ADAPTIVE ALGORITHMS FOR WEIGHTED MYRIAD FILTER OPTIMIZATION *

Sudhakar Kalluri *Gonzalo R. Arce*

Department of Electrical Engineering
University of Delaware, Newark, DE 19716, U. S. A
kalluri@ee.udel.edu

ABSTRACT

Stochastic gradient-based adaptive algorithms are developed for the optimization of *Weighted Myriad Filters*, 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. An implicit formulation of the filter output is used to derive an expression for the gradient of the mean absolute error (MAE) cost function, leading to necessary conditions for the optimal filter weights. An adaptive steepest-descent algorithm is then derived to optimize the filter weights. This is modified to yield an algorithm with a very simple weight update, computationally comparable to the update in the classical LMS algorithm. Simulations demonstrate the robust performance of these algorithms.

I INTRODUCTION

Classical statistical signal processing has been dominated by the assumption of the Gaussian model for the underlying signals. However, a large number of physical processes are impulsive in nature and are better-described by heavy-tailed non-Gaussian distributions. Several robust filtering and estimation techniques have been proposed to combat impulsive noise and outliers. *Weighted median filters* have been widely used in image processing due to their ability to preserve edges and reject outliers [1]. *Weighted myriad filters* have been proposed recently [2] motivated by α-stable distributions, which accurately model impulsive processes [3]. The *characteristic exponent* α ($0 < \alpha \leq 2$) of an α-stable distribution controls the heaviness of its tails; a smaller α signifies heavier tails. The special cases $\alpha = 2$ and $\alpha = 1$ yield the Gaussian and Cauchy distributions, respectively.

In this paper, we derive necessary conditions for optimal weighted myriad filters under the mean absolute error (MAE) criterion [4, 5]. An *implicit formulation* of the filter output is used to find an expression for the gradient of the cost function, leading to adaptive steepest-descent algorithms to optimize the filter weights. Simulations, involving lowpass filtering a quadchirp signal in α-stable noise, demonstrate the robust performance of these algorithms in impulsive environments.

*THIS WORK WAS SUPPORTED IN PART BY THE NATIONAL SCIENCE FOUNDATION UNDER GRANT MIP-9530923.

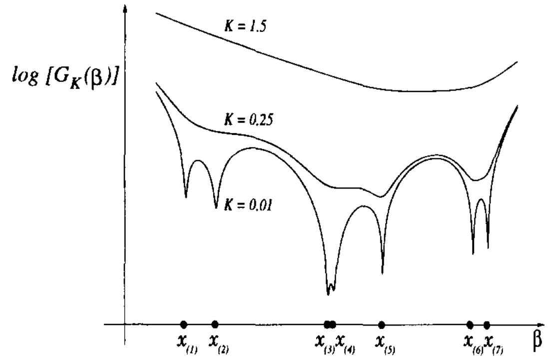

Figure 1. Weighted myriad objective function, input samples $\mathbf{x} = [4.9, 0.0, 6.5, 10.0, 9.5, 4.7, 1.0]$, weights $\mathbf{w} = [0.05, 0.1, 0.6, 0.9, 0.6, 0.1, 0.05]$.

II WEIGHTED MYRIAD FILTERS

Just as the mean and median are based on the Gaussian and Laplacian distribution, respectively, the *myriad* is defined as the Maximum Likelihood estimate of location of the Cauchy distribution [2]. For a set of N independent and identically distributed (i.i.d.) observations $\{x_i\}_{i=1}^N$, drawn from a Cauchy distribution with location parameter β and scaling factor $K > 0$: $f(x;\beta) = \left(\frac{K}{\pi}\right)\frac{1}{K^2+(x-\beta)^2}$, the sample myriad $\hat{\beta}_K$ maximizes the likelihood function $\prod_{i=1}^N f(x_i;\beta)$. Thus, $\hat{\beta}_K \triangleq \arg\min_\beta \prod_{i=1}^N [K^2 + (x_i - \beta)^2]$. By assigning *non-negative* weights to reflect the levels of reliability of the input samples, we obtain the *weighted myriad*. For an observation vector $\mathbf{x} \triangleq [x_1, x_2, \ldots, x_N]^T$ and a weight vector $\mathbf{w} \triangleq [w_1, w_2, \ldots, w_N]^T$, the *weighted myriad filter* (WMyF) output is $\hat{\beta}_K(\mathbf{w},\mathbf{x}) \triangleq \arg\min_\beta G_K(\beta,\mathbf{w},\mathbf{x})$, where

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

is called the *weighted myriad objective function.* It is easy to show using (1) that the filter output depends only on $\frac{\mathbf{w}}{K^2}$, hence there are only N independent filter parameters.

