FREQUENCY SELECTIVE FILTERING USING WEIGHTED ORDER STATISTIC ADMITTING REAL-VALUED WEIGHTS

José Luis Paredes[1], Nelson Pérez García[2]

[1]Universidad de Los Andes, Grupo de Investigación de Telecomunicaciones (GITEL).
e-mail: paredesj@ula.ve

Recibido: diciembre de 2007 Recibido en forma final revisado: diciembre de 2010

ABSTRACT

In this paper, a weighted order statistic (WOS) filtering structure admitting real-valued weights is introduced. The proposed filtering approach can effectively address a number of signal and image processing applications that require robust bandpass or highpass operations where the underlying contamination follows a nonsymmetric heavy-tailed distribution. The effect of negative weighting in the filtering operation is studied under a statistical viewpoint using a weight monotonic test. Furthermore, an adaptive optimization algorithm for the design of this class of WOS filters is also introduced. Several computer simulations show the performance of the proposed filtering structure.

Keywords: Robust signal processing, Nonlinear filters, Impulsive noise, Weigthed order statistic, Median filters.

FILTROS DE FRECUENCIA SELECTIVA BASADOS EN ORDEN ESTADÍSTICO ADMITIENDO PODERACIONES REALES

RESUMEN

En este artículo, se desarrollan filtros de orden estadísticos ponderados admitiendo ponderaciones reales. El enfoque de filtrado propuesto puede ser utilizado eficientemente en innumerables aplicaciones de procesamiento de señales e imágenes que requieren filtrado robusto tipo paso-alto o paso-banda, donde la contaminación de fondo es un ruido de naturaleza impulsiva, modelado por distribuciones de “colas pesadas” asimétricas. El efecto de las ponderaciones negativas en el proceso de filtrado se analiza bajo un punto de vista estadístico empleando pruebas monotónicas ponderadas. Adicionalmente, también se presenta un algoritmo de optimización para el diseño de esta nueva familia de filtros no lineales. Se muestran numerosas simulaciones del desempeño de los filtros propuestos, así como también su correspondiente algoritmo de optimización.

Palabras clave: Procesamiento robusto de señales, Filtros no lineales, Ruido impulsivo, Orden estadístico ponderado, Filtros de mediana.

INTRODUCTION

Nonlinear filters based on weighted rank ordering of the input samples have been proven to outperform linear filters in applications where the underlying contamination has impulsive characteristic or obeys a heavy-tailed distribution (Mitra & Sicuranza, 2001). In particular, weighted order statistic (WOS) filtering structure (Yu & Liao, 1994; Yli-Harja *et al*., 1991), that encompasses weighted median (WM), median, and rank order filters, has received considerable attention since they have demonstrated improved filtering ability compared to other popular non-linear filters.

WOS exploits, in some sense, the information provided by the relative ranking of the input samples as well as in their temporal ordering. More precisely, given an observation vector $\mathbf{X} = [\, \mathbf{X_1, X_2, ..., X_N} \,]^T$, a set of weights $\langle \mathbf{W_1, W, ... , W_N} \rangle$ and a threshold value $\mathbf{W_0}$, the output of a WOS filter is defined as:

$$\mathbf{Y = W_0} : th \text{ largest value of the set}$$

$$[\, \mathbf{W_1 \lozenge X_1, W_2 \lozenge X_2, ..., W_N \lozenge X_N} \,] \qquad (1)$$

where $\lozenge$ denotes the replication operator defined as

$$\mathbf{W_i \lozenge X_i} = \overbrace{\mathbf{X_i, X_i, ..., X_i}}^{\mathbf{W_i}\text{ times}}.$$

In (1), the filter weights are restricted to be nonnegative — a constrain that leads to a filtering structure with “lowpass” filtering behavior. Hence, this filtering structure is discarded for applications that require "bandpass" or "highpass" type filtering characteristics. Since these filters are smoothers they will be referred, hereinafter, as WOS smoothers to differentiate them from the new structure to be proposed in this paper.

The limitations of WOS smoothers have motivated a number of researches to define more powerful order statistic type filters capable of addressing the more demanding filter tasks. For instance, hybrid filtering structure has been developed in (Yin & Neuvo, 1994 and Song & Lee, 1996), whereas Arce *et al*. in (Arce *et al*. 1995) add data-dependent weighting to the structure of WOS smoothers. The shortcomings of all these filtering approaches, however, are their computational complexity and the large number of filter parameters.

In this paper, we propose a more general and powerful WOS filtering structure that is able to address problems that require bandpass or highpass operations. Since the new structure can synthesize any frequency-selective characteristic, they will be referred as WOS filters. In the new framework, WOS filters are allowed to use real-valued weights, and therefore they can be designed to retain or restore some desired signal frequency and reject others, similar to linear filters but with the implicit robustness in the rank-ordering operation.

Much like the class of WOS smoothers belongs to a richer family of nonlinear filters, the so-called stack smoothers (Wendt *et al*. 1986), the proposed filters are part of a wider family of nonlinear filters, the so-called stack filters (Paredes & Arce, 1999). Moreover, like WOS smoothers that are defined by a threshold logic gate (Yli-Harja *et al*. 1991) in the binary domain created by a threshold decomposition operator, the proposed WOS filtering structure admits a similar binary representation, as well. Furthermore, WOS filters include, as special cases, WOS smoothers (Yli-Harja *et al*. 1991), WM smoothers (Yin *et al*. 1996) and WM filters admitting negative-valued weights. This last class of filters has been recently introduced in (Arce, 1998) and proven to be successful in applications that require robust bandpass characteristics. However, their performance degrades when the underlying contamination no longer follows a symmetric distribution. The round trip time in the transmission control protocol (Li, 2000) and non-Rayleigh amplitude distribution in a broadband communication channel (Middleton. 1999) are just two examples that are better modelled by a nonsymmetric heavy-tail distribution.

In general, the real-valued weights of the proposed WOS filter must be determined following some optimal criterion. In this paper, we exploit the binary representation of the proposed filter to develop an iterative optimization algorithm to find the best filter weights that minimize the mean absolute error (MAE) between the filter output and some desired signal. We compare the performances of the optimal WOS filter to those yielded by a WOS smoother, an FIR filter and a WM filter designed for the same tasks.

