Adaptive Weighted Myriad Filter Algorithms for Robust Signal Processing in α-Stable Noise Environments [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

Stochastic gradient-based adaptive algorithms are developed for the optimization of *Weighted Myriad Filters*. *Weighted Myriad Filters* form a class of nonlinear filters, motivated by the properties of α-stable distributions, that have been proposed for robust non-Gaussian signal processing in impulsive noise environments. The *weighted myriad* for an N-long data window is described by a set of non-negative weights $\{w_i\}_{i=1}^N$ and the so-called linearity parameter $K > 0$. In the limit as $K \to \infty$, the filter reduces to the familiar *weighted mean filter* (a constrained linear FIR filter).

In this paper, necessary conditions are obtained for optimality of the filter weights under the mean absolute error criterion. An implicit formulation of the filter output is used to find an expression for the gradient of the cost function. Using instantaneous gradient estimates, an adaptive steepest-descent algorithm is then derived to optimize the weights. This algorithm involves a very simple update term that is computationally comparable to the update in the classical LMS algorithm. The robust performance of this adaptive algorithm is demonstrated through a computer simulation example involving lowpass filtering of a one-dimensional chirp-type signal in impulsive noise.

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

1 Introduction

The traditional approach to a statistical signal processing problem has been to derive the *optimal* solution based on a particular signal and noise model for the problem at hand. This approach relies on the belief that the optimal solution will be adequate under small deviations from the nominal statistical model. Classical statistical signal processing theory has been dominated by the assumption of the Gaussian model for the statistical characteristics of the underlying processes. The Gaussian model is valid for some real-world processes and is justified by the Central Limit Theorem. There are, however, many processes occurring in practice that are decidedly non-Gaussian. For instance, a large number of physical processes are impulsive in nature and are more accurately modelled by heavy-tailed non-Gaussian distributions. Impulsive signals and noise are characterized by sharp spikes or occasional outliers in the data. Examples of impulsive processes include atmospheric noise in radio links, ocean acoustic noise, switching transients in telephone channels, and multiple access interference in radio communication networks [1, 2, 3]. Systems optimized under the Gaussian assumption can suffer severe performance degradation under non-Gaussian noise [4]. It is well-known, for instance, that linear filters perform poorly in the presence of outliers in the data.

Several techniques have been proposed to combat impulsive noise and, more generally, to deal with uncertainties in the assumed statistical models. The presence of outliers can be viewed as arising from an uncertainty regarding the assumed model. Most of these methods are based on the theory of *robust statistics* [5, 6]. *Robust* signal processing techniques [7] are designed to perform well under *nominal* conditions and still be *adequate* when the signal and noise statistics deviate from the nominal model. Median filters, and their generalizations based on *order statistics*, have been widely used in image processing due to their ability to preserve edges and fine detail while rejecting outliers. Median and weighted median filters (WMF) [8] are derived to be optimal under the Laplacian noise distribution, which is more heavy-tailed than the Gaussian distribution and, therefore, more suited to model outliers. These filters are limited by the fact that they are *selection filters* (the output of a selection filter is always, by definition, one of the input samples).

In recent years, there has been considerable interest in signal processing based on α-stable distributions, which have been shown to accurately model impulsive noise processes [9, 10]. These distributions have a parameter α ($0 < \alpha \leq 2$), called the *characteristic exponent*, which controls the heaviness of their tails; a smaller α signifies a heavier-tailed distribution. For $0 < \alpha < 2$, α-stable random variables have infinite variance. The limiting case $\alpha = 2$ leads to the Gaussian distribution, while the case $\alpha = 1$ corresponds to the Cauchy distribution. The Gaussian and Cauchy distributions are the only *symmetric* α-stable distributions having closed-form expressions for their density functions. The use of the α-stable distribution as a statistical model is justified theoretically by two properties [9]. The first is the *stability property*: the sum of two independent stable random variables with the same characteristic exponent is also stable with the same characteristic exponent. The second is the Generalized Central Limit Theorem: if the sum of an infinite number of independent and identically distributed (i. i. d.) random variables (with finite or infinite variance) converges in distribution, the limiting distribution is α-stable. Thus, α-stable random variables can arise in the physical world as the effects of a large number of independent contributing factors, in the same way as Gaussian random variables do. For example, α-stable distributions have been used to model multiple access interference in radio networks where the independent interfering sources are modelled as a Poisson field in space and the superposition of the interfering electromagnetic waves follows an α-stable distribution [3, 11].

Weighted Myriad Filters (WMyF) have been proposed recently as a class of robust, nonlinear filters based on α-stable distributions [12, 13]. They have been used in robust communications and image processing applications [14, 15]. These filters have been derived as extensions of the *sample myriad*, defined as the maximum likelihood estimate (MLE) of the location parameter of the Cauchy distribution (an α-stable distribution with $\alpha = 1$) [13, 16]. The *weighted myriad* for an observation window of length N is described by non-negative weights $\{w_i\}_{i=1}^{N}$ and a *linearity parameter* $K > 0$. As $K \to \infty$, the filter reduces to the familiar *weighted mean filter* (a normalized FIR filter whose weights sum to unity). The term *myriad* was coined because, for small K, the filter tends to favor values near *clusters* of input samples. The case $K \to 0$ leads to a highly robust *selection filter* called the *weighted mode-myriad filter*. The *sample myriad* is intimately related to maximum likelihood estimates (MLEs) of location of α-stable distributions; the MLE of location of an α-stable distribution approaches the *mean* (the sample myriad with $K \to \infty$) as $\alpha \to 2$, while it approaches the *mode-myriad* (the sample myriad with $K \to 0$) as $\alpha \to 0$.

In this paper, we consider the problem of *optimization* of the parameters of weighted myriad filters for the case $K > 0$ [17, 18]. We design the filter to optimally estimate a desired signal according to some statistical error criterion. Two popular criteria in this approach are the mean square error (MSE) and the mean absolute error (MAE). We adopt the MAE criterion in this paper due to its convenience and also because it is more robust against outliers. However, the modifications to our solutions are trivial if the MSE criterion is adopted.

We derive necessary conditions for optimality of weighted myriad filters. These conditions result in a set of highly nonlinear equations that are difficult to solve in closed-form for the optimal filter parameters. The use of nonlinear optimization techniques here is hampered by the fact that we also require knowledge of the statistics of the underlying signals, which enter into the equations in a nonlinear fashion. In applications where the signal statistics are unknown or insufficient, or when the signals are non-stationary, *adaptive signal processing algorithms* have been used with great advantage [19]. We follow this approach and derive stochastic gradient-based adaptive algorithms to optimize the filter parameters. In [20], robust adaptive *linear* filtering algorithms, based on Fractional Lower Order Statistics, have been introduced for impulsive noise environments modelled by α-stable distributions. The present paper, on the other hand, deals with robust adaptive *nonlinear* filtering algorithms for impulsive noise environments.

For the case $K > 0$, we use an *implicit formulation* of the filter output to find an expression for the gradient of the MAE cost function. We then derive an adaptive steepest-descent algorithm, using instantaneous gradient estimates, to optimize the weights. This algorithm involves a very simple update term that is computationally comparable to the update in the classical LMS adaptation algorithm. For the special case $K \to 0$, we are faced with a cost function that is discontinuous in the filter weights. The optimization for this case requires quite a different approach and will be considered in future publications. In the present paper, we confine ourselves to the general case $K > 0$.

The paper is organized as follows. Section 2 introduces the class of weighted myriad filters. In Section 3, we state the optimal filtering problem and derive necessary conditions for optimality. Adaptive algorithms for learning the optimal filter weights are derived in Section 4. In Section 5, we present simulation results involving lowpass filtering a one-dimensional chirp-type signal in α-stable noise.

2 Weighted Myriad Filters

Just as the weighted mean filter and the weighted median filter (WMF) are generalizations of the sample *mean* and the sample *median*, respectively, the class of *weighted myriad filters* (WMyF) is developed from the so-called sample *myriad*. In this section, we first give a brief introduction to the sample myriad (for a detailed treatment, see [12, 13]). We then define weighted myriad filters and describe some of their properties which will be useful in the later sections on filter optimization.

The sample mean and median arise out of maximum likelihood (ML) estimation of the location parameters of the Gaussian and Laplacian distributions, respectively. Analogously, the sample myriad is defined as the ML estimate of location of the Cauchy distribution. Consider a set of N independent and identically distributed (i.i.d.) observations, denoted $\{x_1, x_2, \ldots, x_N\}$, drawn from a Cauchy distribution with location parameter β and scaling factor $K > 0$:

$$f(x;\beta) \;=\; \left(\frac{K}{\pi}\right) \frac{1}{K^2 + (x-\beta)^2}. \tag{1}$$

The sample myriad is the value $\hat{\beta}_K$ that maximizes the likelihood function $L(x_1, x_2, \ldots, x_N; \beta) = \prod_{i=1}^{N} f(x_i;\beta)$ or, equivalently, *minimizes* the expression $\prod_{i=1}^{N}[K^2 + (x_i-\beta)^2]$. Thus,

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

where the last step is because the logarithm is an increasing function. Defining $\rho(x;\beta) \stackrel{\triangle}{=} \log[K^2 + (x-\beta)^2]$, we have $\hat{\beta}_K = \arg\min_{\beta} \sum_{i=1}^{N} \rho(x_i;\beta)$, which defines an M-estimator [5, 6]. For location estimates, $\rho(x;\beta)$ is usually of the type $\rho(x-\beta)$, which is the case here. Interestingly, the myriad includes the mean as a limiting case, converging to the sample mean as $K \to \infty$ [13].