The weighted myriad filter output is the value of β at the global minimum of $G_K(\beta)$, which is a polynomial of degree $2N$. Fig. 1 shows typical plots of $\log(G_K(\beta))$ for a data window size $N = 7$, with the order statistics $\{x_{(m)}\}_{m=1}^N$ (samples sorted in increasing order of magnitude) shown on the horizontal axis. Since all the local minima are in the range $[x_{(1)}, x_{(N)}]$ of the input samples, so is the filter output.

The filter output $\hat{\beta}$ is one of the roots of the $(2N-1)$ degree derivative polynomial $G'(\beta) \triangleq \frac{\partial G_K(\beta,\mathbf{w},\mathbf{x})}{\partial \beta} : G'(\hat{\beta}) = 0$.

The limiting case $K \to \infty$ (holding the weights finite) leads to the linear *weighted mean filter*: $\hat{\beta}_\infty = \sum_{j=1}^N w_j x_j / \sum_{j=1}^N w_j$, hence the name *linearity parameter* for K. When $K \to 0$, we obtain the *weighted mode-myriad filter* [2, 5], a robust selection filter (the output is one of the input samples) that is highly resistant to outliers.

III ADAPTIVE FILTER OPTIMIZATION

Given an input vector $\mathbf{x}$, a weight vector $\mathbf{w}$ and linearity parameter K ($0 < K < \infty$), denote the weighted myriad filter output as $y \equiv y_K(\mathbf{w},\mathbf{x}) \equiv y(\mathbf{w},\mathbf{x})$. The filtering error, in estimating a desired signal d, is $e = y - d$. The MAE cost function is $J(\mathbf{w}, K) \triangleq E\{|e|\} = E\{|y_K(\mathbf{w},\mathbf{x}) - d|\}$, where $E\{\cdot\}$ represents statistical expectation. Since the filter output depends only on $\frac{\mathbf{w}}{K^2}$ [5], the optimal filtering action is independent of the choice of K. The cost function is therefore written simply as $J(\mathbf{w})$. The optimal weights minimize $J(\mathbf{w})$ subject to the constraints $w_i \geq 0$, $i = 1, 2, \ldots, N$. To obtain necessary conditions for optimality, equate the gradient of $J(\mathbf{w})$ to zero:

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

where $\operatorname{sgn}(x)$ is the *sign* function with $\operatorname{sgn}(0) = 0$. The above conditions lead to highly nonlinear equations that are difficult to solve in closed-form for the optimal weights. We therefore minimize the cost function $J(\mathbf{w})$ using the *steepest descent* method. Further, to deal with situations where the statistics of the signals are unknown or time-varying, we use *instantaneous estimates* for the gradient by removing the expectation operator in the gradient expression, just as in the classical LMS algorithm [6]. Including the constraint of non-negative weights and using (2), we have the following adaptive algorithm to update the filter weights:

$$w_i(n+1) = P\left[w_i(n) - \mu \operatorname{sgn}(e(n)) \frac{\partial y}{\partial w_i}(n)\right], \tag{3}$$

where $w_i(n)$ is the ith weight at the nth iteration, $\mu > 0$ is the step-size of the update, $e(n)$ is the error at the nth iteration and $P[u] \triangleq \begin{cases} u, & u > 0 \\ 0, & u \leq 0. \end{cases}$ To find $\frac{\partial y}{\partial w_i}$, note from Section II that $G'(y) = 0$. Using (1) and the fact that $G(y) > 0$, we can derive

$$G'(y) = 2\,G(y) \sum_{j=1}^{N} \frac{w_j(y-x_j)}{K^2 + w_j(y-x_j)^2} \tag{4}$$

and

$$H(y, w_i) \triangleq \sum_{j=1}^{N} \frac{w_j(y-x_j)}{K^2 + w_j(y-x_j)^2} = 0, \tag{5}$$

where the function $H(\cdot,\cdot)$ emphasizes the implicit dependence of y on w_i, holding all other quantities fixed. Implicit differentiation of (5) with respect to w_i yields

$$\left(\frac{\partial H}{\partial y}\right) \cdot \left(\frac{\partial y}{\partial w_i}\right) + \left(\frac{\partial H}{\partial w_i}\right) = 0. \tag{6}$$

As an important aside, we can show by differentiation of (4), combined with the definition of $H(y, w_i)$ in (5), that

$$\frac{\partial H}{\partial y} = \frac{1}{2}\frac{G''(y)}{G(y)} \geq 0. \tag{7}$$

