Generalized Median Filtering and Related Nonlinear Filtering Techniques

YONG HOON LEE, MEMBER, IEEE, AND SALEEM A. KASSAM, SENIOR MEMBER, IEEE

Abstract—We consider some generalizations of median filters which combine properties of both the linear and median filters. In particular, $L$ filters and $M$ filters are considered, motivated by robust estimators which are generalizations of the median as a location estimator. A related filter, which we call the modified trimmed mean (MTM) filter, is also described. The filters are evaluated for their performance on noisy signals containing sharp discontinuities or edges. It is shown that $M$ filters can offer a more favorable combination of the running mean and median filters than can $L$ filters, while MTM filters generally have better characteristics than $M$ filters. We also show that an MTM filter is a data-dependent modification of $L$ filters. The concept of double-window filtering is introduced as a refinement of MTM filtering. One representative set of filtered sequences of a test input using these filters are presented to illustrate the performance characteristics of these filters.

I. INTRODUCTION

In many signal processing applications, the suppression of unwanted components in an input discrete-time signal sequence can be achieved by linear filtering. In some situations, however, linear filtering is inadequate for this purpose. If a desired signal with sharp edges is corrupted by noise, as in some noisy image data, then linear filters designed to remove the noise will also smooth out signal edges. In addition, impulsive noise components cannot be suppressed sufficiently by linear filtering. In such cases, some form of nonlinear or adaptive filtering would be preferable.

A nonlinear scheme called median filtering has been used with some success in these situations. The median filtered version of a discrete-time signal is obtained by replacing the input signal value at each point by the median of the signal values in some finite neighborhood around that point. Note that in linear FIR filtering the input value at a point is replaced instead by a linear combination of the samples in a neighborhood of the point. The simple nonlinear action of the median filter gives it some very interesting characteristics. In particular, it generally preserves edges and monotonic trends while suppressing impulsive noise components quite well. Following the use of the median filter by Tukey [1], it has been applied in several areas of digital signal processing which include speech processing [2], [3] and image enhancement [4]–[6]. In such applications, the underlying signal to be enhanced often has edges and impulsive noise components. To develop some design procedures for median filtering, its properties, both deterministic and statistical, have been investigated [7]–[9]. In addition, fast algorithms and hardware implementations for median filtering have been devised [10]–[12].

Although median filtering can be used to suppress impulsive noise components while preserving sharp edges, it often fails to provide sufficient smoothing of nonimpulsive noise components. In [13], two techniques are introduced in an effort to overcome this limitation. One technique is to do repeated filtering until the resulting output is sufficiently smooth, or until no further changes occur. Recently it has been shown that any sequence of length $L$ is converted under repeated median filtering to a sequence which is invariant to further median filtering after, at most, $(L - 2)/2$ passes [9]. This convergence property of repeated median filtering is useful in some signal processing applications such as in image coding [14]. However, insofar as smoothing is concerned, it is difficult to predict whether the result will be "smooth" or not. In the worst case, as pointed out in [9], the result can be the noise sequence only. The other technique introduced is the use of a median filter in combination with a linear filter. Here, linear filtering is introduced because the averaging operation is more powerful in suppressing nonimpulsive noise components. This technique works fairly well in some applications such as speech processing [2]. It should be noted, however, that low-pass linear filters will always smear signal edges. Therefore, the linear filter for this technique must be fairly low-order system (a 3-point Hann window has been used). It can be expected that some other nonlinear techniques should exist which inherently combine linear and nonlinear operations, and which allow a degree of control over the relative influences of these operations, giving better results.

Bovik, Huang, and Munson [15] have recently introduced a generalization of median filters. The defined an order statistic filter (OSF) in which the input value at a point is replaced by a linear combination of the ordered values in a neighborhood of the point. The class of OSF filters includes as special cases the median filter, the running mean filter, and the min (max) filter which uses an
extreme value instead of the median [6]. Note that an OSF combines a nonlinear operation, i.e., ordering, with a linear operation.

In this paper, we will consider another generalization of the median filter which stems from robust estimation theory. We shall note that the OSF also has a basis as a robust estimation technique. In estimation of a location parameter from a set of observations, the sample mean gives the least-squares estimate. On the other hand, the sample median is the estimate minimizing the sum of absolute deviations of the observations from the estimate. Although many properties of the sample mean are more desirable than those of the sample median, the sample mean performs poorly in the presence of outliers among the observations (impulsive noise) [16],[17]. In order to obtain estimates which enjoy some of the desirable properties of both the sample mean and the sample median, one can consider using an estimator minimizing the sum of a function other than the square or absolute value of the deviations. Suppose a set of observations \( x_i = \theta + w_i, i = 1, \ldots, n \), are given, where \( \theta \) is a location parameter and \( \{w_i\}_{i=1}^n \) is a sequence of i.i.d. zero-mean noise components. Then for some even function \( \rho(x) \) which is nondecreasing in \( |x| \), we can define an estimator \( \hat{\theta} \) of \( \theta \) which minimizes \( \sum_{i=1}^n \rho(x_i - \hat{\theta}) \), or which equivalently satisfies \( \sum_{i=1}^n \Psi(x_i - \hat{\theta}) = 0 \) where \( \Psi(x) = d\rho(x)/dx \) is an odd function. Note that the choice \( \Psi(x) = x \) gives us the sample mean for \( \rho(x) = x^2 \), and \( \Psi(x) = \text{sgn}(x) \) gives us the sample median for \( \rho(x) = |x| \), in the latter case, assuming that the \( x_i, i = 1, \ldots, n \), are all distinct. The class of estimators generalized in this fashion are called \( M \) estimators. Such estimates are generalized forms of maximum-likelihood estimates, for which \( \rho(x) = -\log f(x) \) where \( f \) is the probability density function of the \( w_i \). A filtering procedure which uses a running \( M \) estimator will be called an \( M \) filter. Thus, the median and running mean filter are special cases of \( M \) filters. There exists another class of robust estimators for a location parameter, called \( L \) estimators, which use a linear combination of ordered samples for the location parameter estimate. Therefore, use of a running \( L \) estimator for filtering is exactly the same as the use of an OSF. From now on an OSF will be referred to as an \( L \) filter for consistency in our terminology. Note that some aspects of \( M \) filters has been previously considered in the statistical literature [18], and \( L \) filters have been mentioned in [19] and [20], although not for the applications we have mentioned above.

