A GENERALIZED WEIGHTED MEDIAN FILTER STRUCTURE ADMITTING REAL-VALUED WEIGHTS

Gonzalo R. Arce

Department of Electrical and Computer Engineering
University of Delaware
Newark, Delaware 19716
(302) 831-8030, arce@eecis.udel.edu

ABSTRACT

Weighted median filters (smoothers) have been shown to be analogous to normalized FIR linear filters constrained to have only positive weights. In this paper, it is shown that much like the mean is generalized to the rich class of linear FIR filters, the median can be generalized to a richer class of weighted median *(WM)* filters admitting positive and negative weights. The generalization follows naturaly and is surprisingly simple. In order to analyze and design this class of WM filters, a new threshold decomposition theory admitting real-valued input signals is developed which, in turn, is used to develop fast adaptive algorithms to optimally design the real-valued filter coefficients. The new WM filter formulation leads to significantly more powerful estimators capable of effectively addressing a number of fundamental problems in signal processing which could not adequately be addressed by prior WM filter (smoother) structures.

1. INTRODUCTION

Weighted median filters (smoothers), introduced by Edgemore in the context of least absolute regression over a hundred years ago [1], have received considerable attention in signal processing research over the last two decades [2, 3]. Although these structures are widely known in the signal processing literature as weighted median filters, for reasons that will become apparent shortly, we will refer to these filters as *weighted median smoothers*. During the last few years, the theory behind WM smoothers has been developing quite fast. It is often stated that there are many analogies between WM smoothers and linear FIR filters. In this paper, however, we show that WM smoothers are highly constrained and that they are significantly less powerful than linear FIR filters. In fact, WM smoothers are equivalent to normalized weighted mean filters – a severely constrained subset of linear FIR filter. Admitting only positive filter weights WM and normalized weighted mean filters are, in essence, smoothers having "low-pass" type filtering characteristics.

A large number of engineering applications require "band-pass" or "high-pass" frequency filtering characteristics. Equalization, deconvolution, prediction, and beamforming are example applications where filters having "band-pass" or "high-pass" characteristics are of fundamental importance. Linear FIR equalizers admitting only positive filter weights, for instance, would lead to completely unacceptable results. Thus, it is not surprising that WM smoothers admitting only positive weights lead to unacceptable results in a number of applications.

THIS WORK WAS SUPPORTED IN PART BY THE NATIONAL SCIENCE FOUNDATION UNDER GRANT MIP-9530923 AND THROUGH COLLABORATIVE PARTICIPATION IN THE ADVANCED TELECOMMUNICATIONS/INFORMATION DISTRIBUTION RESEARCH PROGRAM (ATIRP) CONSORTIUM SPONSORED BY THE U.S. ARMY RESEARCH LABORATORY UNDER COOPERATIVE AGREEMENT DAAL01-96-2-0002.

In this paper, based on fundamental principles of parameter estimation, we define a new WM filtering structure that admits positive and negative weights. The generalization follows naturally and is surprisingly simple. As would be expected, WM filters reduce to WM smoothers whenever the filter coefficients are constrained to be positive. In order to analyze the new WM filter class, we first define a new threshold decomposition framework which allows real-valued inputs and which allows negative sample weighting. The new threshold decomposition architecture overcomes the shortcomings associated with prior definitions. Through the use of the new threshold decomposition architecture, an adaptive algorithm is developed for optimizing the new MW filter structure under the mean absolute error (MAE) criterion.

2. WEIGHTED MEDIAN FILTERS WITH REAL-VALUED WEIGHTS

The sample median and sample mean have deep roots in statistical estimation theory. In particular, they are the *Maximum Likelihood* (ML) estimators of location derived from sets of independent and identically distributed (i.i.d.) samples obeying the Laplacian and Gaussian distributions, respectively. The sample mean and median thus play an analogous role in location estimation. While the mean is associated with the Gaussian distribution which often emerges naturally in practice, the median is related to the Laplacian distribution which has heavier tails and can often provide a better model for impulsive-like processes.