In the following, the sample myriad is generalized to the *weighted myriad*. Two cases are treated: the general case, $K > 0$ and the special limiting case as $K \to 0$.

2.1 The Weighted Myriad Filter (WMyF): $K > 0$

Similar to the extension of the sample mean to the weighted mean, the *weighted myriad* is defined by assigning weights to the samples in the ML location estimation. The weights reflect the different levels of reliability of the observed samples. Consider a set of observations $\{x_i\}_{i=1}^N$ and a set of filter weights $\{w_i\}_{i=1}^N$. Define the observation vector $\mathbf{x} \triangleq [x_1, x_2, \ldots, x_N]^T$ and the weight vector $\mathbf{w} \triangleq [w_1, w_2, \ldots, w_N]^T$. For a given $K > 0$, the *weighted myriad filter* (WMyF) output is given by

$$\begin{aligned} \hat{\beta}_K(\mathbf{w}, \mathbf{x}) &\triangleq myriad(K; w_1 \circ x_1, w_2 \circ x_2, \ldots, w_N \circ x_N) \\ &= \arg\min_{\beta} G_K(\beta, \mathbf{w}, \mathbf{x}), \end{aligned} \tag{3}$$

where

$$G_K(\beta, \mathbf{w}, \mathbf{x}) \triangleq \prod_{i=1}^{N} [K^2 + w_i(x_i - \beta)^2] \tag{4}$$

is called the *weighted myriad objective function* since it is minimized by the weighted myriad, and $w_i \circ x_i$ denotes the weighting operation in (4). When the context is clear, we shall refer to $\hat{\beta}_K(\mathbf{w}, \mathbf{x})$ as $\hat{\beta}_K$, or just $\hat{\beta}$. Likewise, we shall compress $G_K(\beta, \mathbf{w}, \mathbf{x})$ to $G_K(\beta)$, or just $G(\beta)$.

It should be pointed out that the formulation of the weighted myriad as a maximum likelihood location estimate from samples of varying reliability constrains the weights to be non-negative. Nevertheless, the weighted myriad could be defined using (3) and (4) with negative weights. However, this results in potential instability of the filter (the output can sometimes be $+\infty$ or $-\infty$). We restrict the weights to be non-negative in this paper: $w_i \geq 0, i = 1, 2, \ldots, N$.

The weighted myriad filter output is the value of β at the global minimum of the *weighted myriad objective function* $G_K(\beta)$. It is easily seen from (4) that, for non-negative weights and $K > 0$, $G(\beta)$ is positive for all β and goes to ∞ as $\beta \to \pm\infty$. Also, it is a well-behaved function since it is in fact a polynomial of degree $2N$. It follows that the filter output $\hat{\beta}$ occurs at one of the local minima of $G(\beta)$. Fig. 1 shows typical plots of $\log(G_K(\beta))$ for a data window of size $N = 7$ (note that either $G(\beta)$ or $\log(G(\beta))$ could be used as an objective function for the weighted myriad). Denote the derivative of $G(\beta)$ as

$$G'(\beta) \triangleq \frac{\partial G(\beta, \mathbf{w}, \mathbf{x})}{\partial \beta}. \tag{5}$$

The filter output $\hat{\beta}$ is one of the roots of $G'(\beta)$:

$$G'(\hat{\beta}) = 0. \tag{6}$$

From (4), we obtain

$$G'(\beta) = \sum_{j=1}^{N} 2w_j(\beta - x_j) \prod_{l=1, l \neq j}^{N} [K^2 + w_l(\beta - x_l)^2]. \tag{7}$$

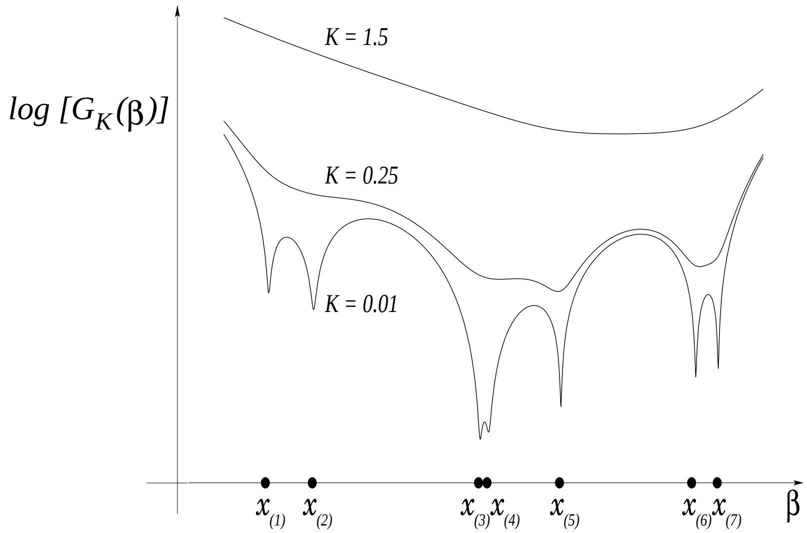

Figure 1: Weighted myriad objective function. Input samples $\mathbf{x} = [4.9, 0.0, 6.5, 10.0, 9.5, 4.7, 1.0]$, weights $\mathbf{w} = [0.05, 0.1, 0.6, 0.9, 0.6, 0.1, 0.05]$.

Note that $G^{'}(\beta)$ is a polynomial of degree $(2N-1)$ and can have as many as $(2N-1)$ real roots. Using (4) again, we can write

$$G^{'}(\beta) \;=\; 2G(\beta)\sum_{j=1}^{N}\frac{w_j(\beta - x_j)}{K^2 + w_j(\beta - x_j)^2}. \tag{8}$$

Noting that $G(\beta) > 0$, we see that the filter output $\hat{\beta}$ satisfies the equation

$$\sum_{j=1}^{N}\frac{w_j(\hat{\beta} - x_j)}{K^2 + w_j(\hat{\beta} - x_j)^2} \;=\; 0. \tag{9}$$

A few simple properties can easily be inferred from (8) and (9). First, it is important to note that the filter has only N independent parameters (even though there are N weights and the parameter K). Using (3) and (4), we can infer that if we change the value of K, we can obtain the same filter output provided the filter weights are appropriately scaled. Thus, we can write

$$\hat{\beta}_K(\mathbf{w}, \mathbf{x}) \;=\; \hat{\beta}_1(\frac{\mathbf{w}}{K^2}, \mathbf{x}) \tag{10}$$

or

$$\hat{\beta}_{K_1}(\mathbf{w}_1, \mathbf{x}) \;=\; \hat{\beta}_{K_2}(\mathbf{w}_2, \mathbf{x}) \quad \text{iff} \quad \frac{\mathbf{w}_1}{K_1^2} \;=\; \frac{\mathbf{w}_2}{K_2^2}. \tag{11}$$

Hence, the filter output depends only on $\frac{\mathbf{w}}{K^2}$. Let $\{x_{(m)}\}_{m=1}^{N}$ denote the order statistics (samples sorted in increasing order of magnitude) of $\mathbf{x}$, with $x_{(1)}$ the smallest and $x_{(N)}$ the largest. By examining the function $G^{'}(\beta)$, it is easily shown [13] that $G(\beta)$ has L local minima and $(L-1)$

local maxima where $1 \leq L \leq N$. Further, for non-negative weights, it can be proved (again, see [13]) that all the local extrema occur within the interval $[x_{(1)}, x_{(N)}]$, the range of the input samples. Thus, we have $x_{(1)} \leq \hat{\beta} \leq x_{(N)}$. This is illustrated in Fig. 1 where the order statistics are shown on the horizontal axis with the smallest $x_{(1)} = 0.0$ and the largest $x_{(N)} = 10.0$.

The weighted myriad is not easy to compute since we have to find the roots of the polynomial $G'(\beta)$, choose the ones that are local minima of $G(\beta)$ and test all the local minima to find the global minimum. In [21], we describe a simple and fast algorithm, using a fixed point search, to compute the filter output approximately.

As K gets larger, the number of local minima of $G(\beta)$ decreases. In fact, it can be proved (by examining the second derivative $G''(\beta)$) that a *sufficient* (but not *necessary*) condition for $G(\beta)$ (and $\log(G(\beta))$) to be convex and, therefore, have a unique local minimum, is that $K > \sqrt{\max\{w_j\}_{j=1}^N}\,(x_{(N)} - x_{(1)})$. In the example of Fig. 1, this condition reduces to $K > 9.49$. As seen from the figure however, this condition is not necessary; the onset of convexity could be at a much lower K. Finally, letting $K \to \infty$ in (9), while holding the weights finite, results in

$$\hat{\beta}_\infty = \frac{\displaystyle\sum_{j=1}^{N} w_j x_j}{\displaystyle\sum_{j=1}^{N} w_j}, \tag{12}$$

which is the limiting case of the *weighted mean filter* ($\hat{\beta}_\infty = 8.07$ in our example). Since the weighted myriad approaches the (linear) weighted mean as K increases, K is referred to as the *linearity parameter.*

2.2 The Weighted Mode-Myriad Filter (WMyF$_0$): $K \to 0$

When the linearity parameter K tends to zero, the weighted myriad reduces to a selection filter that is highly resistant to outliers. As Fig. 1 shows, all the local minima are close to the input samples for very low K ($K = 0.01$). The filter output moves from the weighted mean 8.07 ($K \to \infty$) to 7.77 ($K = 1.5$), 6.43 ($K = 0.25$) and finally to 4.71 ($K = 0.01$), which is near the cluster of samples $x_{(3)} = 4.7$ and $x_{(4)} = 4.9$. Note that, for $K = 0$, the objective function $G_0(\beta)$ is zero whenever β is one of the input samples. In this case, there are N local minima, one at each input sample and it would appear that any of the input samples could be the output (all of them minimize $G_0(\beta)$ to zero). However, we obtain a meaningful result if we define the filter output to be the *limit* of the WMyF output as $K \to 0$. The *weighted mode-myriad filter* (WMyF$_0$) output is given by [12, 13]

$$\hat{\beta}_0(\mathbf{w}, \mathbf{x}) \triangleq \lim_{K \to 0} \hat{\beta}_K(\mathbf{w}, \mathbf{x}). \tag{13}$$

