

Performance Characteristics of the Median LMS Adaptive Filter

Geoffrey A. Williamson, *Member, IEEE*, Peter M. Clarkson, *Senior Member, IEEE*,
and William A. Sethares, *Member, IEEE*

Abstract—The performance of gradient search adaptive filters, such as the least mean squares (LMS) algorithm, may degrade badly when the filter is subjected to input signals which are corrupted by impulsive interference. The median LMS (MLMS) adaptive filter is designed to alleviate this problem by protecting the filter coefficients from the impact of the impulses. MLMS is a modification of LMS, obtained by applying a median operation to the raw gradient estimates of the mean squared error performance surface. An analysis of the MLMS algorithm is provided for the class of independent and identically distributed inputs. For these inputs we establish exponential convergence of the MLMS algorithm. The rate of convergence is shown to depend on order statistics of the input, but unlike that for LMS shows little dependence on characteristics of the impulsive interference. The steady state performance of the LMS and MLMS is also examined. The average deviation of the parameter estimates from their optimal values caused by the arrival of an impulse is assessed. This analysis indicates a significantly improved performance for MLMS compared to LMS. Analytic predictions for both convergence and steady state behavior are supported by simulations.

I. INTRODUCTION

ADAPTIVE signal processing, and particularly adaptive filtering, provides a powerful approach to many signal processing problems [1]. The capacity of adaptive algorithms to operate when limited *a priori* information is available makes them ideally matched to many practical applications. The most widely used adaptive filter is the least mean squares (LMS) algorithm [2]. Although this algorithm suffers from slow and generally nonuniform convergence, it is famous for its simplicity both in concept and implementation. Less well known is that the LMS, and other gradient based adaptive algorithms, degrade badly when the filter is subjected to input signals which are corrupted by impulsive interference. Sparse impulses arise frequently in a variety of practical situations, including speech, image, and biomedical processing, and communications applications. Possible treatments for such problems include preprocessing using simple thresholding

and interpolation schemes, and also more sophisticated nonlinear transformations of the data including median and other order statistic (OS) filters (see, for example, [3]–[5]). In particular, the use of the median has become common in the treatment of impulsive data as in [6]–[8]. Median filters have been found to have the capacity to remove sparse impulses from data, while still preserving sharp edges. One approach which incorporates the robustness of the median operation into an adaptive framework is provided by the idea of a median LMS (MLMS) algorithm [9]. This algorithm is similar in structure to the LMS but is designed to protect the filter coefficients from the impact of impulsive interferences by applying a median filtering operation to the raw gradient estimates. The filter output at each iteration, however, remains a linear combination of past inputs, as in the LMS algorithm.

Simulations produced using various forms of impulsive noise have demonstrated the clear, and in many cases decisive, advantages of this algorithm compared to LMS or other linear gradient estimators [9]. This initial work has been simulation based, however, and only the most rudimentary attempts have been made to provide an analytical framework for the performance of the algorithm. This paper is the first attempt to provide such an organized analysis for the median LMS adaptive filter.

The analysis of median and other OS filters is generally difficult due to the nonlinear nature of the ordering transform (see, for example, [10]–[12]). The incorporation of this nonlinear function into the LMS update, which is itself a nonlinear recursive combination of data vectors, compounds the difficulties. This necessarily restricts consideration to relatively simple cases. Additionally, a number of simplifying assumptions are required to render the analysis tractable. Nevertheless, the results provided, while generally approximate, are supported by simulations, and do represent a significant step towards the quantification of the MLMS algorithm.

The paper is organized as follows: In Section II we define the MLMS algorithm, together with models for impulsive and nonimpulsive signals, and we give a simple example illustrating the potential of the MLMS algorithm. Of primary concern is the performance of the median LMS algorithm as compared to that of the LMS. In particular, we compare the performance of the LMS and MLMS algorithms when nonimpulsive inputs are corrupted by sparse impulsive interference. Of equal impor-

Manuscript received September 24, 1990; revised December 12, 1991. This work was supported in part by the National Science Foundation under Grants MIP-8911676 and MIP-9102620.

G. A. Williamson and P. M. Clarkson are with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616.

W. A. Sethares is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706.

IEEE Log Number 9205095.

tance, however, is the question of how the algorithms compare when no impulses occur. In Section III we examine these issues. We deal almost exclusively with stochastic signals. In particular, we restrict attention to the relatively simple case of zero-mean, independent identically distributed (i.i.d.) inputs. We find that the MLMS algorithm, like the LMS, converges exponentially for the class of i.i.d. inputs. We also demonstrate that, unlike the LMS, the MLMS exhibits "smooth" convergence when the inputs are corrupted by impulsive noise. In particular, we show that the convergence of the MLMS is essentially determined by the properties of the input data, rather than the impulse sequence. This is a major advantage of the MLMS algorithm over LMS. In Section IV we examine the performance of the MLMS algorithm in steady state. In particular, we analyze the effects of a single impulse on the LMS and MLMS algorithms. We use a combination of analysis and simulations to show that the cost for MLMS (in terms of deviation from the optimal filter coefficients) is significantly less than for LMS. Finally, in Section V we examine the behavior of the algorithm when the signal itself is impulsive or contains an impulsive component. We find that, as expected, the algorithm is generally less suited to such data, tending to ignore the impulsive component, and that in a few pathological cases the MLMS algorithm actually becomes unstable.

II. PRELIMINARIES

A. Input Signal Definitions

Consider the adaptive filtering system shown in Fig. 1, where the input signal x_n , and the desired signal d_n are corrupted by impulsive interferences ζ_n and η_n , respectively. The measured input signal (regressor) is an $(L \times 1)$ vector x_n^m with

$$x_n^m = x_n + \zeta_n \quad (2.1a)$$

where $x_n^t = [x_n \ x_{n-1} \ \cdots \ x_{n-L+1}]$ is the modeled component of the input and $\zeta_n^t = [\zeta_n \ \zeta_{n-1} \ \cdots \ \zeta_{n-L+1}]$ is the impulsive component. The measured desired signal d_n^m is

$$d_n^m = d_n + \eta_n. \quad (2.1b)$$

We assume that the signals x_n , d_n are completely modeled by the relation

$$d_n = f^* x_n \quad (2.2)$$

where f^* is the $(L \times 1)$ vector of optimal least-squares coefficients obtained from

$$f^* = R^{-1} g \quad (2.3)$$

with $R = E\{x_n x_n^t\}$ and $g = E\{x_n d_n\}$. Here, we have implicitly assumed x_n , d_n are jointly stationary and that R is positive definite. We also assume that the impulsive interferences ζ_n and η_n are zero-mean i.i.d., that they are drawn from symmetric densities, and that they are mutually uncorrelated and are uncorrelated with both x_n and d_n . The input x_n is assumed to be i.i.d. with variance σ_x^2 . As a working definition we will assume that the non-im-

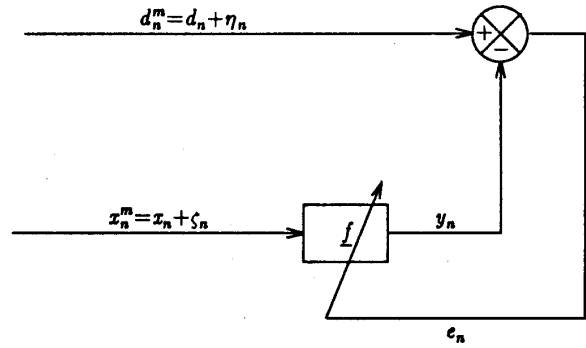


Fig. 1. Adaptive filtering with impulse corrupted signals.

pulsive component has samples drawn from a density with tails no heavier than Gaussian.