The last step is because $G''(y) \geq 0$ since y is a local minimum of $G(\cdot)$, and $G(y) > 0$ always. Substituting the expressions for $\frac{\partial H}{\partial y}$ and $\frac{\partial H}{\partial w_i}$, derived using (5), into (6), we can find $\frac{\partial y}{\partial w_i}$. Using that in (3), we obtain

Adaptive Weighted Myriad Filter Algorithm I

$$w_i(n+1) = P\left[w_i(n) + \mu \operatorname{sgn}(e(n)) \frac{r_i(n)}{R(n)}\right]; \tag{8}$$

$$r_i(n) \triangleq \left\{\frac{(y - x_i)}{\left(1 + \frac{w_i}{K^2}(y-x_i)^2\right)^2}\right\}(n) \tag{9}$$

and

$$R(n) \triangleq \left\{\sum_{j=1}^{N} w_j \frac{1 - \frac{w_j}{K^2}(y-x_j)^2}{\left(1 + \frac{w_j}{K^2}(y-x_j)^2\right)^2}\right\}(n). \tag{10}$$

Note that $R(n)$ is non-negative since it can be shown [5] to be proportional to $\frac{\partial H}{\partial y}(n)$, which is non-negative from (7). To prevent the unstable behaviour due to $R(n)$ becoming very small, we add a *stabilization parameter* $a > 0$ to $R(n)$; the update denominator is then lower bounded by a. This leaves the direction of the gradient estimate unchanged (the update term is proportional to the negative of the gradient estimate). We thus have

Adaptive Weighted Myriad Filter Algorithm Ia

$$w_i(n+1) = P\left[w_i(n) + \mu \operatorname{sgn}(e(n)) \frac{r_i(n)}{a + R(n)}\right]. \tag{11}$$

By removing the denominators from the update terms, we obtain the following greatly simplified algorithm:

Adaptive Weighted Myriad Filter Algorithm II

$$w_i(n+1) = P\left[w_i(n) + \mu \operatorname{sgn}(e(n))\, r_i(n)\right] \tag{12}$$

where $r_i(n)$ is given by (9). The update here is computationally comparable to the least mean absolute deviation (LMAD) algorithm $w_i(n+1) = w_i(n) - \mu \operatorname{sgn}(e(n))\, x_i(n)$. To understand the operation of Algorithm II, rewrite (9):

$$r_i(n) = \Delta[y(n) - x_i(n); \frac{w_i(n)}{K^2}], \Delta[u;\xi] \triangleq \frac{u}{(1+\xi u^2)^2}. \tag{13}$$

Fig. 2 illustrates the operation of the algorithm and the *update magnitude function* $\Delta[u;\ \xi]$ is shown in Fig. 3. Assume that $e(n) > 0$, i.e. $d(n) < y(n)$ at the current iteration, so that $\operatorname{sgn}(e(n)) = +1$. Then, from (12) and (13), the weights $w_i(n)$ are increased ($r_i(n) > 0$) if $x_i(n) < y(n)$ ($i = i_1, i_2$ in Fig. 2). Increasing weight w_i moves the filter output towards x_i. Thus, we can conclude (taking into account the case $e(n) < 0$) from Fig. 2 that the algorithm moves the filter output towards the samples that are on the *same side of* $y(n)$ *as* $d(n)$. That is, *the filter output moves towards the samples that are closer to the desired signal.* Also, Fig. 3 shows that the update is negligible for samples $x_i(n)$ that are far from the current estimate $y(n)$; the algorithm is thus robust to outliers.

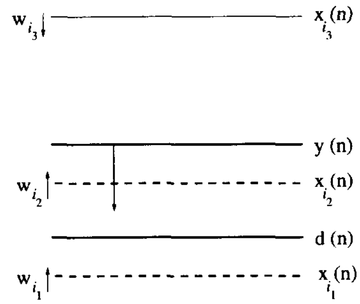

Figure 2. Operation of Adaptive Algorithm II.

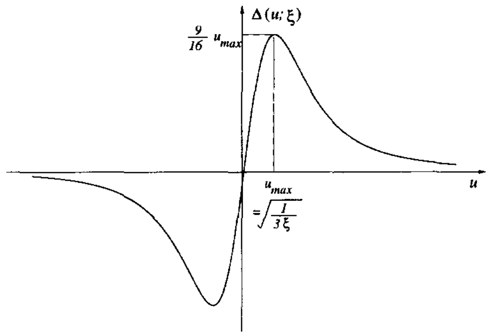

Figure 3. Update magnitude function $\Delta(u;\ \xi)$.

