

# Statistically-Efficient Filtering in Impulsive Environments: Weighted Myriad Filters

**Juan G. Gonzalez**

*Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA*  
*Email: juan.gonzalez@intellep.com*

**Gonzalo R. Arce**

*Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA*  
*Email: arce@ee.udel.edu*

*Received 1 August 2001 and in revised form 15 November 2001*

Linear filtering theory has been largely motivated by the characteristics of Gaussian signals. In the same manner, the proposed *Myriad Filtering* methods are motivated by the need for a flexible filter class with high statistical efficiency in non-Gaussian impulsive environments that can appear in practice. Myriad filters have a solid theoretical basis, are inherently more powerful than median filters, and are very general, subsuming traditional linear FIR filters. The foundation of the proposed filtering algorithms lies in the definition of the *myriad* as a tunable estimator of location derived from the theory of robust statistics. We prove several fundamental properties of this estimator and show its optimality in practical impulsive models such as the  $\alpha$ -stable and generalized- $t$ . We then extend the myriad estimation framework to allow the use of weights. In the same way as linear FIR filters become a powerful generalization of the mean filter, filters based on running myriads reach all of their potential when a weighting scheme is utilized. We derive the “normal” equations for the optimal myriad filter, and introduce a suboptimal methodology for filter tuning and design. The strong potential of myriad filtering and estimation in impulsive environments is illustrated with several examples.

**Keywords and phrases:** weighted myriad filters, weighted median filters, impulsive noise, heavy tails, alpha-stable distributions, Cauchy distribution, phase-locked loop.

## 1. INTRODUCTION

A large number of filtering algorithms used in practical applications are limited to the cases of Gaussian noise and/or linear operation, presenting serious performance degradation in the presence of impulsive contamination. The need for a flexible theory of robust nonlinear filtering that can be efficiently applied in real impulsive environments has been repeatedly acknowledged in the signal processing community. Significant research efforts, especially in the field of image processing, have concentrated on finding suitable alternatives to the linear filter that are *robust* or resistant to the presence of impulsive noise. Among these, the approach that has received considerable attention is that of median filters. Today, due to their sound underlying theory, they are being increasingly used in image processing commercial products. An important shortcoming that has hampered their use in other fields is that their output is always constrained, by definition, to one of the samples in the input window. Although this “selection” characteristic is very desirable in image processing

applications [1], it gives efficiency losses that are unacceptable for many other practical applications. It is well known, for example, that the median loses as much as 40% efficiency over the sample mean when used as a location estimator in Gaussian environments [2].

It can be argued that the inefficiency of median-based methods in practical problems arises from their unsuitability to noise processes that can be found in real life. While most noise processes that appear in practice obey density functions that are bell-shaped, the Laplacian model, for which the median is optimal, has a “peaky” density<sup>1</sup> that makes it, in general, rather artificial.<sup>2</sup>

---

<sup>1</sup>This is a consequence of a general result proved in [1], according to which, selection-type estimators of location cannot be optimal for bell-shaped distributions.

<sup>2</sup>It is interesting to note that Laplace introduced his model artificially as a probability distribution for which the median is optimal [3]. Later, and because of its simplicity, the Laplacian model became known as an example of heavy-tailed distribution.

A different approach to robustness that has been very successful in the field of *robust statistics*, (and along which we will develop the work in this paper), is the use of generalized maximum likelihood estimators or M-estimators [2, 4]. Given a set of samples,  $x_1, x_2, \dots, x_N$ , an M-estimator of location is defined as the parameter  $\hat{\beta}$  that minimizes a sum of the form  $\sum_{i=1}^N \rho(x_i - \beta)$ , where  $\rho$  is the *cost function* associated with  $\hat{\beta}$ . The behavior of the M-estimator is completely characterized by the shape of  $\rho$ . When  $\rho(x) = x^2$ , for example, the associated M-estimator is the least squares estimator, which corresponds to the sample mean. For  $\rho(x) = |x|$ , on the other hand, the M-estimator is equivalent to the sample median. In a more general manner, if  $\rho(x) = -\log f(x)$ , where  $f$  is a density function,  $\hat{\beta}$  corresponds to the maximum likelihood estimator associated with  $f$ . The possibility of easily manipulating the cost function  $\rho$  to control the behavior of the estimator, has been the key for the success of M-estimation methods in different areas of statistics and engineering [2, 4, 5].

Several nonlinear filters based on running M-estimators of location have been introduced in the literature [6, 7]. Although these so-called *M-filters* have been proven useful as signal smoothers in impulsive noise environments, they have not yet been well understood nor sufficiently developed to constitute a powerful alternative for the robust filtering problem. We believe that the reason for this apparent failure is not the inadequacy of M-estimation methods but the lack of an approach suited to the specific problems of filtering applications. Despite its great popularity, the theory of robust statistics has been built for problems fundamentally different from those seen in signal processing. For example, the sample size (filter's window size) in signal processing applications is usually much smaller than that of classical statistics problems. Also, real-time operation is a usual requirement that brings a new dimension to the computational problems. While low computational complexity is not a big plus in statistics, it is an almost necessary requirement for real-time signal processing applications.

The nature of the noise environment can also present fundamental dissimilarities. In robust statistics it is often assumed that the underlying "impulsive" distributions obey an " $\epsilon$ -contamination" law with density  $f(x) = (1 - \epsilon)f_0(x) + \epsilon h(x)$ , where  $f_0$ , the *nominal* density, is usually Gaussian,  $\epsilon$  is a small positive constant, and  $h$  is an arbitrary heavy-tailed density which explains the presence of gross errors in the data [4]. Intuitively, one out of  $1/\epsilon$  samples is allowed to be contaminated by uncertain gross errors. While the  $\epsilon$ -contamination model is very suitable for explaining the presence of typos or data-gathering errors (as it is usual in statistics), many decidedly impulsive phenomena in communications and signal processing are characterized by a *clearly different* structure. This is the case for the important class of processes that arise in practice as the superposition of many small independent and *impulsive* effects. According to a generalized form of the central limit theorem, processes of this nature are well modeled by a class of heavy-tailed distributions with infinite variance known as  $\alpha$ -stable [8, 9, 10]. In the same way as the central limit theorem has called for the use of

the Gaussian model in many practical applications, the generalization of the central limit theorem is a strong theoretical result compelling the use of  $\alpha$ -stable models in real-world problems involving impulsive signals.

In this paper, we introduce a novel class of M-filters motivated by the need for a flexible filtering framework with high statistical efficiency in distribution families which, like the  $\alpha$ -stable, can appear in engineering practice. The foundation of these *Myriad Filters*, as we propose to call them, lies in the definition of the *sample myriad* as an M-estimator derived from tunable cost functions of the form

$$\rho(x) = \log[k^2 + x^2], \quad (1)$$

where  $k$  is the tunable parameter. A rationale for the choice of such cost functions, along with a quick introduction to  $\alpha$ -stable distributions is presented in Section 2.

It is important to note that the use of the above cost functions in robust estimation problems is not new. As we will see later, they can be easily generated from maximum likelihood considerations on well-known heavy-tailed distributions such as the Cauchy and the Student's  $t$ . Since the introduction of the Cauchy maximum likelihood estimator by Fisher in 1922 [11], estimators related to (1) have been used repeatedly and under very different contexts as efficient alternatives to cope with the presence of impulsive noise [12, 13, 14, 15, 16, 17]. However, none of the previously proposed techniques has exploited the potential of tuning  $k$  in order to control the behavior of the estimator. The availability of  $k$  as a tunable parameter is in fact the most powerful feature of the myriad. We show in Section 3 that, depending on  $k$ , the sample myriad can present drastically different behaviors, ranging from *highly resistant* mode-type estimators to the familiar (Gaussian-efficient) sample average. This rich variety of operation modes is the key concept explaining important optimality properties of the myriad in the class of symmetric  $\alpha$ -stable distributions. We discuss and demonstrate these properties in Section 4.

In addition to the differences in noise structure that motivate the introduction of the cost functions in (1), the use of filter weights strive for another fundamental difference between filtering and location estimation problems. To visualize the strong potential that filter weights can add to a filtering framework, it suffices to see the enormous gap in versatility and performance separating the mean filter from its weighted version, the class of linear FIR filters. In Section 5 we introduce the *weighted M-filter*, an intuitively appealing weighted estimator derived from the concepts of M-filtering and maximum likelihood. When the cost function in (1) is used, the *weighted M-filter* gives rise to a powerful structure that inherits the flexibility and statistical efficiency of the sample myriad, namely the *weighted myriad*. The introduction of weights opens the optimal filter problem, that is, the problem of finding the best set of weights for accomplishing a given task. This problem, similar in flavor to the Wiener filter problem of conventional linear theory, is addressed in Section 6, where several possible paths to myriad filter design are introduced and illustrated. We begin our discussion with a glimpse

at the models that motivate the choice of the cost functions in (1).

## 2. THE SAMPLE MYRIAD

### 2.1. A glimpse at $\alpha$ -stable distributions

A wide variety of signals found in engineering practice arise as the superposition of many small and independent effects. Typical examples include atmospheric radio noise, underwater acoustic noise, and multi-user interference in mobile communication systems. The following generalized form of the central limit theorem explains the goodness of the  $\alpha$ -stable model for signals of this type. Informally, *if a random variable  $X$  can be the limit of a normalized sum of (possibly shift corrected) independent and identically distributed (i.i.d.) random variables, then  $X$  is  $\alpha$ -stable.*

The limit here is taken as the number of elements in the sum tends to  $\infty$ . More formal statements of this theorem can be found in most advanced probability theory textbooks, such as [8].

Instead of a theorem, many authors consider the above as the most fundamental definition of an  $\alpha$ -stable random variable. We can note that, thanks to the central limit theorem, the Gaussian distribution is an obvious member of the  $\alpha$ -stable family. The complete class of symmetric  $\alpha$ -stable distributions was first analytically characterized by Lévy in 1925 [18]. He found that *symmetric*  $\alpha$ -stable random variables follow a characteristic function of the form

$$\phi(\omega) = e^{-\gamma|\omega|^\alpha}. \quad (2)$$

The parameter  $\gamma$ , usually called the *dispersion*, is a positive constant related to the scale of the distribution (note that  $\gamma^{1/\alpha}$  is a scale parameter of the distribution). The parameter  $\alpha$  is usually called the *characteristic exponent* or *index*. In order for (2) to define a characteristic function, the values of  $\alpha$  must be restricted to the interval  $(0, 2]$ . Conceptually speaking,  $\alpha$  determines the impulsiveness or tail heaviness of the distribution (smaller values of  $\alpha$  indicate increased levels of impulsiveness). The limit case,  $\alpha = 2$ , corresponds to the zero-mean Gaussian distribution with variance  $2\gamma$ . All other values of  $\alpha$  correspond to heavy-tailed distributions with infinite variance and algebraic tail behavior of the form<sup>3</sup>

$$\Pr(|X| > x) \sim cx^{-\alpha}, \quad \text{as } x \rightarrow \infty, \quad (3)$$

for some positive constant  $c$ . The case  $\alpha = 1$  corresponds to the “zero-centered” Cauchy distribution, which has density

$$f(x) = \frac{\gamma}{\pi} \frac{1}{\gamma^2 + x^2}. \quad (4)$$

<sup>3</sup>The symbol  $\sim$  denotes asymptotic similarity. Formally,  $X$  has algebraic tails if there exist positive constants  $\alpha$  and  $c$  such that  $\lim_{x \rightarrow \infty} x^\alpha \Pr(|X| > x) = c$ . In the terminology of heavy-tailed processes,  $\alpha$  is usually called the *tail index* or *tail constant* of the process.

When  $\alpha \neq 1, 2$ , no closed expressions exist for the density functions, making it necessary to resort to series expansions or integral transforms to describe them [9].