We shall be concerned with three cases:

$$\text{Case I} \quad \eta_n = \zeta_n = 0 \text{ for all } n.$$

$$\text{Case II} \quad \zeta_n = 0 \text{ for all } n.$$

$$\text{Case III} \quad \eta_n = 0 \text{ for all } n.$$

Case I corresponds to the so-called homogeneous problem [13] in which

$$d_n^m = f^* x_n^m. \quad (2.4)$$

Cases II and III are inhomogeneous cases with the root of inhomogeneity residing in d_n^m and x_n^m , respectively.

The impulsive signal, whether present in x_n or in d_n , will be assumed to have the form

$$i_n = a_n A_n \quad (2.5)$$

where a_n is i.i.d. with

$$p(a_n = 1) = c; \quad p(a_n = 0) = 1 - c \quad (2.6)$$

where c is the arrival probability. The distribution of the amplitude A_n is arbitrary subject to the constraint that $\text{var}\{A_n\} \gg E\{x_n^2\}$. Various models could be employed; for example, the amplitude A_n has been modeled as unipolar or bipolar lognormal [8]. To simplify the analysis we shall assume that the amplitudes are independent of the arrival times and are zero mean with a symmetric density. Note that this implies that a_n and A_n are mutually uncorrelated. The variance of i_n is then

$$\text{var}\{i_n\} = c \text{var}\{A_n\} = \sigma_i^2.$$

Note that the occurrence of impulses are Bernoulli trials with

$$\begin{aligned} p(C_{j:N}) &= p(j \text{ impulses in the last } N \text{ trials}) \\ &= \binom{N}{j} c^j (1-c)^{N-j}. \end{aligned} \quad (2.7)$$

As such the arrival density can be considered as a binomial approximation to a Poisson process [14]. We note that this is a common model of impulsive behavior (see, for example, [8]).

Note that the definitions in the paragraphs above represent an artificial delineation of signals into impulsive and non-impulsive. In practice, signals and noise may occur with a continuous spread of distributions from uniform through Gaussian, with increasingly heavy tails, to double exponential and beyond. The models used here are an idealization intended to facilitate the analysis. It is not suggested that the MLMS will only perform well with data drawn from this particular model. In fact, simulations reported in [15], [16] have demonstrated that the MLMS performs well with various forms of impulsive noise.

B. The Median LMS Algorithm

Our concern is the relative performance of the LMS algorithm as characterized by the update equation

$$f_{n+1} = f_n + \mu e_n x_n^m \quad (2.8)$$

and the MLMS algorithm which is defined by the modified update

$$f_{n+1} = f_n + \mu \text{med} \{e_n x_n^m\}_N. \quad (2.9)$$

For both (2.8) and (2.9), e_n is the difference between the desired signal and filter output

$$y_n = f_n^t x_n^m \quad (2.10)$$

with

$$e_n = d_n^m - y_n.$$

In (2.9), the notation $\text{med} \{ \}_N$ acting on a vector sequence s_n is defined as a vector γ_n with elements

$$\gamma_n(i) = \text{med} [s_n(i), s_{n-1}(i), \dots, s_{n-N+1}(i)].$$

Here, $s_n(i)$ is the i th component of the vector s_n for sample index n , and for convenience N is restricted to odd values. The median is defined by

$$\gamma_n(i) = s_{(M)}(i)$$

where $s_{(j)}(i)$ is the j th algebraically ordered element of $\{s_n(i), s_{n-1}(i), \dots, s_{n-N+1}(i)\}$ and where $M = (N + 1)/2$. The incorporation of the median into the LMS update generates a computational overhead for the algorithm. However, efficient algorithms facilitate the median calculation in $O(N)$ operations per coefficient with a comparable storage requirement [17]. Typically, performance considerations motivate the choice of $N < L$ and overall computation is $O(NL)$ which is comparable to LMS.

Just as with LMS, MLMS adjusts the coefficients of a filter with a linear structure. The coefficients f^* of (2.3) remain optimal. We define a vector of filter errors as $\mathbf{v}_n = f_n - f^*$, applicable for both LMS and MLMS. For LMS

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mu e_n \mathbf{x}_n^m \quad (2.11)$$

and for MLMS

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mu \text{med} \{e_n \mathbf{x}_n^m\}_N. \quad (2.12)$$

C. A Simple Example

We illustrate the potential of the MLMS algorithm by considering the effects of impulsive interference in a simple adaptive filtering problem. Consider $x_n = d_n$ where x_n is an i.i.d. Gaussian input. We examine Cases II and III with the impulsive interference defined by (2.5) with $\text{var} \{A_n\} = 10\sigma_x^2$ and the arrival probability c equal to 0.02. Fig. 2 shows the coefficient error $f_n(0)$ for a sample run from Case II. The result for LMS is shown in Fig. 2(a) with the corresponding result for MLMS in Fig. 2(b). Fig. 3 shows a similar example for Case III. The parameters chosen for all these trials were $\sigma_x = 0.1$, filter length $L = 10$, adaptation constant $\mu = 0.1$, and median window length $N = 3$. As can be seen from the figures, for the MLMS in both II and III the parameter converges smoothly towards its optimal value $f^*(0) = 1$, and the convergence appears to be exponential. By contrast, the LMS parameter trajectory is not smooth, being perturbed by the impulsive noise. In Case II, the parameter converges in the mean but is subjected to random fluctuations caused by the impulses. In Case III, the trajectory fails entirely to approach the optimal solution—it shows a bias in addition to the random fluctuations. This is, of course, an idealized example. Practical problems cannot be expected to set out the distinctions between these algorithms so starkly. Even so, the potential for improved performance over the LMS is clear.

III. CONVERGENCE ANALYSIS

A critical aspect of any adaptive algorithm in its convergence properties. Indeed, the fact that the LMS algorithm converges under modest excitation conditions [18] is perhaps its most important feature. Of concern, then, is the effect on the attractive convergence properties of LMS caused by the introduction of the median operation in MLMS. In this section, we analyze the mean convergence of MLMS and compare our results with those for LMS. We consider Cases I–III of Section II, first reviewing the previously developed analysis for LMS and then presenting our new results for MLMS. Convergence rate information for both algorithms in all three cases is summarized in Table I.