The organization of the paper is as follows. In the first section, the new WOS filtering structure admitting real valued weights is introduced. Next, we explore a binary representation of the WOS filtering operation using threshold decomposition. Then, the effect of negative weights in the filtering operation is studied. An adaptive optimization algorithm for the design of the proposed filter is developed and used in the design of a robust highpass filter and a frequency selective filter. Finally, some conclusions are drawn at the end.

WEIGHTED ORDER STATISTIC FILTERS ADMITTING NEGATIVE WEIGHTS

To define the running WOS filters, let $\{X(\cdot)\}$ be a discrete-time continuous-valued signal. The running WOS filter passes a window over the signal $\{X(\cdot)\}$ that selects, at each instant *n*, a set of samples to constitute the observation vector

$$\mathbf{X(n) = [X(n - b), X(n - b + 1),...,X(n),...,X(n + b)\,]^T =}$$
$$\mathbf{= X_1(n)X_2(n),..., X_N(n)]^T}$$

where $\mathbf{X_i(n) = X(n - b + i - 1)}$, $\mathbf{N = 2b + 1}$, is the observation window size and *T* denotes the transpose operation. Following a similar approach to that described in (Arce, 1998), the output of a WOS filter with real-valued weights $\langle \mathbf{W_1, W_2, ..., W_N ; W_0} \rangle$ is defined as:

$$\mathbf{Y(n) = W_0}\text{: th largest value of the set}$$

$$\mathbf{[|W_1| \Diamond sgn(W_1)X_1(n),...,|W_N| \Diamond sgn(W_N)X_N(n)]} \quad (2)$$

where $\mathbf{sgn(\cdot)}$ is the sign function defined as $\mathbf{sgn(x) = +1}$ for $\mathbf{x \geq 0}$, **-1 otherwise.** $\mathbf{W_0}$ is the selection parameter and must be in the range $\left[\mathbf{0}, \sum_{i=1}^{N} |\mathbf{W_i}|\right]$.

Note that for each sample in the observation window the sign of the weight is passed to the corresponding sample to yield the signed sample, $\mathbf{S_i = sgn(W_i)X_i}$, that is weighed through the replication operator by the magnitude of $\mathbf{W_i}$. Thus, weighting in the WOS filter structure is equivalent to uncoupling the weight sign from its magnitude, merging the sign with the observation sample and replicating the "signed" sample according to the magnitude of the weight (Arce, 1998). The filter output is then the $\mathbf{W_0}$ largest value of the "signed" sample set. Although the weights in (2) may seem restricted to integer values, a more general interpretation of the ◊ operator is presented next similar to the one used for WOS smoothers (Yin & Neuvo, 1994).

The output of a WOS filter for non-integer weights can be determined as follows:

1) Sort the "signed" sample set $\left\{\mathbf{S_i = sgn(W_i)X_i}\Big|_{i=1}^{N}\right\}$.

2) Start from the higher end of the sorted "signed" samples and add up the corresponding magnitude weight.

3) The WOS filter output is the signed sample whose magnitude weight causes the sum to become $\geq \mathbf{W_0}$.

As an example, consider the WOS filter with real-valued weights $\langle \mathbf{0.2, -0.4, 0.6, -0.4, 0.1; 0.55} \rangle$ acting on an observation vector $\mathbf{X = [6, -8, 4, 3, 2]^T}$. The filter output is 6 since when starting from the right (largest value of the signed set) adding up the weights, the selection parameter $\mathbf{W_0}$ is not reached until the weight associate with this sample is added.

The three-step procedure described above to compute the output of the WOS filter can be compactly expressed as follows. Let $\mathbf{S_{(i)}}$ be the *i*-th smallest sample in the signed sample set, where $\mathbf{S_{(1)} \leq S_{(2)} \leq ... S_{(N)}}$. Further, let $\left|\mathbf{W_{I(k)}}\right|$ be the absolute valued weights corresponding to the sorted samples, where $\mathbf{I(k)}$ is the location in the observation window of the *k*-th order statistic. Then a necessary and sufficient condition for $\mathbf{S_{(N-k)}}$, $\mathbf{0 \leq k \leq N - 1}$, to be the output of the WOS filter is

$$\mathbf{k = min}\left\{\mathbf{j}\,\left|\,\sum_{i=0}^{j}\left|\mathbf{W_{I(N-i)}}\right| \geq \mathbf{W_0}\right.\right\} \tag{3}$$

For the above example, for $\mathbf{j = 1}$,

$$\left|\mathbf{W_{I(N)}}\right| + \left|\mathbf{W_{I(N-1)}}\right| = \left|\mathbf{W_2}\right| + \left|\mathbf{W_1}\right| = \mathbf{0.4 + 0.2 \geq 0.55}\ ,$$

therefore the filter outputs $\mathbf{S_{(N-1)} = 6}$.

Note that as $\mathbf{W_0} \rightarrow \mathbf{0}$, WOS filter outputs the largest signed sample whereas for $\mathbf{W_0} \rightarrow \sum_{i=1}^{N}\left|\mathbf{W_i}\right|$, the filter output becomes the smallest value in the signed set. Note also that WOS filters include median, rank order, and weighted median (WM) filters as special cases. In particular, if the selection parameter $\mathbf{W_0}$ is set to $\sum_{i=1}^{N}\left|\mathbf{W_i}\right|\mathbf{/2}$, the WOS filter reduces to a weighted median (WM) filter admitting real-valued weights (Arce, 1998). Furthermore, if the weights are restricted to be nonnegative WOS filters become WOS smoothers (Yu & Liao. 1994; Yli-Harja *et al*. 1991).

THRESHOLD DECOMPOSITION AND WOS FILTERS

Threshold decomposition (TD) is a useful tool utilized in the analysis and optimization of stack filters (Arce, 1998) and smoothers (Wendt *et al*. 1986). TD has been initially developed for signal quantized to finite number of levels (Wendt *et al*. 1986) and extended later to real-valued signals (Yin & Neuvo. 1994; Arce. 1998). For the purpose of this paper, we adopt a similar approach to that described in (Arce, 1998). Let $\mathbf{Z} \in \mathfrak{R}^{\mathbf{N}}$ be an *N*-dimensional vector, where $\mathfrak{R}$ denotes the set of real numbers.

