Robust Frequency-Selective Filtering using Weighted Myriad Filters Admitting Real-Valued Weights [1]

Sudhakar Kalluri and Gonzalo R. Arce

Department of Electrical and Computer Engineering
University of Delaware, Newark, Delaware 19716
phone: (302) 831-8030
fax: (302) 831-4316
e-mail: *kalluri@ee.udel.edu* and *arce@ee.udel.edu*

Abstract

Weighted Myriad Smoothers have recently been proposed as a class of nonlinear filters for robust non-Gaussian signal processing in impulsive noise environments. However, weighted myriad smoothers are severely limited, since their weights are restricted to be non-negative. This constraint makes them unusable in bandpass or highpass filtering applications which require negative filter weights. Further, they are incapable of amplifying selected frequency components of an input signal, since the output of a weighted myriad smoother always lies within the dynamic range of its input samples.

In this paper, we generalize the weighted myriad smoother to a richer structure, a *weighted myriad filter* admitting *real-valued* weights. This involves assigning a *pair* of filter weights, one positive and the other negative, to each of the input samples. Equivalently, the filter can be described as a weighted myriad smoother applied to a *transformed* set of samples that includes the original input samples as well as their negatives. The weighted myriad filter is analogous to a normalized linear FIR filter with real-valued weights whose *absolute* values sum to unity. By suitably scaling the output of the weighted myriad filter, we extend it to yield the so-called *scaled weighted myriad filter*, which includes (but is more powerful than) the traditional unconstrained linear FIR filter. Finally, we derive stochastic gradient-based nonlinear adaptive algorithms for the optimization of these novel myriad filters under the mean square error criterion.

Submitted to the IEEE Transactions on Signal Processing

EDICS Paper Category: SP 2.1.5 (Rank-Order and Median Filters)

[1]This research was funded in part by the National Science Foundation under Grant MIP-9530923.

1 Introduction

The statistical signal processing literature has traditionally been dominated by linear estimators, which are optimal under the assumption of the Gaussian model for the signal statistics. However, a large number of real-world processes (including radar clutter, ocean acoustic noise, and multiple-access interference in wireless communication systems) have been found to be *impulsive* in nature, with sharp spikes or occasional outliers present in the data [1, 2]. Such processes are more accurately modeled by distributions having heavier-than-Gaussian tails in their density functions [3]. There is, therefore, a need for the development of *robust nonlinear* estimators for impulsive environments, based on heavy-tailed non-Gaussian distributions for the underlying signals.

The class of M-estimators (Maximum-Likelihood type estimators) of location, developed in the theory of robust statistics [4], has been of fundamental importance in the development of robust signal processing techniques [5]. Given a set of observations (input samples) $\{x_i\}_{i=1}^N$, an M-estimate of their common location is given by

$$\hat{\theta} \stackrel{\triangle}{=} \arg\min_{\theta} \sum_{i=1}^{N} \rho\left(x_i - \theta\right), \tag{1}$$

where $\rho(\cdot)$ is called the *cost function* of the M-estimator. Maximum-Likelihood location estimates form a special type of M-estimators, with the observations being independent and identically distributed (i.i.d) and $\rho(u) \sim -\log f(u)$, where $f(u)$ is the common density function of the input samples. The linear, *weighted median* and the recently proposed *weighted myriad* smoother families, are all derived from Maximum-Likelihood location estimators. Weighted median smoothers, and other nonlinear smoothers based on *order statistics* [6, 7], are derived from the heavy-tailed Laplacian noise model, and are extensively used in robust image processing applications. *Weighted Myriad Smoothers* [8, 9, 10, 11] have been developed based on the so-called α-stable distributions [2, 3], which are more heavy-tailed than the Gaussian as well as the Laplacian distributions, while including the Gaussian distribution as a special limiting case. Myriad smoothers have been successfully utilized in several

applications in robust communications and image processing [12, 13, 14].

The fundamental M-estimators that generate the linear, weighted median and weighted myriad smoother families are the *sample mean*, *sample median* and *sample myriad*, respectively. Using the Gaussian and Laplacian density functions in Maximum-Likelihood location estimation, we obtain the cost functions for the sample mean and the sample median as $\rho(u) = u^2$ and $\rho(u) = |u|$, respectively. The sample myriad [15, 8, 9, 10] is defined using the Cauchy density function, which is the only symmetric (non-Gaussian) α-stable density function that is expressible in closed-form. The cost function thus obtained is $\rho(u) = \log(K^2+u^2)$, where the parameter K controls the robustness of the estimator; a more detailed description is given in Section 2. The first three rows of Table 1 show the cost functions for the sample mean, median and myriad estimators. From a signal processing point of view, these *plain* location M-estimators, using i.i.d observations, do not adequately capture the statistical relationships among the different samples in an input signal. A proper extension of these estimators would be to consider multivariate density functions for the input samples: $\{X_i\}_{i=1}^N \sim f_{\mathbf{X}}(\mathbf{x};\theta) = f_{\mathbf{X}}(x_1-\theta, x_2-\theta, \ldots, x_N-\theta)$. An alternative approach that is more tractable is to consider samples that are independent, but *not* identically distributed, with common location θ, but varying scale factors S_i: $X_i \sim f(x_i-\theta; S_i)$. Using this approach, Maximum-Likelihood location estimators are generalized by introducing *non-negative* weights $\{w_i \geq 0\}_{i=1}^N$ in their cost function expressions; the weight w_i is related to the scale factor S_i, and the varying scales reflect the non-uniform *reliabilities* of the input samples. By extending this principle to the more general M-estimator of (1), we obtain a *weighted* M-estimator:

$$\hat{\theta} \stackrel{\triangle}{=} \arg\min_{\theta} \sum_{i=1}^{N} \rho\left(\sqrt{w_i}\,(x_i-\theta)\right). \tag{2}$$

For reasons to be described shortly, these estimators are also referred to as M-smoothers. The last three rows of Table 1 show the different M-smoothers obtained using the cost functions for the linear, median and myriad smoother families, respectively. The notations $w_i \diamond x_i$ (for the weighted median) and $w_i \circ x_i$ (for the weighted myriad), shown in the last

Smoother	**Cost Function**	**Output $\hat{\theta}$**
Linear	$\sum_{i=1}^{N} (x_i - \theta)^2$	mean $(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{N} x_i \ / \ N$
Median	$\sum_{i=1}^{N} \lvert x_i - \theta \rvert$	median $(x_1, x_2, \ldots, x_N)$
Myriad	$\sum_{i=1}^{N} \log \left[K^2 + (x_i - \theta)^2\right]$	myriad $(x_1, x_2, \ldots, x_N; K)$
Weighted Mean	$\sum_{i=1}^{N} w_i (x_i - \theta)^2$	mean $(w_i \cdot x_i)\lvert_{i=1}^{N} = \sum_{i=1}^{N} w_i x_i \ / \ \sum_{i=1}^{N} w_i$
Weighted Median	$\sum_{i=1}^{N} w_i' \lvert x_i - \theta \rvert; \ w_i' \stackrel{\triangle}{=} \sqrt{w_i}$	median $(w_i' \diamond x_i)\lvert_{i=1}^{N}$
Weighted Myriad	$\sum_{i=1}^{N} \log \left[K^2 + w_i (x_i - \theta)^2\right]$	myriad $(w_i \circ x_i; K)\lvert_{i=1}^{N}$

Table 1: M-estimator and M-smoother (weighted M-estimator) cost functions and outputs for various filter families.

column of the table, reflect the weighting operation in (2).

In spite of their proven robustness properties, weighted M-estimators in general, and weighted median and myriad smoothers in particular, are severely limited in their potential for signal processing and communications applications. This is directly attributable to the constraint of *non-negative weights*, which makes these estimators no more than *smoothers* or "lowpass" type filters; this is precisely why they are called 'M-smoothers'. M-smoothers are thus unusable for bandpass or highpass filtering applications which require negative filter weights. Further, they are incapable of arbitrary amplification of specified frequency components of an input signal, since the output of an M-smoother is restricted to the dynamic range of its input samples. That the weighted myriad smoother in particular is severely constrained is clear from the fact that this filter is analogous to the *weighted mean smoother*, which is a normalized linear FIR filter with non-negative weights summing to unity. It is evident that such a severely handicapped linear filter would be quite useless in most signal processing applications. In order to overcome the limitations of the weighted myriad smoother, it needs to be generalized into a robust filter structure that possesses the full signal processing power of the unconstrained linear FIR filter (with unnormalized and

real-valued weights), while performing efficiently in impulsive environments.

Recently, the weighted median smoother has been successfully generalized to yield a *weighted median filter* structure that admits real-valued weights [16, 17]. This new filter is derived by assigning a *pair* of filter weights, one of them positive and the other negative, to each input sample. In the present paper, we first generalize this approach by extending the class of M-smoothers defined in (2), leading to a new class of *M-filters* that allow for real-valued weights. We then focus on the special case of myriad estimators, generalizing the weighted myriad smoother into a *Weighted Myriad Filter* structure that admits real-valued weights. This filter is analogous to the weighted mean *filter* with real-valued weights. However, the weighted myriad filter cannot achieve arbitrary amplification of selected frequency components of an input signal, since its output is bounded in magnitude by the maximum absolute value of the input samples. By appropriately scaling (multiplying) the filter output, we further extend this structure to yield the so-called *Scaled Weighted Myriad Filters*, which include as a special case the traditional unconstrained linear FIR filter. Myriad filters thus provide for a *robust generalization* of linear signal processing, outperforming linear filters in a variety of applications in impulsive environments.

The rest of the paper is organized as follows. Section 2 introduces the weighted myriad smoother. The class of myriad filters admitting real-valued weights is developed in Section 3. Having defined these filter structures, the important issue of optimization of their parameters (weights) is addressed in Section 4. The optimization of weighted myriad smoothers was considered in [14]. In the present paper, we derive necessary conditions for optimality of the different myriad filters under the mean square error (MSE) criterion, and develop stochastic gradient-based (LMS-type) nonlinear adaptive algorithms for the optimization of the filter parameters. The performance of the different adaptive myriad filters is demonstrated in Section 5 through computer simulations of frequency-selective filtering in impulsive noise environments.

2 Weighted Myriad Smoothers

This section gives a brief introduction to the weighted myriad smoother. For a more detailed treatment, see [8, 9, 10, 11, 14]. It should be noted that in all the previous papers on myriad filtering, the weighted myriad smoother has been referred to as the weighted myriad *filter*. To avoid confusion, we shall henceforth reserve the term *weighted myriad filter* for the more general case when *real-valued* weights are admissible, while the *smoother* itself is restricted to have non-negative weights.

Weighted myriad smoothers are derived from the *sample myriad*, which is defined as the Maximum-Likelihood estimate of location of data following the Cauchy distribution. As mentioned in Section 1, myriad filters are motivated by the properties of α-stable distributions [3], of which the Gaussian and Cauchy distributions are special cases. For simplicity and tractability, the myriad has been defined using the Cauchy distribution, the only (non-Gaussian) α-stable distribution that has a closed-form expression for its density function. However, the myriad in fact turns out to be the optimal estimator of the location of α-stable distributions for the triplet $\alpha = 0, 1, 2$ [9, 10]. Consider now a set of N independent and identically distributed (i.i.d.) random variables $\{X_i\}_{i=1}^N$, each following a Cauchy distribution with location parameter θ and scaling factor $K > 0$. Thus, $X_i \sim \text{Cauchy}(\theta, K)$, with the density function

$$f_{X_i}(x_i; \theta, K) \;=\; \frac{K}{\pi} \cdot \frac{1}{K^2 + (x_i - \theta)^2} \;=\; \frac{1}{K}\, f\left(\frac{x_i - \theta}{K}\right), \tag{3}$$

where $f(v) \stackrel{\triangle}{=} \dfrac{1}{\pi} \cdot \dfrac{1}{1 + v^2}$ is the density function of a *standard* Cauchy random variable: $\dfrac{X_i - \theta}{K} \sim \text{Cauchy}(0, 1)$. Given a set of observations $\{x_i\}_{i=1}^N$, the sample myriad $\hat{\theta}_K$ maximizes the likelihood function $\prod_{i=1}^N f_{X_i}(x_i; \theta, K)$. Using (3) and some manipulation, we can write

$$\begin{aligned} \hat{\theta}_K \;&\stackrel{\triangle}{=}\; \text{myriad}\,(x_1, x_2, \ldots, x_N; K) \;=\; \arg\min_{\theta} \prod_{i=1}^{N} \left[1 + \left(\frac{x_i - \theta}{K}\right)^2\right] \\ &=\; \arg\min_{\theta} \sum_{i=1}^{N} \log\left[K^2 \;+\; (x_i - \theta)^2\right], \end{aligned} \tag{4}$$

using the fact that $\log(\cdot)$ is a strictly increasing function. Comparing (4) with the definition of M-estimators in (1), we see that the M-estimator cost function for the sample myriad is given by $\rho(u) = \log(K^2 + u^2)$. It is interesting to note that the sample myriad includes the sample mean as the special limiting case when $K \to \infty$ [9]: $\lim_{K\to\infty} \hat{\theta}_K = \sum_{i=1}^{N} x_i \,/\, N$.