A. Analysis for LMS

Case I: The analysis in this case derives from standard results on the convergence of LMS. For Case I we have $d_n^m = d_n$, $x_n^m = x_n$. In terms of the parameter errors \mathbf{u}_n , using (2.4), (2.10), and (2.11), the LMS update may be written

$$\mathbf{v}_{n+1} = (I - \mu \mathbf{x}_n \mathbf{x}_n^t) \mathbf{v}_n. \quad (3.1)$$

Taking expectations, we have

$$E\{\mathbf{v}_{n+1}\} = (I - \mu E\{\mathbf{x}_n \mathbf{x}_n^t\}) E\{\mathbf{v}_n\}. \quad (3.2)$$

In obtaining (3.2), we have made the assumption that \mathbf{x}_n and \mathbf{v}_n are independent. This assumption is widely used

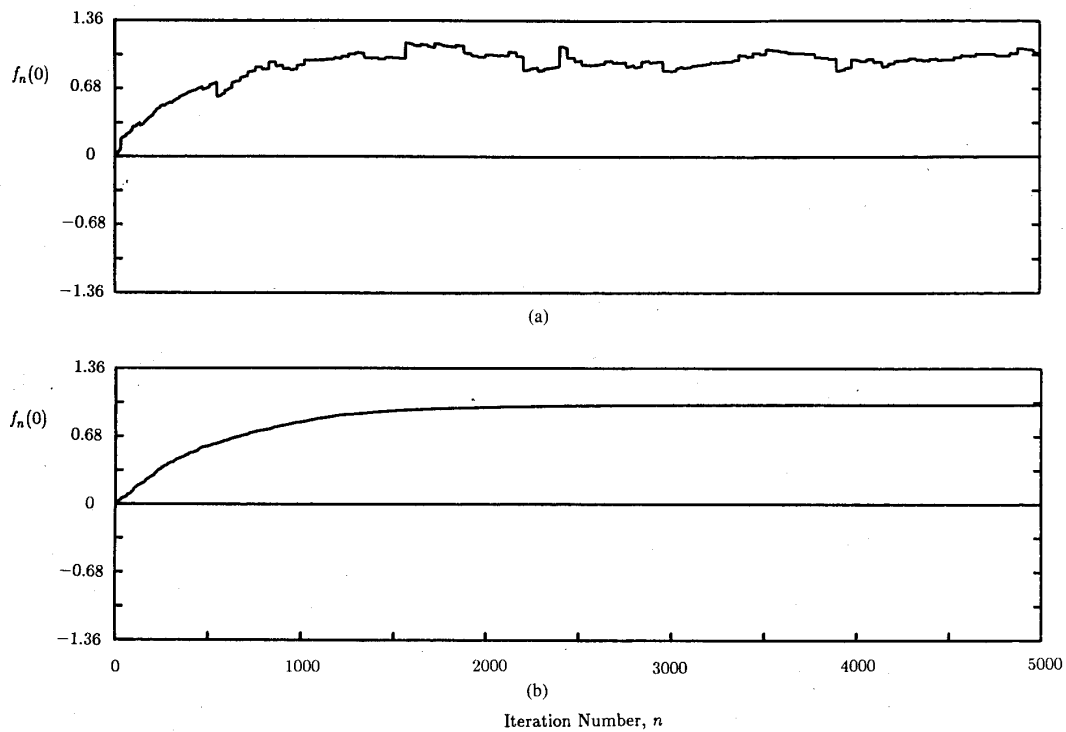


Fig. 2. $f_n(0)$ for Case II with $d_n^m = x_n^m + \eta_n$, x_n i.i.d. $N(0, 0.1)$. Parameters $L = 10$, $\mu = 0.1$, $N = 3$, $f^*(0) = 1$, and $f_0(0) = 0$. (a) LMS, (b) MLMS.

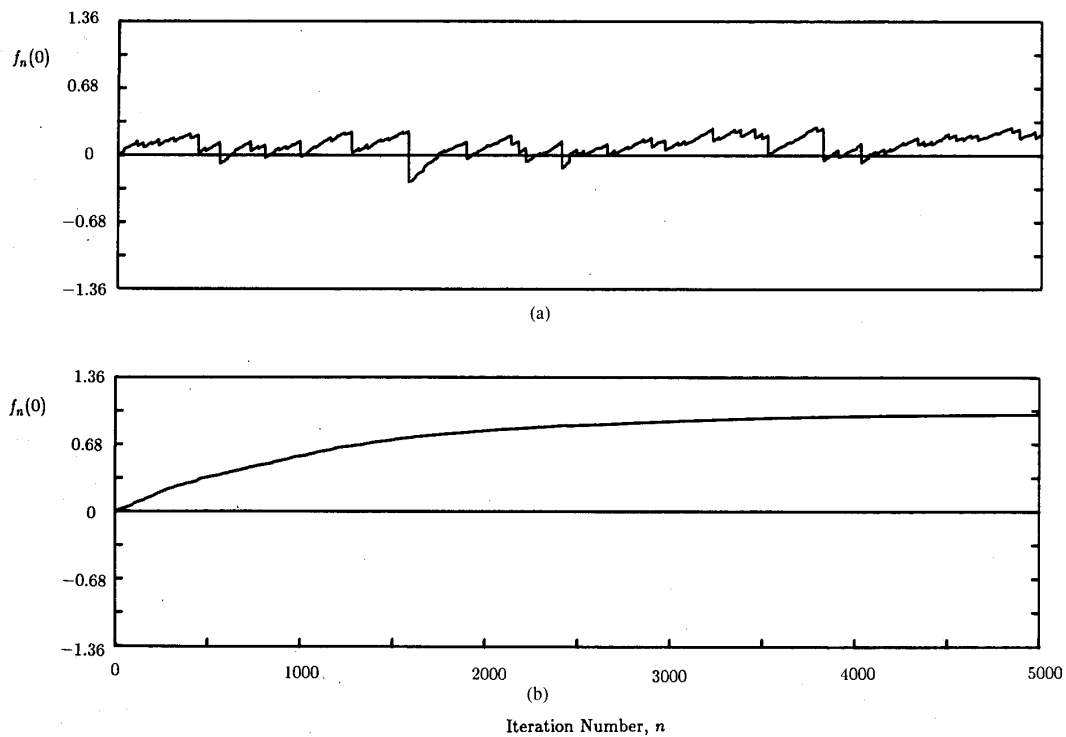


Fig. 3. $v_n(0)$ for Case III with $d_n^m = x_n$, $x_n^m = x_n + \zeta_n$, x_n i.i.d. $N(0, 0.1)$. Parameters $L = 10$, $\mu = 0.1$, $N = 3$, $f^*(0) = 1$, and $f_0(0) = 0$. (a) LMS, (b) MLMS.

TABLE I
COMPARISON OF CONVERGENCE RATES γ^2 IN $E\{v_{n+1}(i)\} = (1 - \mu\gamma^2)E\{v_n(i)\}$ FOR LMS AND MLMS

Case	LMS	MLMS
I	σ_x^2	$\gamma^2 \approx \beta_M^2$
II	σ_x^2	$\gamma^2 \approx \beta_M^2$
III	$\sigma_x^2 + \sigma_f^2$	$\beta_{M-1}^2 \leq \gamma^2 \leq \beta_{M+1}^2$

and has been shown to give excellent agreement with simulation results, even with highly correlated inputs [19]. Equation (3.2) exactly describes the evolution of $E\{v_n\}$ when the update algorithm follows the true gradient of the theoretical error surface, which is known to correspond well with the actual LMS behavior when μ is small [20].

With x_n an i.i.d. sequence as assumed in Section II, we have

$$E\{x_n x_n^t\} = \sigma_x^2 I. \quad (3.3)$$

The parameter estimates are (in the mean) decoupled. The mean parameter errors evolve according to

$$E\{v_{n+1}(i)\} = (1 - \mu\gamma^2)E\{v_n(i)\} \quad (3.4)$$

where the convergence rate constant $\gamma^2 = \sigma_x^2$. Of course, we must select μ to satisfy $0 < \mu < 2/\sigma_x^2$ for convergence in the mean.

Case II: We now treat the case where the signals are $d_n^m = f^{*t} x_n + \eta_n$, $x_n^m = x_n$. For LMS, (2.11) may be written

$$v_{n+1} = [I - \mu x_n x_n^t] v_n + \mu x_n \eta_n. \quad (3.5)$$

Taking expectations we arrive again at (3.2) by the assumptions on the statistics of the signals involved. Hence, in the mean, LMS behaves for Case II just as it does for Case I, although one must expect the variance of $v_n(i)$ to be larger due to the presence of η_n in (3.5).

