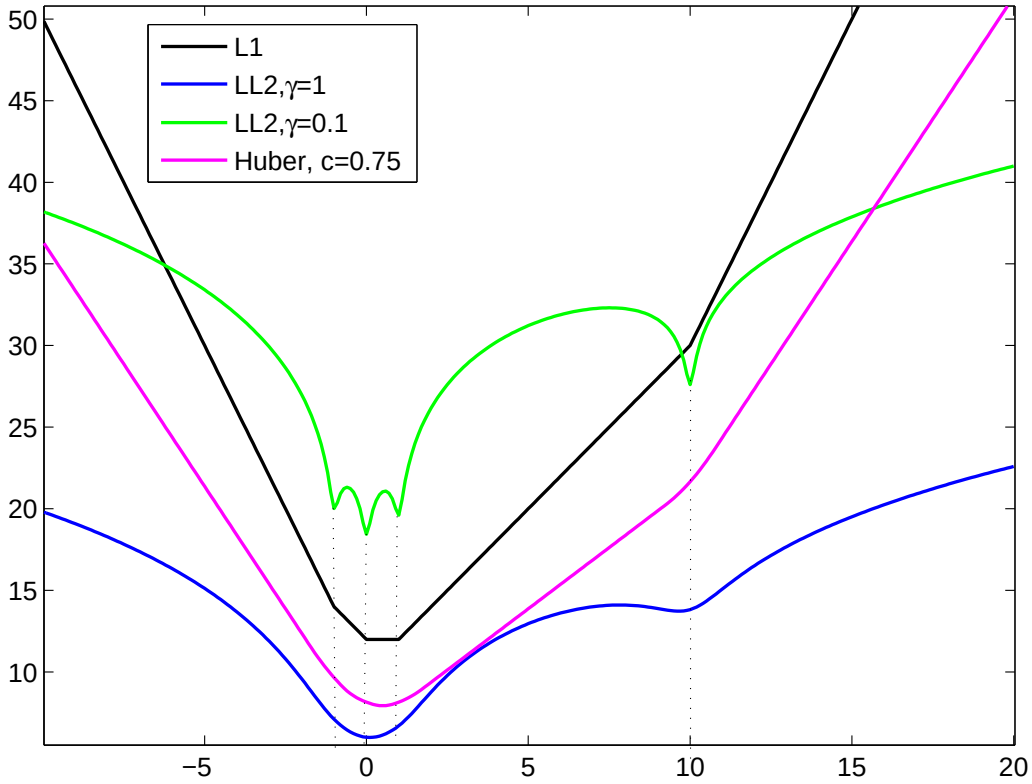


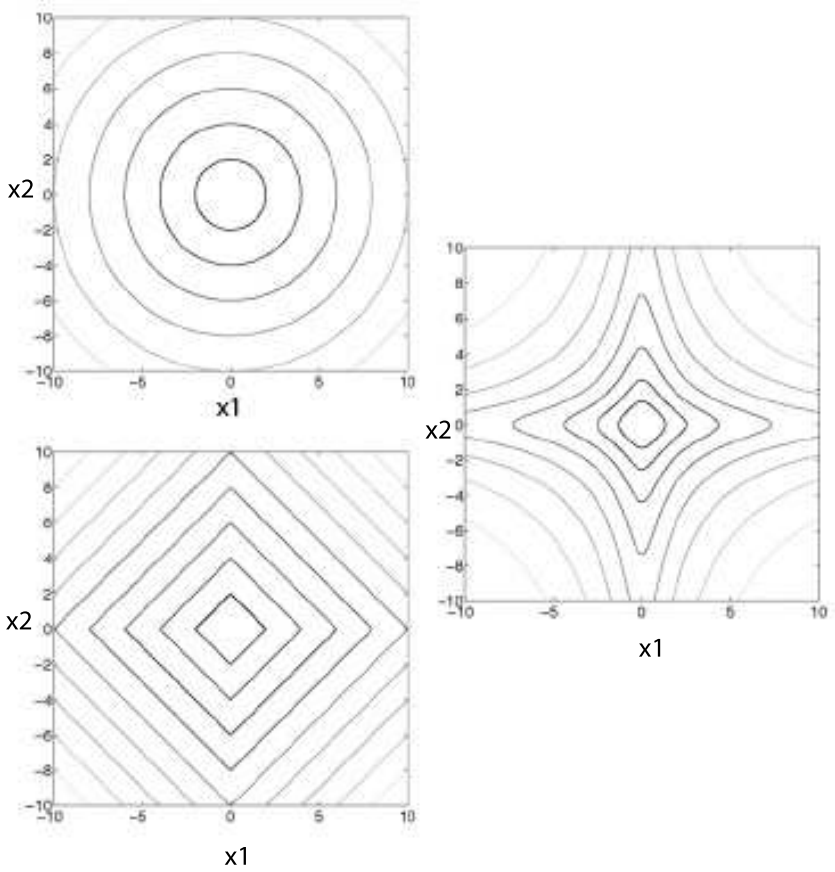
Figure: Comparison of the ℓ_1 **(black)** norm, the Huber cost function with $c = 0.75$ in (magenta) and the LL_2 norm with $\gamma = 1$ (blue) and $\gamma = 0.1$ (green) for the 1D case. The squared ℓ_2 norm (red) is plotted as reference.