Symmetric  $\alpha$ -stable densities maintain many of the features of the Gaussian density. They are smooth, unimodal, symmetric with respect to the mode, and bell-shaped. Figure 1 illustrates the impulsive behavior of symmetric  $\alpha$ -stable processes as the characteristic exponent  $\alpha$  is varied. Each one of the plots shows an independent and identically distributed (i.i.d.) “zero-centered” symmetric  $\alpha$ -stable signal with unitary geometric power.<sup>4</sup> In order to give a better feeling of the impulsive structure of the data, the signals are plotted twice under two different scales. As it can be appreciated, the Gaussian signal ( $\alpha = 2$ ) does not show impulsive behavior. For values of  $\alpha$  close to 2 ( $\alpha = 1.7$  in the figure), the structure of the signal is still similar to the Gaussian, although some impulsiveness can now be observed. As the value of  $\alpha$  is decreased, the impulsive behavior increases progressively.

Alpha-stable models have been successful in a variety of technical fields including hydrology, economics, physics, and engineering. Recently, they have been the subject of increased attention in communications and signal processing [10, 19, 20, 21, 22].

Although several filtering solutions in the presence of  $\alpha$ -stable noise have been proposed, most of these have been concentrated on bringing impulse-resistance to the filter design process, while keeping the filter structure linear [10, 23]. This approach is not useful in many practical applications, where impulse-resistance of the *filter operation* is the principal objective—an objective intrinsically impossible to meet by means of linear structures. In the following we introduce a nonlinear filtering and estimation structure motivated by the properties of  $\alpha$ -stable distributions.

### 2.2. The sample myriad: nonlinear estimation and filtering in the $\alpha$ -stable framework

In developing efficient estimation algorithms for  $\alpha$ -stable distributions, there is usually a trade-off between the optimality of the methods and the computational complexity associated with the lack of closed-form expressions for the density functions. In order to develop efficient and computationally tractable algorithms that are suited to real-time applications, we will introduce the *sample myriad* as a class of location M-estimators derived from optimality conditions in the Cauchy distribution. There are two reasons to do this. First, the Cauchy density function is the *only* heavy-tailed symmetric  $\alpha$ -stable density for which a closed-form expression is available. Second, and more important, we will prove that the resulting class of estimators presents significant optimality properties in the  $\alpha$ -stable family.

Given a symmetric density function  $f(x - \beta)$  with symmetry center  $\beta$ , the most desirable optimality properties of an estimator of  $\beta$  are perhaps encountered in the maximum likelihood (ML) estimator. It is known that, under mild regu-

<sup>4</sup>The geometric power is an indicator of signal strength suited to the class of processes with infinite variance [19].

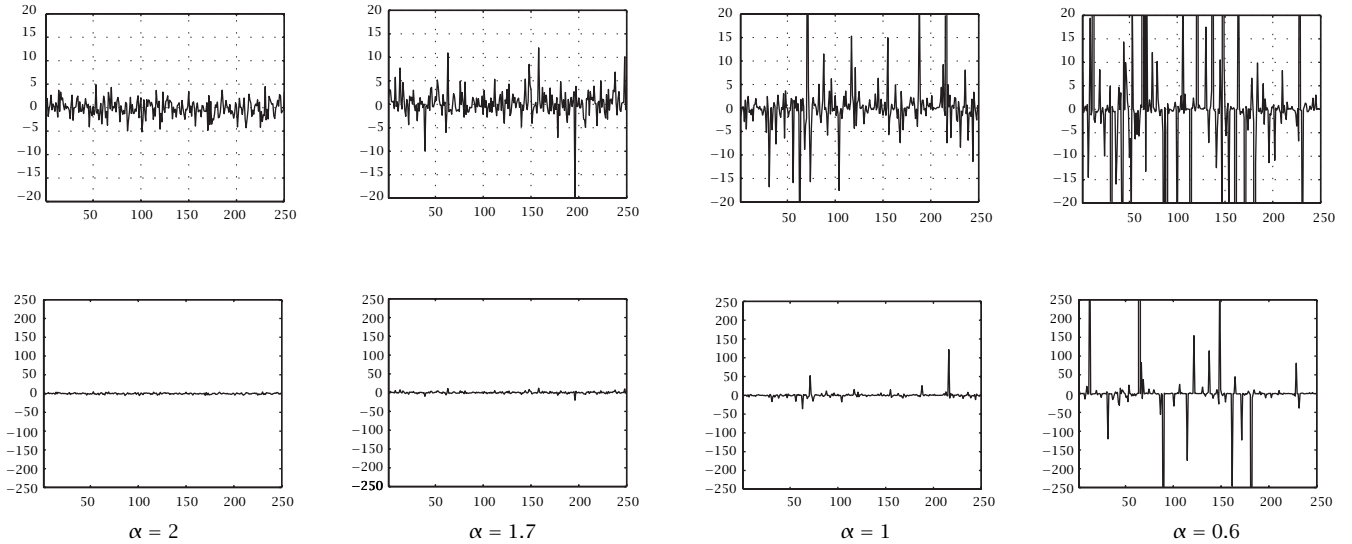


FIGURE 1: Impulsive behavior of i.i.d.  $\alpha$ -stable signals as the tail constant  $\alpha$  is varied. Signals are plotted twice under two different scales.

larity conditions on  $f$ , the ML estimator presents minimum asymptotic variance. In the framework of M-estimation of location, a density function  $f$  generates its maximum likelihood estimator through the cost function  $\rho(x) = -\log f(x)$ . Given the Cauchy density function with dispersion  $k$ ,

$$f(x) = \frac{k}{\pi} \frac{1}{k^2 + x^2}, \quad (5)$$

its associated cost function has thus a logarithmic structure of the form

$$\rho(x) = \log[k^2 + x^2]. \quad (6)$$

For reasons that will become apparent later, we will refer to the M-estimator associated with this cost function as the *sample myriad*. Its formal definition is stated next.<sup>5</sup>

**Definition 1** (sample myriad). Given a set of samples  $x_1, x_2, \dots, x_N$ , and a real parameter  $k > 0$ , the sample myriad of order  $k$  is defined as<sup>6</sup>

$$\begin{aligned} \hat{\beta}_k &= \text{myriad}\{k; x_1, \dots, x_N\} \\ &= \arg \min_{\beta} \sum_{i=1}^N \log[k^2 + (x_i - \beta)^2]. \end{aligned} \quad (7)$$

<sup>5</sup>As it is usual for filters derived from location estimators, we will call “myriad filter” any running window filter outputting the sample myriad of the elements in the window.

<sup>6</sup>For any real function  $G$ , we denote  $\arg \min_{\beta} G(\beta)$  as the value of  $\beta$  for which  $G(\beta)$  is minimum. It is worth noting that, according to this definition, it is possible to find sample sets for which the myriad is not unique. To protect Definition 1 from lack of formalism, we will accept *any* answer to equation (7) as a valid calculation of the myriad. The “degenerate” event of getting more than one myriad is not of critical importance, as its associated probability is either negligible or zero for most cases of interest.

In some situations, the following equivalent expression can be more computationally convenient:

$$\hat{\beta}_k = \arg \min_{\beta} \prod_{i=1}^N [k^2 + (x_i - \beta)^2]. \quad (8)$$

Note that, unlike the sample mean or median, the definition of the sample myriad involves the free-tunable parameter  $k$ . This parameter will play a fundamental role in the theory of myriad filters, and it is the topic of study in the following section. For reasons that will become apparent there, we will refer to  $k$  as the *linearity parameter* of the myriad.

It is important to emphasize that Definition 1 does not make any assumption about the distribution of the samples or the value of the linearity parameter. As we will show in the following two sections, the possibility of freely tuning  $k$  will provide the sample myriad with important optimality properties over a wide variety of non-Cauchy models, including the Gaussian.

Intuitively, since the sample myriad is optimal for such a heavy-tailed distribution like the Cauchy, we should expect a high resistance of the estimator to the presence of impulsive noise. This idea of deriving resistant methods from the Cauchy distribution is not new. It has been proven many times under different circumstances that methods derived from the Cauchy distribution inherit, in general, improved performance in the presence of heavy-tailed noise. Successful examples include applications in communications [12, 22], image processing [16], and geology [15, 24]. Different forms of the sample myriad have also been studied in the past in the context of robust estimation and filtering [11, 14, 15, 17, 25]. From these, it is worth highlighting [17], where the potential of the myriad filter as a signal smoother in  $\alpha$ -stable noise was early recognized. We want to emphasize that, even though the Cauchy distribution and the Myriad itself have been studied

and applied extensively, no previous work has exploited the potential of tuning the parameter  $k$ , nor has shown the optimal behavior of the myriad in the  $\alpha$ -stable family as  $k$  is varied.

### 3. THE LINEARITY PARAMETER $k$ : UNVEILING THE POWER OF THE MYRIAD

#### 3.1. The linear and mode properties

The behavior of the myriad estimator introduced in Definition 1 is markedly dependent on the value of its linearity parameter  $k$ . The following property explains the behavior of the myriad as  $k$  is made large.

*Property 1* (linear property). Given a set of samples,  $x_1, x_2, \dots, x_N$ , the sample myriad  $\hat{\beta}_k$  converges to the sample average as  $k \rightarrow \infty$ . This is,

$$\lim_{k \rightarrow \infty} \hat{\beta}_k = \lim_{k \rightarrow \infty} \text{myriad}\{k; x_1, \dots, x_N\} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (9)$$

*Proof.* First note that<sup>7</sup>  $\hat{\beta}_k \leq x_{(N)}$  by checking that for any  $i$ , and for  $\beta > x_{(N)}$ ,  $k^2 + (x_i - \beta)^2 > k^2 + (x_i - x_{(N)})^2$ . In the same way,  $\hat{\beta}_k \geq x_{(1)}$ . Hence,

$$\begin{aligned} \hat{\beta}_k &= \arg \min_{x_{(1)} \leq \beta \leq x_{(N)}} \prod_{i=1}^N [k^2 + (x_i - \beta)^2] \quad (10) \\ &= \arg \min_{x_{(1)} \leq \beta \leq x_{(N)}} \\ &\quad \times \left\{ k^{2N} + k^{2N-2} \sum_{i=1}^N (x_i - \beta)^2 + O(k^{2N-4}) \right\}, \quad (11) \end{aligned}$$

where  $O$  denotes the usual *asymptotic order* as  $k \rightarrow \infty$ . Since adding or multiplying by constants does not affect the arg min operator, we can transform (11) in

$$\hat{\beta}_k = \arg \min_{x_{(1)} \leq \beta \leq x_{(N)}} \left\{ \sum_{i=1}^N (x_i - \beta)^2 + \frac{O(k^{2N-4})}{k^{2N-2}} \right\}. \quad (12)$$

Letting  $k \rightarrow \infty$ , the term  $O(k^{2N-4})/k^{2N-2}$  becomes negligible, and

$$\hat{\beta}_k \rightarrow \arg \min_{x_{(1)} \leq \beta \leq x_{(N)}} \left\{ \sum_{i=1}^N (x_i - \beta)^2 \right\} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (13) \quad \square$$

Plainly, an infinite value of  $k$  converts the myriad into the sample average. This behavior explains our choice of *linearity* for the name of the parameter: the larger the value of  $k$ , the closer the behavior of the myriad to a linear estimator. As the myriad moves away from the linear region (large values of  $k$ ) to lower linearity values, the estimator becomes more resistant to the presence of impulsive noise. In

the limit, when  $k$  tends to zero, the analysis of the myriad leads to the discovery of a novel location estimator with particularly good performance in the presence of *very* impulsive noise. As we will see below, this estimator treats every observation as a possible outlier, assigning more credibility to the most repeated values in the sample. This “mode-type” characteristic has led us to name this estimator the *mode-myriad*.