As a very interesting and useful variation of median filters, we will also introduce what we will call the modified trimmed mean (MTM) filter. The MTM filter selects the sample median from a window centered at a point, and then averages only those samples inside the window close to the sample median. It will be seen that the MTM filter is closely related to the \( L \) and \( M \) filters. Furthermore, we will show that the MTM filter can be thought of as a data-dependent modification of \( L \) filters.

The window sizes of \( L \), \( M \), and MTM filters are constrained as in median filtering. If the duration of a narrow pulse of the signal to be preserved is \( I \), then the maximum window size allowed is \( 2I - 1 \) in the noise-free case. One may, however, need to use a window size larger than \( 2I - 1 \) for sufficient smoothing of nonimpulsive noise components, but then fine signal details will be filtered out. To overcome this difficulty, a double-window (DW) variation of the MTM filter will be introduced. In the DW MTM filter, a small and a large window are used to produce each output point, the small window allowing retention of fine details of the signal and the large window allowing noise suppression to be adequate. Specifically, the DW MTM filter uses the median computed from observations within a smaller window centered at a point to reject outliers in the large window, from which an average value is then computed. We will show that the DW MTM filter can have very good performance characteristics.

The organization of this paper is as follows. In Sections II and III we consider \( L \) and \( M \) filters, respectively, and discuss their properties. In Section IV the MTM and DW MTM filters are defined and their properties are discussed. We present some simulation results in Section V.

II. The \( L \) Filter

The output \( y_k \) of an \( L \) filter of window size \( 2N + 1 \) at time index \( k \), for an input sequence \( \{x_i\} \) is given by

\[
y_k = \sum_{j=1}^{2N+1} A_j x_{(j)}^k
\]

where \( x_{(j)}^k \) is the \( j \)th smallest sample among the \( 2N + 1 \) samples inside the window centered at \( k \), and where \( \{A_j\}_{j=1}^{2N+1} \) is a set of constant weights with \( \sum_{j=1}^{2N+1} A_j = 1 \). Throughout this paper we assume that the input and output sequences are two-sided infinite. Fig. 1 illustrates the operation of an \( L \) filter, showing the separation of the nonlinear (ordering) and linear (weighted averaging) operations. By proper choice of weights, \( L \) filters are tunable from the linear running mean filter to nonlinear filters such as the median or min (max) filters.

\( L \) filters have properties similar to those of median filters. These include the following properties, where the output sequence \( \{y_i\} \) of a smoother \( S \) is written in general as \( \{y_i\} = S(\{x_i\}) \) for input sequence \( \{x_i\} \).

Property 1 (Scale and Translation Invariance): For any symmetric \( L \) filter, meaning one for which the filter weights satisfy

\[
A_i = A_{2N+1-(i-1)}, \quad i = 1, 2, \ldots, 2N + 1
\]

\[
S(a \{x_i\} + b\{1\}) = aS(\{x_i\}) + b\{1\}
\]

where \( \{1\} \) is the sequence of constant values 1, and \( a, b \) are any constants.

To prove this we need to note only that when \( \{z_i\} = \{a \{x_i\} + b\{1\} \}, \) we have \( z_{(j)}^k = \{a \{x_{(j)}^k\} + b; \) the result follows from symmetry of the filter weights. Note that (2) does not remain valid from any arbitrary nonconstant sequence replacing the sequence \( \{1\} \), so that the \( L \) filter has a limited linearity property (superposition does not hold).
patterns. When an edge enters into a window the median operation is the identity operation. Since a discontinuity that when the input is any will, in general, smear it. The extent to which edges are smeared depends on the set of weights. These should be symmetric weights; the result is obtained by using property

\begin{equation}
S \{ x_i \} = \{ x_i \}
\end{equation}

when \( \{ x_i \} = a \{ i \} + \{ b \} \), for any constants \( a \) and \( b \).

To prove this property, we observe that when the input sequence for the \( L \) filter is \( \{ i \} \) the output at time \( k \)

\begin{align*}
y_k &= \sum_{j=1}^{2N+1} A_j x_{k-j-(N+1)} \\
&= \sum_{j=1}^{2N+1} A_j [k + j - (N + 1)] \\
&= k
\end{align*}

where the third equality comes from the symmetry of the weights; the result is obtained by using property 1. Note that when the input is any monotone sequence, the ordering operations of a symmetric \( L \) filter at each window may be disregarded, so that the output of the \( L \) filter is exactly the same as that of a linear FIR filter with the same weights. Therefore, symmetric \( L \) filter cannot preserve arbitrary monotone trends except in the special case of median filtering. In this case, the corresponding linear FIR operation is the identity operation. Since a discontinuity between two flat portions of an input sequence (an “ideal” edge) constitutes a nonlinear monotone trend, an \( L \) filter will, in general, smear it. The extent to which edges are smeared depends on the set of weights. These should be nonzero only over a central portion of the window of length \( 2N + 1 \) for good edge preservation. The above discussion illustrates that in between linear behavior (constant weighting, running mean filter) and nonlinear edge-preserving behavior (median filter), a variety of weighting patterns can be employed to obtain useful compromise characteristics.

In addition to good edge retention with adequate averaging out of random noise, one often requires that impulsive noise components are also suppressed reasonably well. Impulsive noise results in the input having relatively large spikes, which cannot be suppressed effectively using linear averaging operations. An \( L \) filter for which the weights \( A_j \) are close to zero for values of \( j \) near 1 or \( 2N + 1 \) can be expected to be more effective in this situation. This is because impulsive noise spikes will tend to appear at the ends of an ordered sequence of values inside any window. We observe that preservation of edges and suppression of impulsive noise require similar weighting patterns. When an edge enters into a window the median

will tend to be associated with data from one side of the edge. The data from the other side of the edge should then be treated as impulsive noise components and not be allowed to heavily influence the smoothed output value. Therefore, filter weights which tend to suppress impulsive noise will also tend to preserve edges. However, when the center of the window hits an edge \( N \) out of \( 2N + 1 \) values should be treated as impulsive noise. This explains why the median filter is the only \( L \) filter which can preserve an ideal edge.