Comparison of the ℓ_1 **(black)** norm, the Huber cost function with $c = 0.75$ in (magenta) and the Lorentzian norm with $\gamma = 1$ (blue) and $\gamma = 0.1$ (green) for the location estimation problem, with observation samples located in $x = \{-1, 0, 1, 10\}$



Lorentzian (Myriad) norm (LL_2)

- It is an everywhere continuous function
- It is convex near the origin
- Large deviations are not heavily penalized as in the case of ℓ_1 or ℓ_2



Robust Sparse Signal Reconstruction

Approaches based on replacing the ℓ_2 norm by the robust metrics for the data fidelity term.

1. ℓ_1-Based Methods

The ℓ_2 norm is replaced by the ℓ_1 norm in the data fitting term.

Constrained least absolute deviation (LAD) regression problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}-\mathbf{\Phi x}\|_1 + \tau\|\mathbf{x}\|_0.$$

Optimum under the ML assumption that the noise obeys a Laplacian distribution.

- This problem is combinatorial and computationally expensive.
- Relaxations to the ℓ_0 constraint have been proposed.

2. ℓ_p-Based Methods

If the corrupting noise has heavier tails than the Laplacian distribution, the ℓ_p norm, with $0 < p < 1$ can be used

Recovery optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \mathbf{\Phi x}\|_p^p + \tau\|\mathbf{x}\|_1.$$

- This problem is optimal under the ML criteria for GGD noise.
- It is robust to very impulsive noise.

4. Lorentzian Based Methods

Appropriate for many impulsive environments using the Lorentzian norm as a fitting term.

Lorentzian-Based Basis Pursuit

Signal estimate based on the following non-convex problem:

$$\min_{x \in \mathbb{R}^n} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{\Phi x}\|_{LL_2,\gamma} \leq \rho.$$

Theorem (Upper bound for the reconstruction error)

Let $\mathbf{\Phi}$ be an $m \times n$ sensing matrix such that $\delta_{2s} < \sqrt{2} - 1$. Then for any signal $\mathbf{x_0} \in \mathbb{R}^n$ such that $|supp(\mathbf{x_0})| \leq s$, and observation noise $\mathbf{z}$ with $\|\mathbf{z}\|_{LL_2,\gamma} \leq \rho$, the solution to LBP, $\mathbf{x}^*$, obeys the following bound:

$$\|\mathbf{x_0} - \mathbf{x}^*\|_2 \leq C_s \gamma \sqrt{m(e^{\rho} - 1)},$$

where the constant C_s depends only on δ_{2s}.

Problem: Slow and complex to solve!

Lorentzian-Based Iterative Hard Thresholding Algorithm

Ideal optimization problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}-\mathbf{\Phi x}\|_{LL_2,\gamma} \text{ subject to } \|\mathbf{x}\|_0 \leq s$$

Iterative algorithm:

$$x^{(t+1)} = H_s\left(x^{(t)} + \mu_{(t)}\mathbf{\Phi}^T W_t(\mathbf{\Phi}x^{(t)} - \mathbf{y})\right),$$

where $H_s(\cdot)$ denotes the hard thresholding operator,

$$W_t(i,i) = \frac{\gamma^2}{\gamma^2 + (\mathbf{y}_i - \boldsymbol{\phi}_i^T x^{(t)})^2},$$

and $\boldsymbol{\phi}_i$ denotes the i-th row on $\mathbf{\Phi}$.