The threshold decomposition acting on this vector is defined as:

$$\mathbf{T} : \mathfrak{R}^{\mathbf{N}} \times \mathfrak{R} \rightarrow \aleph^{\mathbf{N}}$$

$$\left.\mathbf{T(Z, q)}\right|_{\mathbf{i}} = \begin{cases} \mathbf{1} & \mathbf{Z_i \geq q} \\ \mathbf{-1} & \mathbf{Z_i < q} \end{cases} \qquad \mathbf{i = 1, 2,..., N} \tag{4}$$

where $q \in \mathfrak{R}$ and $\aleph = \{\mathbf{-1, 1}\}$. Thus, TD maps a real-valued vector to an infinite set of binary vectors. For short notation, TD of a vector $\mathbf{Z}$ at level q is denoted as

$$\mathbf{z^q = [z_1^q,...,z_N^q]^T = [sgn(Z_1 - q),..., sgn(Z_N - q)]^T}.$$

Each component of the original real-valued vector can be reconstructed from its binary representation as:

$$\mathbf{Z_i = \frac{1}{2}\int_{-\infty}^{\infty} z_i^q\, dq} \tag{5}$$

Now, let $\mathbf{X}$ be an observation vector and let s^q be the TD of the signed observation vector at level q, i.e.,

$$\mathbf{s^q = [s_1^q,...,s_N^q]^T = [sgn(sgn(W_1)X_1 - q),...,sgn(sgn(W_N)X_N - q)]^T},$$

for some filter coefficients $\mathbf{W_i}\Big|_{i=1}^{N}$. In the binary domain, the output of the WOS filter with weights $\mathbf{W_i}\Big|_{i=1}^{N}$ and threshold parameter $\mathbf{W_0}$ reduces to

$$\mathbf{f(s^q)} = \begin{cases} \mathbf{1\ \ if\ \sum_{i=1}^{N}\left|W_i\right|(s_i^q + 1) \geq 2\,W_0} \\ \mathbf{-1\ \ otherwise} \end{cases} \triangleq \mathbf{sgn(W^T \xi^q)} \tag{6}$$

where $\mathbf{W = [W_0\,, \left|W_1\right|, \left|W_2\right|,..., \left|W_N\right|\,]^T}$ and

$\xi^q = \left[\xi_0^q, \xi_1^q,..., \xi_N^q\right]^T$ with $\xi_0^q = -1$ and $\xi^q = \dfrac{\mathbf{s_i^q + 1}}{\mathbf{2}}$,

$\mathbf{i = 1,2,..., N}$.

In (6) the compact representation (3) of the filtering operation has been exploited. Furthermore, $\mathbf{f}(\mathbf{s}^q)$ is a special case of Boolean functions, and is called threshold logic gate (Muroga, 1971).

Using (5) the real-valued WOS filter output can be obtained by an integration operation as follows

$$\mathbf{Y} = \frac{1}{2}\int_{-\infty}^{\infty} \mathbf{sgn}\left(\mathbf{W}^T \xi^q\right)\mathbf{dq} \qquad (7)$$

Note that applying WOS filter on a real-valued signal is equivalent to decomposing the real-valued signal using threshold decomposition, applying the binary WOS filter (6) to each binary signal separately, and then adding up the binary outputs to obtain the real-valued output (Paredes & Arce, 1999).

The integral term in (7) seems to be difficult to implement. However, further simplification of this expression can be achieved if we note that for any $\mathbf{q} \in (-\infty, \mathbf{S}_{(1)}]$ or $\mathbf{q} \in (\mathbf{S}_{(i-1)}, \mathbf{S}_{(i)}]$, $\mathbf{i = 2,..., N}$ or $\mathbf{q} \in (\mathbf{S}_{(N)}, \infty]$, TD outputs the same binary vector. Hence, there are at most **N + 1** different binary vectors $\mathbf{s}^q$. Using this fact, Eq. (7) becomes

$$\mathbf{Y} = \frac{1}{2}\left[\int_{-\infty}^{S_{(1)}} \mathbf{sgn}(\mathbf{W}^T \xi^{S_{(1)}})\mathbf{dq}\right] + \frac{1}{2}\left[\sum_{i=2}^{N}\int_{S_{(i-1)}}^{S_{(i)}} \mathbf{sgn}(\mathbf{W}^T \xi^{S_{(i)}})\mathbf{dq} + \int_{S_{(N)}}^{\infty} \mathbf{sgn}(\mathbf{W}^T \xi^{\infty})\mathbf{dq}\right] \qquad (8)$$

As $\xi_i^{\infty} = \frac{s_i^{\infty} + 1}{2} = \mathbf{0}$ and after some simplification, (8) reduces to

$$Y = \frac{\mathbf{S}_{(1)} + \mathbf{S}_{(N)}}{\mathbf{2}} + \frac{\mathbf{1}}{\mathbf{2}}\sum_{i=2}^{N}(\mathbf{S}_{(i)} - \mathbf{S}_{(i-1)})\ \mathbf{sgn}(\mathbf{W}^T \xi^{S_{(i)}}) \qquad (9)$$

where $\mathbf{S}_{(i)}$ is the *i*-th smallest signed sample. This is another form to implement the WOS filtering operation and gives the same result as the one obtained using the three-step procedure showed above. This representation, however, provides us with an interesting interpretation of the WOS filter. It turns out that, the filter output is given by the sum of the midrange of the signed samples $(\mathbf{S}_{(1)} + \mathbf{S}_{(N)})/\mathbf{2}$ and a linear combination of the difference between successive order statistic $(\mathbf{S}_{(i)} - \mathbf{S}_{(i-1)})$ weighted by a factor of $\pm\mathbf{1/2}$.

EFFECTS OF NEGATIVE WEIGHTS IN THE FILTERING OPERATION

In this section, we study the effect that real-valued weights have in the WOS filtering process. The approach followed in this section is different to the one presented by Arce in (Arce, 1998) where the effects of negative weighting in WM filters are studied using the theory of maximum likelihood estimate of location and the analogies between FIR and WM filters. Here, we exploit the concept of weight monotonic property recently introduced in (Marshall, 2002; Marshall, 2004).

As mentioned above, WOS filters belong to a richer class of nonlinear filters, the so-called stack filters (Paredes & Arce, 1999) and, hence, they satisfy the stacking property. More precisely, if two binary vectors $\mathbf{u} \in \aleph^N$ and $\mathbf{v} \in \aleph^N$ stack (i.e. $u_i \leq \upsilon_i$ for $\mathbf{i = 2,..., N}$), their respective outputs stack, $f(\mathbf{u}) \leq f(\mathbf{v})$. This property is a restriction on the type of signal processing tasks that can be carried out by a filter possessing this property.