The *mode-myriad* filter is the special case when all the weights are unity. It can be shown that the weighted mode-myriad is the most repeated input sample, if unique. Thus $\hat{\beta}_0$ is a mode-like estimator, hence the term mode-myriad. When the most repeated sample is not unique, the filter output reduces to [15]

$$\hat{\beta}_0 \;=\; \arg \min_{x_j \,\in\, \mathcal{M}} \prod_{i=1, x_i \neq x_j}^{N} w_i (x_i - x_j)^2, \tag{14}$$

where $\mathcal{M}$ is the set of most repeated values among the input samples. Note that the weights for the weighted mode-myriad have to be *strictly positive*, $w_i > 0$. When the input samples are distinct, the set $\mathcal{M}$ becomes the set of input samples $\{x_i\}_{i=1}^{N}$. In this case, the weighted mode-myriad filter output can be expressed, after a few simple manipulations, as $\mathrm{WMyF}_0(\mathbf{w}, \mathbf{x}) = \arg\min_{x_j} G_0(x_j, \mathbf{w}, \mathbf{x})$, with the weighted mode-myriad objective function $G_0(x_j, \mathbf{w}, \mathbf{x})$ defined as

$$G_0(x_j, \mathbf{w}, \mathbf{x}) \;\equiv\; G_0(x_j) \;\triangleq\; \frac{\prod_{i=1, i\neq j}^{N} |x_i - x_j|}{\sqrt{w_j}}. \tag{15}$$

From (15), we see that $G_0(x_j)$ is small if w_j is large (which means that x_j is being emphasized) or if $\prod_{i=1, i\neq j}^{N} |x_i - x_j|$ is small (which happens when many of the x_i are close to x_j). Since $G_0(x_j)$ has to be the smallest for the filter output to be x_j, it is clearly seen that the filter favors input samples (having significant weights) that are clustered together. For the example of Fig. 1, the WMyF$_0$ output is $x_{(3)} = 4.7$, which is part of the cluster of samples $x_{(3)}$ and $x_{(4)}$.

3 Filter Optimization

In this section, we address the problem of *optimization* of the filter parameters of weighted myriad filters for the case when the linearity parameter K satisfies $K > 0$. The filters are designed to optimally estimate a desired signal according to a *statistical error criterion*. Although we focus on the mean absolute error (MAE) criterion, our solutions are applicable to the mean square error (MSE) criterion with trivial modifications.

3.1 Problem Statement

Given an input (observation) vector $\mathbf{x} \triangleq [x_1, x_2, \ldots, x_N]^T$, a weight vector $\mathbf{w} \triangleq [w_1, w_2, \ldots, w_N]^T$ and linearity parameter K, denote the weighted myriad filter output as $y \equiv y_K(\mathbf{w}, \mathbf{x})$, sometimes abbreviated as $y(\mathbf{w}, \mathbf{x})$. The filtering error, in estimating a desired signal d, is then defined as $e = y - d$. Under the mean absolute error (MAE) criterion, we define the cost function

$$J_1(\mathbf{w}, K) \triangleq E\{|e|\} = E\{|y_K(\mathbf{w}, \mathbf{x}) - d|\}, \tag{16}$$

where $E\{\cdot\}$ represents statistical expectation. The mean square error (MSE) is defined as

$$J_2(\mathbf{w}, K) \triangleq E\left\{e^2\right\} = E\left\{(y_K(\mathbf{w}, \mathbf{x}) - d)^2\right\}. \tag{17}$$

When the error criterion adopted is clear from the context, the cost function is written as $J(\mathbf{w}, K)$. Further, we see from (10) and (11) that the optimal filtering action is independent of K (the filter weights can be scaled to keep the output invariant to changes in K). The cost function is therefore sometimes written simply as $J(\mathbf{w})$, with an assumed arbitrary choice of K. With the constraint of non-negative weights, the optimization problem is stated as follows:

$$\begin{array}{c} \text{minimize } J(\mathbf{w}, K) \\ \text{subject to } w_i \geq 0, \; i = 1, 2, \ldots, N. \end{array}$$

This is a nonlinear optimization problem with inequality constraints. Obtaining conditions for a global minimum that are both necessary and sufficient is quite a formidable task. We restrict ourselves to finding only necessary conditions.

3.2 Conditions for Optimality

The cost functions defined in (16) and (17) appear to be non-convex in the weights and thus are likely to have multiple local minima. Assuming that the optimal weights are at one of the local minima, we derive necessary conditions for optimality by equating the gradient of the cost function, with respect to the weights, to zero. Differentiating the MAE cost function in (16) with respect to the weight w_i results in

$$\begin{aligned} \frac{\partial J_1(\mathbf{w}, K)}{\partial w_i} &= \frac{\partial}{\partial w_i} E\{|y_K(\mathbf{w}, \mathbf{x}) - d|\} \\ &= E\left\{\operatorname{sgn}(y - d) \frac{\partial y}{\partial w_i}\right\}, \end{aligned} \tag{18}$$

where

$$\operatorname{sgn}(x) = \begin{cases} +1, & x > 0 \\ 0, & x = 0 \\ -1, & x > 0 \end{cases}$$

is the *sign* function. For the MSE cost function of (17), we obtain

$$\frac{\partial J_2(\mathbf{w}, K)}{\partial w_i} = 2\, E\left\{ (y - d) \frac{\partial y}{\partial w_i} \right\}. \tag{19}$$

The necessary conditions for filter optimality are then stated for the MAE as

$$E\left\{ \operatorname{sgn}(y - d) \frac{\partial y}{\partial w_i} \right\} = 0,\ w_i \ \geq\ 0,\ i = 1, 2, \ldots, N \tag{20}$$

and can similarly be written for the MSE. We therefore need an expression for $\frac{\partial y}{\partial w_i}$, the partial derivative of the filter output $y_K(\mathbf{w}, \mathbf{x})$ with respect to the weight w_i while holding K, the rest of the weights, and the input vector $\mathbf{x}$, constant.

3.3 Optimal Weighted Myriad Filter

Referring to (5) and (6), the weighted myriad filter output for $K > 0$, $y \equiv y_K(\mathbf{w}, \mathbf{x})$, satisfies

$$G'(y) = \left. \frac{\partial G_K(\beta, \mathbf{w}, \mathbf{x})}{\partial \beta} \right|_{\beta = y} = 0. \tag{21}$$

From (8), we obtain

$$G'(y) = 2\, G_K(y, \mathbf{w}, \mathbf{x}) \sum_{j=1}^{N} \frac{w_j (y - x_j)}{K^2 + w_j (y - x_j)^2}. \tag{22}$$

In order to find $\frac{\partial y}{\partial w_i}$, for a given K, we hold the other weights and the input vector $\mathbf{x}$, constant. To emphasize this, rewrite (22), suppressing the quantities that are held constant, as follows:

$$G'(y) = 2\, G(y, w_i)\, H(y, w_i), \tag{23}$$

where $G(y, w_i) \triangleq G_K(y, \mathbf{w}, \mathbf{x})$ and

$$H(y, w_i) \triangleq \sum_{j=1}^{N} \frac{w_j (y - x_j)}{K^2 + w_j (y - x_j)^2}. \tag{24}$$

Using the above definitions in (21), the filter output satisfies

$$G(y, w_i)\, H(y, w_i) = 0 \tag{25}$$

or, since $G(y, w_i) > 0$,

$$H(y, w_i) = 0. \tag{26}$$

In (25) and (26), we have *implicit* formulations of the filter output y as a function of w_i with the other weights and $\mathbf{x}$ fixed. To obtain an expression for $\frac{\partial y}{\partial w_i}$, we differentiate (26) implicitly with respect to w_i:

$$\left(\frac{\partial H}{\partial y}\right) \cdot \left(\frac{\partial y}{\partial w_i}\right) + \left(\frac{\partial H}{\partial w_i}\right) \;=\; 0. \tag{27}$$

Thus,

$$\frac{\partial y}{\partial w_i} \;=\; -\,\frac{\left(\dfrac{\partial H}{\partial w_i}\right)}{\left(\dfrac{\partial H}{\partial y}\right)}. \tag{28}$$

We therefore need expressions for $\frac{\partial H}{\partial y}$ and $\frac{\partial H}{\partial w_i}$. Before finding these, we digress briefly to investigate the quantity $\frac{\partial H}{\partial y}$; this will be useful later in the paper. Rewrite (23) using compressed notation as

$$G^{'}(y) \;=\; 2\,G(y)\;H(y, w_i). \tag{29}$$

Differentiating this with respect to y, we have

$$\begin{aligned} G^{''}(y) \;&=\; 2\,G^{'}(y)\;H(y, w_i) \;+\; 2\,G(y)\,\frac{\partial H}{\partial y} \\ &=\; 2\,G(y)\,\frac{\partial H}{\partial y} \end{aligned} \tag{30}$$

where the second step is because $G^{'}(y) \;=\; 0$ from (21). Note that $G^{''}(y)$ is defined as

$$G^{''}(y) \;\triangleq\; \left.\frac{\partial^2 G_K(\beta, \mathbf{w}, \mathbf{x})}{\partial \beta^2}\right|_{\beta \,=\, y}. \tag{31}$$

From (30), we have

$$\frac{\partial H}{\partial y} \;=\; \frac{1}{2}\,\frac{G^{''}(y)}{G(y)}. \tag{32}$$

Note that, since y is a local minimum of $G(\cdot)$, the second derivative is non-negative: $G^{''}(y) \geq 0$. Further, $G(y) > 0$ always. Therefore,

$$\frac{\partial H}{\partial y} \;\geq\; 0, \tag{33}$$

a fact of great significance that will be used later in Section 4.2.