The sample myriad can be generalized to the weighted myriad smoother by assigning non-negative weights to the input samples (observations); the weights reflect the varying levels of "reliability". To this end, the observations are assumed to be drawn from independent Cauchy random variables which are, however, *not* identically distributed. Given N observations $\{x_i\}_{i=1}^N$ and non-negative 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 θ, but varying scale factors $\{S_i\}_{i=1}^N$: $X_i \sim \text{Cauchy}(\theta, S_i)$, where

$$S_i \stackrel{\triangle}{=} \frac{K}{\sqrt{w_i}} > 0,\ i = 1, 2, \ldots, N. \tag{5}$$

A larger value for the weight w_i (smaller scale S_i) makes the distribution of X_i more concentrated around θ, thus increasing the reliability of the sample x_i. Note that the special case when all the weights are equal to unity corresponds to the sample myriad at the nominal scale factor K, with all the scale factors reducing to $S_i = K$.

It is important to realize that the location estimation problem being considered here is intimately related to the problem of filtering a time-series $\{x(n)\}$ using a sliding window. The output $y(n)$, at time n, can be interpreted as an estimate of location based on the input samples $\{x(n - N_1), \ldots, x(n-1), x(n), x(n+1), \ldots, x(n+N_2)\}$. Further, the aforementioned model of independent but not identically distributed samples can capture the temporal relationships usually present among the input samples. To see this, note that the output $y(n)$, as an estimate of location, would rely more on (give more weight to) the sample $x(n)$, when compared with samples that are further away in time. By assigning varying scale factors in modeling the input samples, leading to different weights (reliabilities), their

temporal correlations can be effectively accounted for.

The weighted myriad smoother output $\hat{\theta}_K(\mathbf{w}, \mathbf{x})$ can now be defined as the value of θ that maximizes the likelihood function $\prod_{i=1}^{N} f_{X_i}(x_i; \theta, S_i)$. Using (3) for $f_{X_i}(x_i; \theta, S_i)$ leads to

$$\begin{aligned} \hat{\theta}_K(\mathbf{w}, \mathbf{x}) &\triangleq \text{myriad}\,(w_i \circ x_i; K)|_{i=1}^{N} = \arg\min_{\theta} \prod_{i=1}^{N} \left[1 + \left(\frac{x_i - \theta}{S_i}\right)^2\right] \\ &= \arg\min_{\theta} \prod_{i=1}^{N} \left[K^2 + w_i\,(x_i - \theta)^2\right] \triangleq \arg\min_{\theta} P(\theta); \end{aligned} \tag{6}$$

the notation $w_i \circ x_i$ denotes the weighting operation in the above expressions. Alternatively, we can write $\hat{\theta}_K(\mathbf{w}, \mathbf{x}) \equiv \hat{\theta}$ as

$$\hat{\theta} = \arg\min_{\theta} Q(\theta) \triangleq \arg\min_{\theta} \sum_{i=1}^{N} \log\left[K^2 + w_i\,(x_i - \theta)^2\right]; \tag{7}$$

thus $\hat{\theta}$ is the global minimizer of $P(\theta)$ as well as of $Q(\theta) \triangleq \log(P(\theta))$. Therefore, depending on the context, we refer to either of the functions $P(\theta)$ and $Q(\theta)$ as the *weighted myriad smoother objective function*. Note that when $w_i = 0$, the corresponding term drops out of $P(\theta)$ and $Q(\theta)$; thus a sample x_i is effectively ignored if its weight is zero. Now, by comparing (7) and (2), we confirm that the weighted myriad smoother is a weighted M-estimator (or an M-smoother) with the cost function $\rho(u) = \log(K^2 + u^2)$.

As we can see from (6), the objective function $P(\theta)$ is a polynomial in θ of degree $2N$, with well-defined derivatives of all orders. Therefore, $P(\theta)$ (and the equivalent objective function $Q(\theta)$) can have several local minima. This makes the exact computation of $\hat{\theta}$ (by direct minimization of $P(\theta)$ or $Q(\theta)$) a prohibitively expensive task. In [11], we have developed fast algorithms for the computation of the weighted myriad smoother output using an indirect method with a high degree of accuracy.

In order to determine the properties of $\hat{\theta}$, we examine the objective function further, focusing on $Q(\theta)$. Using (6) and (7), the derivative of $Q(\theta)$ can be written as

$$Q'(\theta) = \frac{P'(\theta)}{P(\theta)} = 2 \sum_{i=1}^{N} \frac{w_i\,(\theta - x_i)}{K^2 + w_i\,(x_i - \theta)^2}. \tag{8}$$

We now briefly present a few basic properties of $Q(\theta)$ and $\hat{\theta}$ that will be useful in later sections of the paper. The proofs of these properties can be found in [11, 14]. First, let

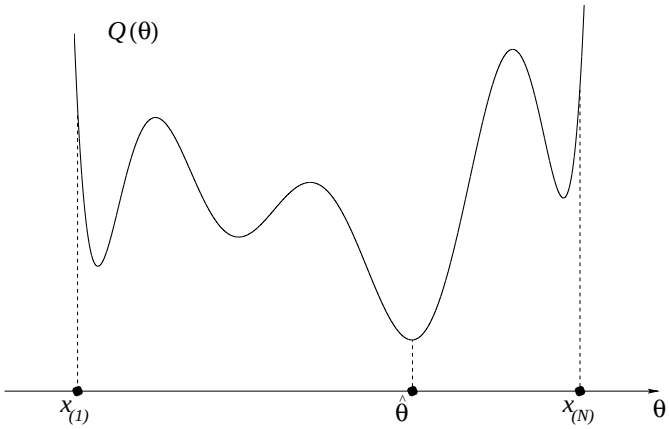


Figure 1: Sketch of a typical weighted myriad objective function $Q(\theta)$.

$\{x_{(j)}\}_{j=1}^{N}$ denote the order statistics (samples sorted in increasing order of amplitude) of the input vector $\mathbf{x}$, with $x_{(1)}$ the smallest and $x_{(N)}$ the largest. The following statements are true:

(a) The weighted myriad smoother has only N independent parameters; the output $\hat{\theta}_K(\mathbf{w}, \mathbf{x})$ depends only on the normalized weight vector $(\mathbf{w}/K^2)$.

(b) The objective function $Q(\theta)$ has a finite number (at most $(2N-1)$) of local extrema, all of which lie within the range $[x_{(1)}, x_{(N)}]$ of the input samples.

(c) The weighted myriad smoother output $\hat{\theta}$ is one of the local minima of $Q(\theta)$:

$$Q'(\hat{\theta}) = 0. \tag{9}$$

(d) $\hat{\theta}$ is always within the range of the input samples: $x_{(1)} \leq \hat{\theta} \leq x_{(N)}$.

Some of the afore-mentioned properties are illustrated by Fig. 1, which shows a sketch of a typical objective function $Q(\theta)$. It is clear from the figure that the output $\hat{\theta}$ is restricted to the dynamic range of the input samples. This is a direct consequence of the constraint of non-negative weights. As a result, the weighted myriad smoother (and in general, any M-smoother) is incapable of highpass or bandpass-type behavior and unable to amplify the dynamic range of an input signal. The severely constrained nature of the weighted myriad

smoother can also be seen by considering the limiting case when $K \to \infty$, with the weights $\{w_i\}$ held constant. Using (8) and (9), we can show that as $K \to \infty$, $Q(\theta)$ reduces to having a single local extremum, and

$$\lim_{K\to\infty} \hat{\theta}_K(\mathbf{w}, \mathbf{x}) \;=\; \hat{\theta}_\infty(\mathbf{w}, \mathbf{x}) \;=\; \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}, \tag{10}$$

which is the (linear) *weighted mean smoother*; K is called the 'linearity parameter' precisely because of this limiting result. Thus, the weighted myriad smoother is analogous to the weighted mean smoother, which is a severely handicapped linear FIR filter, since its weights are constrained to be non-negative, while also summing to unity.

3 Myriad Filters with Real-Valued Weights

In this section, we first extend the class of M-smoothers (weighted M-estimators) defined in (2), leading to the more general class of M-filters that allow for negative weights. We then consider the special case of the family of myriad M-estimators, generalizing the weighted myriad smoother of Section 2 into a variety of myriad filters admitting real-valued weights.

3.1 M-filters

Before we can extend the general class of M-smoothers, it is useful to examine the special cases of the weighted mean and median smoothers. Referring to Table 1, the output of a weighted mean smoother, with input vector $\mathbf{x}$ and a weight vector $\mathbf{w}$ of non-negative weights, is given by

$$\hat{\theta}\,(\mathbf{w}, \mathbf{x}) \;=\; \text{mean}\,(w_i \cdot x_i)|_{i=1}^{N} \;=\; \sum_{i=1}^{N} w_i x_i \;/\; \sum_{i=1}^{N} w_i. \tag{11}$$

Notice from Table 1 that $\hat{\theta}$ minimizes the cost function $\sum_{i=1}^{N} w_i (x_i - \theta)^2$. When the weights are allowed to be both positive and negative, $\{w_i \in \mathcal{R}\}_{i=1}^{N}$, a natural extension of (11) would be to write

$$\tilde{\theta}\,(\mathbf{w}, \mathbf{x}) \;\triangleq\; \sum_{i=1}^{N} w_i x_i \;/\; \sum_{i=1}^{N} |w_i|, \tag{12}$$

for the *weighted mean filter* output. Note that the normalization factor in (12) is now the sum of the *absolute values* of the filter weights. Now, we can rewrite (11) as

$$\tilde{\theta}\,(\mathbf{w},\mathbf{x}) \;=\; \sum_{i=1}^{N} |w_i| \cdot \mathrm{sgn}(w_i)x_i \;/\; \sum_{i=1}^{N} |w_i|; \tag{13}$$

thus the sign of the weight is uncoupled from its magnitude and attached to the corresponding input sample. By comparing with (11), we can rewrite (13) as

$$\tilde{\theta}\,(\mathbf{w},\mathbf{x}) \;=\; \mathrm{mean}\,(|w_i| \cdot \mathrm{sgn}(w_i)x_i)|_{i=1}^{N} \;\stackrel{\triangle}{=}\; \mathrm{mean}\,(g_i \cdot z_i)|_{i=1}^{N} \;=\; \hat{\theta}\,(\mathbf{g},\mathbf{z}), \tag{14}$$

where $\{g_i = |w_i|\}_{i=1}^{N}$ and $\{z_i = \mathrm{sgn}(w_i)x_i\}_{i=1}^{N}$ are represented by the vectors $\mathbf{g}$ and $\mathbf{z}$, respectively. The weighted mean filter can therefore be thought of as a weighted mean *smoother* applied to a modified set of samples $\{z_i\}_{i=1}^{N}$, using the non-negative weights $\{g_i \geq 0\}_{i=1}^{N}$. Referring to the cost function for the weighted mean smoother in Table 1, we can infer that the weighted mean *filter* output $\tilde{\theta}\,(\mathbf{w},\mathbf{x})$ minimizes the cost function $\sum_{i=1}^{N} |w_i| \cdot (\mathrm{sgn}(w_i)x_i - \theta)^2$.