*Definition 2* (sample mode-myriad). Given a set of samples  $x_1, x_2, \dots, x_N$ , the mode-myriad estimator,  $\hat{\beta}_0$ , is defined as

$$\hat{\beta}_0 = \lim_{k \rightarrow 0} \hat{\beta}_k, \quad (14)$$

where  $\hat{\beta}_k = \text{myriad}\{k; x_1, x_2, \dots, x_N\}$ .

The following property explains the behavior of the mode-myriad as a kind of generalized sample mode, and provides a simple method for determining the mode-myriad without recurring to Definition 2.

*Property 2* (mode property). The mode-myriad  $\hat{\beta}_0$  is always equal to one of the most repeated values in the sample. Furthermore,

$$\hat{\beta}_0 = \arg \min_{x_j \in \mathcal{M}} \prod_{i=1, x_i \neq x_j}^N |x_i - x_j|, \quad (15)$$

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

*Proof.* Since  $k$  is a positive constant, the definition of the sample myriad in (8) can be reformulated as  $\hat{\beta}_k = \arg \min_{\beta} H_k(\beta)$ , where

$$H_k(\beta) = \prod_{i=1}^N \left[ 1 + \frac{(x_i - \beta)^2}{k^2} \right]. \quad (16)$$

When  $k$  is very small, it is easy to check that

$$H_k(\beta) = O\left(\frac{1}{k^2}\right)^{N-r(\beta)}, \quad (17)$$

where  $r(\beta)$  is the number of times the value  $\beta$  is repeated in the sample, and  $O$  denotes the asymptotic order as  $k \rightarrow 0$ . In the limit, the exponent  $N - r(\beta)$  must be minimized in order for  $H_k(\beta)$  to be minimum. Therefore, the mode-myriad  $\hat{\beta}_0$  will lie on a maximum of  $r(\beta)$ , or in other words,  $\hat{\beta}_0$  will be one of the most repeated values in the sample. Now, let  $r = \max_j r(x_j)$ . Then, for  $x_j \in \mathcal{M}$ , expanding the product in (16) gives

$$H_k(x_j) = \left\{ \prod_{i, x_i \neq x_j} \frac{(x_i - x_j)^2}{k^2} \right\} + O\left(\frac{1}{k^2}\right)^{N-r-1}. \quad (18)$$

Since the first term in (18) is  $O(1/k^2)^{N-r}$ , the second term is negligible for small values of  $k$ , and  $\hat{\beta}_0$  can be calculated as

<sup>7</sup>Here,  $x_{(i)}$  denotes the  $i$ th-order statistic of the sample.

$$\begin{aligned}
 \hat{\beta}_0 &= \arg \min_{x_j \in \mathcal{M}} H_k(x_j) \\
 &= \arg \min_{x_j \in \mathcal{M}} \prod_{i, x_i \neq x_j} \frac{(x_i - x_j)^2}{k^2} \\
 &= \arg \min_{x_j \in \mathcal{M}} \prod_{i, x_i \neq x_j} |x_i - x_j|. \quad \square
 \end{aligned} \tag{19}$$

An immediate consequence of the mode property is the fact that running-window filters based on the mode-myriad are *selection-type*, in the sense that their output is always, by definition, one of the samples in the input window. This “selection” property, shared also by the median, makes mode-myriad filtering a suitable framework for image processing, where the application of selection-type filters has been shown convenient [1, 26, 27].

### 3.2. Understanding $k$ : how much is large and how much is small?

The linear and mode properties indicate the behavior of the myriad estimator for large and small values of  $k$ . From a practical point of view, it is important to determine if a given value of  $k$  is large (or small) enough for the linear (or mode) property to hold approximately. With this in mind, it is instructive to look at the myriad as the maximum likelihood location estimator generated by a Cauchy distribution with dispersion  $k$  (geometrically,  $k$  is equivalent to half the interquartile range). Given a set of samples, the ML method locates the generating distribution in a position where the probability of the specific sample set to occur is maximum. When  $k$  is large, the generating distribution is highly dispersed, and its density function looks flat (see the density function corresponding to  $k_2$  in Figure 2). If  $k$  is large enough, *all* the samples can be accommodated inside the interquartile range of the distribution, and the ML estimator visualizes them as “well-behaved” (no outliers). In this case, a desirable estimator would be the sample average, in complete agreement with the linear property. From this consideration, it should be clear that a fair approximation to the linear property can be obtained if  $k$  is large enough so that *all* the samples can be seen as “well-behaved” under the generating Cauchy distribution.<sup>8</sup> From our experience, we have found that values of  $k$  on the order of the data range,  $k \sim x_{(N)} - x_{(1)}$ , make often the myriad an acceptable approximation to the sample average.

On the other side, when  $k$  is small, the generating Cauchy distribution is highly localized, and its density function looks similar to a positive impulse. The effect of such a localized distribution is conceptually equivalent to observing the samples through a magnifying lens. In this case, most of the data look like possible outliers, and the ML estimator has trouble locating a large number of observations inside the interquartile range of the density (see the density function corresponding to  $k_1$  in Figure 2). Putting in doubt most of the data at hand, a desirable estimator would tend to maximize the

<sup>8</sup>Of course, the degree to which a sample can or cannot be considered “well-behaved” is as arbitrary as the concept of “fair approximation,” and the conceptual sharpening of this criterion is left to the personal judgment of each user.

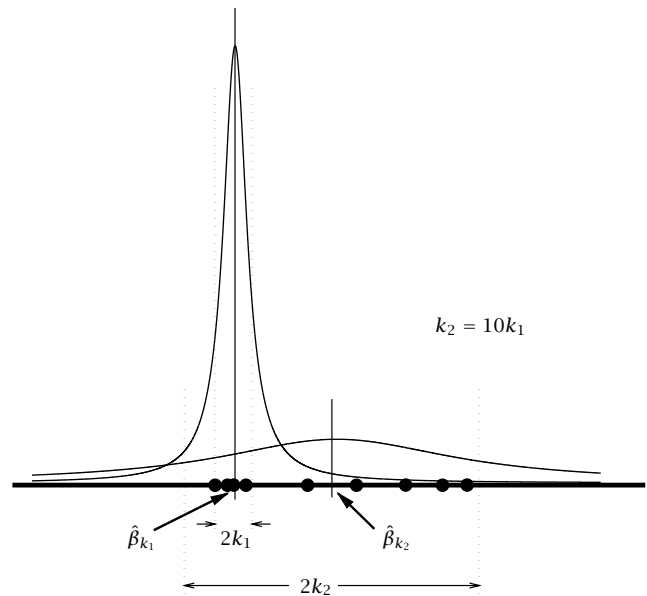


FIGURE 2: The role of the linearity parameter when the myriad is looked as a maximum likelihood estimator. When  $k$  is large, the generating density function is spread and the data are visualized as “well-behaved” (the optimal estimator is the sample average). For small values of  $k$ , the generating density becomes highly localized, and the data are visualized as very impulsive (the optimal estimator is a cluster locator).

number of samples inside the interquartile range, inducing to position the density function in the vicinity of a data cluster. In the limit case, when  $k \rightarrow 0$ , the density function gets infinitely localized, and the only visible clusters will be made of repeated value sets. In this case, one of the most crowded clusters (i.e., one of the most repeated values in the sample) will be located by the estimator, in accordance with the mode property. From this consideration, it should be clear that a fair approximation to the mode property can be obtained if  $k$  is made significantly smaller than the distances between sample elements. From our experience, we have found that  $k$  on the order of

$$k \sim \min_{i \neq j} |x_i - x_j|, \tag{20}$$

is often enough for the myriad to be considered approximately a mode-myriad.

### 3.3. Tuning $k$ : controlling the myriad behavior

The myriad estimator offers a rich class of modes of operation that can be easily controlled by tuning the linearity parameter  $k$ . When the noise is Gaussian, for example, large values of the linearity can provide the optimal performance associated with the sample mean, whereas for highly impulsive noise statistics, the resistance of mode-type estimators can be achieved by using myriads with low linearity. The trade-off between efficiency at the Gaussian model and resistance to impulsive noise can be managed by designing appropriate values for  $k$  (see Figure 3).

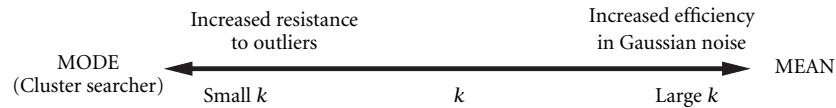


FIGURE 3: Functionality of the myriad as  $k$  is varied. Tuning the linearity parameter  $k$  adapts the behavior of the myriad from impulse-resistant mode-type estimators (small  $k$ ) to the Gaussian-efficient sample mean (large  $k$ ).

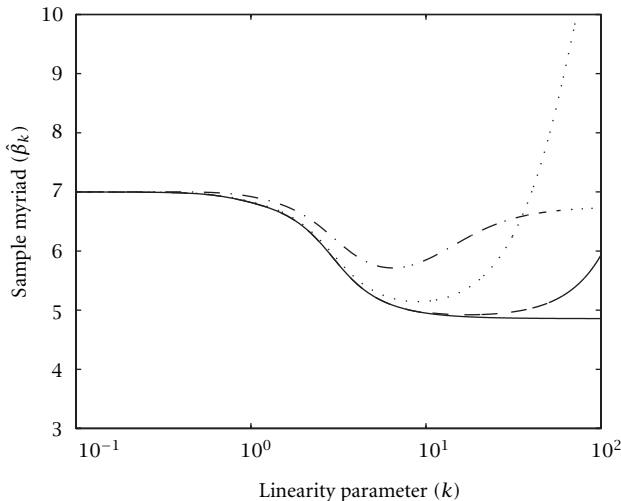


FIGURE 4: Values of the myriad as a function of  $k$  for the following data sets: (solid) original data set = 0, 1, 3, 6, 7, 8, 9; (dash-dot) original set plus an additional observation at 20; (dotted) additional observation at 100; (dashed) additional observations at 800,  $-500$ , and 700.

To illustrate the above, it is instructive to look at the behavior of the sample myriad shown in Figure 4. The solid line shows the values of the myriad as a function of  $k$  for the data set  $\{0, 1, 3, 6, 7, 8, 9\}$ . It can be observed that, as  $k$  increases, the myriad tends asymptotically to the sample average. On the other hand, as  $k$  is decreased, the myriad favors the value 7 which indicates the location of the cluster formed by the samples 6, 7, 8, 9. This is a typical behavior of the myriad for small  $k$ : it tends to favor values where samples are more likely to occur or cluster. We coined the term *myriad* as a result of this behavior.

The dotted line shows how the sample myriad is affected by an additional observation of value 100. For large values of  $k$ , the myriad is very sensitive to this new observation. On the contrary, for small  $k$ , the variability of the data is assumed to be small, and the new observation is considered an outlier, not influencing significantly the value of the myriad.

More interestingly, if the additional observations are the very large data 800,  $-500$ , 700, (dashed curve), the myriad is practically unchanged for moderate values of  $k$  ( $k < 10$ ). This behavior exhibits a very desirable outlier rejection property, not found, for example, in median-type estimators.

### 3.4. Scale-invariant operation

Unlike the sample mean or median, the operation of the sample myriad is not scale invariant, that is, for fixed values of the

linearity parameter, its behavior can vary depending on the units of the data. This is formalized in the following proposition, whose proof is straightforward.