In [15], the optimal weights are obtained for an \( L \) filter, minimizing the output variance, when the input sequence is a constant signal with additive white noise. It is interesting to note that for Gaussian noise, the optimum \( L \) filter is the running mean filter, whereas for the double-exponential noise probability density function (pdf), the optimum weights are appreciably nonzero only over a narrow central portion of the window. This result agrees with our expectations since the double-exponential probability density function is often used to model impulsive noise components.

A particularly simple structure for an \( L \) filter is obtained by imposing the requirement that all weights be equal to a constant within some central portion of the window, with the remaining weights taken to be zero. Suppose that only \( 2(N - T) + 1 \) central weights are taken to have nonzero values \( 1/(2(N - T) + 1) \). Let us define \( T \) in terms of a parameter \( \alpha \) by \( T = \lfloor \alpha (2N + 1) \rfloor \), where \( 0 \leq \alpha \leq 0.5 \) and \( \lfloor X \rfloor \) is the largest integer less than or equal to \( X \). Then the output \( y_k \) for this \( L \) filter is given by

\begin{equation}
y_k = \frac{1}{2N - T + 1} \sum_{j=T+1}^{2N+1} x_{k,j} \tag{5}
\end{equation}

In robust estimation theory, the \( L \) estimator based on these weights is called the \( \alpha \)-trimmed mean estimate. The \( \alpha \)-trimmed mean estimate first deletes or trims a number \( \lfloor \alpha (2N + 1) \rfloor \) of samples from each end of the ordered data set and then calculates the mean of the remaining numbers. When \( \alpha = 0 \), the trimmed mean becomes the usual mean, and when \( \alpha = 0.5 \), it becomes the median. From now on, the \( L \) filter described by (5) will be referred to as the \( \alpha \)-trimmed mean filter \(( \alpha \)-TM filter). The \( \alpha \)-TM filter, as a generalization of the median filter, has also recently been introduced in [21].

The optimal \( T \) (or \( \alpha \)) for a given window size can be obtained when the input is a constant signal plus additive white noise by using the results in [15] and [22]. Suppose the input sequence \( \{ x_i \} = \{ s + n_i \} \) where \( s \) is a constant signal and \( \{ n_i \} \) is a zero-mean i.i.d. noise sequence. Then the output variance of an \( L \) filter is given by

\begin{equation}
E[(y_k - s)^2] = E \left[ \left( \sum_{j=1}^{2N+1} A_j n_{k,j} \right)^2 \right] = \sum_{i=1}^{2N-1} \sum_{j=1}^{2N+1} A_i A_j H_{ij} \tag{6}
\end{equation}

where \( H_{ij} = E[n_{i,j}^2] \). The matrix of \( H_{ij} \)'s for a variety of
noise distributions have been obtained numerically in [22]. We note that the weights of an \( \alpha \)-TM filter are uniquely determined for given \( T \) and \( 2N + 1 \). Thus, by using the \( H_{ij} \) for a particular noise distribution and window size we can calculate the output variances for \( T = 0, 1, \cdots, N \). The optimum \( T \) is then chosen as the one which gives the minimum variance. Tables I and II show the output variances for \( \alpha \)-TM filters driven by white noise with Gaussian and double-exponential distributions, respectively. Window sizes \( 2N + 1 = 3, 5, 7, \) and \( 9 \) were used and the variances were evaluated for \( T = 0, 1, \cdots, N \) for each window size. For Gaussian input distribution, \( \alpha = 0 \) gives the minimum variance, while for the double-exponential distribution, \( \alpha = 0.4 \) gives the minimum variance. These results are as expected, in light of our earlier discussion on the optimum coefficients for \( L \) filtering.

From the definition of the \( \alpha \)-TM filter, we see that it can suppress up to \( T \) impulsive noise components on either side of the sample median within a window. The edge characteristic of an \( \alpha \)-TM filter is illustrated by considering its effect on an ideal edge. Suppose that the input sequence is binary with only one transition from \( 0 \) to \( H > 0 \) at time \( j \), i.e.,

\[
x_k = \begin{cases} 
0, & k \leq j \\
H, & k \geq j.
\end{cases}
\]

(7)

Then the output of the \( \alpha \)-TM filter is given by

\[
y_k = \begin{cases} 
0, & k < j - (N - T) \\
H, & k \geq j + (N - T) \\
[k - (j - N + T) + 1]H/[2(N - T) + 1], & \text{otherwise}.
\end{cases}
\]

(8)

Note that the smearing of the edge is directly affected by the height of the edge \( H \), and the edge becomes a ramp. In fact, for this input the \( \alpha \)-TM filter works exactly like a running-mean filter with window size \( 2N + 1 - 2T \). It can be said, in general, that an \( \alpha \)-TM filter with window size \( 2N + 1 \) smears edges, but the amount of smearing is less than that of the running mean filter with the same window size.

To reduce smearing of edges when an \( \alpha \)-TM filter is applied, it could be useful if \( \alpha \) can be chosen according to a simple adaptive scheme. Such a scheme can be based on the utilization of an edge detector, and use of the value \( \alpha = 0.5 \) when an edge is found to exist with a reasonably high probability at any time index \( k \). Otherwise, an \( \alpha \)-trimmed mean can be used for some fixed choice of \( \alpha \) in between 0 and 0.5, obtained according to a priori information about the relative frequency of impulsive noise components.

We have observed that there exists a clear tradeoff between the linear (noise averaging) and nonlinear (edge preserving and impulsive noise suppression) properties of the \( L \) filter. In the next section, we shall see that \( M \) filters can offer a more favorable combination of these two properties.

### III. The \( M \) Filter

The output \( y_k \) of an \( M \) filter is defined as a solution of the equation

\[
\sum_{i=k-N}^{k+N} \Psi(x_i - y_k) = 0
\]

(9)

where \( \Psi \) is some odd, continuous, and sign-preserving function, so that \( \Psi(x) \) is positive (negative) whenever \( x \) is positive (negative). When \( \Psi \) is the linear function defined by \( \Psi(x) = ax \), for some constant \( a \), the \( M \) filter reduces to a running mean filter. We will see, on the other hand,
that the $M$ filter approaches the median filter in its characteristics as $\Psi$ approaches the hard limiter (signum function) under certain conditions. Before proceeding further let us establish the following basic results.