IV SIMULATION RESULTS

The adaptive algorithms of Section III were tested using a computer simulation example involving lowpass filtering a chirp-type signal in α-stable noise. Fig. 4(a) shows the *clean* quadchirp signal $s(n)$, a sinusoid with quadratically increasing instantaneous frequency. The *desired signal* $d(n)$ (Fig. 4(b)) was obtained by passing $s(n)$ through an FIR lowpass filter of window length $N = 11$. The signal $s(n)$ is corrupted by additive symmetric zero-mean α-stable noise, yielding a noisy observed signal $x(n)$ (the *training signal*). The *characteristic exponent* α and the *dispersion* γ of the noise were chosen as $\alpha = 1.4$ and $\gamma = 0.1$ (the dispersion γ decides the spread of the distribution around the origin [3]). The chosen α-stable noise process simulates low-level Gaussian-type noise along with impulsive interference.

The linear filter was trained using the least mean absolute deviation (LMAD) algorithm (also called the *sign LMS* algorithm): $w_i(n+1) = w_i(n) - \mu\,\mathrm{sgn}(e(n))\,x_i(n)$. The initial filter weights were all chosen to be zero. The weighted median filter adaptation used an adaptive weighted order statistic (WOS) algorithm described in [1]. The initial filter for the weighted median and the weighted myriad algorithms was the identity filter; all the off-center weights were zero, while the center weight was chosen as 10.0. For the weighted myriad algorithms, the linearity parameter was set to $K = 1.0$ and the *stabilization parameter* of Algorithm Ia (see (11)) was set to $a = 1.0$. All the filters had window length $N = 11$. The adaptive algorithms were implemented using multiple passes through the signals.

The trajectories of some of the filter weights for Algorithms Ia and II are shown in Fig. 5. The weight curve $w_8(n)$ is non-monotonic, while the weight $w_6(n)$ (the weight of the center sample) is monotonically decreasing. This is due to the large initial value $w_6(0) = 10.0$ which, in the beginning, pulls the off-center weights (including $w_8(n)$) up from their initial zero values. The other off-center weight curves are similar to $w_8(n)$. The figure shows that Algorithms Ia and II converge to almost the same weight values, but Algorithm II converges significantly faster. The convergence of the various algorithms is confirmed by Fig. 6, which shows the MAE learning curves obtained by time-averaging the absolute filtering error signals $|e(n)|$ (no comparison of convergence speeds is intended in presenting these curves).

The various trained filters were applied to a *test signal* $x'(n)$ obtained by adding a different realization of noise to the clean quadchirp signal of Fig. 4(a). The test signal $x'(n)$ (Fig. 4(c)) was chosen to be more impulsive than the training signal $x(n)$. Fig. 4(d) shows the output of the designed lowpass FIR filter. This filter is severely affected by the impulses in the test signal. The output of the linear filter, trained using the LMAD adaptation, is shown in Fig. 4(e) and is evidently far from the desired signal; the trained linear filter is only marginally better than the lowpass FIR filter. The weighted median filter output of Fig. 4(f) is less affected by the impulses, but exhibits severe distortion, in part because the filter is constrained to be a selection filter. The filter is also unable to remove the high-frequency portions of the quadchirp signal completely. The outputs of the weighted myriad filters from Algorithms Ia and II, shown in Figs. 4(g) and (h), respectively, are visually the

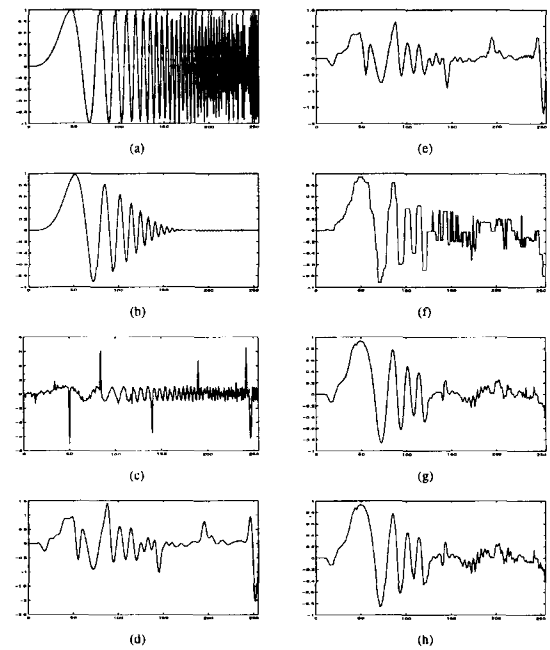

Figure 4. (a) $s(n)$: clean quadchirp signal, (b) $d(n)$: desired signal (lowpass FIR filtering of $s(n)$), (c) $x'(n)$: noisy quadchirp test signal, (d) $y'_{lpfir}(n)$: lowpass FIR filtering of $x'(n)$, (e) $y'_{lmad}(n)$: linear (LMAD) filter output, (f) $y'_{wm}(n)$: weighted median filter output, (g) $y'_{wmyIa}(n)$: weighted myriad filter output (algorithm Ia), (h) $y'_{wmyII}(n)$: weighted myriad filter output (algorithm II).