**Proposition 1.** Let  $\hat{\beta}_k(\mathbf{X})$  denote the myriad of order  $k$  of the data in the vector  $\mathbf{X}$ . Then, for  $c > 0$ ,

$$\hat{\beta}_k(c\mathbf{X}) = c\hat{\beta}_{k/c}(\mathbf{X}). \quad (21)$$

According to (21), a change of scale in the data will be preserved in the myriad only if  $k$  experiences the same change of scale. Thus, the scale dependence of the myriad can be easily overcome if  $k$  carries the units of the data, or in other words, if  $k$  is a *scale parameter* of the data.

**Remark 1.** There is a common misconception about the role played by scale parameters like  $k$  in M-estimation applications. In order to solve the scale dependence problem, robust statisticians usually recommend to replace  $k$  by some resistant estimator of scale such as the MAD<sup>9</sup> or a maximum likelihood estimator [2, 4]. This approach is arbitrary and often leads to inefficient tuning values of the parameter  $k$ . As it will be illustrated in the following section, a good design of the tuning parameter should take into account not only the scale, but the whole structure of the underlying distribution, especially its impulsiveness.

## 4. OPTIMALITY OF THE SAMPLE MYRIAD IN PRACTICAL NOISE MODELS

### 4.1. Optimality in the $\alpha$ -stable model

The following result states the optimality of the mode-myriad as a location estimator for very impulsive  $\alpha$ -stable distributions.

**Proposition 2.** Let  $T_{\alpha, \gamma}(x_1, x_2, \dots, x_N)$  denote the maximum likelihood location estimator derived from a symmetric  $\alpha$ -stable distribution with characteristic exponent  $\alpha$  and dispersion  $\gamma$ . Then,

$$\lim_{\alpha \rightarrow 0} T_{\alpha, \gamma}(x_1, x_2, \dots, x_N) = \text{myriad}\{0; x_1, x_2, \dots, x_N\}. \quad (22)$$

*Proof.* To avoid the notational burden, we assume  $\gamma = 1$  (the generalization of the proof for arbitrary values of  $\gamma$  is straightforward). An integral expression for the zero-centered, unit dispersion, symmetric  $\alpha$ -stable density when  $\alpha < 1$  is given

<sup>9</sup>The median absolute deviation is defined as  $\text{MAD} = \text{median}\{x_i - \text{median}\{x_i\}\}$ .

[9] by

$$f_\alpha(x) = \frac{\alpha}{(1-\alpha)\pi} \frac{1}{|x|^{1/(1-\alpha)}} \times \int_0^{\pi/2} v(\theta) e^{-v(\theta)} |x|^{-\alpha/(1-\alpha)} d\theta, \quad (23)$$

where  $0 < v(\theta) < 1$  is a bounded real function that depends on  $\alpha$  [9, 10]. The value of  $f_\alpha(0)$  can be calculated as the limit when  $x \rightarrow 0$  in the expression above. A convenient rearrangement of (23) gives us

$$f_\alpha(x) = \frac{C_\alpha}{|x| + \xi_\alpha(x)}, \quad (24)$$

where  $C_\alpha$  is a constant given by

$$C_\alpha = \frac{\alpha}{(1-\alpha)\pi} \int_0^{\pi/2} v(\theta) e^{-v(\theta)} d\theta, \quad (25)$$

and  $\xi_\alpha(x)$  is a small function given by

$$\xi_\alpha(x) = |x| \left[ \frac{|x|^{\alpha/(1-\alpha)} \int_0^{\pi/2} v(\theta) e^{-v(\theta)} d\theta}{\int_0^{\pi/2} v(\theta) e^{-v(\theta)} |x|^{-\alpha/(1-\alpha)} d\theta} - 1 \right]. \quad (26)$$

The significance of  $\xi_\alpha$  relies on the fact that, for any finite interval, it converges *uniformly* to 0 as  $\alpha \rightarrow 0$ . Now, based on (24), the ML estimator is given by

$$T_\alpha = \arg \min_{\beta} \prod_i [ |x_i - \beta| + \xi_\alpha(x_i - \beta) ]. \quad (27)$$

Note that the constant  $C_\alpha$  has been dropped out since it is irrelevant for the minimization problem. Developing the products in (27), we get  $T_\alpha = \arg \min_{\beta} H_\alpha(\beta)$ , where

$$H_\alpha(\beta) = \prod_i |x_i - \beta| + \sum_i \left\{ \xi_\alpha(x_i - \beta) \prod_{j, j \neq i} |x_j - \beta| \right\} + \sum_{i,j} \left\{ \xi_\alpha(x_i - \beta) \xi_\alpha(x_j - \beta) \prod_{k, k \neq (i,j)} |x_k - \beta| \right\} + \dots \quad (28)$$

Let  $x_{\min}$  and  $x_{\max}$  denote, respectively, the minimum and maximum values in the sample set. The unimodality of  $f_\alpha$  guarantees that  $H_\alpha(\beta) > H_\alpha(x_{\min})$  for  $\beta < x_{\min}$ , and  $H_\alpha(\beta) > H_\alpha(x_{\max})$  for  $\beta > x_{\max}$ . This restricts the location of  $T_\alpha$  to be bracketed in the closed interval  $[x_{\min}, x_{\max}]$ . Restricted to that interval, all but the first term in (28) can be made arbitrarily small by making  $\alpha \rightarrow 0$ . This tells us that, as  $\alpha \rightarrow 0$ ,  $T_\alpha$  tends to have a “selection” behavior, in the sense that it converges to one of the sample values. The specific point of convergence can be identified by analyzing the behavior of the cost function  $H_\alpha(\beta)$  in the vicinity of each sample value. To do so, let  $r_i$  denote the number of times the value of  $x_i$  is repeated in the sample. Then, for  $\beta$  close to  $x_i$ , the first  $r_i$  terms in (28) can be made arbitrarily small, no matter what the value of  $\alpha$  is. Thus, as  $\alpha \rightarrow 0$ , the magnitude of  $H_\alpha(\beta)$  in the vicinity of  $x_i$  is mainly driven by the  $(r_i + 1)$ th

term in (28), which can be continuously approximated by

$$\lim_{\beta \rightarrow x_i} \sum_{i_1, \dots, i_{r_i}} \xi_\alpha(x_{i_1} - \beta) \cdots \xi_\alpha(x_{i_{r_i}} - \beta) \prod_{j, j \neq (i_1, \dots, i_{r_i})} |x_j - \beta| = [\xi_\alpha(0)]^{r_i} \prod_{j, x_j \neq x_i} |x_j - x_i|. \quad (29)$$

Hence, the point of convergence is the value  $x_i$  with the minimum associated cost (29), for small values of  $\alpha$ . Since  $\xi_\alpha(0) \rightarrow 0$  as  $\alpha \rightarrow 0$ , it is obvious that  $r_i$  must be maximized, or in other words,  $T_\alpha$  must converge to one of the most repeated values in the sample. Taking this fact into account, the minimization of (29) leads to the simplified result

$$\lim_{\alpha \rightarrow 0} T_\alpha = \arg \min_{x_j \in \mathcal{M}} \prod_{i, x_i \neq x_j} |x_i - x_j|, \quad (30)$$

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

In words, the ML estimator of location derived from an  $\alpha$ -stable distribution with small  $\alpha$ , behaves like the sample mode-myriad. Proposition 2 completes what we call the  *$\alpha$ -stable triplet* of optimality points satisfied by the myriad. On one extreme ( $\alpha = 2$ ), when the distributions are very well behaved, the myriad reaches optimal efficiency by making  $k = \infty$ . In the middle ( $\alpha = 1$ ), the myriad reaches optimality by making  $k = \gamma$ , the dispersion parameter of the Cauchy distribution. On the other extreme ( $\alpha \rightarrow 0$ ), when the distributions are extremely impulsive, the myriad reaches optimality again, this time by making  $k = 0$ . The  $\alpha$ -stable triplet demonstrates the central role played by myriad estimation in the  $\alpha$ -stable framework.

As  $\alpha$  is increased from 0 to 2, it is reasonable to expect, somehow, a progressive increase of the optimal  $k$ , from  $k = 0$  to  $k = \infty$ . The following proposition indicates the behavior of the optimal  $k$  for a general value of  $\alpha$ . Its proof is a direct consequence of Proposition 1 and the fact that  $y^{1/\alpha}$  is a scale parameter of the  $\alpha$ -stable distribution.

**Proposition 3.** *Let  $\alpha$  and  $\gamma$  denote the characteristic exponent and dispersion parameter of a symmetric  $\alpha$ -stable distribution as defined in (2). 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 nonnegative linearity parameter. Then,*

$$k_o(\alpha, \gamma) = k_o(\alpha, 1) \gamma^{1/\alpha}. \quad (31)$$

Proposition 3 indicates a “separability” of  $k_o$  in terms of  $\alpha$  and  $\gamma$ , reducing the optimal tuning problem to that of determining the function  $k(\alpha) = k_o(\alpha, 1)$ . This function is of fundamental importance for the proper operation of the myriad in the  $\alpha$ -stable framework, and we will refer to it as the  *$\alpha$ - $k$  curve*. Its form is obviously conditioned to the performance criterion chosen, and it may even depend on the sample size. In general, as discussed above, we should expect the  $\alpha$ - $k$  curve to be monotonically increasing, with  $k(0) = 0$  (very impulsive point) and  $k(2) = \infty$  (Gaussian point).



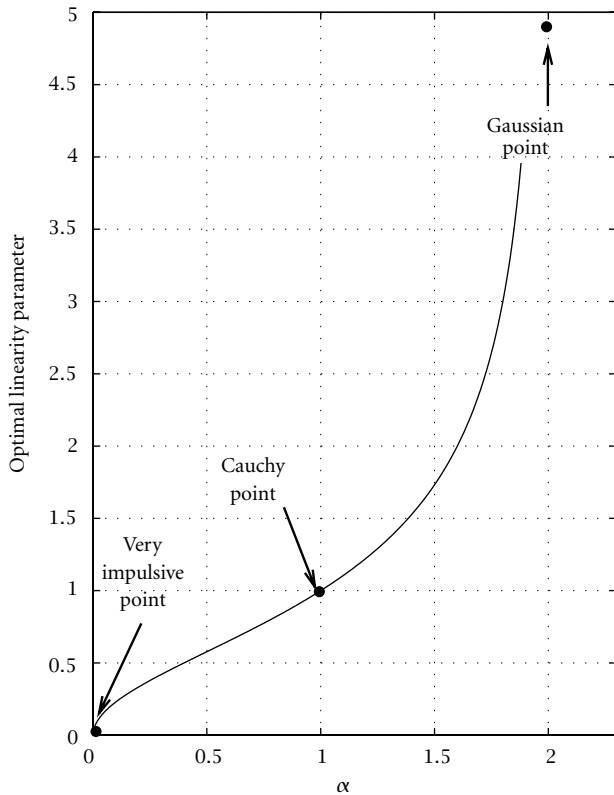


FIGURE 5: Empirical  $\alpha$ - $k$  curve for  $\alpha$ -stable distributions. The curve values at  $\alpha = 0, 1$ , and  $2$  constitute the optimality points of the  $\alpha$ -stable triplet.

If the performance criterion is the asymptotic variance, for example, then  $k(1) = 1$ , corresponding to the Cauchy point of the  $\alpha$ -stable triplet. The exact computation of the  $\alpha$ - $k$  curve for  $\alpha$ -stable distributions<sup>10</sup> is a challenging problem that we leave for further research. A simple empirical form that we have found to give consistently efficient results in a variety of conditions is

$$k(\alpha) = \sqrt{\frac{\alpha}{2-\alpha}} y^{1/\alpha}, \quad (32)$$

which is plotted in Figure 5. The  $\alpha$ - $k$  curve is a valuable tool for estimation and filtering problems that must *adapt* to the impulsiveness conditions of the environment.  $\alpha$ - $k$  curves in the  $\alpha$ -stable framework can be used, for example, to develop myriad-based adaptive detectors for channels with time-varying impulsiveness [28, 29].