Returning to our main task of finding $\frac{\partial y}{\partial w_i}$ in (28), we evaluate $\frac{\partial H}{\partial w_i}$ and $\frac{\partial H}{\partial y}$ using (24) as follows:

$$\begin{aligned} \frac{\partial H}{\partial w_i} \;&=\; \frac{\partial}{\partial w_i}\sum_{j=1}^{N} \frac{w_j (y - x_j)}{K^2 + w_j (y - x_j)^2} \\ &=\; \left(\frac{1}{K^2}\right) \cdot \left[\frac{(y - x_i)}{\left(1 + \dfrac{w_i}{K^2}\,(y - x_i)^2\right)^2}\right] \end{aligned} \tag{34}$$

and

$$\begin{aligned}\frac{\partial H}{\partial y} &= \frac{\partial}{\partial y}\sum_{j=1}^{N}\frac{w_j(y-x_j)}{K^2+w_j(y-x_j)^2}\\ &= \sum_{j=1}^{N}\frac{w_j}{K^2}\cdot\frac{1-\dfrac{w_j}{K^2}(y-x_j)^2}{\left(1+\dfrac{w_j}{K^2}(y-x_j)^2\right)^2}.\end{aligned} \tag{35}$$

Finally, using (34) and (35) in (28), we obtain the following expression for $\frac{\partial y}{\partial w_i}$:

$$\frac{\partial y}{\partial w_i} = \frac{\left[\dfrac{-(y-x_i)}{\left(1+\dfrac{w_i}{K^2}(y-x_i)^2\right)^2}\right]}{K^2\cdot\left[\displaystyle\sum_{j=1}^{N}\frac{w_j}{K^2}\cdot\frac{1-\dfrac{w_j}{K^2}(y-x_j)^2}{\left(1+\dfrac{w_j}{K^2}(y-x_j)^2\right)^2}\right]} \tag{36}$$

which we can now use in (20) to obtain the necessary conditions for the optimal weighted myriad filter under the MAE criterion:

$$E\left\{\operatorname{sgn}(y-d)\,\frac{\left[\dfrac{(y-x_i)}{\left(1+\dfrac{w_i}{K^2}(y-x_i)^2\right)^2}\right]}{\left[\displaystyle\sum_{j=1}^{N}w_j\frac{1-\dfrac{w_j}{K^2}(y-x_j)^2}{\left(1+\dfrac{w_j}{K^2}(y-x_j)^2\right)^2}\right]}\right\} = 0,\ w_i \geq 0,\ i=1,2,\ldots,N. \tag{37}$$

Note that the necessary conditions for the optimal filter under the MSE criterion can be easily found by using (36) in (19); the only change we need to make in (37) is to replace $\operatorname{sgn}(y-d)$ by $(y-d)$. Note also that as $K\to\infty$ in (37), while keeping the weights finite, we obtain

$$E\left\{\operatorname{sgn}(y_\infty-d)\,\frac{(y_\infty-x_i)}{\displaystyle\sum_{j=1}^{N}w_j}\right\} = 0,\ w_i \geq 0,\ i=1,2,\ldots,N,$$

which can be shown to be the conditions for the optimal weighted mean filter under the MAE criterion. This is consistent with the fact, as shown in (12), that the weighted myriad approaches the weighted mean as $K\to\infty$.

4 Adaptive Filtering Algorithms

The necessary conditions for optimality, derived in Section 3 (see (37)), involve expressions that are very complicated. In attempting to solve for the optimal weights, we encounter two problems. First, we require knowledge of the joint statistics of all the signals involved. Even with this knowledge, it is almost impossible to evaluate (in closed-form) the statistical expectations entering into the optimality conditions. Second, even if we could write down the equations in closed-form, solving the resulting highly nonlinear equations for the optimal weights would be a formidable task. We therefore adopt the approach of adaptive optimization of the filter weights. In situations where the statistics of the signals are unknown or time-varying, the use of adaptive algorithms is frequently the only recourse available.

4.1 General Formulation

In order to find the optimal filter weights, we minimize the MAE cost function $J(\mathbf{w})$ using the *steepest descent* method. Noting that the weights are constrained to be non-negative, we obtain the following algorithm to update the filter weights:

$$w_i(n+1) \ = \ P\left[w_i(n) - \mu \frac{\partial J}{\partial w_i}(n)\right], \ i = 1, 2, \ldots, N \tag{38}$$

where $w_i(n)$ denotes the ith weight at the nth iteration, $\mu > 0$ is the step-size of the update, and $P[\cdot]$, defined by

$$P[u] \ \triangleq \ \begin{cases} u, & u > 0 \\ 0, & u \leq 0, \end{cases} \tag{39}$$

projects the updated weight onto the constraint space of the weights. In practice, $P[u]$ is set to a small positive value ϵ if $u \leq 0$. Note that the cost function $J(\mathbf{w})$ could have many local minima and the above algorithm does not guarantee convergence to the global minimum. One way to tackle this problem is to run the algorithm with several different initial weight vectors $\mathbf{w}(0)$ and choose the best final weights from the different runs. The gradient $\frac{\partial J}{\partial w_i}(n)$ is given from (18) as

$$\frac{\partial J}{\partial w_i}(n) \ = \ E\left\{\operatorname{sgn}(y(n) - d(n)) \, \frac{\partial y}{\partial w_i}(n)\right\}. \tag{40}$$

Since the lack of knowledge of the signal statistics precludes the evaluation of the statistical expectation in (40), we use *instantaneous estimates* for the gradient just as in the LMS algorithm [19].

To this end, removing the expectation operator in (40) and substituting into (38), we have

$$w_i(n+1) \;=\; P\left[w_i(n) \;-\; \mu\; \mathrm{sgn}(e(n))\; \frac{\partial y}{\partial w_i}(n)\right],\; i = 1, 2, \ldots, N \tag{41}$$

where $e(n) \;=\; y(n) - d(n)$ is the error at the nth iteration.

4.2 Adaptive Weighted Myriad Filter Algorithms

For the weighted myriad filter, the expression for $\frac{\partial y}{\partial w_i}(n)$ is given by (36). Using this in (41), we obtain the following adaptive algorithm for updating the filter weight w_i:

Adaptive Weighted Myriad Filter Algorithm I

$$w_i(n+1) \;=\; P\left[w_i(n) \;+\; \mu\, \mathrm{sgn}(e(n))\, \frac{\left\{\dfrac{(y - x_i)}{\left(1 + \dfrac{w_i}{K^2}\,(y - x_i)^2\right)^2}\right\}(n)}{K^2 \,.\, \left\{a \;+\; \displaystyle\sum_{j=1}^{N} \frac{w_j}{K^2} \cdot \dfrac{1 - \dfrac{w_j}{K^2}\,(y - x_j)^2}{\left(1 + \dfrac{w_j}{K^2}\,(y - x_j)^2\right)^2}\right\}(n)}\right], \tag{42}$$

where $a > 0$ (not present in (36)) is a *stabilizing constant.* In the following, we explain the rationale behind the introduction of this constant. First, note that, for $a = 0$, the update term in (42) is proportional to an estimate of the gradient, $\frac{\partial J}{\partial w_i}(n)$, of the MAE cost function. Recognizing that, in a gradient descent algorithm, the *direction* of the gradient conveys most of the required update information, we can modify the update term by scaling it by any *positive* factor that is *common* to all the weights. This will change the magnitude of the update without affecting the direction of the gradient estimate. Referring to (35), we see that the denominator of the update term in (42) is equal (for $a = 0$) to the quantity $K^2 \frac{\partial H}{\partial y}(n)$ which, from (33), is non-negative and common to the updates of all the weights. This term can lead to numerical problems in a practical implementation of the algorithm. Specifically, when the term $\frac{\partial H}{\partial y}(n)$ is very small, the weight update becomes very large in magnitude. Adding a constant $a > 0$ to the term $\frac{\partial H}{\partial y}(n)$ ensures that it is bounded away from zero. This operation preserves the direction of the current gradient estimate, leaving the final values of the weights unchanged. To choose the value of the stabilizing constant a, note that the update denominator in (42) is $K^2(a + \frac{\partial H}{\partial y}(n))$. By setting $a = \frac{1}{K^2}$, we ensure that, as $\frac{\partial H}{\partial y}(n) \to 0$, Algorithm I reduces to Algorithm II, a simplified algorithm described later in this section.