The sample mean and median can be generalized by extending the model of Maximum Likelihood estimation. Let the sample set $X_1, \cdots, X_N$ be independent but not identically distributed. In particular, assume the $X_i' s$ obey the same distribution but assume that their variance is not identical for all samples. Under the Gaussian assumption, the Maximum Likelihood estimate of location in this case can be shown to be the value β minimizing

$$G_2(\beta) = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (X_i - \beta)^2, \tag{1}$$

where σ_i^2 is the variance of the $i'th$ sample in the set. The value β minimizing (1) is the normalized weighted average

$$\bar{\beta} = \frac{\sum_{i=1}^{N} W_i \cdot X_i}{\sum_{i=1}^{N} W_i} \tag{2}$$

with $W_i = 1/\sigma^2 > 0$. Likewise, under the Laplacian model, the Maximum Likelihood estimate of location minimizes the sum of weighted absolute deviations

$$G_1(\beta) = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} |X_i - \beta|. \tag{3}$$

The value $\tilde{\beta}$ minimizing (3) is the weighted median originally introduced over a hundred years ago by Edgemore [1] and defined as

$$\tilde{\beta} = \text{MEDIAN}\left(W_1 \diamond X_1, W_2 \diamond X_2, \cdots, W_N \diamond X_N\right), \tag{4}$$

where $W_i = 1/\sigma^2 > 0$, and where $\diamond$ is the replication operator defined as $W_i \diamond X_i = \overbrace{X_i, X_i, \cdots, X_i}^{W_i \text{ times}}$. It should be noted that the weights in (2) and (4) are contrained to take on non-negative values due to their inverse relationship to the variances of the respective observation samples.

Notably, the location parameter estimation problem just described is related to the time-series filtering problem where the output $Y(n)$, at time n, can be thought of as an estimate of location, and where $X(n-N_1), \cdots, X(n), \cdots, X(n+N_2)$ are the set of observation samples. Although these samples, in general, exhibit temporal correlation, the independent but not identically distributed model can be used to capture the mutual correlation. This is possible by observing that the estimate $Y(n)$ can rely more on the sample $X(n)$ than on the other samples that are further away in time. Thus, $X(n)$ is more reliable than $X(n-1)$ or $X(n+1)$, which in turn may be more reliable than $X(n-2)$ or $X(n+2)$, and so on. By assigning different variances (reliabilities) to the independent but not identically distributed location estimation model, the temporal correlation used in time-series filtering is captured. The weighting structures in (2) and (4) reflect the varying reliability of the samples in the observation set.

From the filter structures described in (2) and (4), it can be seen that the class of weighted median filters (smoothers), is equivalent to the class of normalized weighted average filters. Since the former filter class is severely constrained allowing only linear combinations of positively weighted input samples, it follows that weighted median filters (smoothers) are also severely limited in their structure.

Much like the sample mean can be generalized to the rich class of linear FIR filters, there must be a logical way to generalize the median to an equivalently rich class of weighted median filters that admit both positive and negative weights. We next show that this is in fact possible. Perhaps the simplest approach to derive the class of weighted median filters with real-valued weights is by analogy. The sample mean $\bar{\beta} = \text{MEAN}\left(X_1, X_2, \cdots, X_N\right)$ can be generalized to the class of linear FIR filters as

$$\beta = \text{MEAN}\left(W_1 \cdot X_1, W_2 \cdot X_2, \cdots, W_N \cdot X_N\right) \tag{5}$$

where $W_i \in R$. In order to apply the analogy to the median filter structure (5) must be written as

$$\bar{\beta} = \text{MEAN}\left(|W_1| \cdot sgn(W_1)X_1, \cdots, |W_N| \cdot sgn(W_n)X_N\right) \tag{6}$$

where the sign of the weight affects the corresponding input sample and the weighting is constrained to be non-negative. By analogy, the class of weighted median filters admitting real-valued weights emerges as

$$\tilde{\beta} = \text{MEDIAN}\left(|W_1| \diamond sgn(W_1)X_1, \cdots, |W_N| \diamond sgn(W_n)X_N\right), \tag{7}$$