**Lemma 1:** A solution $y_k$ of (9) always exists; if $\Psi$ is严格 increasing, the solution is unique.

The proof of this lemma is quite simple and is omitted. We now present the following two properties of an $M$ filter which are very similar to those for $L$ filters that we considered in Section II.

**Property 1 (Translation Invariance):** For any $M$ filter, there exist solutions for the output sequence satisfying

$$S(\{x_i\} + a\{1\}) = S(\{x_i\}) + a\{1\}$$

for any constant $a$.

This property is easily established. Note that the scale invariance property that holds for $L$ filters no longer holds for $M$ filters because $\Psi(ax) \neq a\Psi(x)$ except when $\Psi(x)$ is linear. In this sense the $M$ filter is, in general, more nonlinear than the $L$ filter.

**Property 2 (Linear Trend Preservation):** For any $M$ filter, there exists a solution for the output sequence satisfying

$$S(\{x_i\}) = \{x_i\}$$

when $\{x_i\} = a\{i\} + \{b\}$, for any constants $a$ and $b$.

This property is proved by using the assumption that $\Psi$ is an odd function $[\Psi(x) = -\Psi(-x)]$. As in $L$ filters, $M$ filters cannot, in general, preserve arbitrary monotone sequences.

By a limiter type $M$ filter (LTM filter) we mean an $M$ filter for which

$$\psi(x) = \begin{cases} g(p), & x > p \\ g(x), & |x| \leq p \\ -g(p), & x < -p \end{cases}$$

where $g(x)$ is a strictly increasing, odd, continuous function, and $p$ is some positive constant. When $g(x) = ax$ we get, in particular, an $M$ filter which will be described as being of the standard type. A standard type $M$ filter (STM filter) characteristic function $\Psi$ is shown in Fig. 2. Without loss of generality, $a = 1/p$ will be used for the STM filter throughout the paper.

Although Lemma 1 cannot be used to establish the uniqueness of the output for the LTM filter, the output is unique as shown below.

**Lemma 2:** For an LTM filter with filter parameter $p$, there exists a unique output $y_k$ at each time $k$. The output always lies in the range

$$m_k - \delta \leq y_k \leq m_k + \delta$$

where $m_k$ is the median in the window, and $\delta = g^{-1}\{[N/(N + 1)]g(p)\}$.

The proof of Lemma 2 is given in Appendix A. Inequality (13) is an important result. It gives a bound for the output of an LTM filter in terms of the output of a median filter, the function $g$, and the parameter $p$. Note that as $N$ goes to infinity, (13) becomes

$$m_k - p \leq y_k \leq m_k + p$$

which is obviously also valid for all $N$. From (14) we can see that as $p$ approaches zero (in the STM filter this means that $\Psi(x)$ approaches the hard limiter), LTM filtering approaches median filtering.

*Fig. 2. $\Psi$ function of STM filters.*
\[ y_k = \frac{1}{2} \left( x_{(3)}^k + m_k + p \right). \]

3) If \( x_{(1)}^k > m_k - p, x_{(3)}^k \geq m_k + p \) and \( \Sigma_{j=1}^{3} \Psi_x^k (x_{(j)}^k - (x_{(3)}^k - p) < 0 \) then \( g(p) + \Sigma_{j=1}^{3} g(x_{(j)}^k - y_k) = 0 \). In this case, for the STM filter, we have
\[ y_k = \frac{1}{2} \left( x_{(1)}^k + m_k + p \right). \]

4) Otherwise, \( \Sigma_{j=1}^{3} g(x_{(j)}^k - y_k) = 0 \), so that for the STM filter, \( y_k = \mu_k \) where \( \mu_k \) is the sample mean in the window.

The proof of the above results is given in Appendix D. Note that Observation 3 allows us to compute the outputs of STM filters quite easily for window size 3.

In \( M \) filtering, the choice of the function \( \Psi \) could be based on a criterion such as output variance for a constant signal in white noise input, as we did for \( L \) filters in Section II. However, it is rather difficult to perform such an analysis for the case of finite \( N \). On the other hand, asymptotic variance expressions are available in the statistical work on robust \( M \) estimation [17]. We find, in particular, that an STM filter does have the property that its asymptotic variance is a minimum for the worst-case \( \epsilon \)-contaminated Gaussian noise probability density function (pdf) \( f(x) = (1 - \epsilon) \eta(x) + \epsilon h(x) \); here \( \eta \) is a zero-mean Gaussian pdf with known variance, \( h \) is an arbitrary symmetric pdf, and \( \epsilon \) is a given degree of contamination. In particular, the worst-case contamination is such as to make the tails of \( f \) have an exponential decay, which is more characteristic of impulsive noise. For this reason and also because it is simpler to implement and analyze, we will now focus on the STM filter. We will show eventually, through simulation results, that it does perform quite well for the filtering problem we described in the Introduction.

From Observation 3, we see that for a fixed value of \( p \), the output of an STM filter with window size 3 can be the sample mean or the sample median or some other value depending on the data. This property holds for an STM filter with any larger window size. We will get the sample median as the output if and only if
\[ \sum_{j=1}^{N} \Psi_x^k (x_{(j)}^k - m_k) = \sum_{j=N+2}^{2N+1} \Psi_x^k (x_{(j)}^k - m_k). \]

On the other hand, the sample mean is obtained when, roughly speaking, all the data in a window are close to the sample median. The following observation gives a sufficient condition for obtaining the sample mean as the output for an STM filter when all \( 2N + 1 \) samples lie in the range \([m_k - p, m_k + p]\).

Observation 4: Suppose
\[ |x_{(1)}^k - m_k| \leq r_1 \leq p \quad \text{and} \quad |x_{(2N+1)}^k - m_k| \leq r_2 \leq p. \]

If either
\[ r_1 \leq \frac{2N + 1}{2N} p - \frac{1}{2} r_2 \]

when
\[ \sum_{j=1}^{N} \Psi_x^k (x_{(j)}^k - m_k) \leq \sum_{j=N+2}^{2N+1} \Psi_x^k (x_{(j)}^k - m_k) \]

then the output of an STM filter is the sample mean.