We note from (10) and (11) that the optimal filtering action is independent of the choice of K; the filter only depends on the value of $\frac{\mathbf{w}}{K^2}$. In this context, we might ask how the algorithm scales as we change the value of K and how we should change the step-size μ and the initial weight vector $\mathbf{w}(0)$ as we vary K. To answer this, let $\mathbf{g}_o \triangleq \mathbf{w}_{o,1}$ denote the optimal weight vector for $K = 1$. Then, from (11), we have $\frac{\mathbf{w}_{o,K}}{K^2} = \frac{\mathbf{g}_o}{(1)^2}$ or $\mathbf{g}_o = \frac{\mathbf{w}_{o,K}}{K^2}$. Now consider two situations. In the first, the algorithm in (42) is used with $K = 1$, step-size $\mu = \mu_1$, weights denoted as $g_i(n)$ and initial weight vector $\mathbf{g}(0)$. This is expected to converge to the weights $\mathbf{g}_o$. In the second, we use the algorithm with a general value of K, step-size $\mu = \mu_K$ and initial weight vector $\mathbf{w}_K(0)$. Rewrite (42) by dividing throughout by K^2 and writing the algorithm in terms of an update of $\frac{w_i}{K^2}$. This is expected to converge to $\frac{\mathbf{w}_{o,K}}{K^2}$ since (42) should converge to $\mathbf{w}_{o,K}$. Since $\mathbf{g}_o = \frac{\mathbf{w}_{o,K}}{K^2}$, we can compare the above two situations and choose the initial weight vector $\mathbf{w}_K(0)$ and the step-size μ_K such that the algorithms have the *same behaviour* in both cases and converge, as a result, to the *same filter*. This means that $g_i(n) = \frac{w_i(n)}{K^2}$ at each iteration n. It can be shown that this results in

$$\mu_K = K^4 \mu_1 \text{ and } \mathbf{w}_K(0) = K^2 \mathbf{w}_1(0). \tag{43}$$

This also implies that if we change K from K_1 to K_2, the new parameters should satisfy

$$\mu_{K_2} = \left(\frac{K_2}{K_1}\right)^4 \mu_{K_1} \text{ and } \mathbf{w}_{K_2}(0) = \left(\frac{K_2}{K_1}\right)^2 \mathbf{w}_{K_1}(0). \tag{44}$$

Considerable simplification of the algorithm in (42) can be achieved by just removing the denominator from the update term; this does not change the direction of the gradient estimate or the values of the final weights. This leads to the following computationally attractive algorithm:

Adaptive Weighted Myriad Filter Algorithm II

$$w_i(n+1) = P\left[w_i(n) + \mu \operatorname{sgn}(e(n)) \left\{\frac{(y - x_i)}{\left(1 + \frac{w_i}{K^2}(y - x_i)^2\right)^2}\right\}(n)\right]. \tag{45}$$

Note that, apart from the effort involved in computing the filter output $y(n)$, the above algorithm involves a very simple update that is computationally comparable to the update in the LMS algorithm or its variant, the LMAD (least mean absolute deviation) algorithm (also called the *sign algorithm* (SA)), which is written as $w_i(n+1) = w_i(n) - \mu \operatorname{sgn}(e(n))\, x_i(n)$ [22]. In our simulations, the simplified algorithm of (45) converged significantly faster than the algorithm of (42). To

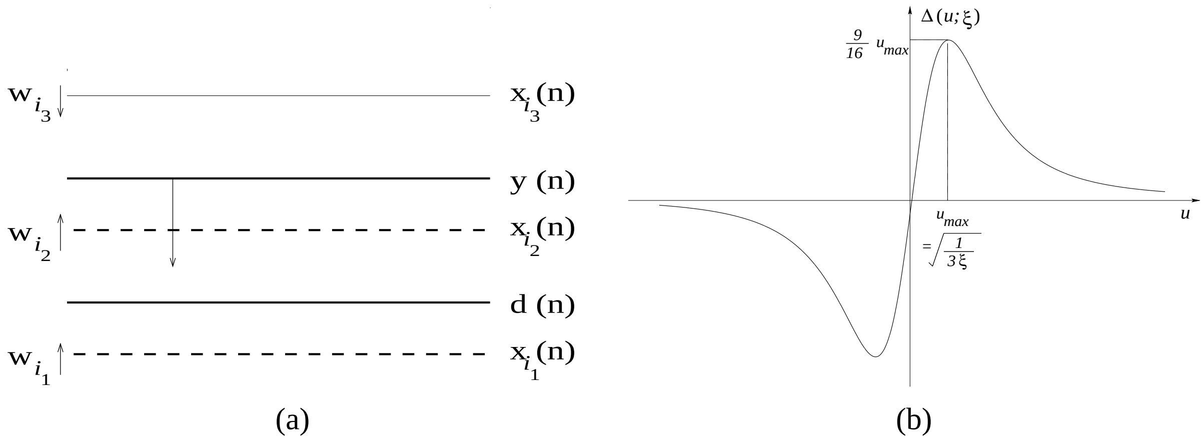

Figure 2: (a) Operation of Adaptive Weighted Myriad Filter Algorithm II, (b) Update magnitude function $\Delta(u;\,\xi)$.

understand the operation of Algorithm II, rewrite (45) as

$$w_i(n+1) \;=\; P\left[w_i(n) \;+\; \mu\,\varepsilon(n)\,\delta(n)\right] \tag{46}$$

where $\varepsilon(n) \;\stackrel{\triangle}{=}\; \mathrm{sgn}(e(n)) \;=\; \mathrm{sgn}\,(y(n)-d(n))$ and

$$\delta(n) \;\stackrel{\triangle}{=}\; \Delta\left[y(n)-x_i(n);\; \frac{w_i(n)}{K^2}\right], \tag{47}$$

with

$$\Delta[u;\,\xi] \;\stackrel{\triangle}{=}\; \frac{u}{(1+\xi u^2)^2}\;. \tag{48}$$

The operation of the algorithm is illustrated in Fig. 2(a). Referring to (46), assume that $e(n) > 0$, i.e. $d(n) < y(n)$ at the current iteration, so that $\varepsilon(n) = +1$. Since $\mu > 0$, we see from (47) and (48) that the weights are increased (positive update: $\delta(n) > 0$) for those i for which $x_i(n) < y(n)$. The remaining weights are reduced. In Fig. 2(a), for example, the weights are increased for $i = i_1$ and $i = i_2$ while the weight w_{i_3} is decreased. Considering the case $e(n) < 0$ also, we can conclude that the filter weights w_i are *increased* for those samples x_i that are on the *same side of the current output estimate as the desired signal.* The effect of increasing a weight w_i is to move the filter output towards x_i. Therefore, referring again to Fig. 2(a), we see that the algorithm *moves the filter output towards the samples that are closer to the desired signal.*

The magnitude of the update is determined by the term $\delta(n) = \Delta(u_i(n);\,\xi_i(n))$ with $u_i(n) = y(n) - x_i(n)$ and $\xi_i(n) = \frac{w_i(n)}{K^2}$. Fig. 2(b) shows the function $\Delta(u;\,\xi)$. This is an odd function of u that is approximately linear for small u and goes to zero for large u. It attains a maximum

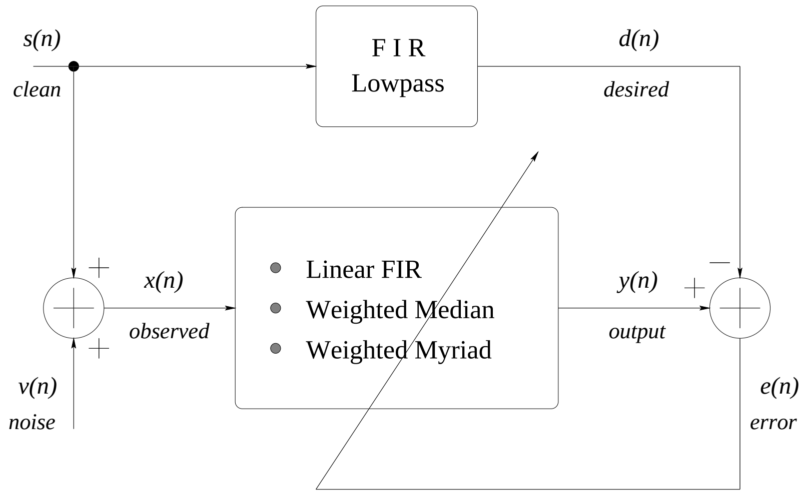

Figure 3: Training system for adaptive filter optimization.

at $u = u_{max} = \sqrt{\frac{1}{3\,\xi}}$ and the peak value is a constant times u_{max}. The value of u_{max} is a rough measure of the *outlier rejection* ability of the function; values much larger than u_{max} are attenuated.

Based on the properties of $\Delta(u;\,\xi)$, we can describe the behaviour of the update magnitude as follows. For samples $x_i(n)$ near the output $y(n)$ ($|u_i(n)| \ll u_{max}$), the update of $w_i(n)$ is approximately linear in $u_i(n)$. Samples far from the current estimate $y(n)$ are treated as outliers and have a negligible effect on the corresponding weights (the update tends to zero for large $|u_i(n)|$). The location of the peak $u_{i,max}(n)$ is inversely related to $\xi_i(n) = \frac{w_i(n)}{K^2}$. When the weights $w_i(n)$ are small in relation to K^2, the $\xi_i(n)$ are small and the $u_{i,max}(n)$ are large. As a result, most of the weights tend to be updated (being in the near-linear portion of $\Delta(u;\,\xi)$). When the weights are large, the $u_{i,max}(n)$ are small. This leads to negligible updates. Thus, the algorithm is robust to outliers and also allows the weights to settle down. It is interesting to note that the function $\Delta(u;\,\xi)$ is related to the *influence function* of the myriad estimator (see [5, 6] for discussions on the influence function of an M-estimator). The influence function determines the robustness of an estimator; this is precisely what we see in the operation of the above adaptive algorithm.