Numerical experiments

Comparison of the performance of the following 8 robust methods:

- ℓ_1-based coordinate descent (ℓ_1-CD)
- ℓ_1 least absolute deviation ℓ_1-LAD solved by ADMM
- Lorentzian-based basis pursuit (LBP)
- Lorentzian-based iterative hard thresholding (LITH)
- Lorentzian-based coordinate descent (L-CD)
- Robust lasso (R-Lasso)
- Huber iterative hard thresholding (HIHT)
- ℓ_1-LS method Traditional CS method for comparisons

Experimental Setup 1

- Sparse signals of length $n = 400$, and sparsity $s = 10$.
- Number of random measurements set to $m = 100$. Gaussian sensing matrices
- α-stable noise models with $\alpha = 1$
- 100 repetitions of each experiment averaged

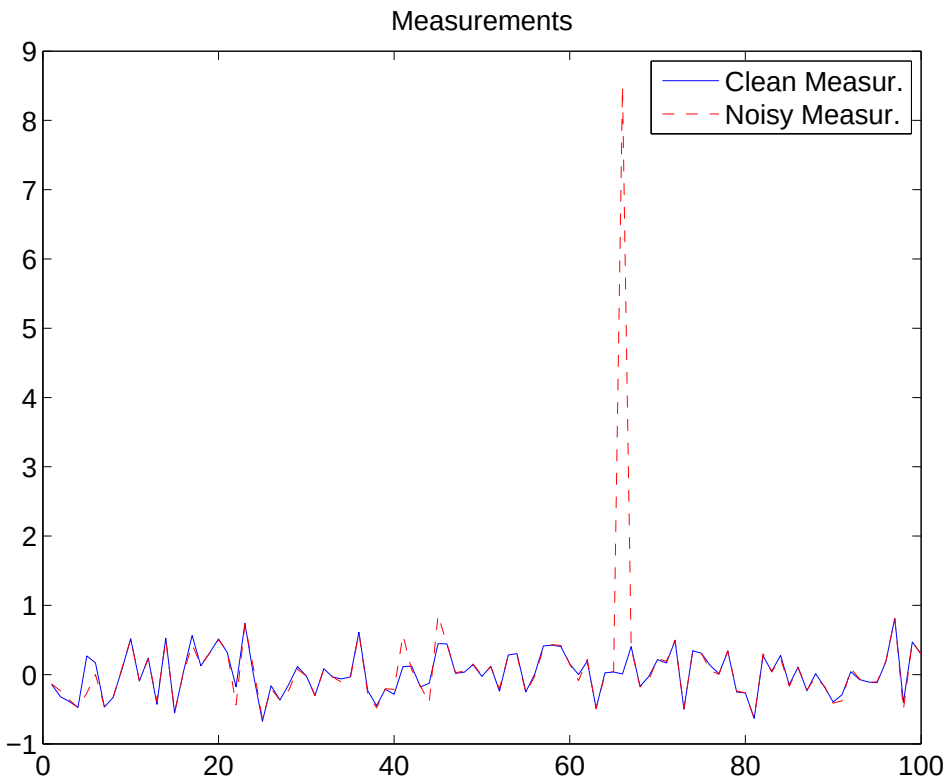
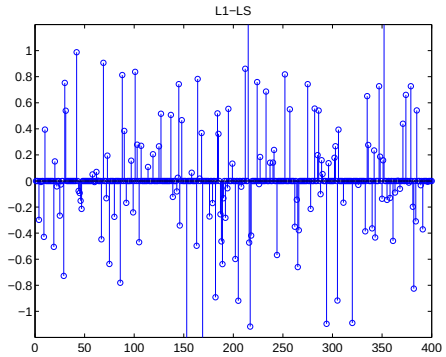
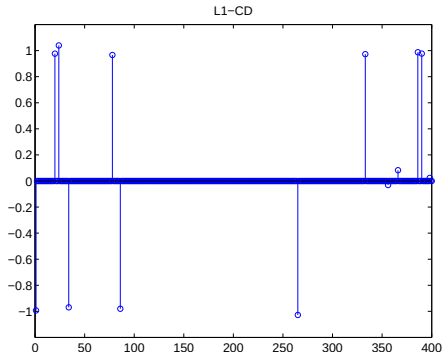