Following the above approach, the weighted median smoother was recently extended to a class of *weighted median filters* with real-valued weights [16]. Thus, referring to Table 1, and by analogy to (14), the output of a weighted median filter with weights $\{w_i \in \mathcal{R}\}_{i=1}^{N}$ was written in [16] as

$$\bar{\theta}\,(\mathbf{w},\mathbf{x}) \;=\; \mathrm{median}\,(|w_i| \diamond \mathrm{sgn}(w_i)x_i)|_{i=1}^{N} \;=\; \arg\min_{\theta} \sum_{i=1}^{N} |w_i| \cdot |\mathrm{sgn}(w_i)x_i - \theta|. \tag{15}$$

However, it was subsequently found in [17] that a *more* general definition of the weighted median filter was possible. Notice in (15) that if a particular weight w_i is negative, the corresponding *modified* sample is $\mathrm{sgn}(w_i)x_i = -x_i$, which is the *mirror* sample of x_i. Thus, depending on the signs of the weights, *some* of the mirror samples $\{-x_i\}$ come into play in the filtering operation. A more general approach might then be to include *all* the mirror samples $\{-x_i\}_{i=1}^{N}$ in the filter definition. An alternative way to achieve this is to assign a *pair* of weights $w_i \geq 0$ and $h_i \leq 0$ to each sample x_i. As before, we uncouple the sign of a weight whenever it is negative, merging the sign with the corresponding sample. Thus, since $h_i \leq 0$, we have the equivalence $h_i \diamond x_i \;=\; |h_i| \diamond \mathrm{sgn}(h_i)x_i \;=\; |h_i| \diamond (-x_i)$. This concept of

double weighting emerges naturally in [17] through the analysis of the so-called stack filters with real-valued weights (weighted median filters are a special case of stack filters). Thus, given the set of weights $\{w_i \geq 0\}_{i=1}^{N}$ and $\{h_i \leq 0\}_{i=1}^{N}$, the weighted median filter output [17] is defined as

$$\bar{\theta} \stackrel{\triangle}{=} \text{median}\left(w_i \diamond x_i, |h_i| \diamond (-x_i)\right)|_{i=1}^{N} = \text{median}\left(\langle w_i, h_i \rangle \diamond x_i\right)|_{i=1}^{N}, \tag{16}$$

where the notation $\langle w_i, h_i \rangle$ reflects the double weight assigned to the sample x_i. Note that the weighted median filter can also be described as a weighted median smoother applied to a modified input vector $\mathbf{z} = [\mathbf{x}^T, -\mathbf{x}^T]^T$ and a modified weight vector of non-negative weights, $\mathbf{g} = [\mathbf{w}^T, |\mathbf{h}|^T]^T$. Referring to the weighted median smoother cost function in Table 1, we can show that the weighted median filter output minimizes the cost function $\sum_{i=1}^{N} w_i \cdot |x_i - \theta| + \sum_{i=1}^{N} |h_i| \cdot |-x_i - \theta|$.

At first glance, the double weighting in (16) seems redundant and the structure in (15), with a single real-valued weight for each sample, appears to be sufficient as in the case of linear FIR filters. However, the peculiarity of redundancy exists only in the case of the linear filter; the weighted mean filter analogous to (16) is given by $\tilde{\theta} = \text{mean}\left(w_i \cdot x_i, |h_i| \cdot (-x_i)\right)|_{i=1}^{N} = \text{mean}\left((w_i - |h_i|) \cdot x_i\right)|_{i=1}^{N}$, the double weight $\langle w_i, h_i \rangle$ thus collapsing to a single real-valued weight $(w_i - |h_i|)$. This conversion into a single weight, due to the superposition property of linear filters, does not occur for the weighted median filter [17], and cannot be expected for the more general class of M-filters to be developed in the following.

Using the afore-mentioned ideas, the extension of the general class of M-smoothers, to allow for real-valued weights, can be achieved in a straightforward manner as follows. The output of the general M-smoother, with weights $\{w_i \geq 0\}_{i=1}^{N}$ and cost function $\rho(\cdot)$, is given from (2) by

$$\hat{\theta}\left(\mathbf{w}, \mathbf{x}\right) = \arg\min_{\theta} \sum_{i=1}^{N} \rho\left(\sqrt{w_i} \cdot (x_i - \theta)\right). \tag{17}$$

Given a set of positive and negative weights, $\{w_i \geq 0\}_{i=1}^{N}$ and $\{h_i \leq 0\}_{i=1}^{N}$, the M-smoother

is generalized to the new class of M-filters described by the filter output

$$\tilde{\theta}\left(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}\right) \stackrel{\triangle}{=} \arg\min_{\theta} \sum_{i=1}^{N} \left[\rho\left(\sqrt{w_i} \cdot (x_i - \theta)\right) + \rho\left(\sqrt{|h_i|} \cdot (-x_i - \theta)\right)\right], \tag{18}$$

where the notation $\langle \mathbf{w}, \mathbf{h} \rangle$ describes the double weight vector with positive weights in $\mathbf{w}$ and negative weights in $\mathbf{h}$. Note that the M-filter $\tilde{\theta}\left(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}\right)$ can also be described as an M-smoother $\hat{\theta}\left(\mathbf{g}, \mathbf{z}\right)$, with a transformed input vector $\mathbf{z} = [\mathbf{x}^T, -\mathbf{x}^T]^T$ and a transformed weight vector of non-negative weights, $\mathbf{g} = [\mathbf{w}^T, |\mathbf{h}|^T]^T$. The rest of this section is devoted to deriving various *myriad filters* from (18).

3.2 Weighted Myriad Filters

Recall from (7) that the output of the weighted myriad smoother with weights $\{w_i \geq 0\}_{i=1}^{N}$ is given by

$$\hat{\theta}_K(\mathbf{w}, \mathbf{x}) = \arg\min_{\theta} Q(\theta) \stackrel{\triangle}{=} \arg\min_{\theta} \sum_{i=1}^{N} \log\left[K^2 + w_i (x_i - \theta)^2\right]. \tag{19}$$

As mentioned in Section 2, the weighted myriad smoother is an M-smoother with the cost function $\rho(u) = \log(K^2 + u^2)$. Using this cost function in the expression for the M-filter output given in (18), we obtain the output of the *weighted myriad filter*, with weights $\{w_i \geq 0\}_{i=1}^{N}$ and $\{h_i \leq 0\}_{i=1}^{N}$, as

$$\begin{aligned} \tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}) &\stackrel{\triangle}{=} \text{myriad}\left(\langle w_i, h_i \rangle \circ x_i; K\right)|_{i=1}^{N} = \text{myriad}\left(w_i \circ x_i, |h_i| \circ (-x_i); K\right)|_{i=1}^{N} \\ &= \arg\min_{\theta} Q_1(\theta), \end{aligned} \tag{20}$$

where the *weighted myriad filter objective function* $Q_1(\theta)$ is given by

$$Q_1(\theta) \stackrel{\triangle}{=} \sum_{i=1}^{N} \left\{\log\left[K^2 + w_i \cdot (x_i - \theta)^2\right] + \log\left[K^2 + |h_i| \cdot (-x_i - \theta)^2\right]\right\}. \tag{21}$$

The weighted myriad filter can be represented as a weighted myriad smoother $\hat{\theta}_K(\mathbf{g}, \mathbf{z})$ with the transformed input and weight vectors $\mathbf{z} = [x_1, x_2, \ldots, x_N, -x_1, -x_2, \ldots, -x_N]^T$ and $\mathbf{g} = [w_1, w_2, \ldots, w_N, |h_1|, |h_2, \ldots, |h_N]^T$, respectively. Consequently, many of the properties of this filter are similar to those of the weighted myriad smoother that were described in

Section 2. Now, let $\{z_{(j)}\}_{j=1}^{2N}$ denote the order statistics (sorted samples) of the transformed input vector $\mathbf{z}$; the smallest and largest order statistics are then given by $z_{(1)} = -M_x = -\max\{|x_i|\}_{i=1}^{N}$ and $z_{(2N)} = +M_x$. The following properties of the filter output $\tilde{\theta}$ and the objective function $Q_1(\theta)$ follow easily from Section 2:

(a) The weighted myriad filter has only $2N$ independent parameters; the filter output $\tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ depends only on the normalized double weight vector $\langle \mathbf{w}/K^2, \mathbf{h}/K^2 \rangle$.

(b) The derivative of the objective function $Q_1(\theta)$ is given by

$$Q_1'(\theta) \;=\; 2 \sum_{i=1}^{N} \left[\frac{w_i \,(\theta - x_i)}{K^2 \;+\; w_i \,(x_i - \theta)^2} \;+\; \frac{|h_i| \,(\theta + x_i)}{K^2 \;+\; |h_i| \,(x_i + \theta)^2} \right]. \tag{22}$$

(c) $Q_1(\theta)$ has not more than $(4N - 1)$ local extrema, all of which lie within the range $[-M_x, +M_x]$ of the *magnitudes* of the input samples.

(d) $\tilde{\theta}$ is one of the local minima of $Q_1(\theta)$:

$$Q_1'(\tilde{\theta}) = 0. \tag{23}$$

(e) $\tilde{\theta}$ is always within the range of the magnitudes of the input samples: $-M_x \leq \tilde{\theta} \leq +M_x$.

Since the weighted myriad filter allows for positive as well as negative weights, it overcomes the limitations of the weighted myriad smoother to the extent that it can accomplish a wide range of frequency-selective filtering operations, including bandpass and highpass filtering. Unlike the weighted myriad smoother output $\hat{\theta}_K(\mathbf{w}, \mathbf{x})$ of (19), which is limited to the range $[x_{(1)}, x_{(N)}]$ of the input samples, the weighted myriad filter output $\tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ is confined to a larger interval $[-M_x, +M_x]$, $M_x = \max\{|x_i|\}_{i=1}^{N}$, which includes the input samples as well as their negatives. However, the fact that the output $\tilde{\theta}$ is restricted to *some* finite interval that depends on the input samples, is in itself a limitation of this filter. Like the weighted myriad smoother, the weighted myriad filter is also incapable of amplifying the dynamic range of an input signal, since the output cannot go beyond the dynamic range of the *magnitudes* of the input samples. To see the constrained nature of this filter, consider

the limiting case when the linearity parameter $K \to \infty$, while holding the weights $\mathbf{w}$ and $\mathbf{h}$ constant. Using (22) and (23), we can show that as $K \to \infty$, $Q_1(\theta)$ has a single local minimum, and

$$\lim_{K\to\infty} \tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}) \;=\; \tilde{\theta}_\infty(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}) \;=\; \frac{\sum_{i=1}^{N} (w_i - |h_i|) \cdot x_i}{\sum_{i=1}^{N} (w_i + |h_i|)}. \tag{24}$$

We can also write the above expression as

$$\tilde{\theta}_\infty(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}) \;=\; \hat{\theta}_\infty(\mathbf{g}, \mathbf{z}) \;=\; \sum_{i=1}^{2N} g_i z_i \;/\; \sum_{i=1}^{2N} g_i, \tag{25}$$

where we have used the modified inputs and weights, $\mathbf{z} = [x_1, \ldots, x_N, -x_1, \ldots, -x_N]^T$ and $\mathbf{g} = [w_1, \ldots, w_N, |h_1|, \ldots, |h_N]^T$, respectively. The restricted nature of the filter is due to the presence of the normalization factor $\sum_{i=1}^{2N} g_i = \sum_{i=1}^{N} (w_i + |h_i|)$ in the above expressions. As (24) shows, the weighted myriad filter is analogous to nothing more than a linear FIR filter with *constrained* real-valued weights.