Based on this fact, Marshall developed in (Marshall, 2002) a test to define whether the kind of filters satisfying the stacking property are useful or not for a specific application. Thus, a simple test is applied to the training data set to determine if it is suitable for processing with a filter having the stacking property. The basic idea in the test is to determine whether the training samples possess the weight monotonic property, defined as follows (Marshall, 2002).

Let $\{\mathbf{X}(\cdot)\}$ be an observed process that is statistically related to a desired process $\{\mathbf{D}(\cdot)\}$. Define a set of training samples $\{(\mathbf{X}(n), \mathrm{D}(n))$ for $\mathrm{n} = \mathrm{n}_0,..., \mathrm{L}, \mathrm{L} > \mathrm{n}_0\}$, where $\mathbf{X}(n)$ is the *N*-dimensional observation vector defined as before. This set of training samples satisfies the weight monotonic property if

$$\mathbf{Pr}(\mathbf{d} = \mathbf{1} \mid |\mathbf{x}| = \mathbf{k}) \geq \mathbf{Pr}(\mathbf{d} = \mathbf{1} \mid |\mathbf{x}| = \mathbf{j}) \qquad (10)$$

for $\mathbf{k} > \mathbf{j}$ with $|\mathbf{x}| = \frac{\mathbf{N}}{\mathbf{2}} + \frac{\mathbf{1}}{\mathbf{2}}\sum_{i=0}^{N}\mathbf{x}_i$ and where $\mathbf{Pr}(\cdot \mid \cdot)$ is the conditional probability, *d* and $\mathbf{x}_i$ are the TD of D and $\mathbf{X}_i$, respectively, at an arbitrary threshold level *q*. Note that $|\mathbf{x}|$ is just the number of 1's in the observation window. Thus, the weight monotonic property holds for distributions where the probability of the output being 1 increases with $|\mathbf{x}|$. If this property holds then stack filters, like WOS filters, will yield good performance in restoring the desired signal from the corrupted one; otherwise it is useless (Marshall, 2004).

For the applications at hand, we are interested in the use of WOS filters in a frequency selection type of filtering process and therefore the weight monotonic property should be held in order to achieve the desired performance. For instance, consider the application of designing a high-pass filter whose main goal is to preserve a high-frequency tone while removing all low-frequency terms. More precisely, let $\mathbf{X(n) = sin(2\pi f_1 n) + sin(2\pi f_2 n + \pi/4) + \eta(n)}$ be the corrupted input signal where $\eta(\cdot)$ is the noise obeying a normal distribution $N(0, 0.1)$ and let $\mathbf{D(n) = sin(2\pi f_2 n + \pi/4)}$ be the desired signal for some $\mathbf{f_1 < f_2}$.

To see if the weight monotonic property is valid for this set of training samples the output conditional probability is determined for two different cases of filter weights. In the first one, all the filter weights are assumed to be positive, i.e. a WOS smoother is used. In the second case, the filter weights are allowed to be negative. In this later case, to compute $|\mathbf{x}|$, the TD of the signed samples is used in place of the TD of the original input sample.

Figure 1(a) depicts the conditional probability function when all the filter weights are positive for an observation window size 7. Note that the values of $\mathbf{Pr\ (d = 1|\ |x|)}$ do not increase monotonically as the number of 1's in the observation window increases. Therefore, the weight monotonic property does not hold and WOS smoothers should be discarded for this application. On the other hand, Fig. 1(b) shows the output conditional probability for the same window size and $\mathbf{sgn(W_2) = -1}$, $\mathbf{sgn(W_6) = -1}$, all the other weights are positive. Note that using negative weights the value of $\mathbf{Pr\ (d = 1|\ |x|)}$ does increase monotonically as $|\mathbf{x}|$ increases, and therefore WOS filters can indeed be used for this application.

In this example, we have shown the effect that negative weights have on the conditional probability. We observe that passing the sign of the weights to the observation samples, indeed, changes $\mathbf{Pr\ (d = 1|\ |x|)}$, forcing the weight monotonic property to hold and therefore enriching the field of application of WOS filters and, overcoming the limitations of WOS smoothers.

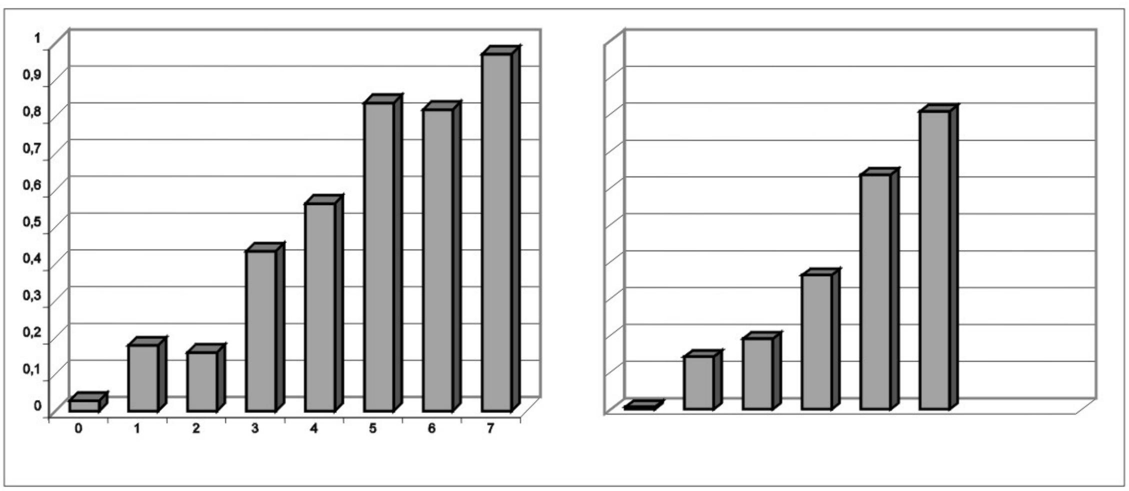

Figure 1. Output conditional probability as a function of number of 1's in the observation window for an observation window size of 7 samples. (a) All filter weights are positive,
(b) $\mathbf{sgn(W_2) = -1}$, $\mathbf{sgn(W_6) = -1}$ the other weights are positive.