with $W_i \in R$ for $i = 1, \cdots, N$. Again, the weight signs are uncoupled from the weight magnitude values and are merged with the observation samples. The weight magnitudes play the equivalent role of positive weights in the framework of weighted median smoothers. It is simple to show that the weighted mean (normalized) and the weighted median operations shown in (6) and (7) respectively minimize $G_2(\beta) = \sum_{i=1}^{N} |W_i| \left(sgn(W_i)X_i - \beta\right)^2$ and $G_1(\beta) = \sum_{i=1}^{N} |W_i||sgn(W_i)X_i - \beta|$. While $G_2(\beta)$ is a convex continuous function, $G_1(\beta)$ is a convex but piecewise linear function whose minima is guaranteed to be one of the "signed" input samples (i.e. sign(W_i) X_i). The WM filter output for non–integer weights can be determined as follows:

1. Calculate the threshold $T_0 = \frac{1}{2}\sum_{i=1}^{N} |W_i|$.
2. Sort the "signed" observation samples $\text{sgn}(W_i)X_i$.
3. Sum the magnitude of the weights corresponding to the sorted "signed" samples beginning with the maximum and continuing down in order.
4. The output is the sample whose magnitude weight causes the sum to become $\geq T_0$.

The following example illustrates this procedure. Consider the window size 5 WM filter defined by the real valued weights $[W_1, \cdots, W_5]^T = [0.1, 0.2, 0.3, -0.2, 0.1]^T$. The output for the observation set $[X_1, \cdots, X_5]^T = [-2, 2, -1, 3, 6]^T$ is found as follows. Summing the absolute weights gives the threshold $T_0 = \frac{1}{2}\sum_{i=1}^{5} |W_i| = 0.45$. The "signed" observation samples, sorted observation samples, their corresponding weight, and the partial sum of weights (from each ordered sample to the maximum) are:

observation samples	$-2,$	$2,$	$-1,$	$3,$	6
corresponding weights	$0.1,$	$0.2,$	$0.3,$	$-0.2,$	0.1
sorted signed samples	$-3,$	$-2,$	$-1,$	$2,$	6
corresponding abs. weights	$0.2,$	$0.1,$	$0.3,$	$0.2,$	0.1
partial weight sums	$0.9,$	$0.7,$	$\underline{0.6},$	$0.3,$	0.1

Thus, the output is -1 since when starting from the right (maximum sample) and summing the weights, the threshold $T_0 = 0.45$ is not reached until the weight associated with -1 is added. The underlined sum value above indicates that this is the first sum which meets or exceeds the threshold. The effect that negative weights have on the weighted median operation is similar to the effect that negative weights have on linear FIR filter outputs. Figure 1 illustrates this concept where $G_2(\beta)$ and $G_1(\beta)$, the cost functions associated with linear FIR and weighted median filters respectively, are plotted as a function of β. Recall that the output of each filter is the value minimizing the cost function. The input samples are again selected as $[X_1, \cdots, X_5] = [-2, 2, -1, 3, 6]$ and two sets of weights are used. The first set is $[W_1, \cdots, W_5]$ $= [0.1, 0.2, 0.3, 0.2, 0.1]$ where all the coefficients are positive, and the second set being $= [0.1, 0.2, 0.3, -0.2, 0.1]$ where W_4 has been changed, with respect to the first set of weights, from 0.2 to -0.2. Figure 1(a) shows the cost functions $G_2(\beta)$ of the linear FIR filter for the two sets of filter weights. Notice that by changing the sign of W_4, we are effectively moving X_4 to its new location $\mathrm{sign}(W_4)X_4 = -3$. This, in turn, pulls the minimum of the cost function towards the relocated sample $\mathrm{sign}(W_4)X_4$. Negatively weighting X_4 on $G_1(\beta)$ has a similar effect as shown in Fig. 1(b). In this case, the minimum is pulled towards the new location of $\mathrm{sign}(W_4)X_4$. The minimum, however, occurs at one of the samples $\mathrm{sign}(W_i)X_i$.