Case III: For this case the input signals are $d_n^m = d_n$, $x_n^m = x_n + \zeta_n$. From (2.11) the parameter error equation becomes

$$v_{n+1} = v_n + \mu(x_n + \zeta_n)[x_n^t f^* - (x_n + \zeta_n)^t f_n].$$

Hence, taking expectations we obtain

$$E\{v_{n+1}\} = [I - \mu(E\{x_n x_n^t\} + E\{\zeta_n \zeta_n^t\})] E\{v_n\} - \mu E\{\zeta_n \zeta_n^t\} f^*. \quad (3.6)$$

This equation decouples to give for the i th component

$$E\{v_{n+1}(i)\} = [1 - \mu(\sigma_x^2 + \sigma_f^2)] E\{v_n(i)\} - \mu \sigma_f^2 f^*(i). \quad (3.7)$$

The parameter error $v_n(i)$ thus converges in the mean to

$$E\{v_n(i)\} \rightarrow -\frac{\sigma_f^2}{\sigma_x^2 + \sigma_f^2} f^*(i) \quad (3.8)$$

at an exponential rate determined by $(1 - \mu(\sigma_x^2 + \sigma_f^2))$. Thus, $E\{f_n(i)\}$ does not converge to $f^*(i)$ as defined by (2.2) and the result is biased. It will be recalled that the

example given in Section II displayed such bias (see also Fig. 3).

B. Analysis for MLMS

Case I: We begin by taking expectations in the MLMS parameter error equation (2.12), where for case I $d_n^m = d_n$, $x_n^m = x_n$. We have

$$E\{v_{n+1}\} = E\{v_n\} + \mu E\{\text{med}\{e_n x_n\}_N\} \quad (3.9)$$

which is given component-wise by

$$E\{v_{n+1}(i)\} = E\{v_n(i)\} + \mu E\{\text{med}\{e_n x_{n-i}\}_N\} \quad (3.10)$$

or

$$E\{v_{n+1}(i)\} = E\{v_n(i)\} - \mu E\{\text{med}\{Z\}\}. \quad (3.11)$$

Z in (3.11) is the set $\{z_1, \dots, z_N\}$,¹ whose elements are given by

$$z_j = -(e_{n-j+1} x_{n-j+1}); \quad j = 1, \dots, N.$$

The nonlinearity of the median operation in (3.11) complicates the analysis for MLMS considerably. We make further progress via two assumptions:

Assumption 1: The parameters $f_n(i)$ are sufficiently slowly varying as to be considered constant over the median window length N .

This is a common assumption in adaptive filtering and control, having parallels in the averaging and time-scale decomposition analysis which has been applied in an adaptive control setting [22], [23]. The assumption is made reliable by constraining μ in the update equation (2.9) to be small.

Assumption 2: If u is a random variable with

$$E\{\text{med}\{u\}_N\} = \kappa$$

and if v is a zero-mean random variable with a symmetric density, then

$$E\{\text{med}\{u + v\}_N\} = \kappa.$$

This assumption is strictly true if u, v are independent and if the density of u is also symmetric. In other cases it represents an approximation whose accuracy depends on the properties of the joint density of u and v . In the following we use a combination of analysis, discussion, and simulation to substantiate the use of this assumption.

For Case I, $e_n x_n$ may be written

$$e_n x_n = -x_n x_n^t v_n.$$

Hence using Assumption 1,

$$z_j = x_{n-i-j+1}^2 v_n(i) + \sum_{\substack{l=0 \\ l \neq i}}^{L-1} x_{n-i-j+1} x_{n-l-j+1} v_n(l); \quad j = 1, \dots, N. \quad (3.12)$$

¹The quantity z_j depends on the time index n and the component i , but to simplify the notation this dependence is not explicitly noted.

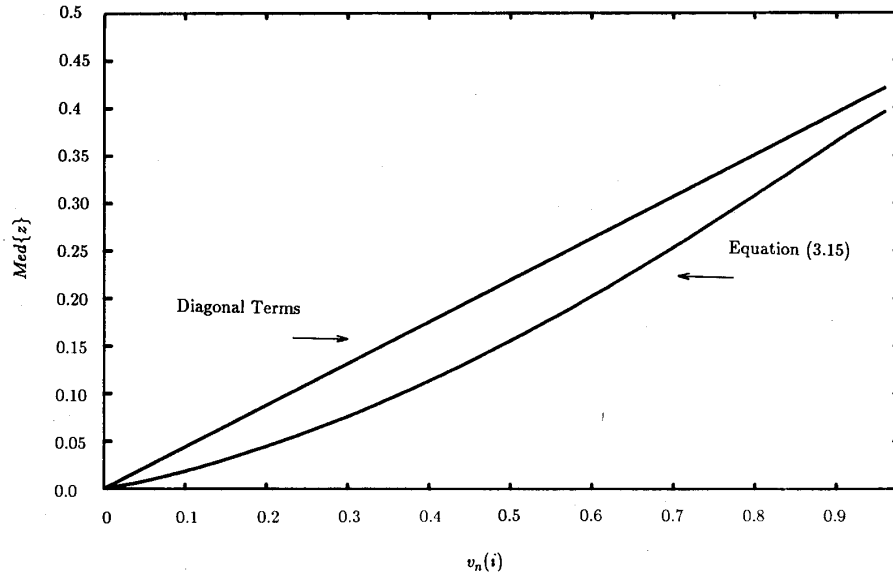


Fig. 4. Median of the distribution for the gradient when the inputs are distributed as $N(0, 1)$.

We refer to $x_{n-i-j+1}^2 v_n(i)$ in (3.12) as the “diagonal term” associated with z_j , and to the remainder as the sum of “off-diagonal terms.” (This terminology stems from the positions of $x_{n-k} x_{n-l}$ products in a correlation matrix for x_n .) Substituting (3.12) into (3.11) and applying Assumption 2 gives

$$E\{v_{n+1}(i)\} = E\{v_n(i)\} - \mu E\{\text{med}\{x_{n-i-j+1}^2 v_n(i)\}_N\}.$$

Finally, Assumption 1 yields

$$E\{v_{n+1}(i)\} = [1 - \mu\beta_M^2]E\{v_n(i)\}. \quad (3.13)$$

Here β_r^2 denotes the expected value of the r th order statistic of x_n^2 from a sample of N .² β_r^2 may be explicitly evaluated [21] as

$$\beta_r^2 = \frac{N!}{(r-1)!(N-r)!} \int_0^\infty y [F(y)]^{r-1} [1-F(y)]^{N-r} f(y) dy \quad (3.14)$$

where $F(y)$ and $f(y)$ are the distribution and density, respectively, for the transformed variable $y = x_n^2$.