Figure: Clean and contaminated reduced random projections

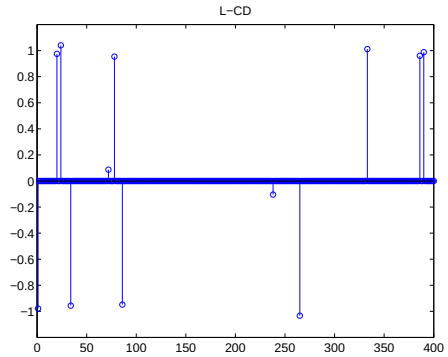
Sparse signal reconstruction



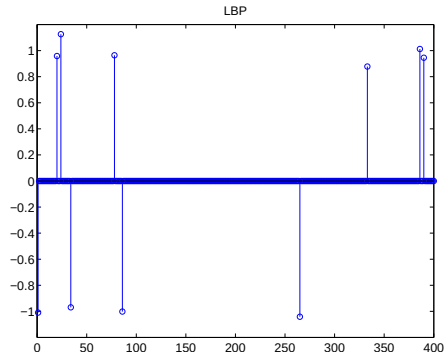
(a) ℓ_1 − *LS* (SER=-6.6 dB)



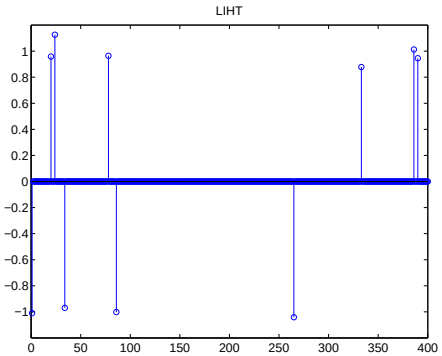
(b) ℓ_1 − *CD* (SER=28.2 dB)



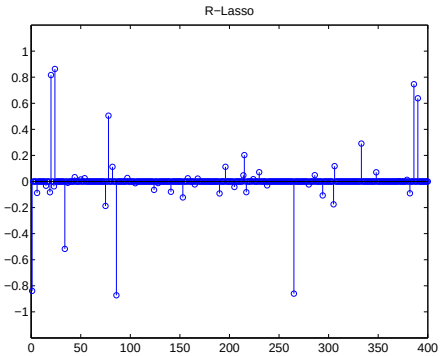
(c) L-CD (SER=25.1 dB)



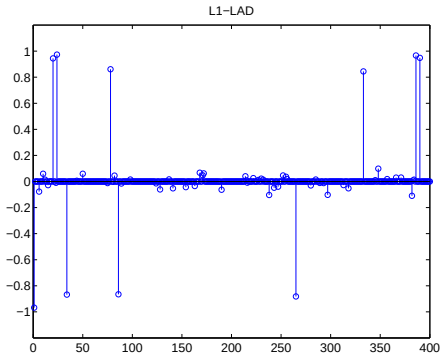
(d) LBP (SER=24.0 dB)



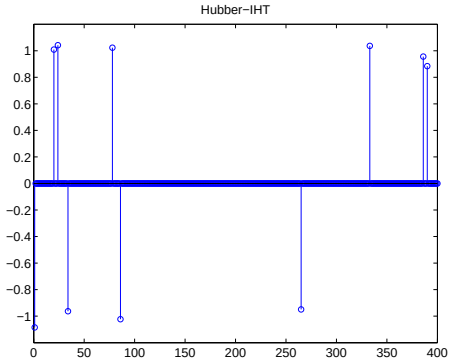
(e) LIHT (SER=24.0 dB)



(f) R-Lasso (SER=8.1 dB)