The proof of Observation 4 is given in Appendix E. It is clear that if the value of \( p \) is made sufficiently large, the STM filter will behave primarily as a running mean filter for slowly varying signals in additive white noise.

Suppose that the input to an STM filter is an ideal edge, defined by (7). Then the output sequence is given by
\[ y_k = \begin{cases} 0, & \text{if } j < j - N \\ \frac{wp}{2N + 1 - w}, & \text{if } j - N \leq k < j \\ H - (N + 1 - w) p/w, & \text{if } j \leq k < j + N \\ H, & \text{if } k \geq j + N \end{cases} \]

for \( w = k - (j - N) + 1 \), when
\[ H > \left( 1 + \frac{N}{N + 1} \right) p. \]

Thus, \( 2N \) output points are different from corresponding input values, as in the case of linear filtering with the same window size, and the edge is smeared.

In Section II, we observed that only \( 2(N - T) \) output points of the \( \alpha \)-TM filter are different from corresponding input values for an ideal edge input. It should be noted, however, that the smearing of the edge described by (17) does not depend on the height of the edge \( H \) but only on the parameters \( p \) and \( N \). Therefore, when \( p \) is reasonably small compared to \( H \), the degree of smearing of edges in STM filtering will be much less than that in \( \alpha \)-TM filtering. In \( \alpha \)-TM filtering, the smearing of edges may be confined to an interval of length \( 2(N - T) \) output points, but the degree of smearing is higher because the output increases linearly between the two extreme values over these points. In STM filtering, the smearing or deviation of the smeared output from the ideal edge occurs over \( 2N \) points, but the magnitude of the deviation is always less than \( p \) as expected from (14). In fact, the STM filter often preserves sharp edges almost as a median filter does, as we shall see in Section V through simulation results. We can see that the STM filter tends to treat as outliers those data points which are reasonably large or small compared to the sample median. Thus, in STM-filtering the ordered data in each window are, in general, treated as
nonsymmetric way depending on the data, unlike \( L \) filtering. When an edge exists whose height is large compared to the noise standard deviation, a properly designed STM filter tends to smooth only those samples lying on one side of the edge, the side which includes the sample median. We conclude that the STM filter can perform like a running-mean filter when neither edges nor impulsive noise exist, and tends to work more like a median filter when edges exist. Impulsive noise can be suppressed effectively by STM filtering because it basically limits the influence of observations deviating substantially from the sample median in any window.

The value of \( p \) in STM filtering should be chosen to get a good compromise between edge-preservation and impulsive noise suppression on the one hand, and nonimpulsive noise suppression on the other. For the latter, \( p \) should be in general at least as large as \( \sigma \), the standard deviation of the noise, whereas for preservation of edges and suppression of impulsive noise, we need \( p \) to have as small a value as is feasible. From (17), using \( k = j - 1 \) or \( j \), it is observed that the maximum amount of smearing of the ideal edge in STM filtering is \( pN/(N+1) \). Thus, for instance, if we want to design an STM filter with maximum amount of smearing approximately \( H/4 \), where \( H \) is the minimum height of edges, \( p \) should be less than \( H(N+1)/4N \). In this case, it should be possible to design an appropriate STM filter if \( \sigma < H(N+1)/4N \).

In this section, we have observed that STM filters can offer more favorable combinations of linear and nonlinear filtering characteristics than can \( L \) filters. We will introduce in the following section some further filtering schemes which are motivated by the desirable characteristics of STM filtering, and which can perform better and are simpler to implement.

Before going to Section IV, let us briefly consider the estimation of \( p \), say \( \hat{p}_k \) at time \( k \), when little information is known \textit{a priori} about the signal and noise. Based on our earlier discussion, it would appear to be reasonable to use the sample standard deviations inside successive windows in setting local values of \( p \). However, the sample standard deviations are heavily influenced by impulsive components and edges inside the windows. For robust \( M \) estimation of a location parameter with an unknown scale parameter, it has been suggested [16], [17] that a useful estimate \( \hat{p} \) of \( p \) is obtained by using the median of the absolute deviations of the data from their sample median (called the MAD estimate). Thus, one choice for the \( \hat{p}_k \) is

\[
\hat{p}_k = C \text{MAD}_k
\]

where \( \text{MAD}_k = \text{median}_{i=2,4,\ldots,2k+1} |x_i - m_k| \), \( m_k \) is the sample median, and \( C \) is a constant. For Gaussian noise \( C \approx 1.5 \) makes \( \hat{p}_k \) a consistent estimate of \( \sigma \).

Several techniques have been mentioned in [17] which can be used to evaluate the output of the STM filter. For our simulation results the iterative Newton's method was used. It was observed that each output value was obtained within five iterations with absolute errors of less than 0.01.

IV. MODIFIED TRIMMED MEAN FILTERS

We have made the observation that in STM filtering the ordered data inside any window are treated in a nonsymmetric way which is inherently data dependent. This allows an STM filter to have better overall performance than does a simple \( \alpha \)-TM filter. In this section, we will use our understanding of the STM filter's characteristics to obtain modified trimmed mean filters (MTM filters) which are very simple to implement and produce results which, in many cases, are at least as good as those obtainable using STM filters.

In \( \alpha \)-TM filtering, a set of \((N - T)\) samples closest to the sample median \( m_k \) is selected from each of the two sets of \( N \) samples on either side of the sample median. Then the average of the \( 2(N - T) \) selected samples and the sample median is used as the output. It can be said that, equivalently, the \( \alpha \)-TM filter averages only those samples in a window inside a range \([m_k - q_1, m_k + q_2]\) where \( q_1 \) and \( q_2 \) depend on the data. A filter which will be called the \textit{modified trimmed mean filter} (MTM filter) is suggested by the above observation. The MTM filter first determines the sample median \( m_k \) inside its window and then chooses an interval \([m_k - q, m_k + q]\) using some prespecified constant \( q \). Within the window, data samples outside this range are discarded and the average value of the rest of the data is used as the output. Note that the number of samples used in the averaging is not fixed \textit{a priori} in STM filtering.