5 Simulation Results

The adaptive algorithms developed in Section 4.2 were evaluated through a computer simulation example involving lowpass filtering of a one-dimensional chirp-type signal corrupted by α-stable noise. Fig. 3 shows the block diagram representing our simulation example. The *desired signal* $d(n)$

is obtained by filtering the *clean signal* $s(n)$ using a linear FIR lowpass filter obtained using standard FIR filter design techniques for a chosen cutoff frequency. The signal $s(n)$ is then corrupted by an additive *noise process* $v(n)$ to yield the *input* or *observed signal* $x(n)$. The objective of the adaptive filtering algorithms is to train the linear FIR, weighted median and weighted myriad filters to converge to filter parameters (weights) so as to minimize the absolute value of the *error signal* $e(n)$ between the filter *output signal* $y(n)$ and the *desired signal* $d(n)$. In this section, we present the results of this training process, using learning curves and filter weight trajectories to demonstrate the convergence of the various adaptive algorithms. We also compare the performance of the trained filters by applying them to a noisy test signal.

In our simulation example, the clean signal $s(n)$ was chosen to be a *chirp-type* signal, a digital sinusoid with quadratically increasing instantaneous frequency. Specifically, the signal, of length $L = 256$, is given by $s(n) = sin(\omega(n)\, n), n = 0, 1, \ldots, L-1$, where the radian frequency is $\omega(n) = \frac{\pi}{3} \cdot \frac{L}{L-1} \left(\frac{n}{L-1}\right)^2$. The desired signal $d(n)$ was obtained by passing $s(n)$ through an FIR lowpass filter of window length $N = 11$, designed for a cutoff frequency $\omega_c = \frac{\pi}{50}$. The weights of the designed filter are shown in Table 1, in the column entitled 'Lowpass FIR'. Fig. 4(a) shows the chirp-type signal $s(n)$ and the desired signal $d(n)$ is shown in Fig. 4(b). The signal $s(n)$ is corrupted by adding a realization of symmetric zero-mean α-stable noise, yielding the noisy observed signal $x(n)$ shown in Fig. 4(c). The additive α-stable noise process simulates low-level Gaussian-type noise along with impulsive interference. The result of lowpass filtering $x(n)$ with the previously designed FIR filter is shown as the signal $y_{lpfir}(n)$ in Fig. 4(d). Clearly, the performance of the FIR lowpass filter is severely affected by the impulses in $x(n)$ and the output $y_{lpfir}(n)$ is far from the desired signal $d(n)$ of Fig. 4(b).

The linear, weighted median and weighted myriad filters were trained using the *training signal* $x(n)$ and the desired signal $d(n)$, each of length $L = 256$. In all cases, the filter window length was chosen to be $N = 11$ in order to ensure a fair comparison with the designed lowpass FIR filter. Since the number of iterations required for convergence was more than $L = 256$, the adaptive algorithms were implemented by multiple passes (140 loops) through the signals, for a total of 34440 iterations. For the linear filter, the following least mean absolute deviation (LMAD) algorithm, also called the *sign algorithm* (SA) [22], was used: $w_i(n+1) = w_i(n) - \mu\, \mathrm{sgn}(e(n))\, x_i(n)$, $i = 1, 2, \ldots, N$, where we use the notation of Section 4. The adaptation of the weighted median filter utilized an adaptive

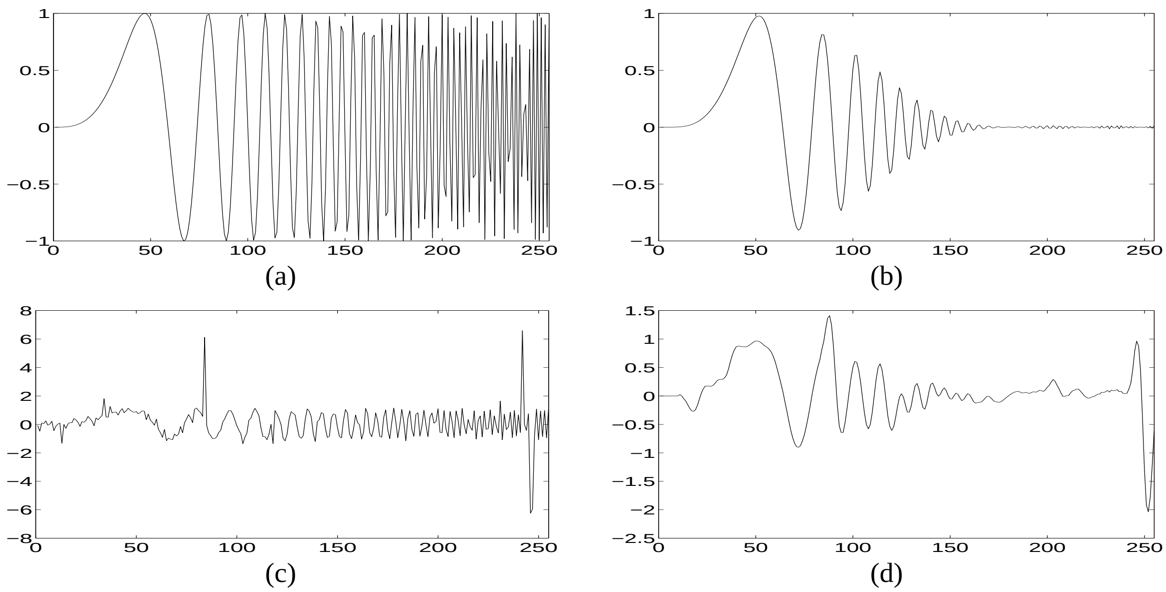

Figure 4: (a) $s(n)$: clean chirp-type signal, (b) $d(n)$: desired signal (lowpass FIR filtering of $s(n)$), (c) $x(n)$: noisy chirp-type signal (additive α-stable noise: characteristic exponent $\alpha = 1.4$, dispersion $\gamma = 0.1$), (d) $y_{lpfir}(n)$: lowpass FIR filtering of $x(n)$.

weighted order statistic (WOS) algorithm from [23]. The parameters of the WOS filter are a set of N non-negative filter weights w_i, $i = 1, 2, \ldots, N$ and a non-negative *threshold* w_0 which, for the weighted median filter, is simply a function of the weights: $w_0 = \frac{\sum_{i=1}^{N} w_i \, + \, 1}{2}$. The adaptive algorithm updates the filter weights as $w_i(n+1) \;=\; P\left[w_i(n) \;-\; 2.0\, \mu\, e(n)\, U(x_i(n) - y(n))\right],\ i = 1, 2, \ldots, N$, where $U(\cdot)$ is the unit step function and $P[\cdot]$ is the projection operator defined in (39). For the weighted myriad filter, Algorithms I and II ((42) and (45)) of Section 4.2 were implemented. The linearity parameter was chosen as $K = 1.0$; recall from (10) and (11) that, in optimizing the filter weights, the choice of K is arbitrary. Algorithm I was implemented both without the so-called stabilization constant (i.e., $a = 0$) and with it ($a = \frac{1}{K^2} = 1.0$ as recommended in Section 4.2). The weighted myriad filter output in all cases was computed using the fixed point search algorithm described in [21].

The initial weights for the linear filter were all chosen to be zero: $\mathbf{w}(0) = \mathbf{0}$. For all the other algorithms, the initial filter was chosen to be the identity filter with all the weights set to zero except the center weight which was set to 10.0: $\mathbf{w}(0) = [0, 0, 0, 0, 0, 10.0, 0, 0, 0, 0, 0]$. The step-sizes of the algorithms were chosen as follows. Among the weighted myriad filter algorithms, the step-sizes were chosen to achieve approximately the same final mean absolute error (MAE). The step-sizes

Weights	Lowpass	Linear	Weighted	Weighted Myriad ($K = 1.0$)		
w_i	FIR	LMAD	Median	Algo. I (a = 0)	Algo. I (a = 1.0)	Algo. II
w_1	0.0144	0.0299	6.4630	0.2997	0.1855	0.2886
w_2	0.0304	0.0338	7.1727	0.2594	0.2768	0.4154
w_3	0.0724	0.0638	7.7844	0.5495	0.4413	0.5227
w_4	0.1245	0.0962	10.9985	0.5031	0.5198	0.6397
w_5	0.1668	0.1356	13.6322	0.7577	0.7152	0.9361
w_6	0.1830	0.1571	14.5779	0.7964	0.7575	0.8760
w_7	0.1668	0.1283	12.3102	0.7960	0.7413	0.9562
w_8	0.1245	0.1145	11.2896	0.5698	0.5511	0.6250
w_9	0.0724	0.0540	7.4682	0.4111	0.3723	0.4537
w_{10}	0.0304	0.0259	7.2095	0.3106	0.2777	0.3902
w_{11}	0.0144	0.0071	5.8129	0.3095	0.2084	0.2818

Table 1: Filter weights obtained by the adaptive algorithms.

were $\mu = 1.0 \times 10^{-1}$ for Algorithm I (both $a = 0$ and $a = 1.0$) and $\mu = 5.0 \times 10^{-2}$ for Algorithm II. The step-sizes for the linear filter ($\mu = 1.0 \times 10^{-4}$) and the weighted median filter ($\mu = 5.0 \times 10^{-3}$) were chosen so that these algorithms converged in approximately the same number of iterations as the fastest weighted myriad filter algorithm (which was Algorithm II).

The final filter weights obtained by the various algorithms are shown in Table 1. The three weighted myriad filter algorithms converged to approximately the same weight vectors. Hence, they achieved almost the same final MAEs (the step-sizes were chosen to ensure this). This permits a meaningful comparison of their convergence speeds, all other factors being equal.