#### 4.2. Optimality in the generalized- $t$ model

The family of generalized- $t$  distributions was introduced by Hall in 1966 as an empirical model for atmospheric radio noise [13]. Generalized- $t$  density functions can be conveniently parameterized as

$$f_x(x) = \frac{c}{(\alpha\sigma^2 + x^2)^{(1+\alpha)/2}}, \quad (33)$$

where  $\sigma > 0$ ,  $\alpha > 0$ , and  $c$  is a normalizing constant given by

$$c = \frac{\Gamma((1+\alpha)/2)}{\sqrt{\pi}\Gamma(\alpha/2)} \alpha^{\alpha/2} \sigma^\alpha. \quad (34)$$

It is easy to check that the distribution defined by (33) is algebraic-tailed, with tail constant  $\alpha$  and scale parameter  $\sigma$ . Although  $\alpha$  may take values larger than 2, its meaning is conceptually equivalent to the characteristic exponent of the  $\alpha$ -stable framework. At one extreme, when  $\alpha \rightarrow \infty$ , the generalized- $t$  distribution is equivalent to a zero-mean Gaussian distribution with variance  $\sigma^2$ . As it is the case with  $\alpha$ -stable distributions, decreased values of  $\alpha$  correspond to increased levels of impulsiveness. For values of  $\alpha \leq 2$ , the impulsiveness becomes high enough to make the variance infinite, and when  $\alpha = 1$ , the model corresponds to the Cauchy distribution. At the other extreme, when  $\alpha \rightarrow 0$ , the distribution exhibits the highest levels of impulsiveness.

The generalized- $t$  model has been shown to provide excellent fits to different types of atmospheric noise found in practice. Thanks to its simplicity and parsimony, it has been used by Middleton as a mathematically tractable approximation to his widely accepted models of electromagnetic radio noise [30]. It is worth noting that long before the introduction of the model by Hall, the generalized- $t$  distributions have been known in statistics as a family of heavy-tailed distributions categorized under the *type VII* of Pearson's distributional system [31].

It is easy to see that the maximum likelihood estimator of the location derived from the  $t$  density in (33) is precisely the sample myriad with linearity parameter

$$k = \sqrt{\alpha}\sigma. \quad (35)$$

The optimality of the myriad for *all* the distributions in the generalized- $t$  family indicates, as in the  $\alpha$ -stable case, its adequateness along a wide variety of noise environments, from the very impulsive ( $\alpha \rightarrow 0$ ) to the well-behaved Gaussian ( $\alpha = \infty$ ). Expression (35) gives the optimal tuning law as a function of  $\alpha$  and  $\sigma$  (note the close similarity with expression (31) for  $\alpha$ -stable distributions). Making  $\sigma = 1$ , we obtain the  $\alpha$ - $k$  curve for generalized- $t$  distributions,  $k(\alpha) = \sqrt{\alpha}$ . Like the  $\alpha$ - $k$  curve of  $\alpha$ -stable distributions, this curve is also monotonically increasing, and contains the optimality points of the  $\alpha$ -stable triplet, namely the Gaussian point ( $k(\infty) = \infty$ ), the Cauchy point ( $k(1) = 1$ ), and the very impulsive point ( $k(0) = 0$ ).

The generalized- $t$  model provides a simple framework to assess the performance of the sample myriad as the impulsiveness of the distributions is changed. It can be proven that the normalized asymptotic variance<sup>11</sup> of the optimal sample

<sup>10</sup>The concept of  $\alpha$ - $k$  curve can also be extended to distribution families other than the  $\alpha$ -stable, as it is discussed in Section 4.2.

<sup>11</sup>Let  $V_T(N)$  be the variance of the estimator  $T$  when the sample size is  $N$ . Then, the normalized asymptotic variance  $V$  is defined as  $V = \lim_{N \rightarrow \infty} NV_T(N)$ .

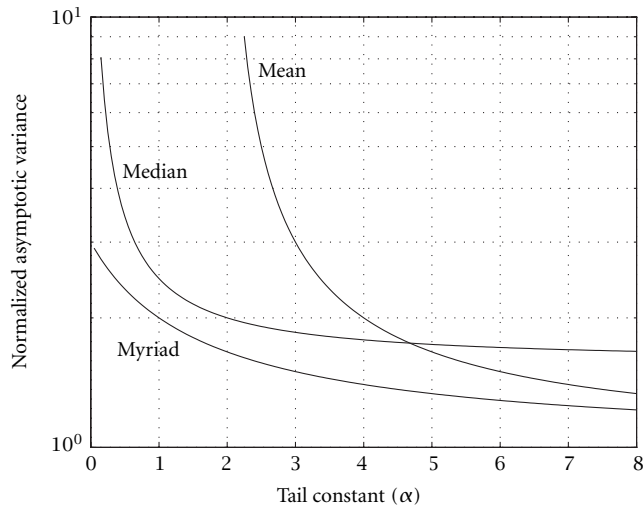


FIGURE 6: Normalized asymptotic variance of the sample mean, sample median, and optimal sample myriad in generalized- $t$  noise. The myriad outperforms the mean and median for any level of impulsiveness.

myriad at the generalized- $t$  model is (for a derivation, see, for example, [29]):

$$V_{\text{myr}} = \frac{\alpha + 3}{\alpha + 1} \sigma^2. \quad (36)$$

A plot of  $V_{\text{myr}}$  versus  $\alpha$  is shown in Figure 6 for  $\sigma = 1$ . The asymptotic variances of the sample mean ( $V_{\text{mean}}$ ) and sample median ( $V_{\text{med}}$ ) are also included for comparison [29]. The superiority of the sample myriad over both mean and median is evident from the figure.

## 5. INTRODUCING WEIGHTS TO THE MYRIAD ESTIMATOR: THE WEIGHTED MYRIAD

### 5.1. M-filters, maximum likelihood, and weights

A filtering framework cannot be considered complete until an appropriate form of *weighted* operation is defined. Filter weights or “coefficients” are extremely important for most applications in which signal correlations are to be exploited. To visualize the potential gains that a weighting framework can add to a filter structure, it suffices to see the enormous gap in versatility and performance separating the mean filter from its weighted version, the class of linear FIR filters. In this section, we introduce a powerful weighting framework for the class of M-filters, directly derived from the concept of maximum likelihood.

Consider the samples  $x_1, x_2, \dots, x_N$ , and a set of nonnegative weights  $W_1, W_2, \dots, W_N$ . Given the cost function  $\rho$ , the goal is to define a “weighted” operator

$$\hat{\beta}_{\mathbf{W}} = \hat{\beta}_{W_1, W_2, \dots, W_N}(x_1, x_2, \dots, x_N), \quad (37)$$

satisfying the following “desirable” *weighting conditions*:<sup>12</sup> (1) when  $W_1 = W_2 = \dots = W_N = 1$ ,  $\hat{\beta}_{\mathbf{W}}$  reduces to the standard M-estimator associated with  $\rho$ ; (2) making  $W_i = 0$  is equivalent to discarding the sample  $x_i$ ; (3) making  $W_i = \infty$  while keeping the other weights finite, forces  $\hat{\beta}_{\mathbf{W}}$  to be equal to  $x_i$ ; and (4)  $\hat{\beta}_{\mathbf{W}}$  is at least piecewise continuous on each of the  $W_i$ .

Conditions (2) and (3) suggest the interpretation of  $W_i$  as a reliability measure associated with the sample  $x_i$ . Intuitively, the weighted filter output gets closer to  $x_i$  as the value of  $W_i$  increases in relation to the other weights. Hence, although the intended application of the weighted filter might not be location estimation,<sup>13</sup> its behavior is intrinsically related to that of a location estimator in which the samples are assigned different levels of reliability.

This leads to a natural weighting framework in the context of maximum likelihood estimation: instead of considering the samples identically distributed, assume that they are generated from the same distribution, but allowing the scale associated with each sample to be different. In this way, an observation assigned to a small scale  $\sigma_i$ , can be related to a highly localized (very reliable) density function,  $(1/\sigma_i)f[(x_i - \beta)/\sigma_i]$ , where  $f$  is the generating “unweighted” pdf. At the other extreme, a large value of  $\sigma_i$  indicates a dispersed density function, which reflects the low reliability of the sample. Given a set of scales,  $\sigma_1, \sigma_2, \dots, \sigma_N$ , we can thus define the *weighted maximum likelihood estimator* as the value  $\hat{\beta}$  that maximizes a “weighted” likelihood function of the form

$$L(\beta) = \prod_{i=1}^N \frac{1}{\sigma_i} f\left[\frac{x_i - \beta}{\sigma_i}\right]. \quad (38)$$

Making  $W_i = 1/\sigma_i$ , and letting  $\rho(x) = -\log f(x)$  as in the “unweighted” M-estimation problem, expression (38) leads in a natural way to the following definition.

*Definition 3* (weighted M-estimator). Given a cost function  $\rho$ , and a set of samples  $x_1, x_2, \dots, x_N$ , we define the *weighted M-estimator* of location as

$$\hat{\beta}_{\mathbf{W}} = \arg \min_{\beta} \sum_{i=1}^N \rho(W_i(x_i - \beta)), \quad (39)$$

where  $W_1, \dots, W_N \geq 0$ , represent the *weights* or “coefficients” of the estimator. A running-window filter based on a weighted M-estimator of location will be called, accordingly, a *weighted M-filter*.

Definition 3 is a powerful generalization of the M-filtering framework. When  $\rho$  is continuous, even, and monotonic increasing on  $[0, \infty)$ , it can be easily proven that  $\hat{\beta}_{\mathbf{W}}$  satisfies the *weighting conditions* (1) through (4). When

<sup>12</sup>Note that these conditions are naturally satisfied by the weighting frameworks of both median and (normalized) linear FIR filters.

<sup>13</sup>In fact, most applications of interest would be both deterministic and statistical in nature; an important example is a low-pass, frequency selective filter, with impulse resistance.

$\rho(x) = x^2$ , for example,  $\hat{\beta}_{\mathbf{w}}$  defines the *weighted mean filter*,  $\hat{\beta}_{\mathbf{w}} = \sum_{i=1}^N W_i^2 x_i / \sum_{i=1}^N W_i^2$ , which can be seen as a “normalized” linear FIR filter with nonnegative coefficients. When  $\rho(x) = |x|$ , on the other hand,  $\hat{\beta}_{\mathbf{w}}$  reduces to the familiar *weighted median* [26].

*Remark 2.* The nonnegativity of the weights in Definition 3 hampers the filter capability to mimic “bandpass” and “high-pass” operations such as those seen in FIR filters with negative weights. This problem has been widely acknowledged, for example, as one of the main limitations of weighted median filters, limiting their applications to problems of the “low-pass” type [26]. Extensions of Definition 3 to allow weights of any sign can be found in [32].

## 5.2. The weighted myriad

The weighted myriad is straightforwardly defined by using the cost function (6) in Definition 3.

*Definition 4* (weighted myriad). Let  $\mathbf{w} = [w_1, w_2, \dots, w_N]$  be a vector of nonnegative weights. Given  $k > 0$ , the weighted myriad of order  $k$  for the data  $x_1, x_2, \dots, x_N$ , is defined as

$$\begin{aligned} \hat{\beta}_{k,\mathbf{w}} &= \text{myriad} \{k; w_1 \circ x_1, \dots, w_N \circ x_N\} \\ &= \arg \min_{\beta} \sum_{i=1}^N \log [k^2 + w_i(x_i - \beta)^2], \end{aligned} \quad (40)$$

where  $w_i \circ x_i$  represents the weighting operation in (40). In some situations, the following equivalent expression can be computationally more convenient

$$\hat{\beta}_{k,\mathbf{w}} = \arg \min_{\beta} \prod_{i=1}^N [k^2 + w_i(x_i - \beta)^2]. \quad (41)$$

As in the unweighted case, we will call  $k$  the *linearity parameter* of the filter. Notice that we have preferred to use  $w_i = W_i^2$  (instead of  $W_i$ ) as the primitive definition of the filter weights. The convenience of this parameterization is evident in the following generalization of the linear property, whose proof is a straightforward generalization of the unweighted case.