As a final note on weighted myriad filters, we shall case consider a special type of weighted myriad filter which has only N weights, instead of the general case of $2N$ weights. This special case is derived by analogy with the weighted mean filter of (14) and the special weighted median filter (15), proposed in [16]. We define a weighted myriad filter *without* double weighting, with real-valued weights $\{w'_i \in \mathcal{R}\}_{i=1}^{N}$, as

$$\bar{\theta}_K(\mathbf{w}', \mathbf{x}) \;\stackrel{\triangle}{=}\; \text{myriad}\,(|w'_i| \circ \text{sgn}(w'_i) x_i)|_{i=1}^{N} \;=\; \arg\min_{\theta}\, Q_2(\theta), \tag{26}$$

where

$$Q_2(\theta) \;\stackrel{\triangle}{=}\; \sum_{i=1}^{N} \log\left[K^2 \;+\; |w'_i| \cdot (\text{sgn}(w'_i) x_i - \theta)^2\right] \tag{27}$$

is the objective function of this special weighted myriad filter. Although this filter does not possess the full power of the general double-weighted filter, it is simpler to design and implement, and it might be sufficient in some applications. To see how the filter of (26) is in fact a special case of the double-weighted myriad filter of (20), consider the general

weighted myriad filter output $\tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ with $\{w_i \geq 0\}_{i=1}^N$ and $\{h_i \leq 0\}_{i=1}^N$, and let $\{w_i' \stackrel{\triangle}{=} w_i + h_i = w_i - |h_i|\}_{i=1}^N$, with the added constraint that $\{w_i \cdot h_i = 0\}_{i=1}^N$. Thus, only one of w_i and h_i is allowed to be non-zero, and w_i' is equal to that non-zero weight. It is easy to show, using (20) and (21) together with the above constraints, that the general filter output $\tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ reduces to $\bar{\theta}_K(\mathbf{w}', \mathbf{x})$ of (26). Now, by comparing (26) and (27) with (19), we see that $\bar{\theta}_K(\mathbf{w}', \mathbf{x})$ can also be expressed as the output of a weighted myriad smoother $\hat{\theta}_K(\mathbf{g}', \mathbf{z}')$, with weights $\{g_i' = |w_i'|\}_{i=1}^N$ and transformed input samples $\{z_i' = \mathrm{sgn}(w_i')x_i\}_{i=1}^N$. The filter output $\bar{\theta}_K$ is seen to be always confined to the interval $[z_{(1)}', z_{(N)}']$, the range of the transformed input samples. Considering the limiting case $K \to \infty$, we can use (24) and the relation between $\mathbf{w}'$ and $\langle \mathbf{w}, \mathbf{h} \rangle$ to show that

$$\lim_{K \to \infty} \bar{\theta}_K(\mathbf{w}', \mathbf{x}) = \bar{\theta}_\infty(\mathbf{w}', \mathbf{x}) = \sum_{i=1}^N w_i' \, x_i \; / \; \sum_{i=1}^N |w_i'|, \tag{28}$$

which is a weighted mean filter with real-valued weights.

3.3 Scaled Weighted Myriad Filters

The weighted myriad filters described in Section 3.2 have the drawback that the output is always constrained to belong to a certain *finite* interval that is a function of the input samples. These filters are therefore not able to amplify the dynamic range of the input signal. Consider the weighted myriad filter output $\tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ of (20), which always lies within the interval $[-M_x, +M_x]$, where $M_x = \max\{|x_i|\}_{i=1}^N$. In order to remove this constraint on $\tilde{\theta}_K$, we examine its limiting case given in (24): $\tilde{\theta}_\infty = \sum_{i=1}^N (w_i - |h_i|) \cdot x_i \; / \; \sum_{i=1}^N (w_i + |h_i|)$. Note that if we removed the normalization factor $S \equiv S(\langle \mathbf{w}, \mathbf{h} \rangle) \stackrel{\triangle}{=} \sum_{i=1}^N (w_i + |h_i|)$ from the expression for $\tilde{\theta}_\infty$, we would obtain the full power of the linear FIR filter in the expression $\sum_{i=1}^N (w_i - |h_i|) \cdot x_i$. Extending this idea to the case of a *finite* value of K, we define the output of the so-called *scaled weighted myriad filter*, with weights $\{w_i \geq 0\}_{i=1}^N$ and $\{h_i \leq 0\}_{i=1}^N$, as

$$\begin{aligned} \tilde{\theta}_K^{(S)}(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}) &\stackrel{\triangle}{=} S(\langle \mathbf{w}, \mathbf{h} \rangle) \cdot \tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}) \\ &= \left[\sum_{i=1}^N (w_i + |h_i|) \right] \cdot \mathrm{myriad}\left(\langle w_i, h_i \rangle \circ x_i; K \right)|_{i=1}^N, \end{aligned} \tag{29}$$

where the weighted myriad filter output $\tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ is given by (20). Now, since we know that $\tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ lies within the interval $[-M_x, +M_x]$, we see that the scaled weighted myriad filter output is constrained by $|\tilde{\theta}_K^{(S)}(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})| \leq S(\langle \mathbf{w}, \mathbf{h} \rangle) \cdot M_x$. However, this interval can be made as large as desired, by simply increasing the magnitudes of the filter weights (thus increasing the factor S). Thus, the scaled filters are capable of input signal amplification, unlike the weighted myriad filters of Section 3.2. Further, since $\tilde{\theta}_\infty^{(S)}(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}) = \sum_{i=1}^{N} (w_i - |h_i|) \cdot x_i$, the scaled weighted myriad filter is analogous to (but more robust than) the traditional linear FIR filter. Hence, the full signal processing potential of myriad filters in impulsive environments can be realized with this new class of scaled filters. It is important to note that, unlike the weighted myriad filter which has only $2N$ independent parameters (the output depending only on the normalized weights $\langle \mathbf{w}/K^2, \mathbf{h}/K^2 \rangle$), the scaled weighted myriad filter depends on $(2N+1)$ parameters, consisting of the N pairs of weights $\{\langle w_i, h_i \rangle\}_{i=1}^{N}$ as well as the linearity parameter K. Finally, it is easily shown that the scaled weighted myriad filter is BIBO (bounded-input bounded-output) stable. Consider a bounded input signal $\{x(n)\}$ with $\{|x(n)| < B\}_{n=-\infty}^{+\infty}$. From the previous discussion, it is clear that the filter output, at time n, is bounded in magnitude by $S(\langle \mathbf{w}, \mathbf{h} \rangle) \cdot M_{x(n)} < S \cdot B$ since $M_{x(n)} = \max\{|x_i(n)|\}_{i=1}^{N} < B$; hence the output signal is bounded as desired.

In Section 3.2, we introduced a special kind of weighted myriad filter with only N weights (*without* double weighting). We can develop a scaled version of this filter following the same approach as above. Referring to the definition of this special filter in (26) and to its limiting case (when $K \to \infty$) in (28), we define the scaled weighted myriad filter with N real-valued weights, $\{w_i' \in \mathcal{R}\}_{i=1}^{N}$, as

$$\begin{aligned} \bar{\theta}_K^{(S)}(\mathbf{w}', \mathbf{x}) &\triangleq \left(\sum_{i=1}^{N} |w_i'| \right) \cdot \bar{\theta}_K(\mathbf{w}', \mathbf{x}) \\ &= \left(\sum_{i=1}^{N} |w_i'| \right) \cdot \text{myriad}\left(|w_i'| \circ \text{sgn}(w_i') x_i \right)_{i=1}^{N}, \end{aligned} \tag{30}$$

where $\bar{\theta}_K(\mathbf{w}', \mathbf{x})$ is given by (26). In the limiting case, we see from (28) that $\bar{\theta}_\infty^{(S)}(\mathbf{w}', \mathbf{x}) = \sum_{i=1}^{N} w_i' \, x_i$, which is the familiar linear FIR filter. The filter described in (30) is a special

case of the scaled filter of (29), with $\{w_i' \triangleq w_i + h_i = w_i - |h_i|\}_{i=1}^N$ and the constraints $\{w_i \cdot h_i = 0\}_{i=1}^N$. Thus, the stability of this special scaled filter and its ability to amplify the range of its input signals, all follow from the previous discussion of the general scaled filter. Note that the filter of (30) has $(N+1)$ independent parameters, consisting of the N real-valued weights $\{w_i'\}_{i=1}^N$ as well as the linearity parameter K. This filter is simpler to design and implement than the general scaled filter, since it requires about half the number of parameters.

4 Adaptive Myriad Filtering Algorithms

In this section, we address the issue of *optimization* of the filter parameters for the different myriad filters that were defined in Section 3. Consider first the general case of an arbitrary nonlinear filter with an N-long input vector $\mathbf{x}$, a *parameter vector* $\mathbf{q}$ of length M, and filter output y. The filtering error in estimating a *desired signal* d is given by $e = y - d$. Under the mean square error (MSE) criterion, the filter parameters are designed to minimize the cost function $J(\mathbf{q}) \triangleq E\{e^2\} = E\{(y-d)^2\}$, where $E\{\cdot\}$ denotes statistical expectation. By setting the gradient of the cost function equal to zero, we obtain the *necessary* conditions to be satisfied by the optimal filter parameters:

$$\frac{\partial J(\mathbf{q})}{\partial q_i} = 2\,E\left\{e\,\frac{\partial y}{\partial q_i}\right\} = 0, \;\; i = 1, 2, \ldots, M. \tag{31}$$

The nonlinear nature of the equations in (31) prevents a closed-form solution for the optimal parameters. We therefore turn to the *method of steepest descent*, which continually updates the filter parameters in an attempt to converge to the global minimum of the cost function $J(\mathbf{q})$:

$$q_i(n+1) = q_i(n) - \frac{1}{2}\,\mu\,\frac{\partial J}{\partial q_i}(n), \;\; i = 1, 2, \ldots, M, \tag{32}$$

where $q_i(n)$ is the ith parameter at iteration n, $\mu > 0$ is the *step-size* of the update, and the gradient at the nth iteration is given by

$$\frac{\partial J}{\partial q_i}(n) = 2\,E\left\{e(n)\,\frac{\partial y}{\partial q_i}(n)\right\}, \;\; i = 1, 2, \ldots, M. \tag{33}$$

When the underlying signal statistics are unavailable and/or rapidly changing, we use *instantaneous estimates* for the gradient, since the expectation in (33) cannot be evaluated. Thus, removing the expectation operator in (33) and using the result in (32), we obtain the following nonlinear adaptive filtering algorithm:

$$q_i(n+1) \;=\; q_i(n) \;-\; \mu\, e(n)\, \frac{\partial y}{\partial q_i}\,(n), \;\; i = 1, 2, \ldots, M. \tag{34}$$

So far, we have imposed no constraints on the filter parameters: $\{q_i \in \mathcal{R}\}_{i=1}^{M}$. Suppose now that a particular filter parameter q_i is restricted to be non-negative, $q_i \geq 0$, a situation that will be encountered in the optimization of some of the filters of Section 3. In this case, the algorithm of (34) has to be modified to ensure that $q_i(n+1) \geq 0$ at each iteration step. The adaptive filtering algorithm for a non-negative parameter is therefore given by

$$q_i(n+1) \;=\; P\left[q_i(n) \;-\; \mu\, e(n)\, \frac{\partial y}{\partial q_i}(n)\right], \tag{35}$$

where $P[\cdot]$, defined by

$$P[u] \;\stackrel{\triangle}{=}\; \begin{cases} u, & u \geq 0 \\ 0, & u < 0, \end{cases} \tag{36}$$

projects the updated value onto the constraint space $\mathcal{R}^+$ of this filter parameter. Depending on whether a particular parameter is constrained or not, the appropriate algorithm out of (34) and (35) can be used. From the expressions in the above two algorithms, we see that the adaptive algorithm for any specific filter directly follows once we evaluate $\{\frac{\partial y}{\partial q_i}\}_{i=1}^{M}$ for that filter; we now proceed to accomplish this for the various myriad filters defined in Section 3.