The trajectories of the filter weights for the various weighted myriad filter algorithms are shown in Fig. 5. Out of the $N = 11$ filter weights, we have chosen the weights w_i, $i = 2, 5, 6$ and 8 to illustrate the weight trajectories for all the algorithms. Note that, in all cases, our choice of initial weights implies $w_2(0) = w_5(0) = w_8(0) = 0.0$ and the center weight $w_6(0) = 10.0$. We see from Fig. 5 that, in all three algorithms, the weight curves $w_2(n)$, $w_5(n)$ and $w_8(n)$ are non-monotonic, while the weight $w_6(n)$ is monotonically decreasing. The reason for this behaviour is the initial large value $w_6(0) = 10.0$ of the center weight, which pulls the other weights up from their initial zero values. The off-center weights continue to increase until the center weight decreases sufficiently; after that, the off-center weights also decrease monotonically (except for the isolated jumps, explained later, in the case of Fig. 5(A)). If *all* the weights were initialized to zero, the weight curves would all be monotonically increasing.

Referring to Fig. 5(A) (Algorithm I, $a = 0$), notice the jumps in the weight trajectories around

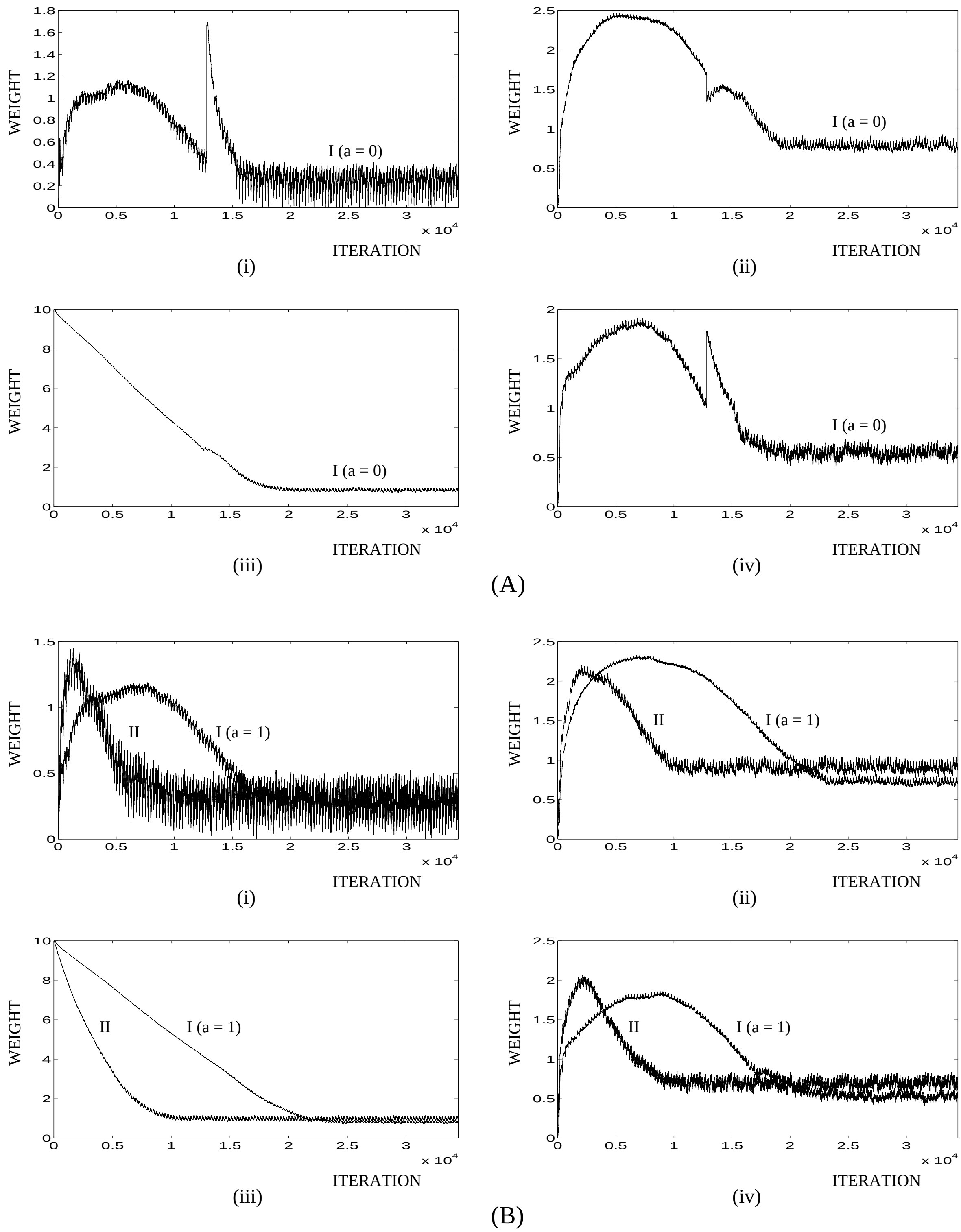

Figure 5: Weight trajectories $w_i(n)$ for Adaptive Weighted Myriad Filter Algorithms ((A): Algorithm I (a = 0), (B) Algorithms I (a = 1.0) and II): (i) $w_2(n)$, (ii) $w_5(n)$, (iii) $w_6(n)$, (iv) $w_8(n)$.

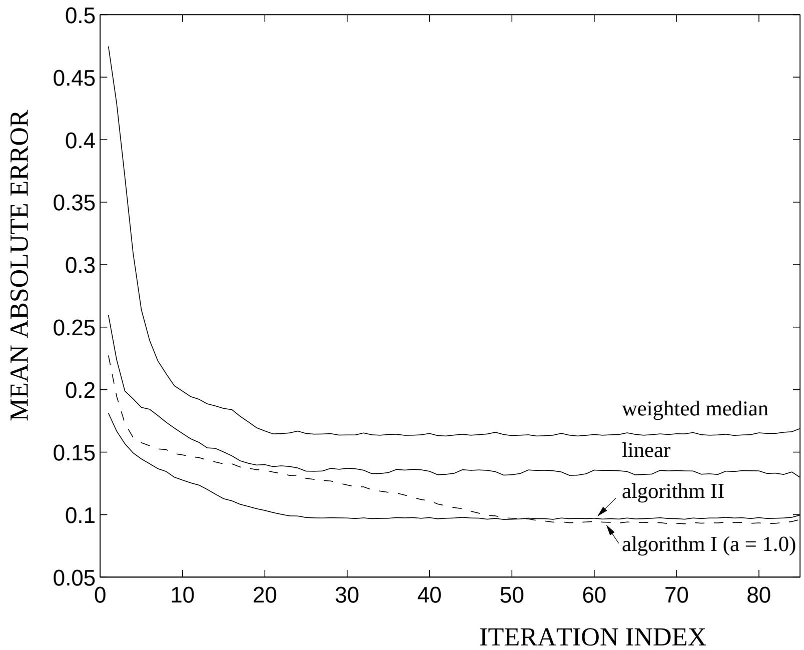

Figure 6: Time-averaged MAE learning curves.

iteration 13000. As mentioned in Section 4.2, the jumps are due to the very small value of the denominator of the update term in (42). A closer investigation reveals that this iteration step corresponds to an observation window in the signal that contained an impulse. The jumps in the weight curves occur due to a combination of two factors: the very small magnitude of the update denominator, and the occurrence of the impulse. Note that without the impulse, the denominator term could become very small, but not abruptly; there would be no jumps in the weight curves in such a case. In practice, the presence of the stabilizing constant $a > 0$ is essential, in order to avoid the occurrence of these jumps. As Fig. 5(B) shows, Algorithm I ($a = 1.0$) completely eliminates the jumps in the weight curves, while converging at a slightly slower rate (24000 iterations, compared to 20000 for Algorithm I ($a = 0$)). We also see from Fig. 5(B) that Algorithm II converges in about 10000 iterations, significantly faster than Algorithm I. Algorithm II also has no jumps in the weight curves. From the above results, we see that Algorithm II, which is computationally the most efficient among the three algorithms, also has superior convergence at comparable MAEs.

Fig. 6 shows time-averaged learning curves in terms of the MAE for the different adaptive algorithms. The MAE learning curve of an adaptive algorithm is a plot of the evolution of the absolute value of the filtering error as a function of the training iterations. Usually, ensemble-averaged

Filter Type		MAE		MSE	
		Training	Test	Training	Test
Lowpass FIR		0.1380	0.1993	0.0971	0.1547
Linear LMAD		0.1282	0.1813	0.0668	0.1141
Weighted Median		0.1593	0.1625	0.0541	0.0554
Weighted Myriad	Algo. I (a = 0)	0.0968	0.0962	0.0194	0.0193
	Algo. I (a = 1.0)	0.0910	0.0903	0.0162	0.0160
	Algo. II	0.0959	0.0947	0.0187	0.0185

Table 2: Mean absolute error (MAE) and mean square error (MSE) incurred in filtering the training and test signals.

learning curves are plotted by averaging the learning curves of a large number of independent trials of the adaptation experiment. We have chosen to *time-average* and further smooth the learning curves of a single trial in order to obtain the learning trends. Thus, in Fig. 6, each iteration index corresponds to averaging (and further smoothing) the absolute error over 400 iterations. We see that the weighted median algorithm has the highest final MAE. The weighted median filter, being a selection filter, is not well-suited to the present application; the linear filter has a better performance, as the figure shows. The weighted myriad filter algorithms I ($a = 1.0$) and II achieve almost the same, and the smallest, final MAEs, demonstrating the robustness of this filter in impulsive noise. The figure also shows that the linear, weighted median and weighted myriad filter algorithm II, all converge in about the same time (recall that the step-sizes were chosen to ensure this). Thus, Algorithm II (the simplest, fastest and most practical weighted myriad filter algorithm) achieves a lower MAE than the linear and weighted median algorithms, at comparable convergence speeds.