Now consider the behavior of the STM filter. Suppose that the output at time \( k \) is \( y_k \), and \( x_i \)'s are the samples inside the window. We can say that the STM filter modifies those samples for which \( x_i - y_k \leq -p \) and \( x_i - y_k \geq p \) by replacing them with \( y_k - p \) and \( y_k + p \), respectively; it then averages the new set of samples to result in \( y_k \). Since \( y_k - p \leq m_k \leq y_k + p \) from (14), it can be said that, equivalently, the STM filter replaces the samples outside a range \([m_k - q_1, m_k + q_2]\) with \( m_k - q_1 \) or \( m_k + q_2 \), depending on whether \( x_i \leq m_k - q_1 \) or \( x_i \geq m_k + q_2 \), respectively, where \( q_1 = m_k - y_k + p \) and \( q_2 = y_k + p - m_k \). Note that \( q_1 \) and \( q_2 \) also depend on the data through \( y_k \). Also, the number of samples inside the range \([m_k - q_1, m_k + q_2]\) is not fixed. Thus, there is an interesting similarity between MTM and STM filters; the main difference is that the MTM filter may reject certain data values, whereas the STM filter effectively only limits the influence of some data values.

The MTM filter is able to reject impulsive noise components because for each output point it starts by obtaining the median value inside the corresponding window. In fact, the size of the window is constrained exactly as in median filtering, so as to allow signal features of a specified minimum width not to be treated as impulses. For any particular window size, a very small value for \( q \) makes the MTM filter behave as a median filter, whereas for very large \( q \) values the filter behaves as a running-mean filter. To get good performance in white Gaussian noise suppression, and simultaneously for edge preservation, the value of \( q \) should be picked to be the maximum value for which edges
can be expected to be preserved. This means that we should pick \( q \) to be approximately \( H-2\sigma \), where \( H \) is the minimum height of edges and \( \sigma \) is the noise standard deviation, for \( H \) approximately equal to or larger than \( 4\sigma \). When \( H \) is relatively small or when information on edge heights is not available, \( q \) may be chosen from knowledge of the noise standard deviation \( \sigma \) only; for example, one can use for \( q \) the value \( 2\sigma \). A smaller value of \( q \) will result in better edge retention but will give poor smoothing, and vice versa. Here again, \( \sigma \) may be estimated using the MAD estimator of (19).

The MTM filter preserves an ideal edge as does a median filter, provided \( q < H \). We can show that the MTM filter can be expected to have better edge preservation for noisy signals. Suppose the ideal edge defined by (7) is corrupted by additive white Gaussian noise with zero mean and standard deviation \( \sigma \). When the window is centered at the edge, that is, at time \( j-1 \) (or \( j \)), the largest (or smallest) sample among the samples to the left (or right) of the center point will, in general, be chosen as the sample median, which therefore causes a smearing of the edge. We can say that the median filter smears a noisy edge because data on both sides of an edge in a window affect the ordering operation. In MTM filtering, however, the output is the averaged value of observations which are in some neighborhood of the sample median. Therefore, for an appropriately chosen \( q \) value, edges will be better preserved in MTM filtering.

It is quite significant that it is possible to view the MTM filter as a data-dependent \( L \) filter with a weighting pattern not constrained to be symmetric. Suppose at some time \( k \), a number \( \gamma \) of data samples among the \( 2N+1 \) samples in the window are within the range \([m_k - q, m_k + q]\). Clearly, averaging only over the \( \gamma \) data points is equivalent to weighting these \( \gamma \) samples by \( 1/\gamma \) and assigning weights 0 to the remaining 2\( N+1-\gamma \) samples. If the data inside the window are assumed to have been ordered, these \( \gamma \) samples occupy consecutive positions in the ordered set.

Thus, the weighting pattern \( \{A_1, A_2, \cdots, A_{2N+1}\} \), which is effectively applied to the ordered data, is a sequence of \((1/\gamma)\)'s flanked by zeros on one or both sides. Neither \( \gamma \) nor the location of the segment of successive \((1/\gamma)\)'s in the weights is fixed, but depends on the data. For instance, consider the ideal edge defined by (7) as the input to a MTM filter with window size 3 and \( q < H \). Then one of the three weighting patterns is automatically selected depending on the data; the weighting is \((1/2, 1/2, 0)\) when \( k = j - 1 \), it is \((0, 1/2, 1/2)\) when \( k = j \), and it is \((1/3, 1/3, 1/3)\) otherwise.

For the type of filtering schemes we have been discussing, the use of a large window size will generally imply a loss of fine details or narrow pulses in a signal. In addition, the degree of edge smearing will generally increase (except for the median and MTM filters) as window size increases. On the other hand, we may need a filter with a large window size to suppress additive nonimpulsive noise sufficiently.

Once the median \( m_k \) within some window has been picked in MTM filtering, one may select to average data values which fall inside the range \([m_k - q, m_k + q]\) from among data in a larger window centered at \( k \). This leads us to the double-window modified trimmed mean (DW MTM) filter which is defined more explicitly as follows. Let windows of length \( 2N+1 \) and \( 2L+1 \), with \( L > N \), be centered at \( k \) (see Fig. 3). First, the sample median \( m_k \) is computed from the small window of size \( 2N+1 \). For some positive number \( q \), an interval \([m_k - q, m_k + q]\) is chosen. Then the mean of points lying within the interval \([m_k - q, m_k + q]\) among the samples in the large window of size \( 2L+1 \) is computed as the output.

The DW MTM filter suppresses nonimpulsive noise effectively by averaging data samples in a large window while it retains narrow signal pulses by choosing the sample median from a smaller window. The behavior of the DW MTM filter varies from that of the median filter with window size \( 2N+1 \) to that of the running mean filter with window size \( 2L+1 \), depending on the data.

The size \( 2N+1 \) of the small window in DW MTM filtering is constrained as in median filtering. This is to prevent loss of narrow signal pulses of a specified minimum width and more generally to retain signal details of such widths. If \( q \) is made very small, the DW MTM behaves almost as a median filter, irrespective of the large window size \( 2L+1 \). On the other hand, a large value of \( q \) produces, in the limit, a running mean filter of window size \( 2L+1 \). The value of \( q \) is generally chosen as in MTM filtering \((q = H - 2\sigma \) for minimum edge height \( H \) and noise standard deviation \( \sigma \)). Thus, the size of the large window is governed by the bandwidth of the low-pass part of the signal; a running mean filter would use a similar window size if no edges were present in the signal. We find that the DW MTM filter performs best relative to other filters we have discussed for signals which are close to being piecewise constant.