OPTIMIZATION OF WOS FILTERS

The main objective in an optimization algorithm is to find the best filter coefficients such that a performance criterion is minimized. A criterion widely used in the design of WOS smoothers (Yin *et al*. 1996) is the mean absolute error (MAE) between the filter's output and the desired signal. If $\{D(\cdot)\}$ and $\{X(\cdot)\}$ are jointly stationary, then the cost function to be minimized is

$$J(W) = E\{|D(n) - Y(n)|\} \tag{11}$$

where $E\{\cdot\}$ denotes the statistical expectation, $W = [W_0, W_1,..., W_N]^T$ and $Y(n)$ is the WOS filter output given by (2) for an input observation vector $X(n)$. Using TD, it is easy to prove that Eq. (11) reduces to

$$J(W) = \frac{1}{2}\int_{-\infty}^{\infty} E|sgn(D(n) - q) - sgn(W^T\xi^q(n))|dq \tag{12}$$

$$= \frac{1}{4}\int_{-\infty}^{\infty} E(sgn(D(n) - q) - sgn(W^T\xi^q(n)))^2 dq \tag{13}$$

where the fact that, for any instant n, the argument inside the absolute value operator can only take on values either in the set $\{-2, 0\}$ or in the set $\{0, 2\}$, has been used to interchange the absolute value and the integral operators in (12) and to replace the absolute value operator by a properly scaled second power operator in (13).

Since it is not possible to find a closed form expression for the filter coefficients that minimize (13), we resort to the steepest descend algorithm to form the iterative optimization algorithm

$$W(n + 1) = W(n) + 2\mu[- \nabla J] \tag{14}$$

where the gradient of the cost function (∇J) has to be computed to update the filter weights. However, since the sgn function has a discontinuity at the origin, its derivative introduces the *dirac* function which is not convenient for further analysis. To overcome this difficulty, the sgn function is approximated by a differentiable function. We approximate sgn(x) by tanh(x) whose derivative is sech(x). With this approximation, the gradient of the cost function becomes

$$\frac{\partial J}{\partial W_0} = \frac{1}{2}\int_{-\infty}^{\infty} E\left\{e^q(n)\sec h^2(W^T\xi^q(n))\right\}dq \tag{15}$$

$$\frac{\partial J}{\partial W_\ell} = -\frac{1}{2}\int_{-\infty}^{\infty} E\left\{e^q(n)\sec h^2(W^T\xi^q(n))sgn(W_\ell)\xi_\ell^q(n)\right\}dq \tag{16}$$

for $\ell = 1,..., N$, and where $e^q(n)$ is the TD of the error at threshold level q, i.e. $e^q(n) = sgn(D(n) - q) - sgn((W^T\xi^q(n))$. Upon closer examination, this error term takes on nonzero values only if q is between the desired output $D(n)$ and the actual filter output $Y(n)$. Assuming that the desired output is one of the signed samples, say $S_{(m)}$, and that the actual output $Y(n)$ is $S_{(j)}$; $e^q(n) \neq 0$ only for $q \in (\min(S_{(m)}, S_{(j)}), \max(S_{(m)}), S_{(j)}]$. Furthermore, as mentioned in section III, for any instant n, $s^q(n)$ takes at most $N + 1$ different values, so does $\xi^q(n)$. Those vectors $\xi^q(n)$ are for $q = S_{(i)}(n)$, $i = 1, 2,..., N$ and $q = \infty$. Hence, combining these two facts, the adaptive optimization algorithm reduces to:

$$W_0(n + 1) = P[W_0(n)] - P\left[\mu \sum_{i = \min(m, j) + 1}^{\max(m, j)} (S_{(i)} - S_{(i - 1)})\, e^{S(i)} \sec h(W^T\xi^{S(i)})\right] \tag{17}$$

$$W_\ell(n + 1) = W_\ell(n) + \mu\, sgn(W_\ell) \sum_{i = \min(m, j) + 1}^{\max(m, j)} (S_{(i)} - S_{(i - 1)})\, e^{S(i)} \sec h(W^T\xi^{S(i)})\xi_\ell^{S(i)} \tag{18}$$

for $\ell = 1,..., N$, and where the instantaneous estimate for the gradient has been used. $P[\cdot]$ is a projection operator defined as $P(x) = x$, if $x \geq 0$ and 0 otherwise. This ensures that the selection parameter takes on nonnegative values (Yin & Neuvo. 1994).

Further simplifications of this optimization algorithm can be achieved following the same arguments used in (Arce. 1998) and (Yin & Neuvo. 1994). It turns out that each term in the summations of (17) and (18) contributes to the weight updating by an amount that basically depends on the function $\sec h(W^T\xi^q)$, which reaches a maximum when its argument is zero. Therefore, for some value of q, $W^T\xi^q$ takes its closest value to zero. It is easy to show that for $q = Y(n)$, $W^T\xi^q$ is closest to zero (Yin & Neuvo. 1994; Arce. 1998). Thus, $q = Y(n)$ produces the largest update contribution. Using this fact, the optimization algorithm reduces to:

$$W_0(n + 1) = P[W_0(n) - \mu(D(n) - Y(n))] \tag{19}$$

$$W_\ell(n + 1) = W_\ell(n) + \mu sgn(W_\ell)(D(n) - Y(n))\xi_\ell^{Y(n)} \tag{20}$$

for $\ell = 1, 2,..., N$, and where

$$\xi_\ell^{Y(n)} = \frac{1}{2}(sgn(sgn(W_\ell)X_\ell - Y(n)) + 1).$$

The principle of the adaptive optimization algorithm can be explained as follows. When the actual output of the WOS filter is smaller than the desired signal, the selection parameter (W_0) is decremented whereas the magnitude of those weights corresponding to the signed samples that are larger than the filter output are incremented (they become more positive if they are positive or more negative if they are negative). Note that the new values of $\mathbf{W_0}$ and $\mathbf{W_i}|_{i=1}^{N}$ try to push the estimate toward the desired signal. A similar analysis can be done if the output of the WOS filter is larger than the desired output.