4.1 Adaptive Weighted Myriad Filters

Given a set of weights $\{w_i \geq 0\}_{i=1}^{N}$ and $\{h_i \leq 0\}_{i=1}^{N}$, and a linearity parameter $K > 0$, the weighted myriad filter output $y \;=\; \tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ is given by (20). As mentioned in Section 3.2, this filter can also be expressed as a weighted myriad *smoother* $\hat{\theta}_K(\mathbf{g}, \mathbf{z})$ with the transformed input and weight vectors $\mathbf{z} \;=\; [x_1, x_2, \ldots, x_N, -x_1, -x_2, \ldots, -x_N]^T$ and $\mathbf{g} \;=\; [w_1, w_2, \ldots, w_N, |h_1|, |h_2, \ldots, |h_N]^T$, respectively. Thus,

$$y \;=\; \tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}) \;=\; \hat{\theta}_K(\mathbf{g}, \mathbf{z}). \tag{37}$$

It was also stated in Section 3.2 that $\tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ depends only on the normalized double weight vector $\langle \mathbf{w}/K^2, \mathbf{h}/K^2 \rangle$. The optimal filtering action is therefore independent of K, which can be chosen arbitrarily (the optimal weights would scale according to the value of K used). From the above discussion, we see that optimizing the weighted myriad filter is equivalent to finding the optimal transformed weights $\{g_i\}_{i=1}^{2N}$. Since these weights are non-negative, the appropriate adaptive algorithm to be used is given by (35), with the parameter vector $\mathbf{q} = \mathbf{g}$ and $M = 2N$. The task that remains is to find an expression for $\{\frac{\partial y}{\partial g_i}\}_{i=1}^{2N}$. From (37), we have

$$\frac{\partial y}{\partial g_i} \;=\; \frac{\partial}{\partial g_i}\,\hat{\theta}_K(\mathbf{g}, \mathbf{z}),\; i = 1, 2, \ldots, 2N. \tag{38}$$

However, the above quantity has already been derived in [14], in the context of optimization of the weighted myriad smoother, and is given by

$$\frac{\partial}{\partial g_i}\,\hat{\theta}_K(\mathbf{g}, \mathbf{z}) \;=\; \frac{\left[\dfrac{-(\hat{\theta} - z_i)}{\left(1 + \dfrac{g_i}{K^2}\left(\hat{\theta} - z_i\right)^2\right)^2}\right]}{\left[\displaystyle\sum_{j=1}^{2N} g_j \cdot \frac{1 - \dfrac{g_j}{K^2}\left(\hat{\theta} - z_j\right)^2}{\left(1 + \dfrac{g_j}{K^2}\left(\hat{\theta} - z_j\right)^2\right)^2}\right]}, \tag{39}$$

where $\hat{\theta} \;\equiv\; \hat{\theta}_K(\mathbf{g}, \mathbf{z})$. Using (37), (38) and (39) in (35), we finally obtain the following adaptive algorithm for the weights $\{g_i\}_{i=1}^{2N}$:

Adaptive Weighted Myriad Filter Algorithm

$$g_i(n+1) \;=\; P\left[g_i(n) \;+\; \mu\, e(n)\, \frac{\left\{\dfrac{(y - z_i)}{\left(1 + \dfrac{g_i}{K^2}\,(y - z_i)^2\right)^2}\right\}(n)}{\left\{\displaystyle\sum_{j=1}^{2N} g_j \cdot \frac{1 - \dfrac{g_j}{K^2}\,(y - z_j)^2}{\left(1 + \dfrac{g_j}{K^2}\,(y - z_j)^2\right)^2}\right\}(n)}\right], \tag{40}$$

where the transformed samples $\{z_i(n)\}_{i=1}^{2N}$ are found from the input samples $\{x_i(n)\}_{i=1}^{N}$ using

$$z_i(n) \;=\; \begin{cases} x_i(n), & i = 1, 2, \ldots, N \\ -x_{i-N}(n), & i = N+1, N+2, \ldots, 2N. \end{cases} \tag{41}$$

Note that the original filter weights $\{w_i(n) \geq 0\}_{i=1}^{N}$ and $\{h_i(n) \leq 0\}_{i=1}^{N}$ can be recovered from the transformed weights $\{g_i(n) \geq 0\}_{i=1}^{2N}$ at any iteration using

$$w_i(n) \;=\; g_i(n), \quad h_i(n) \;=\; -|h_i(n)| \;=\; -g_{i+N}(n), \;\; i = 1, 2, \ldots, N. \tag{42}$$

In Section 3.2, we introduced a special type of weighted myriad filter having only N weights, instead of the general case of $2N$ weights. We now consider the problem of optimizing the parameters of this filter. Given a set of real-valued weights $\{w_i' \in \mathcal{R}\}_{i=1}^{N}$, the output of this filter, $y \;=\; \bar{\theta}_K(\mathbf{w}', \mathbf{x})$, is given by (26). It was mentioned in Section 3.2 that $\bar{\theta}_K(\mathbf{w}', \mathbf{x})$ could be expressed as a special case of the general double-weighted myriad filter output $\tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ of (20). Indeed, given the weights $\{w_i \geq 0\}_{i=1}^{N}$ and $\{h_i \leq 0\}_{i=1}^{N}$, we have $y \;=\; \bar{\theta}_K(\mathbf{w}', \mathbf{x}) \;=\; \tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$, with $\{w_i' \;\triangleq\; w_i + h_i \;=\; w_i - |h_i|\}_{i=1}^{N}$ and the added constraints $\{w_i \cdot h_i \;=\; 0\}_{i=1}^{N}$. Thus, only one of w_i and h_i is allowed to be non-zero, and w_i' is equal to that non-zero weight. Due to the peculiar nature of this relationship between $\langle \mathbf{w}, \mathbf{h} \rangle$ and $\mathbf{w}'$, however, the general adaptive algorithm of (40) cannot be directly used for the optimization of the weights $\{w_i'\}_{i=1}^{N}$; an entirely different approach is therefore necessary. Now, since the weights $\{w_i'\}_{i=1}^{N}$ are all real-valued, the appropriate adaptive algorithm to be used is given by (34), with the parameter vector $\mathbf{q} = \mathbf{w}'$ and $M = N$. All that remains is to find an expression for

$$\frac{\partial y}{\partial w_i'} \;=\; \frac{\partial}{\partial w_i'} \bar{\theta}_K(\mathbf{w}', \mathbf{x}), \; i = 1, 2, \ldots, N; \tag{43}$$

this is derived in Appendix A and is given by

$$\frac{\partial}{\partial w_i'} \bar{\theta}_K(\mathbf{w}', \mathbf{x}) \;=\; \frac{-\left[\dfrac{K^2 \operatorname{sgn}(w_i') \left(\bar{\theta} - \operatorname{sgn}(w_i') x_i\right)}{\left(K^2 \;+\; |w_i'| \cdot \left(\bar{\theta} - \operatorname{sgn}(w_i') x_i\right)^2\right)^2}\right]}{\left[\displaystyle\sum_{j=1}^{N} |w_j'| \frac{K^2 \;-\; |w_j'| \cdot \left(\bar{\theta} - \operatorname{sgn}(w_j') x_j\right)^2}{\left(K^2 \;+\; |w_j'| \cdot \left(\bar{\theta} - \operatorname{sgn}(w_j') x_j\right)^2\right)^2}\right]}, \tag{44}$$

where $\bar{\theta} \equiv \bar{\theta}_K(\mathbf{w}', \mathbf{x})$. Using (34), we finally obtain the following adaptive algorithm to update the weights $\{w_i'\}_{i=1}^N$:

$$w_i'(n+1) = w_i'(n) - \mu\, e(n)\, \frac{\partial \bar{\theta}}{\partial w_i'}(n), \tag{45}$$

with $\frac{\partial \bar{\theta}}{\partial w_i'}(n)$ given by (44).

4.2 Adaptive Scaled Weighted Myriad Filters

The output of the scaled weighted myriad filter with weights $\{w_i \geq 0\}_{i=1}^N$ and $\{h_i \leq 0\}_{i=1}^N$, and linearity parameter $K > 0$, is given from (29) by

$$y = \tilde{\theta}_K^{(S)}(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}) = \left[\sum_{j=1}^{N} (w_j + |h_j|) \right] \cdot \tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}), \tag{46}$$

where the (unscaled) weighted myriad filter output $\tilde{\theta}_K(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x})$ is defined in (20). Using the interpretation of the weighted myriad filter as a weighted myriad smoother, as shown in (37), we can rewrite (46) as

$$y = \tilde{\theta}_K^{(S)}(\langle \mathbf{w}, \mathbf{h} \rangle, \mathbf{x}) = \left(\sum_{j=1}^{2N} g_j \right) \cdot \hat{\theta}_K(\mathbf{g}, \mathbf{z}), \tag{47}$$

with the transformed inputs $\mathbf{z} = [x_1, x_2, \ldots, x_N, -x_1, -x_2, \ldots, -x_N]^T$ and transformed weights $\mathbf{g} = [w_1, w_2, \ldots, w_N, |h_1|, |h_2, \ldots, |h_N]^T$. As we mentioned in Section 3.3, the scaled weighted myriad filter has $(2N+1)$ independent parameters represented by the N pairs of weights $\{\langle w_i, h_i \rangle\}_{i=1}^N$, and the linearity parameter K. Equivalently, from (47), the filter can be described by the modified weights $\{g_i \geq 0\}_{i=1}^{2N}$ and the linearity parameter $K > 0$. Since the parameters are all non-negative in this second formulation of the filter, the appropriate adaptive algorithm for their optimization is given by (35). It would appear that the parameter vector $\mathbf{q}$ to be used in (35) should consist of the weights $\{g_i\}_{i=1}^{2N}$ and the linearity parameter K. However, it makes more sense to optimize the *squared linearity parameter* $\kappa \stackrel{\triangle}{=} K^2$ rather than K itself, for two reasons. First, in the definitions of all the different myriad filters, K always occurs only in the form of its squared value. Second, an adaptive algorithm for K faces an ambiguity in the sign of K and might converge to a

negative value (that is, to $-\sqrt{\kappa}$). Therefore, the parameter vector used in (35) is chosen to be $\mathbf{q} = [\mathbf{g}^T, \kappa]^T$, with a length $M = 2N + 1$. Our task is then to derive expressions for $\{\frac{\partial y}{\partial g_i}\}_{i=1}^{2N}$, and for $\frac{\partial y}{\partial \kappa}$. After optimizing κ, the linearity parameter can be recovered as $K = +\sqrt{\kappa}$.

Differentiating with respect to the weight g_i in (47), we have

$$\frac{\partial y}{\partial g_i} \;=\; \hat{\theta}_K(\mathbf{g}, \mathbf{z}) \;+\; \left(\sum_{j=1}^{2N} g_j\right) \cdot \frac{\partial}{\partial g_i}\, \hat{\theta}_K(\mathbf{g}, \mathbf{z}); \tag{48}$$

note that we already know $\frac{\partial}{\partial g_i}\, \hat{\theta}_K(\mathbf{g}, \mathbf{z})$ from (39). Now, differentiating (47) with respect to the $\kappa = K^2$, holding the weights $\mathbf{g}$ fixed, we obtain

$$\frac{\partial y}{\partial \kappa} \;=\; \left(\sum_{j=1}^{N} g_j\right) \cdot \frac{\partial}{\partial \kappa}\, \hat{\theta}_K(\mathbf{g}, \mathbf{z}). \tag{49}$$

An expression for $\frac{\partial}{\partial \kappa}\, \hat{\theta}_K(\mathbf{g}, \mathbf{z})$ is derived in Appendix B, and is given by

$$\frac{\partial}{\partial \kappa}\, \hat{\theta}_K(\mathbf{g}, \mathbf{z}) \;=\; \frac{\left[\displaystyle\sum_{j=1}^{2N} \frac{g_j\,(\hat{\theta} - z_j)}{\left(\kappa \;+\; g_j\,(\hat{\theta} - z_j)^2\right)^2}\right]}{\left[\displaystyle\sum_{j=1}^{2N} g_j\, \frac{\kappa - g_j\,\left(\hat{\theta} - z_j\right)^2}{\left(\kappa + g_j\,\left(\hat{\theta} - z_j\right)^2\right)^2}\right]}, \tag{50}$$

where $\hat{\theta} \;\equiv\; \hat{\theta}_K(\mathbf{g}, \mathbf{z})$. Substituting the expressions for $\frac{\partial y}{\partial g_i}$ and $\frac{\partial y}{\partial \kappa}$ (from (48) and (49)) into (35) results in the following algorithms to update the weights $\{g_i\}_{i=1}^{2N}$ and the squared linearity parameter, $\kappa = K^2$:

Adaptive Scaled Weighted Myriad Filter Algorithms

$$g_i(n+1) \;=\; P\left[g_i(n) \;-\; \mu\, e(n) \cdot \left\{\hat{\theta}\,(n) \;+\; \left(\sum_{j=1}^{2N} g_j(n)\right) \cdot \frac{\partial \hat{\theta}}{\partial g_i}\,(n)\right\}\right] \tag{51}$$

and

$$\kappa(n+1) \;=\; P\left[\kappa(n) \;-\; \mu\, e(n) \cdot \left(\sum_{j=1}^{2N} g_j(n)\right) \cdot \frac{\partial \hat{\theta}}{\partial \kappa}\,(n)\right], \tag{52}$$

where $\frac{\partial \hat{\theta}}{\partial g_i}\,(n)$ and $\frac{\partial \hat{\theta}}{\partial \kappa}\,(n)$ are given by (39) and (50), respectively. Note that the transformed inputs and weights $\{z_i(n), g_i(n)\}_{i=1}^{2N}$ that occur in the above algorithms can be mapped to the original samples and weights $\{x_i(n), \langle w_i(n), h_i(n)\rangle\}_{i=1}^{N}$ when desired, using (41) and (42).