V. SIMULATED PERFORMANCE

In this section, we present one representative set of results illustrating the performance characteristics of the various filters we have discussed so far. Fig. 4 shows the...
Fig. 4. Test input consisting of signal and additive white Gaussian noise, with three impulses.

Fig. 5. Outputs of (a) median, (b) running mean, and (c) combination filters, $2N + 1 = 7$.

Fig. 6. Outputs of (a) $\alpha$-TM filter with weights $(0, 0, 1/3, 1/3, 1/3, 0, 0)$, and STM-filters with (b) $p = \sigma$ and (c) $2\sigma$, $2N + 1 = 7$.

particular test input that was used for this set of results. This test input was obtained by adding white Gaussian noise of standard deviation $\sigma = 2$ and three impulsive noise components to an original signal shown as the dashed curve in Fig. 4. The original signal has flat portions, a sinusoidal part, and a narrow pulse of duration 4 units; the minimum edge height was 10.

Fig. 5(a)-(c) shows the results of filtering the test input by the median, the mean, and the combination filter, respectively, where the latter is the median filter followed by a 3-point Hanning window with weights $(1/4, 1/2, 1/4)$. A window size $2N + 1 = 7$ was used for all single window filters in this section because the duration of the narrow pulse to be preserved is $I = 4$. It is seen that the running mean filter smears edges severely and fails to suppress impulsive noise sufficiently, while the median filter preserves edges and suppresses impulsive noise reasonably well. However, the running mean filter performs better than the median filter for nonimpulsive noise suppression (see the center portion of the sine wave). It is conspicuous that the portion of the median filtered output where the first two impulsive noise spikes exist before filtering is quite noisy. This is because the ordering operation of the median filter is influenced by the impulsive noise, so that the underlying white Gaussian noise there is not adequately suppressed. The result of combination filtering shows one reasonably good compromise between median and running mean filtering.

Fig. 6(a)–(c) exhibits the results of generalized median filtering, $\alpha$-TM filtering with $2/7 \leq \alpha < 3/7$, and STM-filtering with $p = \sigma$ and $p = 2\sigma$, respectively. Both the $\alpha$-TM and the STM filter are better in suppression of noise but worse with regard to edge preservation, compared to
the median filter. However, especially in STM filtering with \( p = \sigma \), edges are preserved almost as in median filtering. Comparing the results of STM filtering to \( p = \sigma \) and \( p = 2\sigma \), we can see that the larger \( p \) is preferred for noise suppression but not for edge preservation. The STM filter with \( p = 2\sigma \) is seen to degrade considerably the last edge, the original height of which was 10.

Fig. 7(a)–(b) shows the results of MTM filtering with \( q = 2\sigma \) and \( 3\sigma \), respectively. The MTM filter is seen to be better in noise suppression than the median filter without smearing out edges. In fact, as we discussed in Section IV, the MTM filter preserves edges better than does the median filter. We find that the MTM filter with \( q = 3\sigma \) is better than one with \( q = 2\sigma \), as we would expect; note that \( H - 2\sigma = 3\sigma \) here. For STM filtering, our simulation shows that the choice \( p = 2\sigma \) [see Fig. 6(c)] results in good noise suppression but causes smearing out of the last edge of height 10. The noise suppression with \( p = 2\sigma \) is the STM filter was better than that in MTM filtering with \( q = 3\sigma \). In Section IV, we made the observation that in STM filtering the samples outside the range \( [m_k - q_1, m_k - q_2] \) are replaced by \( y_k - p \) or \( y_k + p \), while in MTM filtering they are discarded. This explains the behavior we observe in the simulations: the STM filter suppresses white Gaussian noise slightly better than does MTM filtering even when \( p \) is less than \( q \).

Fig. 7(c) shows the results of DW MTM filtering, again with \( q = 3\sigma \). The size of the large window and the small window were 11 and 7, respectively. The output is better than any other filtered output shown so far, with respect jointly to noise suppression and edge preservation.

The results of adaptive STM, MTM, and DW MTM filtering using the MAD estimator are shown in Fig. 8 (a)–(c), respectively. For STM filtering, the constant \( C \) in (19) was chosen to be 1.5 (making the \( \hat{p}_k \) consistent estimators of \( \sigma \) for Gaussian noise), and for MTM and DW MTM filtering the value of \( C \) was chosen to be 3.0. The results of adaptive filtering are not as good as those of corresponding nonadaptive filtering methods which depend on a priori knowledge of the value of \( \sigma \). Nonetheless, the adaptive filters may be said to perform somewhat better than the median filter. More work remains to be done on
adaptive versions of the filtering techniques we have considered in this paper. For example, one promising approach to setting parameter values is to use estimates of the noise standard deviation obtained from portions of the input signal which are identified as being free of impulses and edges.

VI. CONCLUSION

We have discussed two classes of nonlinear filtering techniques, the $L$ and $M$ filters, and a specific useful modification of such filters, the MTM filter. All of these filters include the median and running mean filters as special cases. It was observed that $M$ filters can offer a more favorable combination of the characteristics of running mean and median filters than can $L$ filters. Nonetheless, both $L$ and $M$ filters have characteristics of linear filters at the expense of those of median filters. The MTM filters were shown to provide better overall characteristics. In fact, we have seen that MTM filters can preserve noisy edges better than can median filters. Thus, repeated use of an MTM filter is possible and may be useful in filtering out severe nonimpulsive noise.

The DW MTM filters were introduced as a variation of MTM filters, and were also indicated to be very useful for filtering an input with relatively low signal-to-noise ratio because of their use of a larger window size for nonimpulsive noise suppression.