COMPUTER SIMULATIONS

To illustrate the performance of the proposed filters and their corresponding adaptive optimization algorithm, two frequency selective applications are carried out. In the first example, we have designed a high-pass WOS filter using the simplified version of the adaptive optimization algorithm, and compare it to the performances yielded by a linear FIR filter, a WOS smoother (Yin & Neuvo, 1994) and a WM filter (Arce, 1998). As performance measures we use the MSE and MAE as well as a visual inspection of the output signals. The second example concerns with the design of a robust frequency selective filter. This last application is quite helpful in biomedical signal processing since it can be used to reduce the effect of the undesired, but always present, 60 Hz frequency component in a biomedical signal.

A. Design of a high-pass WOS filter

The first application at hand is the design of a 33-tap high-pass WOS filter with a cut-off frequency of 0.15 Hz (normalized Nyquist frequency equal to 1). As a training signal we use the two-tone signal described in Section IV with $\mathbf{f_1 = 0.03Hz}$ and $\mathbf{f_2 = 0.25Hz}$, where the desired signal is the higher frequency component. Whereas as test signals the two-tone signal is contaminated with nonsymmetric α-stable noise simulating noise with impulsive characteristic (Nikias & Shao, 1995).

The simplified adaptive optimization algorithm is used to design the WOS filter where the filter weights were initialized to small random numbers (on the order of 10^{-1}) and the step-size has been fixed to 10^{-3}.

For comparative purposes, a 33-tap linear FIR filter is designed using Matlab's fir1 function with a cut-off frequency of **0.15Hz**. On the other hand, a WOS smoother is designed for the same task using Yin's fast adaptive algorithms (Yin & Neuvo, 1994). Likewise, the fast optimization algorithm introduced by Arce in (Arce, 1998) is used to design a WM filter. For both algorithms, the same training data, weight initial condition and step-size, as described above, are used.

Figure 2 shows the performance of the various filters for the application at hand on noise-free signal (left half) and on impulsive noise signal (right half). Note, in Fig. 2(e), that the WOS smoother fails to remove the low-frequency component, as expected since those filters can only synthe-

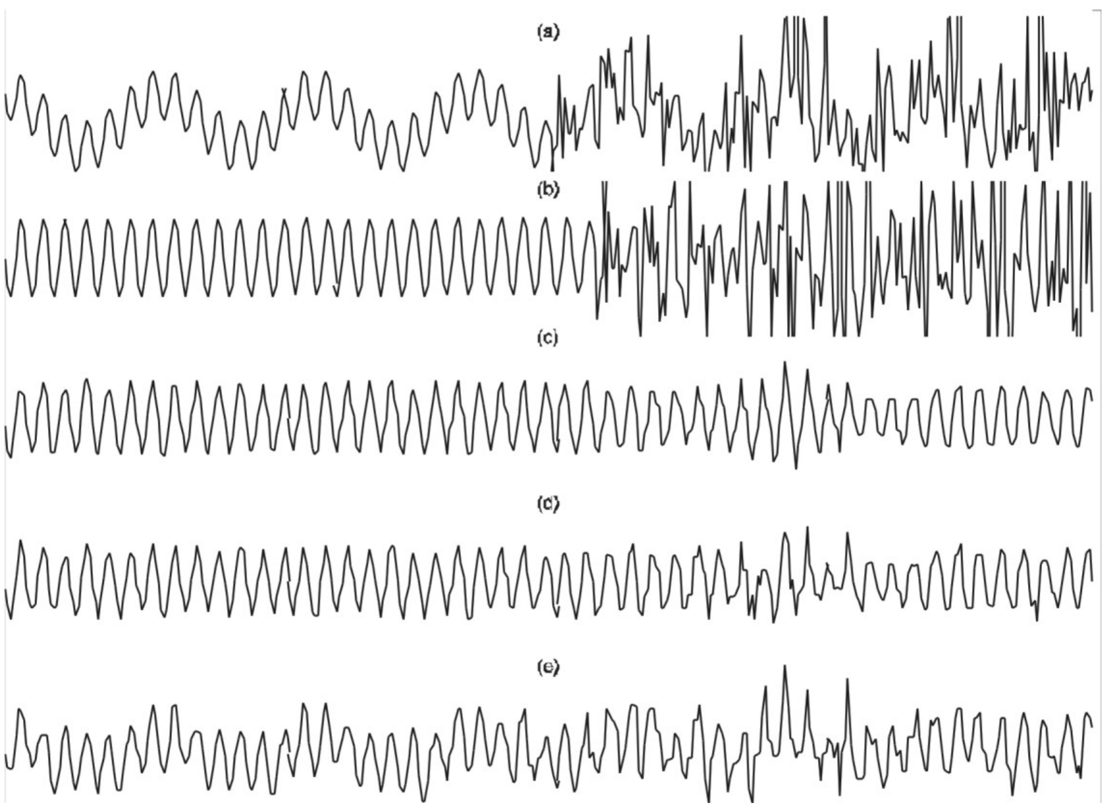

Figure 2. High-pass filter design. (a) Input test signal, (b) Linear high-pass FIR filter, (c) Optimal WOS filter, (d) Optimal WM filter, (e) Optimal WOS smoother with non-negative weights.

size low-pass behavior. Note also that WM filter as well as WOS filter yield similar performance, at least, visually for the noise-free part of the signal.

In order to test the robustness of the different filters, the two-tone signal is contaminated with additive nonsymmetric impulsive noise. The additive noise was chosen to have a zero-mean nonsymmetric α-stable distribution (Nikias & Shao, 1995) with a *characteristic exponent* $\alpha = \mathbf{1.5}$, a dispersion $\delta = \mathbf{0.5}$ and a *skewness parameter* $\beta = \mathbf{0.75}$. Impulsive noise is well-modelled by the heavy-tailed class of α-stable distributions which include the Gaussian distribution as special case when $\alpha = \mathbf{2}$ and $\beta = \mathbf{0}$. The characteristic exponent ($0 < \alpha \leq 2$) measures the heaviness of the tails (a smaller value indicates heavier tails), whereas the dispersion decides the spread of the distribution around the origin. The skewness parameter ($-1 \leq \beta \leq 1$) is a shape parameter that defines the degree of symmetric around the origin of the distribution. Thus, when $\beta = \mathbf{0}$ the distribution is symmetric whereas as $\beta \to \mathbf{1}$ ($\beta \to \mathbf{-1}$) the probability that a negative (positive) impulse occurs is decreased.