*Property 3* (linear property). Let the weights  $w_1, \dots, w_N > 0$ , be assigned to the samples  $x_1, \dots, x_N$ , then

$$\lim_{k \rightarrow \infty} \hat{\beta}_{k,\mathbf{w}} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i}. \quad (42)$$

Equation (42) provides a fundamental link between myriad and linear FIR filters. Informally, a linear FIR filter with positive weights  $w_i$ , can be seen as the serial concatenation of an amplifier of gain  $\sum_{i=1}^N w_i$ , and a myriad filter with  $k = \infty$  and the same weight values.

A generalization of the mode property also holds for the weighted myriad.

*Property 4* (mode property). Let the weights  $w_1, \dots, w_N > 0$  be assigned, respectively, to the samples  $x_1, \dots, x_N$ . Then,

$$\begin{aligned} \hat{\beta}_{0,\mathbf{w}} &= \lim_{k \rightarrow 0} \hat{\beta}_{k,\mathbf{w}}(x_1, \dots, x_N) \\ &= \arg \min_{x_j \in \mathcal{M}} \left( \frac{1}{w_j} \right)^{r/2} \prod_{i=1, x_i \neq x_j}^N |x_i - x_j|, \end{aligned} \quad (43)$$

where  $\mathcal{M}$  is the set of most repeated values in the sample, and  $r$  is the number of repetitions associated with a member of  $\mathcal{M}$ .

*Proof.* Following the steps of the proof for the unweighted version, it is straightforward that

$$\hat{\beta}_{0,\mathbf{w}} = \arg \min_{x_j \in \mathcal{M}} \prod_{i=1, x_i \neq x_j}^N w_i (x_i - x_j)^2. \quad (44)$$

Dividing by  $\prod_{i=1}^N w_i$ , and applying the square root to the expression to be minimized, we get the desired result.  $\square$

Weighted myriad filters inherit the optimal performance of the sample myriad in both  $\alpha$ -stable and generalized- $t$  environments. In addition to the linear and mode properties, the following are important basic properties of the weighted myriad.

*Property 5* (no undershoot/overshoot). The output of a weighted myriad filter is always bracketed by the inequalities

$$x_{(1)} \leq \hat{\beta}_{k,\mathbf{w}}(x_1, x_2, \dots, x_N) \leq x_{(N)}, \quad (45)$$

where  $x_{(1)}$  and  $x_{(N)}$  denote the minimum and maximum samples in the input window.

*Proof.* Note that for  $\beta < x_{(1)}$ ,

$$k^2 + w_i(x_i - x_{(1)})^2 < k^2 + w_i(x_i - \beta)^2 \quad (46)$$

and hence

$$\prod_{i=1}^N [k^2 + w_i(x_i - x_{(1)})^2] < \prod_{i=1}^N [k^2 + w_i(x_i - \beta)^2]. \quad (47)$$

This implies that any  $\beta < x_{(1)}$  is “beat” by  $x_{(1)}$  for the minimization of the weighted myriad objective function. Therefore, the weighted myriad cannot be less than  $x_{(1)}$ . A similar argument on  $x_{(N)}$ , leads to the conclusion that the weighted myriad cannot be larger than  $x_{(N)}$ .  $\square$

*Property 6* (shift and sign invariance). Let  $z_i = x_i + b$ . Then, for any  $k$  and  $\mathbf{w}$ ,

- (i)  $\hat{\beta}_{k,\mathbf{w}}(z_1, \dots, z_N) = \hat{\beta}_{k,\mathbf{w}}(x_1, \dots, x_N) + b$ ;
- (ii)  $\hat{\beta}_{k,\mathbf{w}}(-z_1, \dots, -z_N) = -\hat{\beta}_{k,\mathbf{w}}(z_1, \dots, z_N)$ .

The proof is trivial from the definition of the weighted myriad in (40).

*Property 7* (unbiasedness). Let  $X_1, X_2, \dots, X_N$  be all independent and symmetrically distributed around their symmetry center  $c$ . Then,  $\hat{\beta}_{k,\mathbf{w}} = \hat{\beta}_{k,\mathbf{w}}(X_1, X_2, \dots, X_N)$  is also symmetric

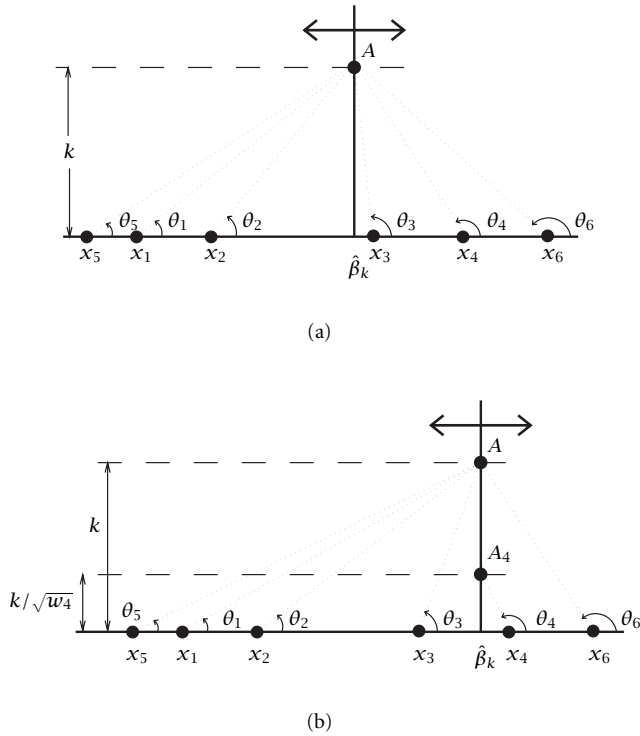


FIGURE 7: (a) The unweighted sample myriad,  $\hat{\beta}_k$ , indicates the position of a moving bar such that the product of distances from point  $A$  to the sample points  $x_1, x_2, \dots, x_N$  is minimum. (b) If the weight  $w_4 > 1$  is introduced, the product of distances is more sensitive to the variations of the segment  $x_4 A_4$ , very likely resulting in a weighted myriad  $\hat{\beta}_{k,w}$  closer to  $x_4$ .

rically distributed around  $c$ . In particular, if  $E\hat{\beta}_{k,w}$  exists, then  $E\hat{\beta}_{k,w} = c$ .

*Proof.* If  $X_i$  is symmetric around  $c$ , then  $2c - X_i$  has the same distribution as  $X_i$ . Thus,  $\hat{\beta}_{k,w}(X_1, X_2, \dots, X_N)$  has the same distribution as  $\hat{\beta}_{k,w}(2c - X_1, 2c - X_2, \dots, 2c - X_N)$ , which, according to Property 6, is equal to  $2c - \hat{\beta}_{k,w}(X_1, X_2, \dots, X_N)$ . It follows that  $\hat{\beta}_{k,w}$  is symmetric around  $c$ .  $\square$

### 5.3. Geometrical interpretation

The operation of the sample myriad as defined in (41) can be interpreted in a more intuitive manner. We allow a vertical bar to run horizontally through the real line as depicted in Figure 7a. Then, the sample myriad,  $\hat{\beta}_k$ , indicates the position of the bar for which the product of distances from the point  $A$  to the sample points  $x_1, x_2, \dots, x_N$  is minimum. If weights are introduced, each sample point  $x_i$  is assigned a different point  $A_i$  in the bar, as illustrated in Figure 7b.

The geometrical interpretation of the myriad is intuitively insightful. When  $k$  approaches 0, it gives a conceptually simple pictorial demonstration of the mode-myriad formula in (15). It can also lead to alternative algorithms for calculating the sample myriad, as stated in the following rather interesting proposition. The proof is not difficult, and we leave it as an exercise to the interested reader.

Proposition 4. Let  $\theta_i$  denote the (signed) angle between the real line and the segment  $x_i A_i$  as depicted in Figure 7b. Then, the geometric position of the sample myriad satisfies

$$\sum_{i=1}^N \sin(2\theta_i) = 0. \quad (48)$$

## 6. MYRIAD FILTER DESIGN AND OPTIMIZATION

### 6.1. Optimal weighted M-filters: the M-filter “normal” equations

A problem of fundamental importance in statistical signal processing is the design of the coefficients of a filter in order to meet certain optimality requirements [33]. In this section, we develop a set of implicit equations that characterize the optimal weighted M-filter for a general cost function  $\rho$ . Let  $X(n)$  be an observed process which is assumed to be statistically related to some desired process  $D(n)$  of interest.  $X(n)$  is typically a transformed or corrupted version of  $D(n)$ . Furthermore, we assume that these processes are jointly stationary. A window of width  $N$  slides across the input process pointwise estimating the desired sequence. The vector containing the  $N$  samples in the window at time  $n$  is  $\mathbf{X}(n) = [X(n - N_1), \dots, X(n), \dots, X(n + N_2)] = [X_1(n), X_2(n), \dots, X_N(n)]$ , with  $N = N_1 + N_2 + 1$ . The running weighted filter output estimates the desired signal as

$$\hat{D}_w(n) = \arg \min_{\beta} \sum_{i=1}^N \rho(W_i(X_i(n) - \beta)). \quad (49)$$

The problem is to find the weight values  $W_1, W_2, \dots, W_N \geq 0$  which will minimize the estimation error. The following proposition characterizes these optimal weights by means of a set of implicit equations.

Proposition 5 (normal equations). Let  $\mathbf{W} = [W_1, \dots, W_N]$  be the weight vector of a weighted M-filter associated with the cost function  $\rho$  as in Definition 3. Let  $L$  define the error measure to be used as the design criterion in the sense that the optimal filter must minimize the loss function

$$J(\mathbf{W}) = E \{L(D - \hat{D}_w)\}. \quad (50)$$

Then, the optimal filter weights satisfy a set of  $N$  “normal equations” of the form<sup>14</sup>

$$E \left\{ L'(D - \hat{D}_w) \frac{\partial \hat{D}_w}{\partial W_i} \right\} = 0, \quad \text{for } i = 1, \dots, N, \quad (51)$$

where

$$\begin{aligned} \frac{\partial \hat{D}_w}{\partial W_i} &= \frac{\rho'(z_i) + z_i \rho''(z_i)}{\sum_{j=1}^N W_j^2 \rho''(z_j)}, \\ z_i &= W_i(X_i - \hat{D}_w). \end{aligned} \quad (52)$$

<sup>14</sup>The prime symbol, ( $'$ ), is used to denote the derivative of the superscripted function.