APPENDIX A

Proof for Lemma 2: Since $0 \leq \delta < p$ and $N$ is a positive integer, $g(p)/2 \leq g(\delta) < g(p)$. Suppose a solution $y_k$ is less than $m_k - \delta$. Then $\delta < m_k - y_k$, so that

$$g(\delta) = \Psi(\delta) \leq \Psi(m_k - y_k) \quad (A.1)$$

because when $m_k - y_k \leq p$, $\Psi \equiv g$ and when $m_k - y_k > p$, $\Psi(m_k - y_k) = g(p)$. Note that (A.1) can be rewritten as

$$\Psi(x_{(N+1)}^k - y_k) > \Psi(x_{(N-1)}^k - (m_k - \delta)). \quad (A.2)$$

Since $y_{k} < m_{k} - \delta$ and $\Psi$ is nondecreasing

$$\Psi(x_{(j)}^k - y_k) \geq \Psi(x_{(j)}^k - (m_k - \delta)) \quad (A.3)$$

for $j = N + 2, N + 3, \cdots, 2N + 1$. Combining (A.1) and (A.2) gives

$$\sum_{j=N+1}^{2N+1} \Psi(x_{(j)}^k - y_k) > \sum_{j=N+1}^{2N+1} \Psi(x_{(j)}^k - (m_k - \delta))$$

$$\geq (N + 1) \Psi(\delta)$$

$$= N \cdot g(p)$$

where the second inequality comes from the fact that $x_{(j)}^k \geq m_k$ for all $j = N + 1, \cdots, 2N + 1$. But $\sum_{j=1}^{N} \Psi(x_{(j)}^k - y_k) \geq -N \cdot g(p)$ so that $\sum_{j=1}^{N} \Psi(x_{(j)}^k - y_k) > 0$, which is a contradiction. Therefore, $y_k \geq m_k - \delta$. Similarly, $y_k \leq m_k + \delta$, so that (A.3) has been proved. To prove the uniqueness of $y_k$, it is enough to show that $\sum_{j=1}^{N} \Psi(x_{(j)}^k - y_k) - \sum_{j=1}^{N} \Psi(x_{(j)}^k - (m_k - \delta))$ is strictly decreasing with respect to $y$ when $-p \leq m_k - y_k \leq p$, from (14). In this range, $\Psi(x_{(j)}^k - y_k) = g(x_{(j)}^k - y_k)$, which is strictly decreasing with respect to $y$. Since all other terms $\Psi(x_{(j)}^k - y_k)$ are nonincreasing with respect to $y$, $\sum_{j=1}^{N} \Psi(x_{(j)}^k - y_k)$ is strictly decreasing for $m_k - p \leq y_k \leq m + p$.

APPENDIX B

Proof for Observation 1: Define the sets $L_a = \{x: m_k - p \leq x \leq m_k + p\}$ and $L_b = \{x: y_k - p \leq x \leq y_k + p\}$. Assume that $-\sum_{j=1}^{N} \Psi(x_{(j)}^k - m_k) < \sum_{j=1}^{2N+1} \Psi(x_{(j)}^k - m_k)$. Then clearly, $x_{(N+1)}^k \in L_b$ whenever $x_{(N+2)}^k \in L_a$, and $m_k < y_k$, so that $\Psi(m_k - y_k) < 0$. Suppose that $x_{(N)}^k \in L_b$ when $x_{(N)}^k \in L_a$. Then $\Psi(x_{(j)}^k - y_k) = -g(p)$ for all $1 \leq j \leq N$. Hence,

$$\sum_{j=1}^{N+1} \Psi(x_{(j)}^k - y_k)$$

$$= \sum_{j=1}^{2N+1} \Psi(x_{(j)}^k - y_k)$$

where the first inequality comes from $\Psi(m_k - y_k) < 0$. Therefore, $\sum_{j=1}^{N} \Psi(x_{(j)}^k - y_k) \neq 0$, which is a contradiction. A similar proof holds for the case when $-\sum_{j=1}^{N} \Psi(x_{(j)}^k - m_k) > \sum_{j=1}^{2N+1} \Psi(x_{(j)}^k - m_k)$. The case 2) can be proved similarly. The case 3) and 4) can be proved in a similar way.

APPENDIX C

Proof for Observation 2: Proof for 1) is obvious. Consider 2). Now $x_{(N)}^k \leq m_k - p$ and $x_{(N+2)}^k < m_k + p$ implies that $-\sum_{j=1}^{N+1} \Psi(x_{(j)}^k - m_k) > \sum_{j=1}^{2N+1} \Psi(x_{(j)}^k - m_k)$, so that $y_k < m_k$. Since $x_{(N+2)}^k \in L_b$, we have $x_{(N+2)}^k \in L_b$ from Observation 1, where $L_a$ and $L_b$ are defined in Appendix B. Thus, $y_k \geq x_{(N+2)}^k - p$. The statement 3) and 4) can be proved in a similar way.

APPENDIX D

Proof for Observation 3: Statement 1) is obvious; let us consider 2). The conditions in 2) imply that $y_k \geq x_{(N)}^k + p$, and the result follows from Observation 1. Statement 3) can be proved similarly. The case 4) is obtained directly from observation 1.

APPENDIX E

Proof for Observation 4: Suppose $y_k = m_k + e$. Then $\sum_{j=1}^{N} \Psi(x_{(j)}^k - m_k) = \sum_{j=1}^{2N+1} \Psi(x_{(j)}^k - m_k)$, we have $y_k \geq m_k$ so that $e \geq 0$. In this case, from (14) and the assumption that $r_1 \leq p$ and $r_2 \leq p$, the necessary and sufficient condition for $y_k$ being the sample mean is $|x_{(j)}^k - y_k| \leq p$, that is $r_1 \leq p - e$. Among the data sequences of length $2N + 1$ which have the same $r_1$ and $r_2$, a data sequence with the following ordered values will yield the maximum $e$: $x_{(j)}^k = m_k - r_1$, $x_{(j)}^k = m_k$ for $j = 2, 3, \cdots, N + 1$, and $x_{(j)}^k = m_k + r_2$ for $j = N + 2, \cdots, 2N + 1$. For such a data sequence, the value of the sample mean is
\[ m_k + (N r_2 - r_1)/(2N + 1), \text{ so that } \epsilon = (N r_2 - r_1)/(2N + 1) \text{ when } y_k \text{ is the sample mean. Note that } \epsilon^* = \text{ the maximum among the } \epsilon's \text{ which is obtained when } y_k \text{ is the sample mean, for given } r_1 \text{ and } r_2. \text{ Using } \epsilon^* \text{ in the condition } r_1 \leq p - \epsilon \text{ yields the result. The remaining part of the observation can be proved in a similar way.}

\textbf{REFERENCES}