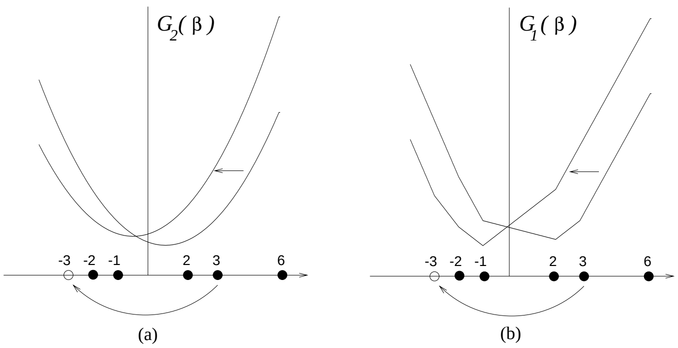

Figure 1: Effects of negative weighting on the cost functions $G_2(\beta)$ and $G_1(\beta)$. The input samples are $[X_1, \cdots, X_5]^T = [-2, 2, -1, 3, 6]^T$ which are filtered by the two set of weights $[0.1, 0.2, 0.3, 0.2, 0.1]^T$ and $[0.1, 0.2, 0.3, -0.2, 0.1]^T$.

3. THRESHOLD DECOMPOSITION FOR REAL-VALUED SIGNALS

In this section we further extend threshold decomposition allowing the decomposition of real-valued signals. This decomposition, in turn, can be used to analyze weighted median filters having real-valued weights. Consider the set of real-valued samples $X_1, \cdots, X_N$ and define a WM filter by the corresponding real valued weights $W_1, \cdots, W_N$. Decompose each sample X_i as

$$x_i^q = \mathrm{sign}\,(X_i - q) \tag{8}$$

where $-\infty < q < \infty$, and where

$$\mathrm{sign}\,(X_i - q) = \begin{cases} 1 & \text{if } X_i > q; \\ 0 & \text{if } X_i = q; \\ -1 & \text{if } X_i < q. \end{cases} \tag{9}$$

Thus, each sample X_i is decomposed into an infinite set of binary points taking values in $[-1, 1]$, and a single point equal to 0 obtained for $X_i = q$. It can be shown that the original real-valued samples X_i can be perfectly reconstructed from the infinite set of thresholded signals as

$$X_i = \frac{1}{2}\int_{-\infty}^{\infty} x_i^q \, dq = \frac{1}{2}\int_{-\infty}^{\infty} \mathrm{sign}\,(X_i - q)\; dq. \tag{10}$$

With this threshold decomposition, the WM filter operation can be implemented as

$$\begin{aligned} \hat{\beta} &= \mathrm{MED}\left(|W_i| \diamond \mathrm{sign}\,(W_i)\, X_i|_{i=1}^{N}\right) \\ &= \mathrm{MED}\left(|W_i| \diamond \frac{1}{2}\int_{-\infty}^{\infty} \mathrm{sign}\left[\mathrm{sign}\,(W_i)\, X_i - q\right]\; dq|_{i=1}^{N}\right). \end{aligned}$$

The expression above represents the median operation of a set of weighted integrals, each synthesizing a signed sample. Note that the same result is obtained if the weighted median of these functions, at each value of q, is taken first and the resultant signal is integrated over its domain. Thus, the order of the integral and the median operator can be interchanged without affecting the result leading to

$$\hat{\beta} = \frac{1}{2}\int_{-\infty}^{\infty} \mathrm{MED}\left(|W_i| \diamond \mathrm{sign}\left[\mathrm{sign}\,(W_i)\, X_i - q\right]|_{i=1}^{N}\right)\; dq.$$

In this representation, the “signed” samples play a fundamental role; thus, we define the “signed” observation vector $\mathbf{S}$ as

$$\mathbf{S} = [\mathrm{sign}(W_1)X_1, \cdots, \mathrm{sign}(W_N)X_N)]^T = [S_1, \cdots, S_N]^T\,.$$

The threshold decomposed signed samples, in turn, form the vector $\mathbf{s}^q$ defined as

$$\begin{aligned} \mathbf{s}^q &= \left[\mathrm{sign}\left[\mathrm{sign}(W_1)X_1 - q\right], \cdots, \mathrm{sign}\left[\mathrm{sign}(W_N)X_N - q\right]\right]^T \\ &= \left[s_1^q, s_2^q, \cdots, s_N^q\right]^T\,. \end{aligned} \tag{11}$$