5 Simulation Results

The adaptive myriad filtering algorithms developed in Section 4 are investigated in this section using three computer simulation examples. The first example illustrates the inadequacy of the weighted myriad *smoother* in bandpass filtering a clean (noiseless) chirp-type signal. In the second example, the weighted myriad *filters* of Section 3.2 are trained to bandpass filter a chirp-type signal corrupted by alpha-stable noise. The third example consists of adaptively designing the *scaled* weighted myriad filters of Section 3.3 to retain the high frequency tones in a sum of sinusoids, followed by selective amplification of the extracted components. Thus, it is demonstrated that the scaled weighted myriad filters possess the full signal processing power of the traditional linear FIR filters.

Example 1:

In this example, we illustrate the fact that the weighted myriad smoother is unusable in bandpass filtering applications, due to its constraint of non-negative weights. Fig. 2(a) shows a clean (noiseless) *chirp-type* test signal $s(n)$, which is a digital sinusoid with a quadratically increasing instantaneous frequency. Specifically, the test signal of length L is given by $s(n) = sin(\omega(n)\, n),\ n = 0, 1, \ldots, L-1$, where the radian frequency $\omega(n) = 0.2\pi \cdot \left(\frac{n}{L-1}\right)^2$ increases quadratically with n from 0 to a maximum of 0.2π. The desired signal $d(n)$, shown in Fig. 2(b), was obtained by passing $s(n)$ through an FIR bandpass filter of window length $N = 21$, designed using MATLAB's **fir1** function with the passband cutoff frequencies $(\omega_1, \omega_2) = (0.16\pi, 0.18\pi)$.

The observed signal $s(n)$ and desired signal $d(n)$ of length $L = 300$ were used to train the linear FIR filter and the weighted myriad smoother (of Section 2), using a filter window length $N = 21$ in each case. The linear filter was trained using the standard LMS algorithm, with the initial weights being all zero: $\mathbf{w}(0) = \mathbf{0}$, while the weighted myriad smoother was trained using the adaptive LMS-type algorithm developed in [14], with the initial weights being all identical and normalized to sum to unity: $w_i(0) = (1/N) = 0.0476,\ i = 1, 2, \ldots, N$. The linearity parameter was arbitrarily chosen to be $K = 1.0$; recall from Section 2 that

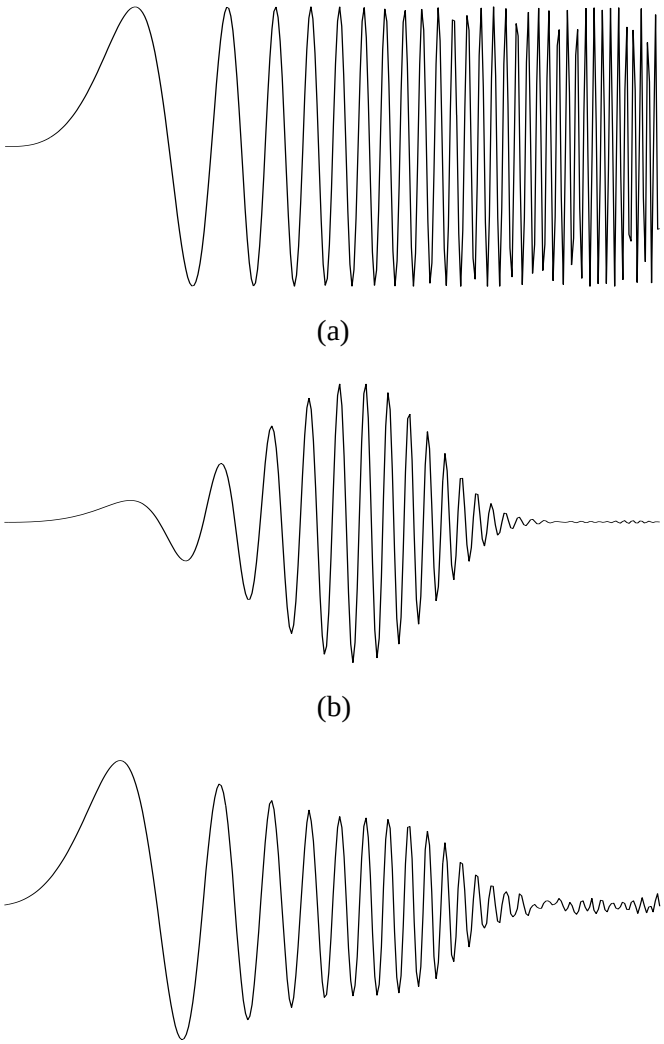


Figure 2: Example 1 (bandpass filtering a chirp-type signal), (a) $s(n)$: clean chirp-type test signal, (b) $d(n) \approx y_{lin}(n)$: desired signal (linear FIR filter output), (c) $y_{wmy}(n)$: weighted myriad smoother output.

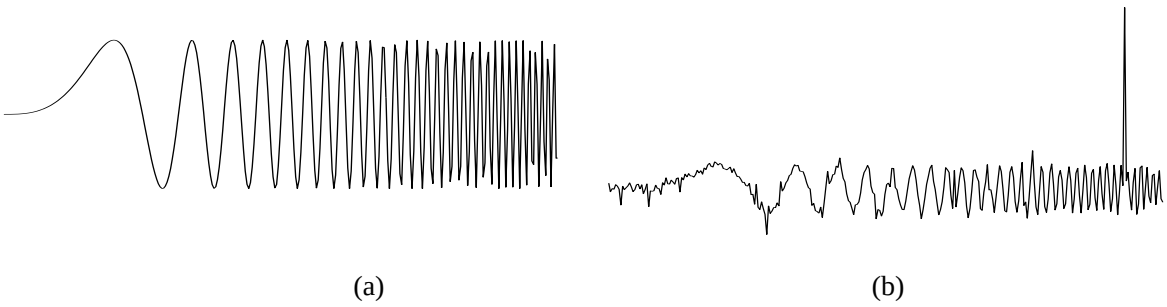
the filter output depends only on $(\mathbf{w}/K^2)$. Both the adaptive algorithms were implemented using 150 passes through the data, for a total of 45000 iterations. The step-size used was $\mu = 0.01$ in both cases, and was chosen to obtain as small an MSE as possible.

The linear filter output $y_{lin}(n)$, using the trained FIR filter on the test signal $s(n)$, was very close to the desired signal $d(n)$ (an MSE of 8.7×10^{-10}) as expected, and is therefore not shown separately in Fig. 2. However, the output $y_{wmy}(n)$ of the optimized weighted myriad smoother, shown in Fig. 2(c), is clearly quite different from the desired signal (an MSE of 0.0846). It is interesting to note from the figure that $y_{wmy}(n)$ is close to a *lowpass* filtered version of the test signal $s(n)$, with a cutoff frequency corresponding to the upper cutoff $\omega_2 = 0.18\pi$ of the designed bandpass linear FIR filter. In attempting to bandpass filter $s(n)$, the weighted myriad smoother thus does the best that it can, which is to lowpass filter $s(n)$ with the upper cutoff frequency. This example clearly shows that the weighted myriad smoother is limited to a lowpass type behavior.

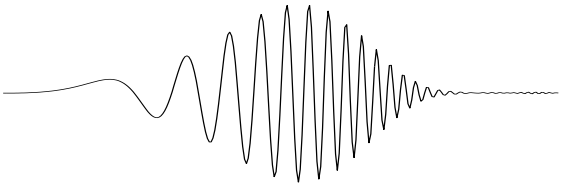
Example 2:

This example demonstrates that, unlike the weighted myriad smoother, the weighted myriad filters of Section 3.2 can perform bandpass filtering operations, since the weights can be both positive and negative. Fig. 3(a) shows a clean (noiseless) chirp-type signal $s(n)$, which is the same signal that was used in Example 1 (see Fig. 2). The desired signal $d(n)$, shown in Fig. 3(c), is also the same as the one obtained in Example 1 through bandpass filtering $s(n)$ with a linear FIR filter of window length $N = 21$.

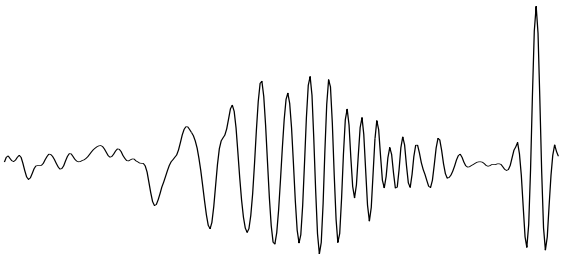
The noiseless signal $s(n)$ and the desired signal $d(n)$ of length $L = 300$ were used to train the linear FIR filter, and the general and special weighted myriad filters defined in Section 3.2, with a filter window of length $N = 21$ in each case. As in Example 1, the linear filter was trained using the LMS algorithm, with initial weights given by $\mathbf{w}(0) = \mathbf{0}$. Recall from Section 4.1 that a general weighted myriad filter with weights $\{w_i \geq 0\}_{i=1}^{N}$ and $\{h_i \leq 0\}_{i=1}^{N}$, and linearity parameter $K > 0$, can be optimized by finding the best transformed weights $\{g_i\}_{i=1}^{2N}$, where $\mathbf{g} = [w_1, w_2, \ldots, w_N, |h_1|, |h_2, \ldots, |h_N]^T$. The gen-



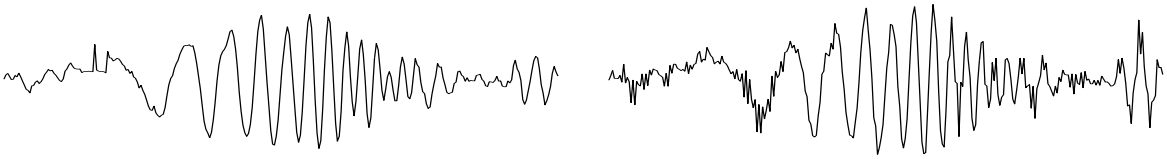
(a) (b)



(c)



(d)



(e) (f)

Figure 3: Example 2 (bandpass filtering a chirp-type signal), (a) clean chirp-type training signal, (b) noisy chirp-type test signal, (c) desired signal, (d) linear FIR filter output, (e) weighted myriad filter output, (f) special weighted myriad filter output.

eral weighted myriad filter with $2N$ weights was optimized using the adaptive algorithm of (40), with the initial transformed weights being all identical and normalized to unity: $g_i(0) = (1/2N) = 0.0238$, $i = 1, 2, \ldots, 2N$. The special weighted myriad filter with weights $\{w_i' \in \mathcal{R}\}_{i=1}^{N}$ was trained using the adaptive algorithm of (45), with the initial weights again being identical and normalized to unity: $w_i'(0) = (1/N) = 0.0476$, $i = 1, 2, \ldots, N$. All the adaptive algorithms were implemented using 100 passes through the data, for a total of 30000 iterations. A step-size $\mu = 0.01$ was chosen in all cases. For both these myriad filters, the optimal filter is independent of K; the linearity parameter in this example was therefore arbitrarily chosen to be $K = 1.0$.

Table 2 shows the final filter weights obtained by the different adaptive algorithms. The noiseless training signal $s(n)$ was filtered using the optimized linear and myriad filters; the mean square errors between these outputs and the desired signal $d(n)$ are shown in the second column of Table 3. Since the filter outputs are very close to the desired signal in all cases, there are not shown here.

Fig. 4 shows the MSE learning curves (squared error versus algorithm iterations) for the adaptive algorithms for the general and special myriad filters. Notice that the algorithm for the special myriad filter converges much faster than in the general case. This is to be expected since the general filter has twice the number of parameters and is naturally more complex. The special filter also achieves a lower MSE than the general case, as seen from Table 3; this is mainly due to the ease of convergence. It is not surprising that the linear filter achieves the lowest MSE in this noiseless case; it also converges the fastest in this case (the learning curve for the linear filter is not shown here).