We may obtain further insight into Assumption 2 and its relation to the relative magnitudes of the parameter error components by considering an explicit example of the density of the elements of Z . In particular, if x_n is an i.i.d. sequence with standardized Gaussian distribution, then using standard results for nonlinear combinations of Gaussian variables [24] the density of z_j as given by (3.12)

can be shown to have the form

$$f(z_j) = \frac{\exp\left\{\frac{\rho z_j}{1-\rho^2}\right\}}{\pi(1-\rho^2)^{1/2}} K_0\left(\frac{|z_j|}{1-\rho^2}\right); \quad j = 1, \dots, N \quad (3.15)$$

where $\rho = v_n(i)$ and where

$$K_0(u) = \frac{1}{2} \int_0^\infty \frac{1}{t} \exp\left[\frac{-u^2}{4t} - t\right] dt \quad (3.16)$$

is the modified Bessel function of the second type of order zero. In (3.15) it has been assumed that

$$\sum_{i=0}^{L-1} v_n^2(i) = 1. \quad (3.17)$$

(This assumption simplifies the form of the constants in (3.15) but is not crucial to the result.) The median of the distribution of z_j (that is the value of z , for which $F(z) = 0.5$) may be obtained by numerical integration of (3.15). Evaluation of this result as a function of the ratio of the diagonal to off-diagonal elements is given in Fig. 4, which shows the median value as $v_n(i)$ varies from 0 to 1 (subject to the constraint (3.17)). For comparison, the median value obtained from Assumption 2, that is by setting all the off-diagonal terms in (3.12) to zero, is also plotted. It is apparent that the approximation gives reasonable agreement at all ratios. It should be kept in mind, however, that the median of the distribution represents a limit for the sample moment as $N \rightarrow \infty$. Therefore, Fig. 4 may

²A more complete notation reflecting the dependency on N is $\beta_{r,N}^2$ as in [21], but except where confusion might arise we will use the simpler notation β_r^2 .

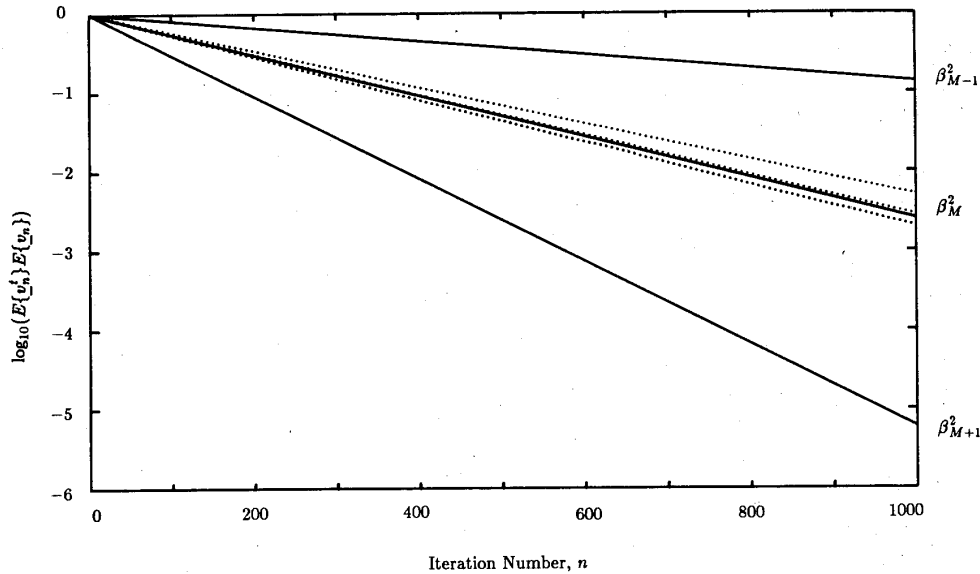


Fig. 5. Exponential convergence for the MLMS algorithm and the impact of Assumption 2. --- Predicted convergence rate (constant β_M^2), curves corresponding to β_{M+1}^2 and β_{M-1}^2 are included for reference. ····· Trial results for several f^* 's. Each trial result represents an average of 100 runs. x_n i.i.d. uniform $[-1, 1]$. Parameters $L = 5$, $\mu = 0.01$, $N = 3$.

have only limited relevance when N is small. Practical experience suggests that Assumption 2 certainly holds when $v_n(i)$ is the dominant component of the parameter error, that is, for the i th component when $|v_n(i)| \gg |v_n(j)|$ for $(j \neq i)$. One consequence of this is that if any one coefficient converges more slowly than the others, its behavior will increasingly be governed by Assumption 2.

A series of simulations has been conducted to examine the performance of the MLMS algorithm compared to that predicted by equations (3.13). The performance of the MLMS was evaluated by calculating $\log_{10}(v_n^T v_n)$ for 1000 iterations. Each trial result was obtained by averaging $(v_n^T v_n)$ over an "ensemble" of 100 runs. Fig. 5 shows a family of plots obtained using iid data, uniformly distributed over $[-1, 1]$, for x_n . Each trial corresponds to a different optimal filter vector f^* , though all were subject to a unit norm constraint. The trials shown in Fig. 5 were obtained using $\mu = 0.01$, $L = 5$, $N = 3$ and for reference curves corresponding to convergence constants γ^2 equal to β_M^2 , β_{M-1}^2 , and β_{M+1}^2 are also plotted. We note that in all cases the results are reasonably close to the convergence predicted by (3.13). The behavior illustrated in these examples is typical of that observed in many similar trials.

Case II: We have

$$\begin{aligned} E\{v_{n+1}(i)\} &= E\{v_n(i)\} - \mu E\left\{\text{med}\left\{x_{n-i}\left[\sum_{l=0}^{L-1} x_{n-l}v_n(l) - \eta_n\right]\right\}_N\right\} \\ &= E\{v_n(i)\} - \mu E\left\{\text{med}\left\{(z_1 - x_{n-i}\eta_n), (z_2 - x_{n-i-1}\eta_{n-1}), \dots\right.\right. \\ &\quad \left.\left.(z_N - x_{n-i-N+1}\eta_{n-N+1})\right\}\right\} \end{aligned} \quad (3.18)$$

where z_j is defined by (3.12). Since the term $x_{n-i}\eta_n$ is zero mean and symmetric, it may be neglected under Assumption 2. Then we have

$$E\{v_{n+1}(i)\} = [1 - \mu\beta_M^2]E\{v_n(i)\}. \quad (3.19)$$

Therefore, in the analysis of convergence in the mean, the addition of η_n to d_n has no effect, and in the mean we have the same behavior here in Case II as we had in Case I. This has been verified by simulations which, for the same parameters, produced results which are in very close agreement with those for Case I. As was illustrated by the example given in Section II, MLMS does enjoy a significant advantage over LMS in Case II. This is manifested as reduced variability for the filter coefficients about the expected trajectory. This subject is explored further in Section IV.

Case III: When impulses occur in the measured data x_n^m , the analysis becomes more intricate. Hence in what follows, we will rely more heavily on Assumption 2. To begin, we substitute (2.1a) and (2.2) into (2.12), and taking expectations we have

$$\begin{aligned} E\{v_{n+1}\} &= E\{v_n\} - \mu E\left\{\text{med}\left\{(v_n^T x_n^m + f^{*T} \zeta_n)\right.\right. \\ &\quad \left.\left.(x_n + \zeta_n)\right\}_N\right\}. \end{aligned} \quad (3.20)$$

For the i th element the update is

$$E\{v_{n+1}(i)\} = E\{v_n(i)\} - \mu E \left\{ \text{med} \left[\sum_{j=0}^{L-1} v_n(j)(x_{n-j} + \zeta_{n-j}) + \sum_{j=0}^{L-1} f^*(j)\zeta_{n-j} \right] (x_{n-i} + \zeta_{n-i}) \right\};$$

$$i = 0, 1, \dots, L-1. \quad (3.21)$$

The probability of r impulses in any N values is defined by $p(C_{r:N})$ in (2.7). For ζ sparse, we assume that $p(C_{2:N})$, $p(C_{3:N})$, \dots , $p(C_{N:N})$ may be neglected. That is, we assume that a maximum of one impulse occurs in any N samples. The analysis of the MLMS update then reduces to consideration of the occurrence of 0 or 1 impulses.