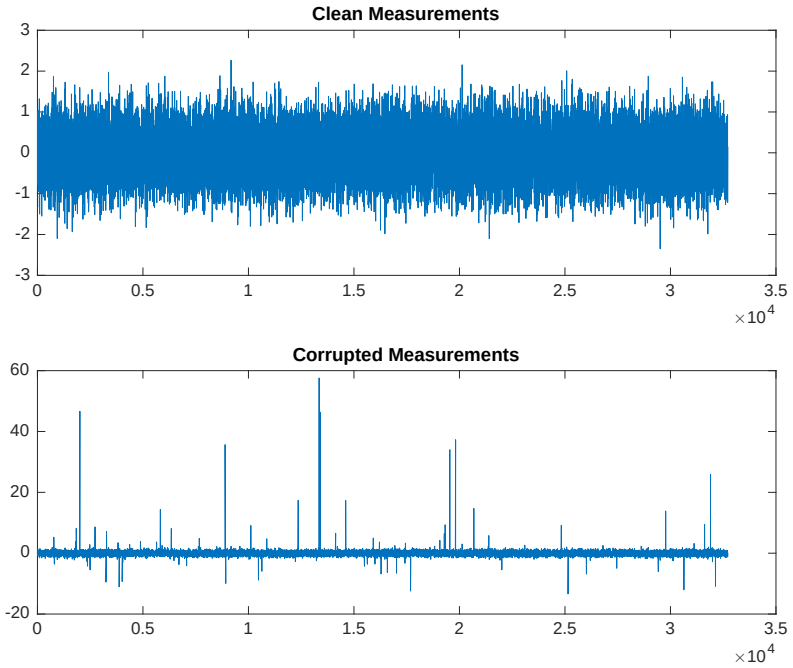
(g) ℓ_1-LAD (SER=16.9 dB)



(h) Huber-IHT (SER=25.1 dB)

Experimental Setup 3

- Reconstruction of the Camera man image of size 256×256
- Image sampled by a random DCT ensambled (50%)
- Number of random measurements $m = 32718$
- Sparsity representation basis: Daubechies *db*4 wavelet
- α-stable noise with $\alpha = 1$
- Scale parameter of the noise $\sigma = 0.01$



(a) ℓ_1 – *LS* (SER=7.3 dB)

(b) ℓ_1 – *LAD* (SER=19.6 dB)

(c) R-Lasso (SER=18.1 dB)

(d) LBP (SER=20.7 dB)

(e) LIHT (SER=19.4 dB)

(f) $\ell_1 - CD$ (SER=20.3 dB)

(g) L-CD (SER=19.2 dB)

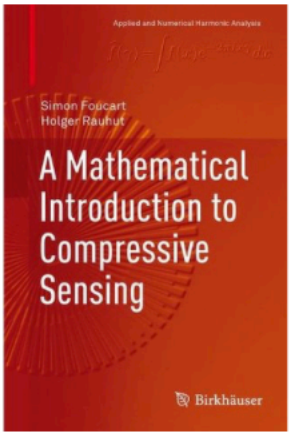
(h) Huber-IHT (SER=19.4 dB)

Robust Sparse Reconstruction Summary

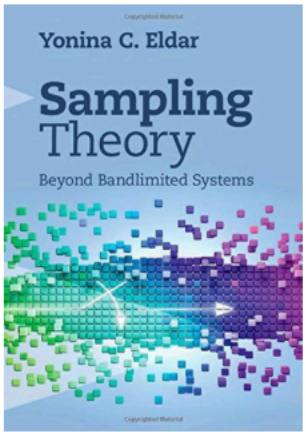
Method	Optimization problem	SER for signal [dB]	SER for image [dB]	Time* [s]
LBP	$\|\mathbf{y}-\mathbf{\Phi x}\|_{LL_2,\gamma}+\lambda\|\mathbf{x}\|_1$	24.0	20.7	10.58
LIHT	$\|\mathbf{y}-\mathbf{\Phi x}\|_{LL_2,\gamma}$ s.t. $\|\mathbf{x}\|_0 \leq s$	24.0	19.4	2.13
R-Lasso	$\frac{1}{2}\|\mathbf{y}-\mathbf{\Phi x}-\mathbf{r}\|_2^2+\tau_x\|\mathbf{x}\|_1+\tau_r\|\mathbf{r}\|_1$	8.1	18.1	7.23
L-CD	$\|\mathbf{y}-\mathbf{\Phi x}\|_{LL_2,\gamma}+\tau\|\mathbf{x}\|_0$	25.1	19.2	6522.7
ℓ_1-CD	$\|\mathbf{y}-\mathbf{\Phi x}\|_1+\tau\|\mathbf{x}\|_0$	28.2	20.3	3814.2
ℓ_1-LS	$\frac{1}{2}\|\mathbf{y}-\mathbf{\Phi x}\|_2^2+\lambda\|\mathbf{x}\|_1$	-6.6	7.3	4.73
ℓ_1-LAD	$\|\mathbf{y}-\mathbf{\Phi x}\|_1+\tau\|\mathbf{x}\|_1$	16.9	19.6	7.05
HIHT	$\sum_{i=1}^{M}\rho\left(\frac{y_i-\Phi_i x}{\sigma}\right)$ s.t. $\|\mathbf{x}\|_0 \leq s$	25.1	19.4	90.78