Letting $\mathbf{W}_a$ be the vector whose elements are the magnitude weights, $\mathbf{W}_a = [|W_1|, |W_2|, \cdots, |W_N|]^T$, the WM filter operation can be expressed as

$$\hat{\beta} = \frac{1}{2}\int_{-\infty}^{\infty} \mathrm{sign}\left(\mathbf{W}_a^T \mathbf{s}^q\right)\; dq. \tag{12}$$

The WM filter representation using the new threshold decomposition is compact although it may seem that the integral term may be difficult to implement in practice. It is shown in [4] that the expression in (12) can be simplified to

$$\hat{\beta} = \frac{S_{(1)} + S_{(N)}}{2} + \frac{1}{2}\sum_{i=1}^{N-1}\left(S_{(i+1)} - S_{(i)}\right)\mathrm{sign}\left(\mathbf{W}_a^T \mathbf{s}^{S_{(i)}^+}\right).$$

The computation of WM filters with the new threshold decomposition is efficient requiring only $N-1$ threshold logic (sign) operators, it allows the input signals to be arbitrary real-valued signals, and it allows positive and negative filter weights.

4. OPTIMAL WEIGHTED MEDIAN FILTERING

In many applications it is desirable to design the weights of a filter in some optimal fashion. In this section we develop adaptive algorithms to find the optimal real-valued weights of WM filters. We assume that the observed process $\{X(n)\}$ is statistically related to some desired process $\{D(n)\}$ of interest. We assume that these processes are jointly stationary. A window of width N slides across the input process pointwise estimating the desired sequence. The vector containing the N samples in the window at time n is $\mathbf{X}(n) = [X_1(n), X_2(n), \cdots, X_N(n)]^T$. The running WM filter output estimates the desired signal as

$$\hat{D}(n) = MED\left[|W_i| \diamond \mathrm{sign}(W_i)X_i(n)|_{i=1}^{N}\right],$$

where both the weights and samples take on real values. The goal is to determine the weight values in $\mathbf{W} = [W_1, \cdots, W_N]^T$ which will minimize the estimation error. Under the Mean Absolute Error (MAE) criterion, the cost to minimize is

$$\begin{aligned} J(\mathbf{W}) &= E\left\{\left|D(n) - \hat{D}(n)\right|\right\} \qquad (13)\\ &= E\left\{\frac{1}{2}\left|\int_{-\infty}^{\infty} \operatorname{sign}(D-q) - \operatorname{sign}\left(\mathbf{W}_a^T \mathbf{s}^q\right) dq\right|\right\}, \end{aligned}$$

where the threshold decomposition representation of the signals was used. The absolute value and integral operators can be interchanged since the integral acts on a strictly positive or a strictly negative function. This results in

$$J(\mathbf{W}) = \frac{1}{2}\int_{-\infty}^{\infty} E\left\{\left|\operatorname{sign}(D-q) - \operatorname{sign}\left(\mathbf{W}_a^T \mathbf{s}^q\right)\right|\right\} dq$$

In [4] we develop an adaptive algorithm to minimize the above cost function. The resultant *fast LMA WM adaptive* algorithm leads to following recursion:

$$\begin{aligned} W_j(n+1) &= W_j(n) + \mu\left(D(n) - \hat{D}(n)\right) \operatorname{sign}(W_j(n))\\ &\quad \times \operatorname{sign}\left(\operatorname{sign}(W_j(n)) X_j(n) - \hat{D}(n)\right), \end{aligned}$$

for $j = 1, \cdots, N$. This algorithm is similar to that of Yin and Neuvo's [5] except that their algorithm is applicable to the design of weighted median smoothers which do not admit negative weight values thus a projection operator mapping all negative weights to zero is needed in their update. Moreover, updates in Yin and Neuvo's algorithm contain thresholded signals at levels determined by the sample order-statistics, not at the "signed" order statistics.