When no impulse occurs the behavior of the algorithm is identical to that for case I. When an impulse does occur the effect on the algorithm depends on the precise location of the impulse within the update "window." Let us examine the effects that an impulse of amplitude A arriving at n_0 has on the succeeding iterations of the update for coefficient i at $n = n_0 + k$ for $k = 0, 1, \dots, N + L - 1$. We divide this into two regions: For $k < i$, the only impulsive terms which occur in the median window are linear in A and in the filter coefficients. Recall from Section II that the amplitude of A is zero mean and has a symmetric density. Hence, using Assumption 1 to factor out $E\{v_n(i)\}$, and Assumption 2 so that for $k < i$ all impulsive terms in the update may be neglected, we have

$$E\{v_{n+1}(i)\} = E\{v_n(i)\} - \mu E\{\text{med}\{x_{n-i}^2\}_N\} E\{v_n(i)\}$$

which is identical to case I.

For $k = i, i + 1, \dots, i + N - 1$ by employing Assumption 2 to remove all of the terms from (3.21) which are symmetric we obtain

$$E\{v_{n+1}(i)\} = E\{v_n(i)\} - \mu E\{\text{med}\{v_n(i)x_{n-i}^2, \dots, v_n(i)x_{n-k}^2 + f_n(i)A^2, \dots, v_n(i)x_{n-i-N+1}^2\}_N\}. \quad (3.22)$$

Hence there is a single term $f_n(i)A^2$ present in the median window, which cannot be simply neglected using the symmetry assumption. Examining (3.22) we see that convergence depends on the relative magnitudes and signs of $v_n(i)x_{n-k}^2$ and $f_n(i)A^2$. These depend on the initial conditions and on the form of f^* , and may be expected to vary during adaptation. In spite of these difficulties we may bound the impact of the impulsive term $f_n(i)A^2$ on the update. Consider two extreme cases:

$$\text{a) } |f_n(i)A^2| \ll |v_n(i)x_{n-k}^2|.$$

Here the impulsive term has little impact on the update and convergence reduces to that of Case I as in (3.13).

$$\text{b) } |f_n(i)A^2| \gg |v_n(i)x_{n-i}^2|.$$

Suppose that $v_n(i) > 0$ (a similar argument to that which follows can be constructed for $v_n(i) < 0$). Then the sequence $\{y_l\} = \{v_l(i)x_{l-i}^2\}$ is nonnegative. If $f_n(i)A^2$ is negative, the addition of this negative term to one of the members of the sequence y_l can only change the median value selection if the element corrupted is $y_{(r)}$, with $r \geq M$, and if $y_{(r)} + f_n(i)A^2 < y_{(M)}$. In this case the median of the i th update corresponds to $y_{(M-1)}$ rather than the usual $y_{(M)}$ for impulse free convergence. If $f_n(i)A^2$ is positive, the argument works in reverse—the addition of this positive term to one of the members of the sequence y_l only changes the median value selection if the element corrupted is $y_{(r)}$, with $r < M$, and if $y_{(r)} + f_n(i)A^2 > y_{(M)}$. In this case the median of the i th update contains diagonal term $y_{(M+1)}$ rather than $y_{(M)}$. Note that in either case the change from $y_{(M)}$ to $y_{(M-1)}$ or $y_{(M+1)}$ can only affect the magnitude of the update (3.22), not its sign. Thus the impulse influences only the rate of convergence. Given the two extremes indicated in a) and b), we see that overall, convergence for this case is governed by (3.4) with

$$\beta_{M-1}^2 \leq \gamma^2 < \beta_{M+1}^2. \quad (3.23)$$

These results are illustrated in Fig. 6, which shows a typical example with $\text{Var}\{A\} = 10\sigma_v^2$, and the impulse arrival probability equal to 0.02. The data and filter parameters employed in this trial were identical to those for the trial of Fig. 5. For ease of comparison the corresponding result from Case I is superimposed. In both cases, the trial commences from $f_0 = \mathbf{0}$. As convergence progresses (and the impulsive term grows in significance) the Case III convergence slows slightly, but remains at all times within the bound (3.23). We can conclude that the performance of the MLMS algorithm in the presence of these sparse impulses is dependent on the distribution of the input data, not that of the impulses, and that in most cases (3.13) remains the best available predictor for MLMS convergence. Finally, we note that this analysis suggests that MLMS enjoys unbiased convergence to f^* . In practice some bias is likely as a consequence of the terms neglected in the analysis. Simulation results suggest that any such bias is generally small and certainly much less than in the comparable result for LMS.

IV. STEADY STATE PARAMETER ERROR ANALYSIS

Our motivation for introducing the median operation into the LMS updates is the mitigation of deleterious effects arising from impulsive noise in the adaptive filter. It is thus useful to quantify the improvement in this regard of MLMS over LMS. Here, as our measure of performance, we focus on the effect of a single impulse arrival on the parameter errors, when the adaptive filter is operating in steady state with zero parameter error. Thus, we assume that $v_0 = 0$, and we consider the arrival of a single impulse at $n = 0$, either in η_0 or ζ_0 . We then quantify the effects of that impulse in terms of the value of $E\{v_T^* v_T\}$, where T denotes the time at which the impulse no longer impinges directly on the algorithm updates. We

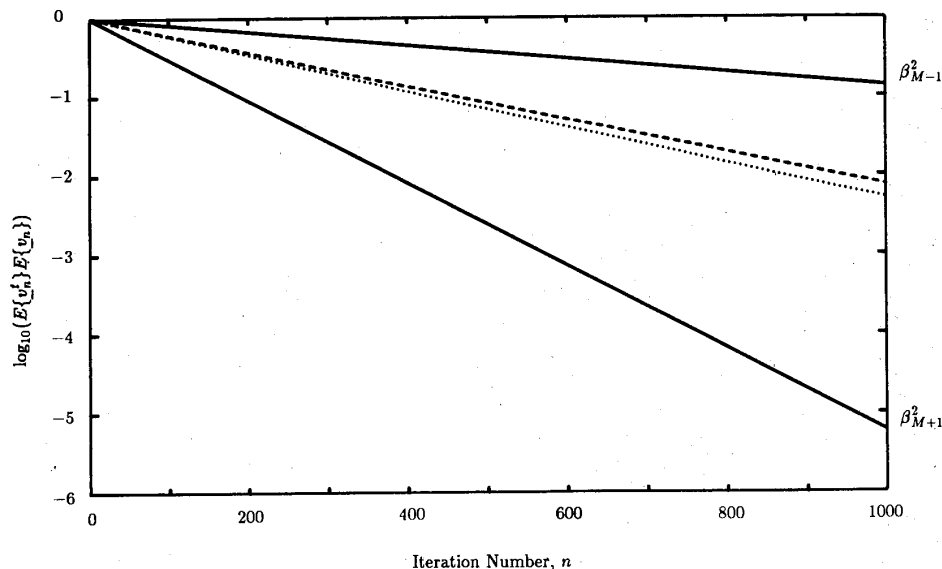


Fig. 6. MLMS convergence: Typical Case III versus Case I result. --- Case III result; ···· Case I result. Trial results represent an average of 100 runs for a typical parameter vector. x_n i.i.d. uniform $[-1, 1]$. Parameters $L = 5$, $\mu = 0.01$, $N = 3$.

assume that the impulse, whether present in d_n or in x_n , has amplitude A .