Figure 2 (right half) depicts the filter outputs of the various filter for a noise input signal. Note the poor performance of the linear FIR filter. On the other hand, WOS and WM filter outputs, shown in Fig. 2(c) and Fig. 2(d), respectively, are not severely degraded by the impulsive noise. However, several minor artifacts appear at the output of the WM filter that are not present in the WOS filter output.

The MSE and MAE for several values of parameter α are computed and shown in Table 1. Each entry in the table was obtained by averaging out the MSE and MAE of 500 realizations of the contaminated two-tone signal. The results in Table I indicate that WOS filters outperform the others for noise signals. Furthermore, the linear FIR filter outperforms the nonlinear structures in the noise-free case, as expected.

Finally, Fig. 3 shows the ensemble-averaged absolute error for the simplified iterative algorithm. This curve plots the absolute error, averaged over 100 trials, as a function of algorithm iterations. As this curve indicates, the adaptive algorithm converges in about 350 iterations.

Table I. Comparison of MSE and MAE for each filter

	MSE				MAE			
Filter type	noise free	$\alpha = 1.25$	$\alpha = 1.50$	$\alpha = 1.75$	noise free	$\alpha = 1.25$	$\alpha = 1.50$	$\alpha = 1.75$
Identity	0.5000	12.0275	2.8458	1.3756	0.6363	1.3712	1.0285	0.8845
FIR filter	10^{-5}	9.1951	1.8528	0.7082	0.0026	1.2414	0.7742	0.5936
WOS filter	0.0547	0.0940	0.0878	0.0804	0.1792	0.2417	0.2336	0.2230
WM filter	0.0621	0.1054	0.0981	0.0896	0.1963	0.2568	0.2476	0.2365
WOS smoother	0.1687	0.2295	0.1941	0.1941	0.3305	0.3805	0.3599	0.3529

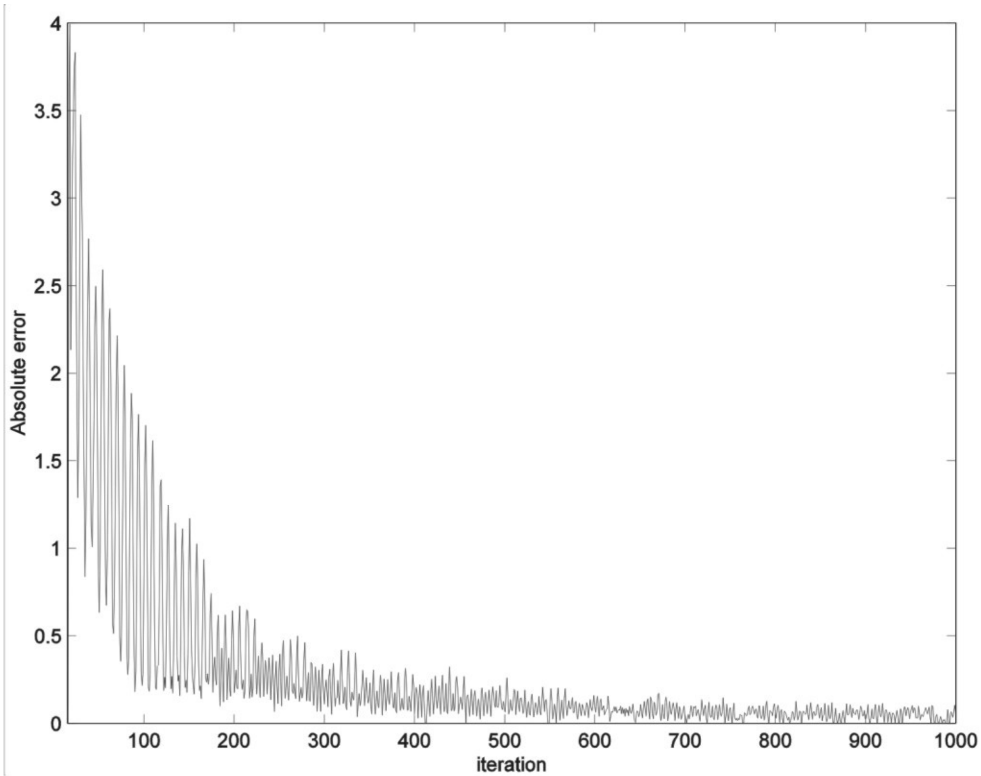

Figure 3. Ensemble-averaged absolute error learning curve for the adaptive WOS filter algorithm.

B. Design of a robust frequency selective WOS filter

To test the frequency selective capabilities of the proposed filter, the adaptive optimization algorithm described in Section V is used to design a 43-tap WOS filter to remove the middle-frequency tone from a signal compounded of three sinusoidal signals corrupted by impulsive noise. The training signals are: the observed signal $\mathbf{X(n)} = \sum_{k=0}^{2} \mathbf{\sin(2\pi f_k n)} + \eta\mathbf{(n)}$ and the desired signal $\mathbf{D(n)} = \mathbf{\sin(2\pi f_0 n) + \sin(2\pi f_2 n)}$ where $\mathbf{f_0} = \mathbf{0.015}$, $\mathbf{f_1} = \mathbf{0.12}$, and $\mathbf{f_2} = \mathbf{0.20}$ (normalized Nyquist frequency equal to 1). The additive noise is nonsymmetric α-stable noise with $\alpha = \mathbf{1.5}$, $\beta = \mathbf{0.5}$ and $\delta = \mathbf{0.5}$. The test signal, shown in Fig. 4(a), is a 1000-sample sequence where the first 500 samples are noise-free whereas the last 500 samples are corrupted with impulsive noise.

Figure 4(d) shows the output of the optimal WOS filter. It can be seen that the performance of WOS filters is not severely affected by the presence of impulsive noise. In contrast, a 43-tap FIR filter designed using Matlab's fir1 function with passed bands $\mathbf{0 \le f \le 0.09}$ and $\mathbf{0.15 \le f \le 0.25}$ performances poorly in the part of the input signal corrupted by impulsive noise. Figure 4(c) shows the FIR filter output for the test signal.