*Proof.* Expression (51) holds straightforwardly after forcing  $\partial J(\mathbf{W})/\partial \mathbf{W} = 0$  as a condition for  $J(\mathbf{W})$  to be a minimum. Now, according to its definition,

$$\hat{D}_{\mathbf{w}} = \arg \min_{\beta} \sum_{i=1}^N \rho(W_j(X_j - \beta)). \quad (53)$$

Taking the derivative with respect to  $\beta$  and making it equal to zero, we obtain the following relation for  $\hat{D}_{\mathbf{w}}$ :

$$\sum_{i=1}^N W_j \rho'(W_j(X_j - \hat{D}_{\mathbf{w}})) = 0. \quad (54)$$

Differentiating with respect to  $W_i$  we get

$$\begin{aligned} \frac{\partial}{\partial W_i} \left[ W_i \rho'(W_i(X_i - \hat{D}_{\mathbf{w}})) \right] \\ + \sum_{j=1, j \neq i}^N \frac{\partial}{\partial W_i} \left[ W_j \rho'(W_j(X_j - \hat{D}_{\mathbf{w}})) \right] = 0, \end{aligned} \quad (55)$$

which leads to

$$\begin{aligned} \rho'(W_i(X_i - \hat{D}_{\mathbf{w}})) \\ + W_i \rho''(W_i(X_i - \hat{D}_{\mathbf{w}})) \left[ (X_i - \hat{D}_{\mathbf{w}}) - W_i \frac{\partial \hat{D}_{\mathbf{w}}}{\partial W_i} \right] \\ - \sum_{j=1, j \neq i}^N W_j^2 \rho''(W_j(X_j - \hat{D}_{\mathbf{w}})) \frac{\partial \hat{D}_{\mathbf{w}}}{\partial W_i} = 0. \end{aligned} \quad (56)$$

Rearranging terms and replacing  $W_i(X_i - \hat{D}_{\mathbf{w}})$  by  $z_i$  for convenience of notation, we get

$$\begin{aligned} \rho'(z_i) W_i \rho''(z_i) (X_i - \hat{D}_{\mathbf{w}}) \\ = -W_i \rho''(z_i) \left[ -W_i \frac{\partial \hat{D}_{\mathbf{w}}}{\partial W_i} \right] + \sum_{j=1, j \neq i}^N W_j^2 \rho''(z_j) \frac{\partial \hat{D}_{\mathbf{w}}}{\partial W_i}. \end{aligned} \quad (57)$$

Finally, solving for  $\partial \hat{D}_{\mathbf{w}}/\partial W_i$ , we obtain

$$\frac{\partial \hat{D}_{\mathbf{w}}}{\partial W_i} = \frac{\rho'(z_i) + z_i \rho''(z_i)}{\sum_{j=1}^N W_j^2 \rho''(z_j)}, \quad (58)$$

which concludes the proof.  $\square$

It is worth noting that expression (51) is the M-filtering equivalent of the *principle of orthogonality* of Wiener filter theory [33]. Thus, Proposition 5 by itself, can lead to direct solutions of the optimal weights when the cost function  $\rho$  is simple. In effect, when  $\rho(x) = x^2$ , that is, when the filter is linear, expression (51) becomes

$$E \left\{ 2(d - \hat{d}_{\mathbf{w}}) \left( \frac{2z_i + 2z_i}{\sum_{i=1}^N 2W_j^2} \right) \right\} = 0, \quad (59)$$

or equivalently,

$$E \left\{ (d - \hat{d}_{\mathbf{w}})(x_i - \hat{d}_{\mathbf{w}}) \right\} = 0, \quad (60)$$

which is one of the forms of the principle of orthogonality. In general, a direct solution to equation (51) might be troublesome, especially when the cost function  $\rho$  is nonconvex (which is the case for the myriad). In such cases, other alternatives such as nonlinear least squares or adaptive LMS algorithms can be developed [34]. In the following section we introduce a suboptimal and very simple technique for myriad filter design derived from the linear property.

## 6.2. Myriadization: bringing impulse resistance to linear filters

The linear property indicates that for very large values of  $k$ , the weighted myriad filter reduces to a constrained linear FIR filter. The meaning of  $k$  suggests that we can provide a linear filter with resistance to impulsive noise by simply reducing the linearity parameter from  $k = \infty$  to a finite value. This would transform the linear filter into a myriad filter with the same weights. In the same way as the term *linearization* is commonly used to denote the transformation of an operator into a linear one, we will refer to the above transformation as *myriadization*.

Myriadization is a simple but powerful technique that brings impulse resistance to constrained linear filters. It provides also a simple methodology to design suboptimal myriad filters in impulsive environments. Basically, we can design a constrained linear filter for Gaussian or noiseless environments using linear filter design techniques, and then provide the filter with impulse resistance capabilities by means of myriadization. The value to which  $k$  is to be reduced can be designed according to the impulsiveness of the environment, for example, by means of an  $\alpha$ - $k$  curve.

*Remark 3.* It must be taken into account that a linear filter has to be in “constrained form” before myriadization can be applied. This means that the filter coefficients  $w_i$  must be nonnegative and satisfy the normalization condition  $\sum_{i=1}^N w_i = 1$ . A filter for which  $\sum_{i=1}^N w_i \neq 1$ , must be first “decomposed” into the cascade of its normalized version with an amplifier of gain  $\sum_{i=1}^N w_i$ .

We illustrate the potential of myriadization in the following two examples.

*Example 1.* Figure 8a depicts a unit-amplitude linearly swept-frequency cosine signal spanning instantaneous frequencies ranging from 0 to 400 Hz. The chirp was generated with MatLab’s `chirp` function having a sampling interval of 0.0005 seconds. Figure 8b shows the chirp immersed in additive Cauchy noise ( $\gamma = 1$ ). The plot is truncated to the same scale as the other signals in the figure. A low-pass linear FIR filter with 30 coefficients processes the chirp with the goal of retaining its low-frequency components. The FIR low-pass filter weights were designed with MatLab’s `fir1` function with a normalized frequency cutoff of 0.05. Under ideal, no-noise conditions, the output of the linear filter would be that of Figure 8c. However, the impulsive nature of the noise introduces severe distortions to the actual output, as depicted in Figure 8d. Myriadizing the linear filter by reducing  $k$

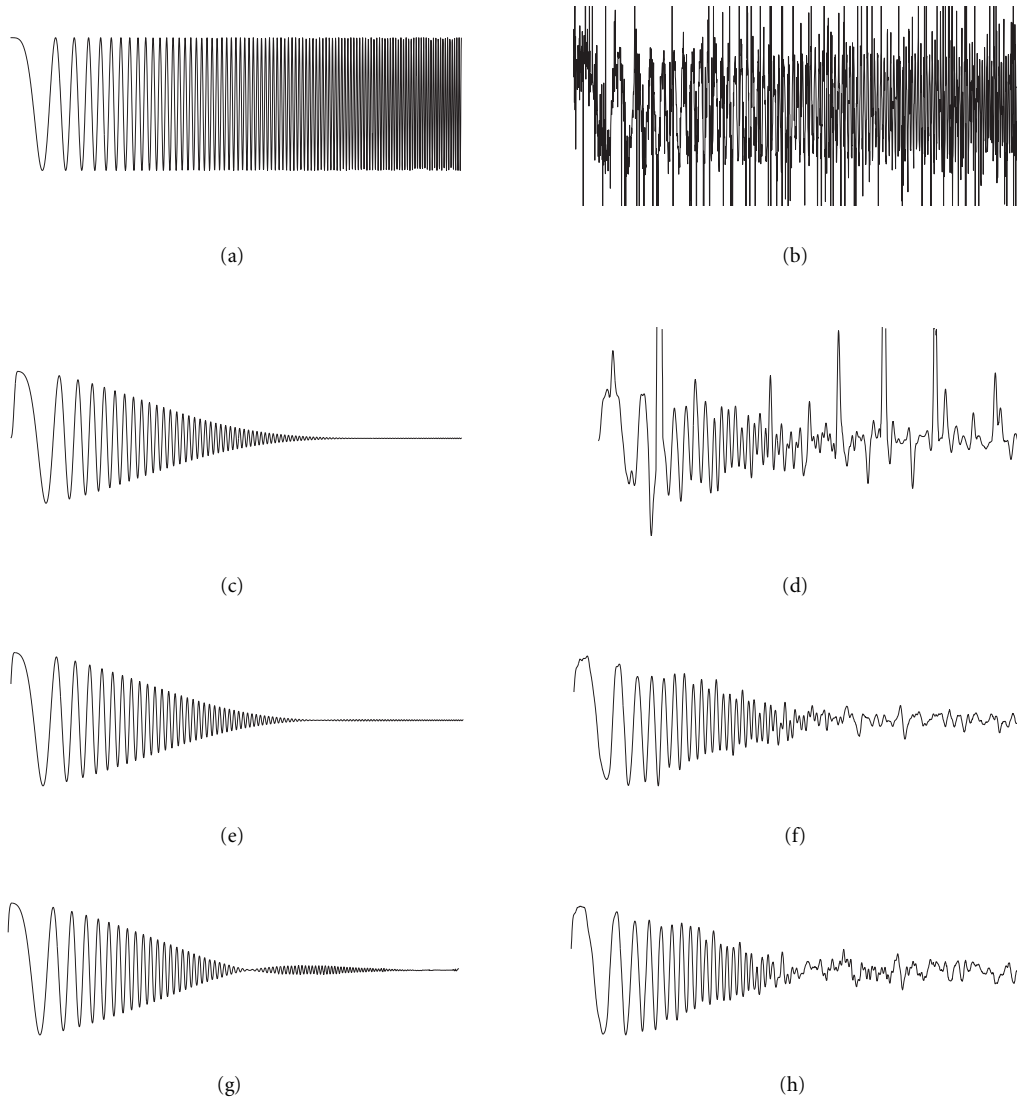


FIGURE 8: Myriadizing a linear low-pass filter in an impulsive environment: (a) chirp signal, (b) chirp in additive impulsive noise, (c) ideal (no-noise) myriad filter output with  $k = \infty$ , (e)  $k = 0.5$ , and (g)  $k = 0.2$ ; (d) Myriad filter output in the presence of noise with  $k = \infty$ , (f)  $k = 0.5$ , and (h)  $k = 0.2$ . An interactive Java demo in which the user can play with different noise processes and linearity parameters is available at <http://www.ece.udel.edu/signals/robust/myriad/myriaddemo/index.html>.

to a finite value of 0.5, significantly improves the filter performance (see Figures 8e, 8f). Further reduction of  $k$  to 0.2 drives the myriad closer to a selection mode where some distortion on the filter output under ideal conditions can be seen (see Figure 8g). The output under the noisy conditions is not improved by further reducing  $k$  to 0.2, or lower, as the filter in this case is driven to a “selection” operation mode.