A. Analysis for Impulse in d_n^m ($\eta_0 = A$)

For LMS, the arrival of an impulse in η_0 directly affects the parameter estimate updates only at time $n = 0$. Therefore $T = 1$, and one may calculate

$$E\{v_T^T v_T\} = \mu^2 A^2 (L\sigma_x^2). \quad (4.1)$$

In MLMS, an occurrence of the single impulse in e_0 will be completely ignored by the median operation in the parameter updates, as $e_n = 0$, and hence $\mu x_n e_n = 0$, for $n \neq 0$. Therefore,

$$E\{v_T^T v_T\} = 0. \quad (4.2)$$

In this simple situation, the performance advantages of MLMS over LMS are clear. MLMS simply does not react when a single interference impulse arrives in the desired signal d_n . Consequently, the parameter estimates are unperturbed from their optimal values. For the LMS algorithm, however, the parameters are moved from their optimal values by a distance $(E\{v_T^T v_T\})^{1/2}$ proportional to the impulse amplitude.

B. Analysis for Impulse in x_n^m ($\zeta_0 = A$)

When the impulse arrives in the measured data x_0^m , the effects of the impulse on the parameter estimate are more intricate, as the impulse remains in the data vector x_n^m over L time steps. We have $x_n^m = x_n + \zeta_n$, where

$$\zeta_n = \begin{cases} [0 \cdots 0 \zeta_0 0 \cdots 0]^T; & n = 0, 1, \dots, L-1 \\ \mathbf{0}; & \text{otherwise} \end{cases} \quad (4.3)$$

with ζ_0 appearing in the $(n+1)$ th position in (4.3).

In LMS, the effect of the impulse disappears only after L time steps, at which point the impulse leaves the measured data vector x_n^m . The parameter error of interest is thus $v_T = v_L$. We obtain an expression for v_T by first substituting for x_n^m in (2.11). Beginning with $v_0 = 0$ and using repeated substitution finally yields

$$\begin{aligned} v_T &= - \left[\sum_{i=0}^{L-2} \sum_{j=i+1}^{L-1} [I - \mu(x_j + \zeta_j)(x_j + \zeta_j)'] \right. \\ &\quad \cdot \mu(x_i + \zeta_i) \zeta_i' f^* \left. \right] - \mu(x_{L-1} + \zeta_{L-1}) \zeta_{L-1}' f^* \\ &= - \sum_{i=0}^{L-1} \mu(x_i + \zeta_i) \zeta_i' f^* + O(\mu^2). \end{aligned} \quad (4.4)$$

The norm squared of v_T is then

$$\begin{aligned} v_T^T v_T &= \sum_{l=0}^{L-1} \sum_{i=0}^{L-1} \mu^2 (f^*)' \zeta_l \\ &\quad \cdot (x_l + \zeta_l)' (x_l + \zeta_l) \zeta_l' f^* + O(\mu^3). \end{aligned} \quad (4.5)$$

Evaluating the expectation of (4.5) with respect to the data x_n , keeping in mind that ζ_n here is deterministic (with $\zeta_0 = A$), and neglecting $O(\mu^3)$, yields

$$E\{v_T^T v_T\} \approx \mu^2 A^2 (L\sigma_x^2 + A^2) \|f^*\|^2. \quad (4.6)$$

With the MLMS algorithm, the computation of $E\{v_T^T v_T\}$ is more difficult due to the nonlinearity of the median operation in the update law (2.12). As in Case III in Section III we may substitute (2.1a) and (2.2) into

(2.12) to obtain

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mu \text{med} \{-(\mathbf{x}_n + \boldsymbol{\zeta}_n) \cdot [(\mathbf{x}_n + \boldsymbol{\zeta}_n)' \mathbf{v}_n + \boldsymbol{\zeta}_n' \mathbf{f}^*]\}_N$$

For convenience, we write this as

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mu \text{med} \{\boldsymbol{\xi}_n\}_N \quad (4.7)$$

with $\boldsymbol{\xi}_n$ defined by

$$\boldsymbol{\xi}_n = -(\mathbf{x}_n + \boldsymbol{\zeta}_n)[(\mathbf{x}_n + \boldsymbol{\zeta}_n)' \mathbf{v}_n + \boldsymbol{\zeta}_n' \mathbf{f}^*]. \quad (4.8)$$

In the expression for $\boldsymbol{\xi}_n$ in (4.8), we will neglect the contribution of $(\mathbf{x}_n + \boldsymbol{\zeta}_n)' \mathbf{v}_n$, since it is typically small in comparison with $\boldsymbol{\zeta}_n' \mathbf{f}^*$ for $n \in \{0, \dots, L-1\}$ given that A is large and \mathbf{v}_n is small.

For the first $(N-1)/2$ updates, $\text{med} \{\boldsymbol{\xi}_n\}_N = 0$ in (4.7), as $\boldsymbol{\zeta}_n = 0$ for $n < 0$. At $n = L + (N-1)/2$, fewer than half of $\boldsymbol{\xi}_n, \boldsymbol{\xi}_{n-1}, \dots, \boldsymbol{\xi}_{n-N+1}$ have impulsive components present, so that we expect the original impulse ζ_0 will no longer directly affect the coefficient updates. In other words, for MLMS T is given by $T = L + (N-1)/2$, with nonzero updates of \mathbf{v}_n occurring over the L times $n = (N-1)/2, \dots, L + (N-3)/2$. We will denote the first update index by $L_1 = (N-1)/2$ and the last by $L_2 = L + (N-3)/2$. We are thus interested in

$$\begin{aligned} \mathbf{v}_T &= \mathbf{v}_{L_2+1} \\ &= \mu \text{med} \{\boldsymbol{\xi}_{L_1}\}_N + \mu \text{med} \{\boldsymbol{\xi}_{L_1+1}\}_N \\ &\quad + \dots + \mu \text{med} \{\boldsymbol{\xi}_{L_2}\}_N. \end{aligned} \quad (4.9)$$

Taking the expectation of the norm squared of \mathbf{v}_T , we then have

$$E\{\mathbf{v}_T' \mathbf{v}_T\} = \mu^2 \sum_{j=L_1}^{L_2} \sum_{k=L_1}^{L_2} E\{\text{med} \{\boldsymbol{\xi}_j\}'_N \text{med} \{\boldsymbol{\xi}_k\}_N\}. \quad (4.10)$$

For $L_1 \leq n \leq L_2$, the component updates of $\mathbf{v}_n(i)$ will be chosen from among the components of $\boldsymbol{\xi}_j, j = L_1 - N + 1, \dots, L_2$. We will assume, in a worst case analysis, that only updates containing nonzero ζ_j will be chosen. This further restricts the possible choices of update time index to $j = 0, 1, \dots, L-1$. Hence, since the $\mathbf{v}_n(i)$ are small, each component of $\text{med} \{\boldsymbol{\xi}_n\}_N$ will be chosen from components of

$$\begin{aligned} \xi_j &\approx -(\mathbf{x}_j + \boldsymbol{\zeta}_j)[\boldsymbol{\zeta}_j' \mathbf{f}^*]; \\ j &= 0, 1, \dots, L-1. \end{aligned} \quad (4.11)$$