Having designed the different bandpass filters in a noiseless environment, we now test their performance when applied to signals in impulsive noise. To this end, alpha-stable noise was added to the signal $s(n)$, with $\alpha = 1.4$ and the so-called *dispersion* of the noise being given by $\gamma = 0.1$. Fig. 3(b) shows the resulting *test signal* $x(n)$; note the heavy distortion in the signal due to the impulsive component. The noisy test signal $x(n)$ was filtered using the

Index	Linear FIR	General Myriad		Special Myriad
i	w_i	w_i	$\|h_i\|$	w_i'
1	0.0084	0.0340	0.0006	0.0095
2	0.0020	0.0156	0.0148	0.0048
3	-0.0134	0.0001	0.0511	-0.0040
4	-0.0410	0.0147	0.1603	-0.0402
5	-0.0732	0.0217	0.2297	-0.0036
6	-0.0891	0.0064	0.4526	-0.1153
7	-0.0677	0.0057	0.2079	-0.0082
8	-0.0046	0.0250	0.0030	0.0013
9	0.0815	0.2941	0.0019	0.0099
10	0.1561	0.9450	0.0210	0.2238
11	0.1857	1.1756	0.0319	0.0249
12	0.1561	0.8962	0.0112	0.2271
13	0.0814	0.3576	0.0073	0.0052
14	-0.0047	0.0143	0.0131	0.0021
15	-0.0676	0.0099	0.2009	-0.0086
16	-0.0891	0.0150	0.4976	-0.1115
17	-0.0732	0.0154	0.1682	-0.0044
18	-0.0409	0.0108	0.2445	-0.0362
19	-0.0133	0.0098	0.0193	-0.0026
20	0.0020	0.0111	0.0125	0.0057
21	0.0085	0.0221	0.0079	0.0068

Table 2: Optimized filter parameters in Example 2.

various filters that had been trained in a noiseless environment; the mean square errors are shown in the last column of Table 3. From the table, we see that the MSEs for the myriad filters are still quite low, while there is a severe degradation in the MSE of the linear filter, since it is not robust to changes in the noise environment. This is also seen in the linear filter output shown in Fig. 3(d), where the severe distortion in the high frequency region is due to the presence of the impulse which affects the linear filter. On the other hand, we see from Figs. 3(e) and (f) that the myriad filters outputs are relatively closer to the desired signal and are not affected as much by the impulse in the signal. It is evident that these filters are more robust to changes in the noise environment. This example thus shows the potential of weighted myriad filters in robust bandpass and highpass filtering applications.

Example 3:

Filter Type	Training (noiseless)	Test (noisy)
Linear FIR	1.6×10^{-9}	0.0544
General Myriad	0.0060	0.0173
Special Myriad	0.0033	0.0229

Table 3: Example 2, Mean square error (MSE) in filtering the training and test signals.

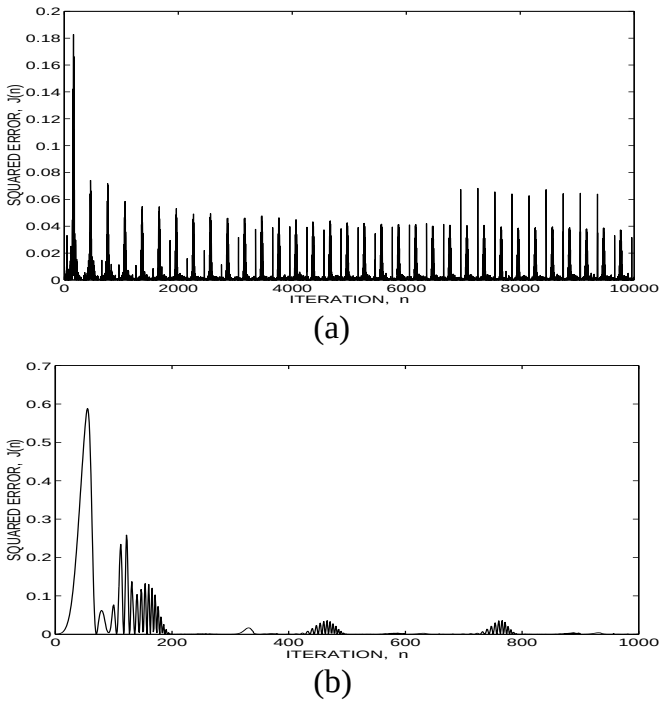
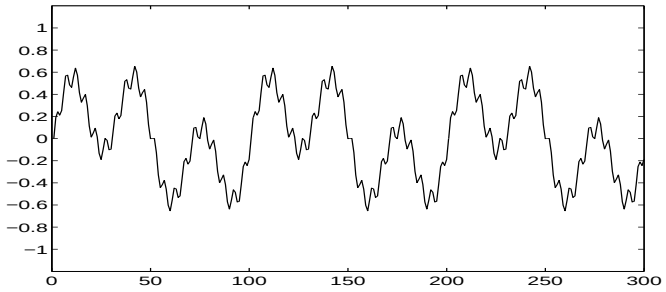
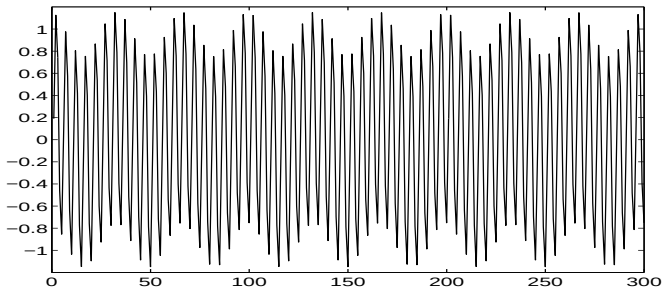


Figure 4: Example 2, MSE learning curves, (a) general weighted myriad filter, (b) special weighted myriad filter.

We show in this example that the scaled weighted myriad filter of Section 3.3 is capable of highpass filtering as well as amplification/attenuation of selected frequency components. The chosen observed signal is a clean (noiseless) sum of three sinusoids: $s(n) = \sum_{k=0}^{2} a_k \sin(2\pi f_k n)$. Fig. 5(a) shows a segment of $s(n)$, with the digital frequencies $f_0 = 0.01$, $f_1 = 0.03$, and $f_2 = 0.2$, and corresponding amplitudes $(a_0, a_1, a_2) = (0.3, 0.4, 0.1)$. The desired signal $d(n)$, shown in Fig. 5(b), is a sum of the two higher frequency sinusoids: $d(n) = \sum_{k=1}^{2} b_k \sin(2\pi f_k n)$, with the amplitudes chosen as $(b_1, b_2) = (0.2, 1.0)$. Note that it is required to *delete* the lowest frequency tone at f_0, *attenuate* the component at f_1 by a factor $(a_1/b_1) = 2.0$, and *amplify* the highest frequency tone at f_3 by a factor $(b_2/a_2) = 10.0$. Further, we can easily verify that the dynamic ranges of the signals are given by $|s(n)| \leq 0.6527$



(a)



(b)

Figure 5: Example 3 (highpass filtering a sum of sinusoids), (a) $s(n)$: clean sum of three sinusoids, (b) $d(n)$: desired highpass component with selective amplification/attenuation.

and $|d(n)| \leq 1.1495$. It is obvious that a weighted myriad filter *without* scaling would not succeed in achieving the amplification of the dynamic range evident in the figure, since its output is restricted to the dynamic range of its input.

The linear FIR filter and the scaled weighted myriad filter were both trained using 6000 samples of the observed signal $s(n)$ and the desired signal $d(n)$, using a window size $N = 21$ in each case. For the linear filter, the standard LMS algorithm was used, with the initial weights being all zero: $\mathbf{w}(0) = \mathbf{0}$. Recall from Section 4.2 that, for a scaled weighted myriad filter with weights $\{w_i \geq 0\}_{i=1}^{N}$ and $\{h_i \leq 0\}_{i=1}^{N}$, and linearity parameter $K > 0$, it is convenient to optimize the modified parameter vector $\mathbf{q} = [\mathbf{g}^T, \kappa]^T$, where the transformed weights are given by $\mathbf{g} = [w_1, w_2, \ldots, w_N, |h_1|, |h_2, \ldots, |h_N]^T$, and $\kappa = K^2$ is the squared linearity parameter. Notice that, unlike the filters considered so far where we could choose an arbitrary value for K, we have to optimize K here in addition to the weights $\{g_i\}$. The scaled weighted myriad filter was trained in this example using the algorithms of (51) and (52), with the

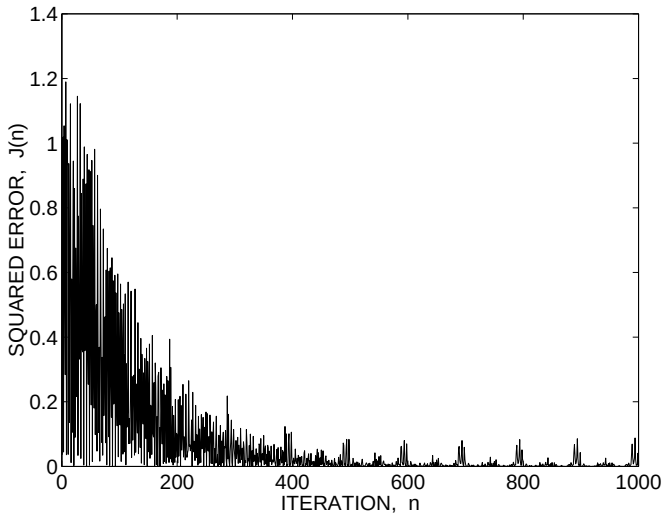


Figure 6: Example 3, MSE learning curve for scaled myriad filtering.

initial value for the linearity parameter being $K(0) = 0.5$, and the initial transformed weights being all identical and normalized to unity: $g_i(0) = (1/2N) = 0.0238$, $i = 1, 2, \ldots, 2N$. A step-size $\mu = 0.1$ was used in all the algorithms.

Fig. 6 shows the MSE learning curve (squared error versus algorithm iterations) for the adaptive scaled myriad filtering algorithms, demonstrating their convergence. The final values of the different filter weights are shown in Table 4, with the optimized value of the linearity parameter given by $K = 0.8236$. The observed signal $s(n)$ was filtered using the optimized linear and scaled myriad filters. Since the resulting outputs were very close to the desired signal (with MSEs equal to 7.7×10^{-8} and 0.0073, respectively), they are not shown here. Thus, the trained scaled myriad filter was successful in emulating the frequency selective behavior of the linear highpass filter. This example clearly shows that the scaled weighted myriad filter is analogous to the linear FIR filter in its ability to arbitrarily shape the spectrum of an input signal, as well amplify the signal if desired.

6 Conclusion

In this paper, we generalize the weighted myriad smoother, proposed for robust nonlinear "lowpass" filtering applications in impulsive environments, into a richer class of *Weighted*

Index	Linear FIR	Scaled Myriad ($K = 0.8236$)	
i	w_i	w_i	$\lvert h_i \rvert$
1	0.6142	0.5913	0.0271
2	0.0248	0.0795	0.2909
3	-0.9666	0.0189	0.9142
4	-0.8975	0.0152	0.7974
5	0.2289	0.3840	0.0820
6	0.9450	1.2599	0.0190
7	0.3432	0.6118	0.1499
8	-0.6728	0.0175	0.8607
9	-0.6398	0.0140	0.8212
10	0.4407	0.5756	0.0222
11	1.1027	0.8490	0.0167
12	0.4408	0.5255	0.0115
13	-0.6392	0.0164	1.0291
14	-0.6715	0.0184	0.9543
15	0.3448	0.5417	0.0399
16	0.9462	1.1263	0.0167
17	0.2296	0.6270	0.0371
18	-0.8972	0.0192	0.7638
19	-0.9665	0.0216	0.8441
20	0.0240	0.0189	0.2996
21	0.6119	0.6732	0.1451

Table 4: Optimized filter parameters in Example 3.