Table 2 shows the mean absolute errors (MAEs) and mean square errors (MSEs) incurred in filtering the noisy chirp-type training signal $x(n)$ of Fig. 4(c) with the various trained filters (see the columns labelled 'Training'). The weighted myriad filters (from all three algorithms) have the best performance in terms of the MAE as well as the MSE. The linear filter, trained on the noisy signal, has a lower MAE and MSE than the designed lowpass FIR filter. The trained weighted median filter has a higher MAE than even the lowpass FIR filter, but achieves a slightly lower MSE than the linear LMAD-trained filter.

In order to test the performance of the various trained filters, they were applied to another noisy chirp-type signal, different from the training signal $x(n)$ of Fig. 4(c). This *test signal* $x^{'}(n)$, shown in Fig. 7(a), was obtained by adding a different realization of noise to the clean chirp-type signal of

Fig. 4(a). The MAE and MSE values, incurred in filtering the test signal $x^{'}(n)$, are listed in Table 2 in the columns labelled 'Test'. The mean errors for the linear (lowpass FIR as well as LMAD-trained) filters increase significantly from the training to the test signal. On the other hand, the weighted median and weighted myriad filters are hardly affected by the change in the additive noise. Fig. 7 shows the results of applying the various filters to the test signal $x^{'}(n)$; the desired signal $d(n)$ of Fig. 4(b) is reproduced in Fig. 7(c) for reference. The output of the designed lowpass FIR filter is shown in Fig. 7(b) and the output of the linear LMAD-trained filter is shown in Fig. 7(d). It is evident that these outputs are quite different from the desired signal; the linear filters are greatly affected by the impulsive nature of the noise. The weighted median filter output of Fig. 7(e) is less affected by the impulses. However, the output is quite distorted, partly because the filter is constrained to be a selection filter. The weighted median filter is also unable to completely remove the high-frequency portions of the chirp-type signal. The outputs of the weighted myriad filters, trained using Algorithm I ($a = 1.0$) and Algorithm II, are shown in Figs. 7(f) and 7(g), respectively. These outputs are visually the closest to the desired signal, especially in the low-frequency portions of the chirp-type signal. The weighted myriad filters have the best outputs and lowest mean errors, while being highly robust to changes in the noise environment.

6 Conclusion

The optimization of *Weighted Myriad Filters* was considered in this paper. Necessary conditions were derived for optimality under the mean absolute error criterion. A stochastic gradient-based adaptive algorithm was developed for learning the optimal filter weights. This was further modified to yield an adaptive algorithm involving a very simple update equation. The performance of the adaptive filters in impulsive environments was investigated through a simulation example involving lowpass filtering a chirp-type signal in α-stable noise. Learning curves and filter weight trajectories served to demonstrate the convergence of the adaptive algorithms. The trained weighted myriad filters achieved lower mean absolute errors than the adaptive linear and weighted median filters, while being highly robust to changes in the noise environment. Theoretical analysis of the convergence of the adaptive weighted myriad filter algorithms is the subject of current research. The optimization of the special case of the *weighted mode-myriad filter* is considered in future publications.

References

[1] M. P. Shinde and S. N. Gupta, "Signal detection in the presence of atmospheric noise in tropics," *IEEE Transactions on Communications*, vol. COM-22, pp. 1055–1063, Aug. 1974.

[2] B. Stuck and B. Kleiner, "A statistical analysis of telephone noise," *Bell Systems Technical Journal*, vol. 53, pp. 1263–1320, Sept. 1974.

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

[4] S. C. Schwartz and J. B. Thomas, "Detection in a non-Gaussian environment," in *Statistical Signal Processing* (E. J. Wegman and J. G. Smith, eds.), New York: Marcel Dekker, 1984.

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

[6] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions.* New York: Wiley, 1986.

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

[8] 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.

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

[10] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes.* New York: Chapman & Hall, 1994.

[11] E. S. Sousa, "Performance of a spread spectrum packet radio network link in a Poisson field of interferers," *IEEE Transactions on Information Theory*, vol. 38, pp. 1743–1754, Nov. 1992.

[12] 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, Georgia), 1996.

[13] J. G. Gonzalez and G. R. Arce, "Weighted myriad filters: A powerful framework for efficient filtering in impulsive environments," *IEEE Transactions on Signal Processing*, 1997. Submitted.

[14] 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.

[15] 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, Georgia), 1996.

[16] 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.

[17] S. Kalluri and G. R. Arce, "Adaptive weighted myriad filter optimization for robust signal processing," in *Proc. of the 1996 CISS*, (Princeton, NJ), 1996.

[18] S. Kalluri and G. R. Arce, "Adaptive algorithms for weighted myriad filter optimization," in *Proc. of the 1997 IEEE ICASSP*, (Munich, Germany), 1997.

[19] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, 1985.

[20] G. Çiydem and A. E. Çetin, "Robust adaptive filtering algorithms for α-stable random processes," in *Proc. of the 1996 CISS*, (Princeton, NJ), 1996.

[21] S. Kalluri and G. R. Arce, "Approximate weighted myriad computation by fixed point search," Tech. Rep. 96-12-2, Department of Electrical Engineering, University of Delaware, 1996.

[22] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "Comparison of the convergence of two algorithms for adaptive FIR digital filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, June 1981.

[23] L. Yin and Y. Neuvo, "Fast adaptation and performance characteristics of FIR–WOS hybrid filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 42, July 1994.

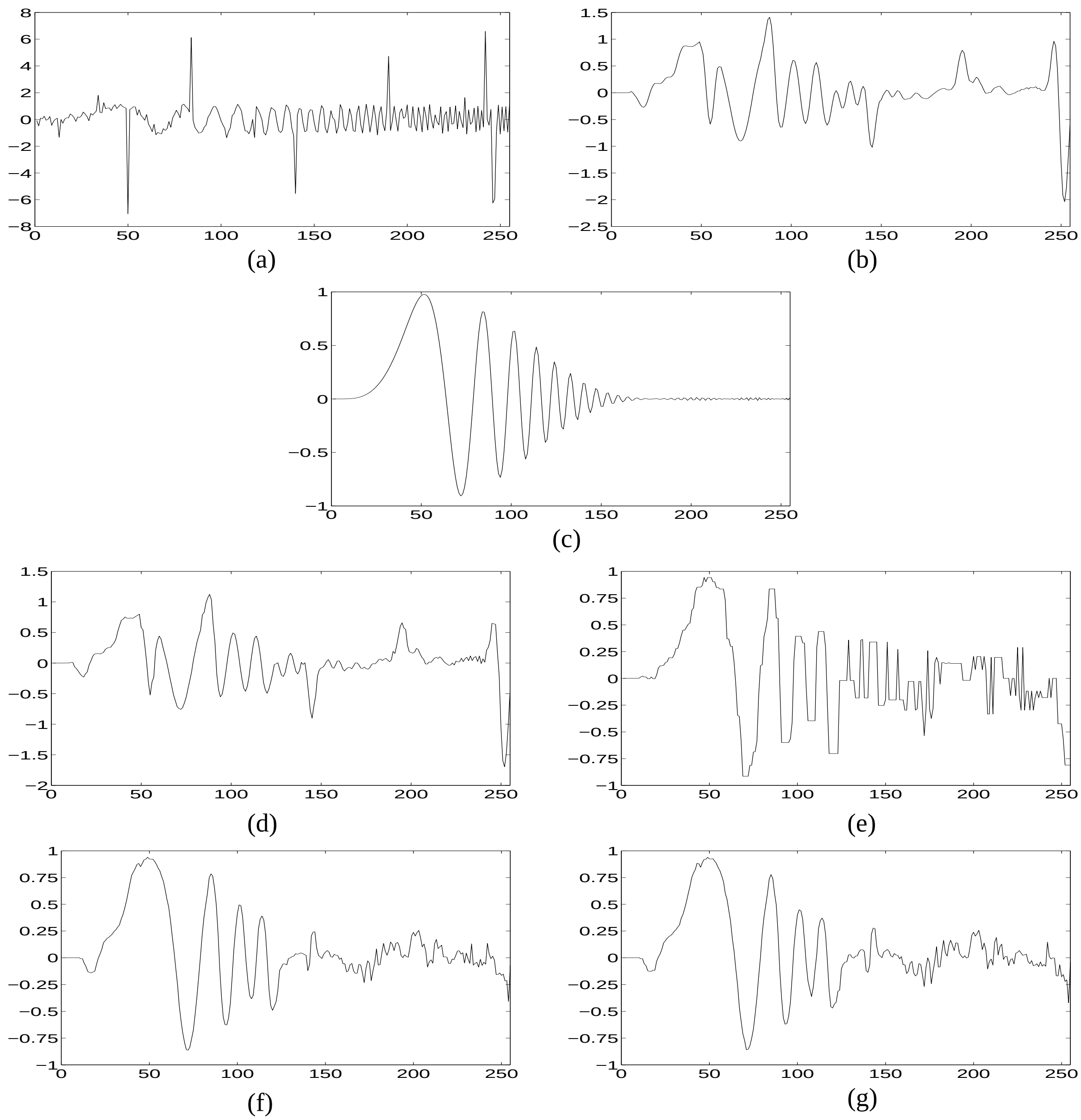

Figure 7: (a) $x^{'}(n)$: noisy chirp-type test signal, (b) $y^{'}_{lpfir}(n)$: lowpass FIR filtering of $x^{'}(n)$, (c) $d(n)$: desired signal, (d) $y^{'}_{lmad}(n)$: linear (LMAD) filter output, (e) $y^{'}_{wm}(n)$: weighted median filter output, (f) $y^{'}_{wmyI}(n)$: weighted myriad filter output (algorithm I (a = 1.0)), (g) $y^{'}_{wmyII}(n)$: weighted myriad filter output (algorithm II).