Table: Summary of eight sparse reconstruction methods.

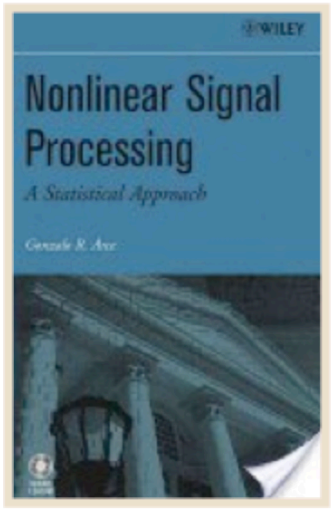
*Execution time required to reconstruct the cameraman image.



A Mathematical Introduction to Compressive Sensing (Applied and Numerical…
› Simon Foucart
Gebundene Ausgabe
EUR 64,99 Prime



Sampling Theory: Beyond Bandlimited Systems
› Yonina C. Eldar
Gebundene Ausgabe
EUR 94,59 Prime

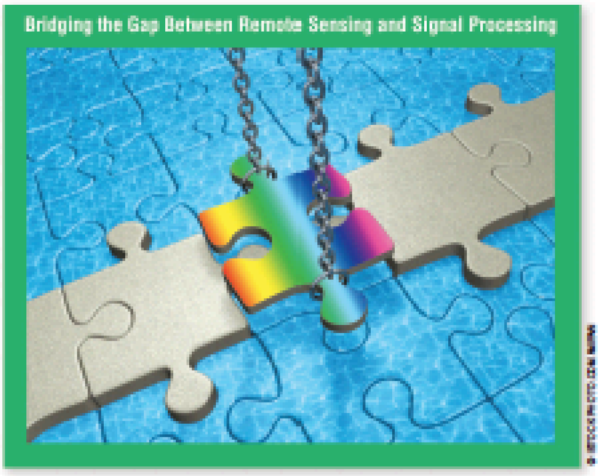


Nonlinear Signal Processing: A Statistical Approach

Gonzalo R. Arce

Gonzalo R. Arce, David J. Brady, Lawrence Carin, Henry Arguello, and David S. Kittle

Compressive Coded Aperture Spectral Imaging



An introduction

Imaging spectroscopy involves the sensing of a large amount of spatial information across a multitude of wavelengths. Conventional approaches to hyperspectral sensing scan adjacent zones of the underlying spectral scene and merge the results to construct a spectral data cube. Push broom spectral imaging sensors, for instance, capture a spectral cube with one focal plane array (FPA) measurement per spatial line of the scene [1], [2]. Spectrometers based on optical bandpass filters sequentially scan the scene by tuning the bandpass filters in steps. The disadvantage of these techniques is that they require scanning a number of zones linearly in proportion to the desired spatial and spectral resolution. This article surveys compressive coded aperture spectral imagers, also known as *coded aperture snapshot spectral imagers (CASSI)* [1], [3], [4], which naturally embody the principles of compressive sensing (CS) [5], [6]. The remarkable advantage of CASSI is that the entire data cube is sensed with just a few FPA measurements and, in some cases, with as little as a single FPA shot.

Digital Object Identifier 10.1109/MSP.2013.2278763
Date of publication: 5 December 2013

INTRODUCTION

CS dictates that one can recover spectral scenes from far fewer measurements than that required by conventional linear scanning spectral sensors. To make this possible, CS relies on two principles: sparsity, which characterizes the spectral scenes of interest, and incoherence, which shapes the sensing structure [5], [7]. Sparsity indicates that spectral images found in nature can be concisely represented in some basis Ψ with just a small