Myriad Filters admitting real-valued weights. These filters are developed by assigning a pair of weights, one positive and the other negative, to each of the input samples. This procedure can be applied to the generalization of any weighted M-estimator of location ('M-smoother'), leading to the class of so-called M-filters. With this new filter structure, weighted myriad filters can now be employed in a variety of applications that require "bandpass" or "highpass" type filtering. By suitably scaling the outputs of these filters, we develop the class of *Scaled Weighted Myriad Filters*. These novel filters constitute a robust generalization of the traditional linear FIR filter, outperforming linear filters in a variety of signal processing and communications applications in impulsive environments. We develop nonlinear adaptive algorithms for the optimization of the various myriad filters. The performance of these adaptive filters is demonstrated through computer simulation examples involving robust

frequency-selective filtering in impulsive noise environments.

A Evaluation of $\frac{\partial}{\partial w_i'}\,\bar{\theta}_K(\mathbf{w}', \mathbf{x})$

From (26) in Section 3.2, the output of the special type of weighted myriad filter with N real-valued weights $\{w_i' \in \mathcal{R}\}_{i=1}^N$ is $\bar{\theta}_K(\mathbf{w}', \mathbf{x}) = \arg\min_{\theta} Q_2(\theta)$, where $Q_2(\theta)$ is given by (27):

$$Q_2(\theta) \;=\; \sum_{i=1}^{N} \log\left[K^2 \;+\; |w_i'| \cdot (\mathrm{sgn}(w_i')x_i - \theta)^2\right]. \tag{53}$$

In this appendix, we shall evaluate the derivative of $\bar{\theta}_K(\mathbf{w}', \mathbf{x})$ with respect to the weight w_i', holding all other quantities constant. Since $\bar{\theta}_K(\mathbf{w}', \mathbf{x}) \equiv \bar{\theta}$ is one of the local minima of $Q_2(\theta)$, it follows that

$$Q_2'(\bar{\theta}) \;=\; 0. \tag{54}$$

Differentiating (53) and substituting into (54), we obtain

$$G(\bar{\theta}, w_i') \;\stackrel{\triangle}{=}\; Q_2'(\bar{\theta}) \;=\; 2\sum_{j=1}^{N} \frac{|w_j'| \cdot (\bar{\theta} - \mathrm{sgn}(w_j')x_j)}{K^2 \;+\; |w_j'| \cdot \left(\bar{\theta} - \mathrm{sgn}(w_j')x_j\right)^2} \;=\; 0, \tag{55}$$

where we have introduced the function $G(\cdot, \cdot)$ to emphasize the *implicit* dependency of the output $\bar{\theta}$ on the weight w_i', since we are interested in evaluating $\frac{\partial \bar{\theta}}{\partial w_i'}$ while holding all other quantities constant. By implicit differentiation of (55) with respect to w_i', we obtain

$$\left(\frac{\partial G}{\partial \bar{\theta}}\right) \cdot \left(\frac{\partial \bar{\theta}}{\partial w_i'}\right) \;+\; \left(\frac{\partial G}{\partial w_i'}\right) \;=\; 0, \tag{56}$$

from which we can find $\frac{\partial \bar{\theta}}{\partial w_i'}$ once we evaluate $\frac{\partial G}{\partial \bar{\theta}}$ and $\frac{\partial G}{\partial w_i'}$. Using (55), it is straightforward to show that

$$\frac{\partial G}{\partial \bar{\theta}} \;=\; 2\sum_{j=1}^{N} |w_j'| \frac{K^2 \;-\; |w_j'| \cdot \left(\bar{\theta} - \mathrm{sgn}(w_j')x_j\right)^2}{\left(K^2 \;+\; |w_j'| \cdot \left(\bar{\theta} - \mathrm{sgn}(w_j')x_j\right)^2\right)^2}. \tag{57}$$

Evaluation of $\frac{\partial G}{\partial w_i'}$ from (55) is, however, a more tricky matter. The difficulty is due to the term $\mathrm{sgn}(w_i')$ that occurs in the expression for $G(\bar{\theta}, w_i')$. It would seem impossible to differentiate $G(\bar{\theta}, w_i')$ with respect to w_i', since this would involve the quantity $\frac{\partial}{\partial w_i'}\mathrm{sgn}(w_i')$,

which clearly cannot be found. Fortunately, this problem can be circumvented by rewriting the expression for $G(\bar{\theta}, w_i')$ in (55), expanding it as follows:

$$G(\bar{\theta}, w_i') \;=\; 2 \sum_{j=1}^{N} \frac{|w_j'|\,\bar{\theta} \;-\; w_j' x_j}{K^2 \;+\; |w_j'|\,\left(\bar{\theta}^2 + x_j^2\right) \;-\; 2 w_j'\,\bar{\theta} x_j}, \tag{58}$$

where we have used the fact that $|w_j'| \cdot \mathrm{sgn}(w_j') = w_j'$. We see that (58) presents no mathematical difficulties in differentiating with respect to w_i', since the term $\mathrm{sgn}(w_i')$ is no longer present. Differentiating with respect to w_i', and performing some straightforward manipulations, we then obtain

$$\frac{\partial G}{\partial w_i'} \;=\; \frac{2\,K^2\,\mathrm{sgn}(w_i')\,\left(\bar{\theta} - \mathrm{sgn}(w_i') x_i\right)}{\left(K^2 \;+\; |w_i'| \cdot \left(\bar{\theta} - \mathrm{sgn}(w_i') x_i\right)^2\right)^2}. \tag{59}$$

Substituting (57) and (59) into (56), we finally obtain the following expression for $\frac{\partial \bar{\theta}}{\partial w_i'}$:

$$\frac{\partial}{\partial w_i'}\,\bar{\theta}_K(\mathbf{w}', \mathbf{x}) \;=\; \frac{-\left[\dfrac{K^2\,\mathrm{sgn}(w_i')\,\left(\bar{\theta} - \mathrm{sgn}(w_i') x_i\right)}{\left(K^2 \;+\; |w_i'| \cdot \left(\bar{\theta} - \mathrm{sgn}(w_i') x_i\right)^2\right)^2}\right]}{\left[\displaystyle\sum_{j=1}^{N} |w_j'|\,\frac{K^2 \;-\; |w_j'| \cdot \left(\bar{\theta} - \mathrm{sgn}(w_j') x_j\right)^2}{\left(K^2 \;+\; |w_j'| \cdot \left(\bar{\theta} - \mathrm{sgn}(w_j') x_j\right)^2\right)^2}\right]}. \tag{60}$$

B Evaluation of $\frac{\partial}{\partial \kappa}\,\hat{\theta}_K(\mathbf{g}, \mathbf{z})$

In this appendix, we shall obtain an expression for the derivative of the weighted myriad smoother output $\hat{\theta}_K(\mathbf{g}, \mathbf{z})$ with respect to its squared linearity parameter $\kappa = K^2$, while holding the weights $\mathbf{g}$ and inputs $\mathbf{z}$ constant. Noting that the vectors $\mathbf{g}$ and $\mathbf{z}$ are of length $2N$, and referring to to Section 2, we conclude from (8) and (9) that the output $\hat{\theta}_K(\mathbf{g}, \mathbf{z}) \equiv \hat{\theta}$ satisfies

$$F(\hat{\theta}, \kappa) \;=\; \sum_{j=1}^{2N} \frac{g_j\,(\hat{\theta} - z_j)}{\kappa \;+\; g_j\,(\hat{\theta} - z_j)^2} \;=\; 0, \tag{61}$$

where the function $F(\cdot, \cdot)$ emphasizes the *implicit* dependency of the output $\hat{\theta}$ on κ, holding all other quantities constant. By implicit differentiation of (61) with respect to κ, we obtain

$$\left(\frac{\partial F}{\partial \hat{\theta}}\right) \cdot \left(\frac{\partial \hat{\theta}}{\partial \kappa}\right) \;+\; \left(\frac{\partial F}{\partial \kappa}\right) \;=\; 0, \tag{62}$$

from which we can extract $\frac{\partial \hat{\theta}}{\partial \kappa}$ once we evaluate $\frac{\partial F}{\partial \hat{\theta}}$ and $\frac{\partial F}{\partial \kappa}$. It is straightforward to find these quantities using the expression for $F(\hat{\theta}, \kappa)$ given in (61):

$$\frac{\partial F}{\partial \hat{\theta}} = \sum_{j=1}^{2N} g_j \frac{\kappa - g_j \left(\hat{\theta} - z_j\right)^2}{\left(\kappa + g_j \left(\hat{\theta} - z_j\right)^2\right)^2} \tag{63}$$

and

$$\frac{\partial F}{\partial \kappa} = -\sum_{j=1}^{2N} \frac{g_j \left(\hat{\theta} - z_j\right)}{\left(\kappa + g_j \left(\hat{\theta} - z_j\right)^2\right)^2}. \tag{64}$$

Substituting (63) and (64) into (62), we finally obtain the following expression for $\frac{\partial \hat{\theta}}{\partial \kappa}$:

$$\frac{\partial}{\partial \kappa} \hat{\theta}_K(\mathbf{g}, \mathbf{z}) = \frac{\left[\sum_{j=1}^{2N} \frac{g_j \left(\hat{\theta} - z_j\right)}{\left(\kappa + g_j \left(\hat{\theta} - z_j\right)^2\right)^2}\right]}{\left[\sum_{j=1}^{2N} g_j \frac{\kappa - g_j \left(\hat{\theta} - z_j\right)^2}{\left(\kappa + g_j \left(\hat{\theta} - z_j\right)^2\right)^2}\right]}. \tag{65}$$

References

[1] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-19, no. 3, pp. 106–127, 1977.

[2] J. Ilow, *Signal Processing in Alpha-Stable Noise Environments: Noise Modeling, Detection and Estimation.* PhD thesis, University of Toronto, Canada, Dec. 1995.

[3] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications.* New York: Wiley, 1995.

[4] P. J. Huber, *Robust Statistics.* New York: Wiley, 1981.

[5] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing," *Proceedings of the IEEE*, vol. 73, Mar. 1985.

[6] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, "Weighted median filters: A tutorial," *IEEE Transactions on Circuits and Systems–II*, vol. 43, Mar. 1996.

[7] I. Pitas and A. Venetsanopoulos, "Order statistics in digital image processing," *Proceedings of the IEEE*, vol. 80, Dec. 1992.

[8] J. G. Gonzalez and G. R. Arce, "Weighted myriad filters: A robust filtering framework derived from α-stable distributions," in *Proc. of the 1996 IEEE ICASSP*, (Atlanta, GA), 1996.

[9] J. G. Gonzalez and G. R. Arce, "Weighted myriad filters: A powerful framework for efficient filtering in impulsive environments," *IEEE Transactions on Signal Processing*. Manuscript in review.

[10] J. G. Gonzalez, *Robust Techniques for Wireless Communications in NonGaussian Environments*. PhD thesis, Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware, U.S.A., Dec. 1997.

[11] S. Kalluri and G. R. Arce, "Fast algorithms for weighted myriad computation by fixed point search," *IEEE Transactions on Signal Processing*. Manuscript in review.

[12] J. G. Gonzalez, D. W. Griffith, and G. R. Arce, "Matched myriad filtering for robust communications," in *Proc. of the 1996 CISS*, (Princeton, NJ), 1996.

[13] P. Zurbach, J. G. Gonzalez, and G. R. Arce, "Weighted myriad filters for image processing," in *Proc. of the 1996 IEEE Int. Symp. on Circuits and Systems*, (Atlanta, GA), 1996.

[14] S. Kalluri and G. R. Arce, "Adaptive weighted myriad filter algorithms for robust signal processing in α-stable noise environments," *IEEE Transactions on Signal Processing*, vol. 46, pp. 322–334, Feb. 1998.

[15] S. Ambike and D. Hatzinakos, "A new filter for highly impulsive α-stable noise," in *Proc. of the 1995 Int. Workshop on Nonlinear Signal and Image Processing*, (Halkidiki, Greece), 1995.

[16] G. R. Arce, "A general weighted median filter structure admitting negative weights," *IEEE Transactions on Signal Processing*, vol. 46, pp. 3195–3205, Dec. 1998.

[17] J. L. Paredes and G. R. Arce, "Stack filters, stack smoothers, and mirrored threshold decomposition," *IEEE Transactions on Signal Processing*. Submitted for publication.