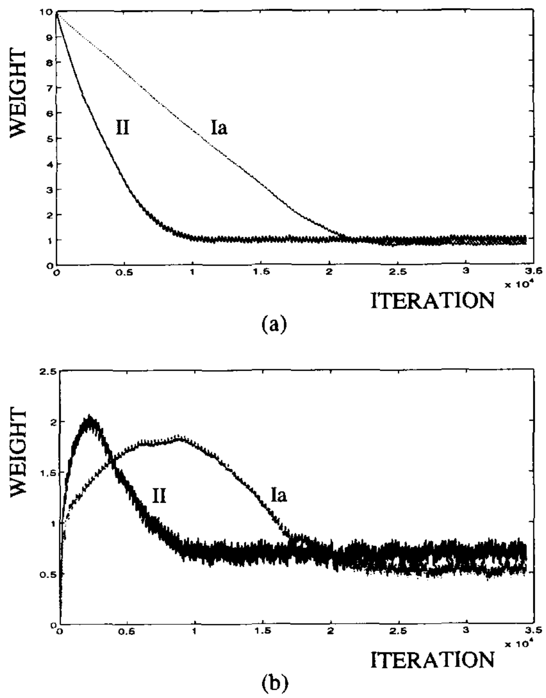

Figure 5. Weight trajectories $w_i(n)$ for Adaptive Weighted Myriad Filter Algorithms Ia and II: (a) $w_6(n)$, (b) $w_8(n)$.

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		0.1563	0.1594	0.0504	0.0516
Weighted Myriad	I	0.0968	0.0962	0.0194	0.0193
	Ia	0.0910	0.0903	0.0162	0.0160
	II	0.0959	0.0947	0.0187	0.0185

Table 1. Mean absolute error (MAE) and mean square error (MSE) in filtering the training and test signals.

closest to the desired signal, especially in the low-frequency portions of the quadchirp signal.

These results are confirmed by Table 1, which shows mean absolute errors (MAEs) and mean square errors (MSEs) incurred in filtering the noisy quadchirp signals $x(n)$ (Training) and $x'(n)$ (Test) with the various trained filters. The weighted myriad filters have the smallest (and similar) MAEs as well as MSEs in all cases. The linear filters are adversely affected by the change from the training to the test signal. The weighted median filter is more robust but has high MAEs. The weighted myriad filters are hardly affected by the change in the noise, demonstrating their high robustness.

V CONCLUSION

Necessary conditions for optimal *Weighted Myriad Filters* were derived under the mean absolute error (MAE) criterion. Stochastic gradient-based adaptive algorithms were developed to optimize the filter weights. The adaptive algorithms were investigated through simulations 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 MAEs than the adaptive linear and weighted median filters, and were more robust to changes in the noise environment.

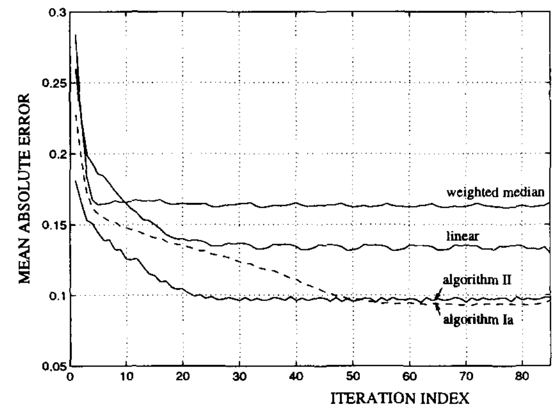

Figure 6. Time-averaged learning curves.

Theoretical analysis of the convergence of the adaptive algorithms is being pursued. Design rules for the choice of algorithm parameters (step-size and initial filter weights) are being developed. Optimizations of some special cases of weighted myriad filters are considered in future publications.

REFERENCES

[1] 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.

[2] 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.

[3] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," *Proceedings of the IEEE*, vol. 81, July 1993.

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

[5] 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*. Submitted for publication.

[6] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 1991.