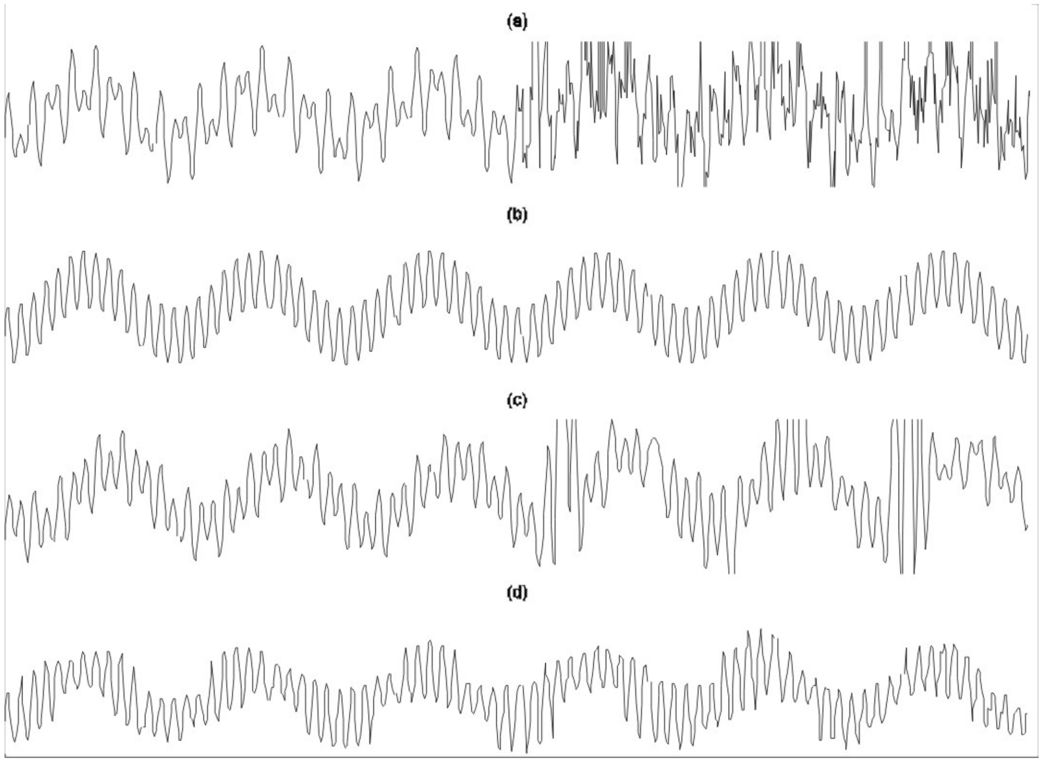

Figure 4. Frequency selective filter design. (a) Input test signal. (b) Desired signal. (c) Linear FIR filter output. (d) WOS filter output.

CONCLUSION

In this paper, we have presented a new class of robust nonlinear filters based on order statistic. The proposed filter can be suitable in applications where robustness as well as high-pass or bandpass behavior are needed in nonsymmetric impulsive noise environments. An adaptive optimization algorithm for the design of this kind of filters was also introduced. The performance of the proposed filters was compared to the performances of the linear FIR filters, WOS smoothers and WM filters through computer-based simulations. It was shown that the proposed filters outperform the other filters not only in the MSE and MAE performance measures but also visually.

ACKNOWLEDGMENT

This work was supported by the University of Los Andes Council of Scientific, Humanistic and Technological Development (CDCHT) under project I-768-04-02-B. Mérida. Venezuela.

REFERENCES

Arce, G.R., Hall, T.A. & Barner, K.E. (Aug. 1995), Permutation weighted order statistic filter lattice. IEEE Transactions Image Processing, Vol. 4, No. 8, pp. 1070–1083.

Arce, G.R. (Dic. 1998), A general weighted median filter structure admitting negative weights. IEEE Transactions Signal Processing, Vol. 46, No. 12, pp. 3195–3205.

Li, Q. (May. 2000), Delay characterization and performance control of wide-area networks," Ph.D. dissertation, University of Delaware, Newark [Online]. Available in: www.ece.udel.edu/~qli.

Marshall, S. (Sep. 2002), Optimum non linear image restoration through linear greyscale operations. Presented at EUSIPCO. Toulouse, France.

Marshall, S. (Feb. 2004), New direct design method for weighted order statistic filters. Vision, Image and Signal Processing IEE Proceedings, Vol. 151, No. 1, pp. 1–8.

Middleton, D. (May. 1999), Non-gaussian noise models in signal processing for telecommunications: new methods an results for class A and class B noise models. IEEE Transactions on Information Theory, Vol. 45, No. 4, pp. 1129 – 1149.

Mitra, S.K. & Sicuranza, G. (2001), Nonlinear Image Processing, (Communications, Networking and Multimedia). Academic Press.

Muroga, S. (1971), Threshold Logic and its Applications. New York, USA: Wiley-Interscience.

Nikias, C. & Shao, M. (1995) Signal Processing with Alpha-Stable Distributions and Applications. New York: Wiley.

Paredes, J.L. & Arce, G.R. (Oct. 1999), Stack filters, stack smoothers and mirrored threshold decomposition. IEEE Transactions Signal Processing, Vol. 47, No. 10, pp. 2757–2767.

Song, J. & Lee, Y.H. (May. 1996), Linear combination of weighted order statistic filters: Canonical structure and optimal design. IEEE Transactions Circuits Syst. II, Vol. 43, No. 5, pp. 349–362.

Wendt, P., Coyle, E. & Gallagher, N.J. (Aug. 1986), Stack filter. IEEE Transactions Speech Audio Processing, Vol. 34, No. 4, pp. 898–911.

Yin, L. & Neuvo, Y. (Jul. 1994), Fast adaptation and performance characteristics of FIR-WOS hybrid filters. IEEE Transactions Signal Processing, Vol. 42, No. 7, pp. 1610–1628.

Yin, L., Yang, R., Gabbouj, M. & Neuvo, Y. (Mar. 1996), Weighted median filters: A tutorial. IEEE Transactions Circuits Syst. II, Vol. 47, pp. 157–192.

Yli-Harja, Astola, J., & Neuvo, Y. (Feb. 1991), Analysis of the properties of median and weighted median filters using threshold logic and stack filter representation. IEEE Transactions Signal Processing, Vol. 39, No. 2, pp. 395–410.

Yu, P. & Liao, W. (Oct. 1994), Weighted order statistics filters — their classification, some properties, and conversion algorithm. IEEE Transactions Signal Processing, Vol. 42, No. 10, pp. 2678–2691.