*Example 2.* We tested the *myriadization* concept on the phase synchronization problem of the first-order Phase-Locked Loop (PLL) depicted in Figure 9. Systems of this type are widely used for recovering carrier phase in coherent demodulators [35]. The PLL has a linear FIR low-pass filter with 13 normalized coefficients intended to let pass only the low frequencies generated by the multiplier. The output of the

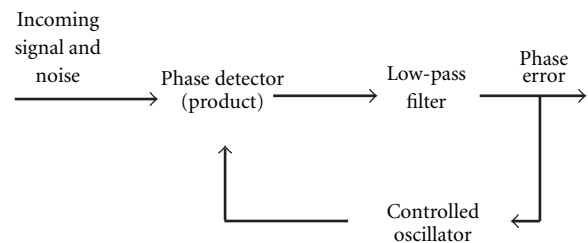


FIGURE 9: Block diagram of the Phase-Locked Loop analyzed in the paper.

low-pass filter represents the phase error between the incoming carrier and the recovered tone provided by the controlled

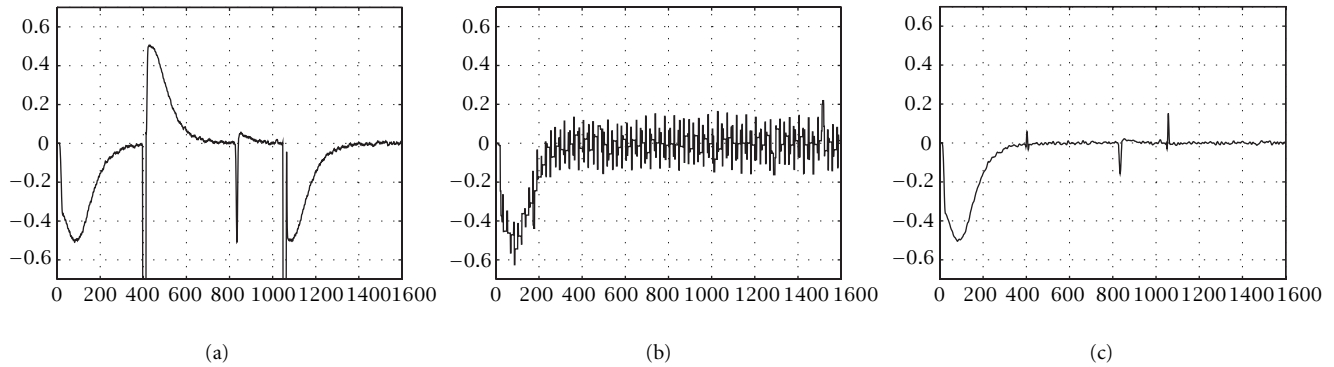


FIGURE 10: Phase error plot for the PLL with (a) a linear FIR filter; (b) an optimal weighted median filter; and (c) a *myriadized* version of the linear filter. The superiority of the myriadized filter is clear.

oscillator. Hence, we can say that the system is working properly (i.e., achieving synchronism), whenever the output of the low-pass filter is close to zero. The incoming signal is a sinusoid of high frequency and unitary amplitude, immersed in additive white Gaussian noise of variance  $10^{-3}$ , yielding a signal-to-noise ratio (SNR) of 30 dB. The parameters of the system, including (linear) filter weights and oscillator gain, were manually adjusted so that the error signal had minimum variance. The details and source code of the simulation can be downloaded, in Simulink and MATLAB formats, from [32]. Three different scenarios, corresponding to three different low-pass filter structures were simulated. The incoming and noise signals were identical for the three systems. At three arbitrary time points ( $t \approx 400, 820, 1040$ ), short bursts of high power Gaussian noise were added to the noise signal. The length of the bursts was relatively short (between 4 and 10 sampling times) compared to the length of the filter impulse response (12 sampling times). The SNR during burst periods was very low (about  $-10$  dB), making the noise look heavy impulsive. Figure 10 shows the phase error in time when the standard linear filter was used. It is evident from the figure that this system is very likely to lose synchronism after a heavy burst. Figure 10b shows the phase error of a second scenario in which a weighted median filter has been designed to imitate the low-pass characteristics of the original linear filter [26, 36]. Although the short noise bursts do not affect the estimate of the phase, the variance of the estimate is very large. This “noise amplification” behavior can be explained from the inefficiency introduced by the “selection” property of the median, that is, the fact that the filter output is always constrained to be one of its inputs. Finally, Figure 10c shows the phase after the low-pass filter has been myriadized using a parameter  $k$  equal to half the carrier amplitude. Although the phase error is increased during the bursts, the performance of the myriadized PLL is not degraded, and the system does not lose synchronism.

## 7. CONCLUSIONS

Weighted myriad filtering is a flexible filtering framework that derives important robustness properties from the im-

pulsive characteristics of symmetric  $\alpha$ -stable distributions. In the same way as linear and median filters are related to Gaussian and Laplacian distributions, respectively, myriad filter theory is based on the definition of the sample myriad as the maximum likelihood location estimator of the Cauchy distribution—the only non-Gaussian symmetric  $\alpha$ -stable distribution for which a closed-form density is available. When weights are introduced in the definition, the weighted myriad filter appears as a rich and flexible class of filters that can range, by simply varying a tuning parameter, from highly robust mode-like filter forms to simple and Gaussian-efficient linear FIR filters. Interactive illustrations of the power of the weighted myriad framework are also available on the Internet at <http://www.ee.udel.edu/signals/robust/myriad/myriaddemo/index.html>. Many interesting and important issues remain open. More research must be done on weighted myriad filter optimization and design, algorithm development for myriad calculation, and the extension of the weighted myriad to allow bandpass-type (negative weights) and vector operations.

## ACKNOWLEDGEMENT

This research was supported in part by the National Science Foundation under Grant MIP-9530923, and through collaborative participation in the Advanced Telecommunications/Information Distribution Research Program (ATIRP) Consortium sponsored by the U.S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0002.

## REFERENCES

- [1] J. G. Gonzalez, D. L. Lau, and G. R. Arce, “Towards a general theory of robust nonlinear filtering: selection filters,” in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, April 1997, Munich, Germany.
- [2] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics. The Approach Based on Influence Functions*, Wiley, New York, 1986.
- [3] P. S. Laplace, “Mémoire sur la probabilité des causes par les évènements,” *Mémoires de Mathématique et de Physique*, vol. 6, 1774.



- [4] P. J. Huber, *Robust Statistics*, John Wiley & Sons, New York, 1981.
- [5] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: a survey," *Proceedings of the IEEE*, vol. 73, pp. 433–481, March 1985.
- [6] I. Pitas and A. N. Venetsanopoulos, *Non-linear Filters*, Kluwer, 1989.
- [7] Y. H. Lee and S. A. Kassam, "Generalized median filtering and related nonlinear filtering techniques," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, no. 5, pp. 672–683, 1985.
- [8] R. Durrett, *Probability: Theory and Examples*, Duxbury Press, New York, 2nd edition, 1996.
- [9] V. M. Zolotarev, *One-Dimensional Stable Distributions*, vol. 65, American Mathematical Society, Rhode Island, 1986.
- [10] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, John Wiley & Sons, New York, 1995.
- [11] R. A. Fisher, "On the mathematical foundations of theoretical statistics," in *Philosophical Transactions of the Royal Society of London*, vol. 222 of A, pp. 309–368, 1922.
- [12] S. Rappaport and L. Kurz, "An optimal nonlinear detector for digital data transmission through non-Gaussian channels," *IEEE Trans. Communications*, vol. 14, no. 3, 1966.
- [13] H. M. Hall, "A new model for impulsive phenomena: application to atmospheric-noise communication channels," Technical reports 3412-8 and 7050-7, Stanford Electronics Laboratories, Stanford University, Stanford, California, 1966.
- [14] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton, New Jersey, 1972.
- [15] F. Steiner, "Most frequent value and cohesion of probability distributions," *Acta Geod., Geophys et Mont. Acad. Sci Hung.*, vol. 8, no. 3-4, pp. 381–395, 1973.
- [16] K. L. Boyer, M. J. Mirza, and G. Ganguly, "The robust sequential estimator: a general approach and its application to surface organization in range data," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 10, pp. 987–1001, 1994.
- [17] S. Ambike and D. Hatzinakos, "A new filter for highly impulsive  $\alpha$ -stable noise," in *Proc. of the 1995 Int. Workshop on Nonlinear Signal and Image Processing*, Halkidiki, Greece, June 1995.
- [18] P. Lévy, *Calcul des Probabilités*, Gauthier Villars, Paris, 1925.
- [19] J. G. Gonzalez, D. W. Griffith, and G. R. Arce, "Zero-order statistics: a signal processing framework for very impulsive processes," in *Proc. IEEE Signal Processing Workshop on Higher Order Statistics*, Banff, Alberta, Canada, July 1997.
- [20] E. Souza, "Performance of a spread spectrum packet radio network in a Poisson field of interferences," *IEEE Transactions on Information Theory*, vol. IT-38, pp. 1743–1754, November 1992.
- [21] J. Ilow and D. Hatzinakos, "Analytic alpha-stable noise modeling in a Poisson field of interferers or scatterers," *IEEE Trans. Signal Processing*, vol. 46, no. 6, pp. 1601–1611, 1998.
- [22] P. Tsakalides and C. L. Nikias, "Maximum likelihood localization of sources in noise modeled as a stable process," *IEEE Trans. Signal Processing*, vol. 43, no. 11, pp. 2700–2713, 1995.
- [23] J. S. Bodenschatz and C. L. Nikias, "Symmetric  $\alpha$ -stable filter theory," *IEEE Trans. Signal Processing*, vol. 45, no. 9, pp. 2301–2306, 1997.
- [24] J. B. C. Silva and A. O. Cutrim, "A robust maximum likelihood method for gravity and magnetic interpretation," *Geoplotation*, pp. 1–31, 1989.
- [25] D. A. Freedman and P. Diaconis, "On inconsistent  $m$ -estimators," *Ann. Statist.*, vol. 10, no. 2, pp. 454–461, 1982.
- [26] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, "Weighted median filters: a tutorial," *IEEE Trans. Circuits and Systems II*, vol. 43, no. 3, 1996.
- [27] G. R. Arce, J. G. Gonzalez, and P. Zurbach, "Weighted myriad filters in imaging," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, November 1996.
- [28] J. G. Gonzalez, D. W. Griffith, and G. R. Arce, "Matched myriad filtering for robust communications," in *Proc. of the 1996 CISS*, Princeton, New Jersey, 1996.
- [29] J. G. Gonzalez, *Robust techniques for wireless communications in Non Gaussian environments*, Ph.D. thesis, Department of Electrical Engineering, University of Delaware, Newark, 1997.
- [30] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagn. compat.*, vol. EMC-19, pp. 106–127, August 1977.
- [31] H. S. Sichel, "The method of frequency-moments and its application to type VII populations," *Biometrika*, vol. 36, pp. 404–425, 1949.
- [32] S. Kalluri and G. R. Arce, "Adaptive weighted myriad filter algorithms for robust signal processing in  $\alpha$ -stable noise environments," *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 322–334, 1998.
- [33] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, New Jersey, 2nd edition, 1991.
- [34] S. Kalluri and G. R. Arce, "Adaptive weighted myriad filter algorithms for robust signal processing in  $\alpha$ -stable noise environments," *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 322–334, 1998.
- [35] H. Myer and G. Aschied, *Synchronization in Digital Communications*, John Wiley & Sons, New York, 1990.
- [36] G. R. Arce, "A general weighted median filter structure admitting negative weights," *IEEE Trans. Signal Processing*, vol. 46, no. 12, pp. 3195, 1998.

**Juan G. Gonzalez** holds several advanced degrees in electrical engineering and applied mathematics. His doctoral work introduced a novel theory of signal processing for communications over impulsive and bursty channels. For this work, he received the Allan P. Colburn Best Dissertation Prize, given to the best doctoral dissertation in the sciences and engineering at the University of Delaware. He has held several senior research positions in the areas of telecommunications and hardware technology at Bell Laboratories, Lucent Technologies and the University of Delaware. His technical interests include wireless systems technology, communication theory, statistical signal processing, and the interactions between algorithm design and hardware implementation. He received the 1995 Whitaker Best Paper Award from the Rehabilitation Engineering Society of North America. He received the Universidad Pontificia Bolivariana Honor Degree Distinction in 1995. Currently, Dr. Gonzalez spends most of his professional time directing Intellectual Property Systems, LLC, a technology start up dedicated to innovation in the telecommunications and Internet arenas.





**Gonzalo R. Arce** received the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN. He then joined the Department of Electrical and Computer Engineering at the University of Delaware, where he is currently Professor and Chairman. He is a frequent consultant to federal and industrial research centers. His research wireless communications, multimedia communications and security, and electronic imaging. He holds three U.S. patents and is a Fellow of the Center for Advanced Studies at the University of Delaware. Dr Arce has served as Associate Editor of the IEEE Transactions on Signal Processing and as Guest Editor in the IEEE Transactions on Image Processing. He is a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society. He is currently a Guest Editor of Optics Express of the Optical Society of America, and a founding member of the Nonlinear Signal and Image Processing Board. He has served as Chair, Co-Chair, and in the advisory committee of several conferences in nonlinear signal processing. Dr Arce is a Fellow of the IEEE for contribution to the theory and applications to nonlinear signal processing.