5. APPLICATIONS OF WM FILTERS WITH REAL VALUED WEIGHTS

The added flexibility provided by negative weights in WM filters is illustrated in this section. Consider the design of a "high-pass" WM filter whose objective is to preserve a high frequency tone while remove all low frequency terms. Figure 2(a) depicts a two-tone signal with normalized frequencies of 0.04 and 0.4 Hz. Figure 2(b) shows the multi-tone signal filtered by a 28-tap linear FIR filter designed by MATLAB's `fir1` function with a normalized cutoff frequency 0.2 Hz. The fast adaptive LMA algorithm was used to optimize a MW filter with 28 weights. The step size used in all adaptive optimization experiments was 10^{-3}. These weights, in turn, were used to filter the multi-tone signal resulting in the estimate shown in Fig. 2(c). The low-frequency components have been clearly filtered out. If the linear FIR filter weights are used with the WM filter, we obtain the output shown in Fig. 2(d). Next, Yin et. al's fast adaptive LMA algorithm was used to optimize a MW filter (smoother) with 28 (positive) weights. The filtered signal attained with the optimized weights is shown in Fig. 2(e). The weighted median smoother clearly fails to remove the low frequency components as expected. The weighted median smoother output closely resembles the input signal as it is the closest output to the desired signal it can produce.

Having designed the various high-pass filters in a noiseless environment, we next test their performance on signals embedded in noise. Stable noise with $\alpha = 1.4$ was added to the two-tone signal. Rather than training the various filters to this noisy environment, we used the same filter coefficients as in the noise-free simulations. Figure 2(f)-(i) illustrates the results. As expected the outputs of the weighted median filter and smoother are not affected, whereas the output of the linear filter is severely degraded as the linear high-pass filter amplifies the high frequency noise. Table I summarizes the MAE values attained by the various filters tested.

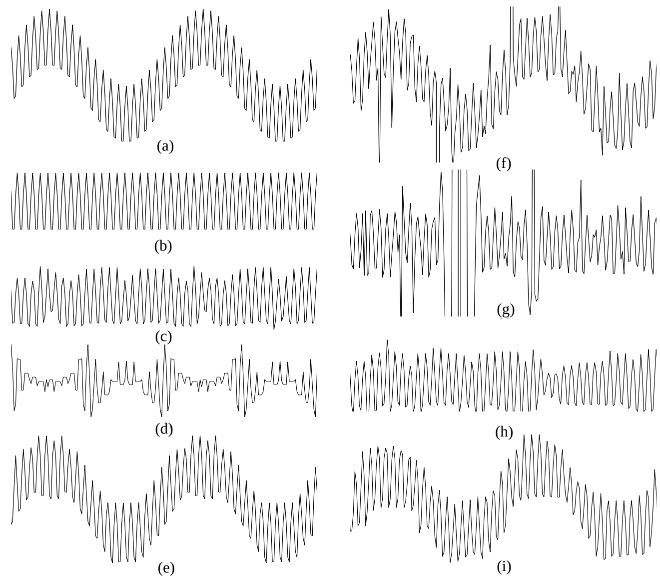

Figure 2: (a) Two-tone signal, outputs from (b) linear FIR filter, (c) WM filter, (d) WM filter with linear FIR weights, (e) WM smoother with positive weights, (f) two-tone signal in stable noise, noisy signal filter with (g) linear FIR filter, (h) WM filter, (i) WM smoother with positive weights.

Table 1: Mean Absolute Filtering Errors

filter	noise free	with stable noise
linear FIR	0.012	0.979
WMF with FIR weights	0.501	0.530
optimal WMF smoother	0.688	0.692
optimal WMF (fast alg.)	0.191	0.209

6. REFERENCES

[1] F. Y. Edgeworth, "A new method of reducing observations relating to several quantities," *Phil. Mag. (Fifth Series)*, vol. 24, 1887.

[2] D. R. K. Brownrigg, "The weighted median filter," *Commun. Assoc. Comput. Mach.*, vol. 27, Aug. 1984.

[3] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, "Weighted median filters: a tutorial," *IEEE Transactions on Circuits and Systems*, vol. 41, May 1996.

[4] G. R. Arce, "A generalized weighted median filter structure admitting real-valued weights," *IEEE Transactions on Signal Processing*, June 1997. submitted.

[5] 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.