The median value of the i th component will not be $\xi_j(i)$, when the i th component of $\boldsymbol{\zeta}_j$ is nonzero. For those values of $\xi_j(i)$, the multiplication of the impulse components in both \mathbf{x}_j^m and e_j results in a term with a factor A^2 , where A is the impulse amplitude. As we have assumed A is large, this particular $\xi_j(i)$ is an extreme value within the median operator's window and will not be the median value. Hence, the components of (4.10) selected by the median

operation have the form

$$\xi_j(i) = x_j(i)[\boldsymbol{\zeta}_j' \mathbf{f}^*]. \quad (4.12)$$

Let us replace ξ_j in (4.10) component-wise with $\xi_j(i)$ from (4.12). This yields

$$\begin{aligned} E\{\mathbf{v}_T' \mathbf{v}_T\} &\approx \mu^2 L \sum_{j=L_1}^{L_2} \sum_{k=L_1}^{L_2} E\{\text{med} \{x_j(i)\boldsymbol{\zeta}_j' \mathbf{f}^*\}_N \\ &\quad \cdot \text{med} \{x_k(i)\boldsymbol{\zeta}_k' \mathbf{f}^*\}_N\}. \end{aligned} \quad (4.13)$$

Let $\chi_j = \text{med} \{x_j(i)\boldsymbol{\zeta}_j' \mathbf{f}^*\}_N \cdot E\{\chi_j \chi_k\}$ in (4.13) evaluates to 0 when j, k are such that χ_j, χ_k are independent. In particular, for those values of j, k for which the median windows do not overlap, we have such independence. One may show that the number of j, k pairs, $j, k \in [L_1, L_2]$, for which there is overlap is $(N-1)(L-N) + NL$. An upper bound for this expression is $2NL$. Since $E\{\chi_j \chi_k\} \leq E\{\chi_j^2\}$, we may then bound $E\{\mathbf{v}_T' \mathbf{v}_T\}$ in (4.13) as

$$E\{\mathbf{v}_T' \mathbf{v}_T\} \leq 2\mu^2 NL^2 E\{\text{med} \{x_j(i)\boldsymbol{\zeta}_j' \mathbf{f}^*\}_N^2\}. \quad (4.14)$$

Also, noting that

$$\begin{aligned} |\boldsymbol{\zeta}_j' \mathbf{f}^*| &\leq A \left[\max_m |f^*(m)| \right] \equiv M \\ E\{\text{med} \{x_j(i)\boldsymbol{\zeta}_j' \mathbf{f}^*\}_N^2\} &\leq E\{\text{med} \{M \text{sgn}(\boldsymbol{\zeta}_j' \mathbf{f}^*) x_j(i)\}_N^2\} \\ &= M^2 E\{\text{med} \{x_j(i)\}_N^2\} \end{aligned} \quad (4.15)$$

where the second line of (4.15) follows because $x_j(i)$ has a symmetric density. Hence, combining (4.14) and (4.15), and using

$$\begin{aligned} \left[\max_m |f^*(m)| \right]^2 &\leq \|f^*\|^2 \\ E\{\mathbf{v}_T' \mathbf{v}_T\} &\leq 2\mu^2 NL^2 A^2 \|f^*\|^2 E\{\text{med} \{x_n\}_N^2\}. \end{aligned} \quad (4.16)$$

From (4.6) and (4.16), comparing root mean square values we have

$$\text{LMS: } (E\{\mathbf{v}_T' \mathbf{v}_T\})^{1/2} \approx \mu A \|f^*\| [L\sigma_x^2 + A^2]^{1/2}$$

$$\begin{aligned} \text{MLMS: } (E\{\mathbf{v}_T' \mathbf{v}_T\})^{1/2} &\leq \mu A \|f^*\| \\ &\quad \cdot [2NL^2 E\{\text{med} \{x_n\}_N^2\}]^{1/2}. \end{aligned} \quad (4.17)$$

In both LMS and MLMS, the impulse arrival in x_n moves the parameter estimates \mathbf{f}_n from their optimal values \mathbf{f}^* . A comparison between the cost incurred hinges on the relative sizes of the final factors in the expressions in (4.17): $2NL^2 E\{\text{med} \{x_n\}_N^2\}$ for MLMS versus $(L\sigma_x^2 + A^2)$ for LMS. However, it is difficult to draw general conclusions about their relative sizes as they depend on several different variables. The critical difference in the formulas is that the cost for LMS is approximately quadratic in A , while the bound on the cost for MLMS is linear in A . For

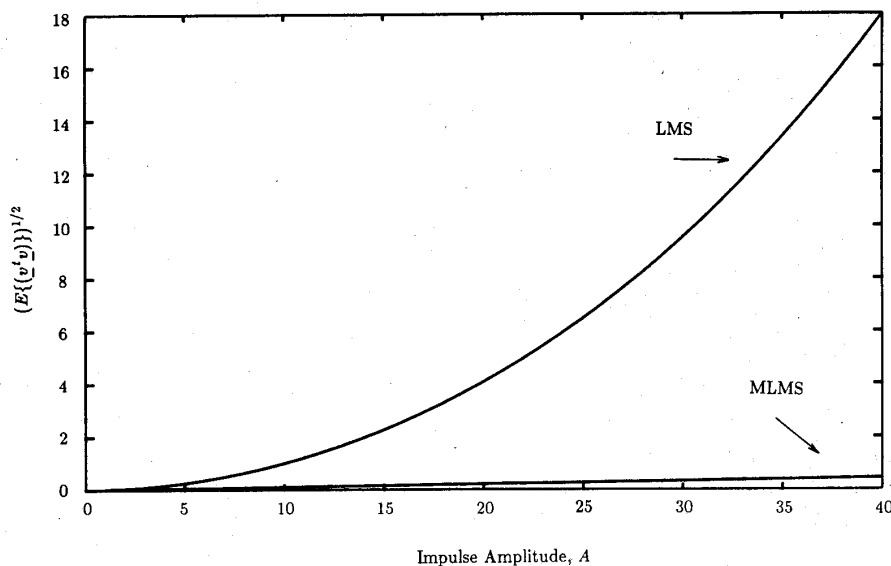


Fig. 7. $\{E\{v_T v_T^t}\}^{1/2}$ for LMS and MLMS subjected to an impulsive interference. x_n i.i.d. uniform $[-1, 1]$. Parameters $L = 10$, $\mu = 0.01$, $N = 3$.

sufficiently large impulses, LMS thus reacts more strongly to the arrival of a single impulse than does MLMS. Therefore, using MLMS reduces the performance cost associated with the impulsive interference. In practice we find that the actual MLMS cost is much lower than the given bound, and that even for small impulses, MLMS outperforms LMS. This is not surprising in view of the derivation of (4.17). We recall that the term NL^2 is obtained by assuming the worst possible combination of dependencies between successive med $\{\cdot\}_N$ operations. Equally, by using $\max_m |f^*(m)|$ we have assumed the highest possible cost for MLMS. In practice these two factors combine to make the bound (4.17) for MLMS extremely conservative. These assertions have been confirmed by simulations. For example, Fig. 7 shows a comparison of $\{E\{v_T v_T^t}\}^{1/2}$, for the two methods, as the amplitude of the impulse is increased. The trial used uniformly distributed $[-1, 1]$ input data and filter parameters of $L = 5$, $\mu = 0.01$ and, for the MLMS $N = 3$. Each point in the curves of Fig. 7 was obtained by averaging over an ensemble of 1000 trials. The result shows the linear dependence on A for the MLMS and quadratic dependence on A for LMS as predicted by (4.17). The LMS curve is comparable with that suggested by (4.17), whereas the MLMS is well below the predicted limit. The performance advantages of MLMS are very clear. Again, this behavior is typical of that observed in similar trials.

V. BEHAVIOR WITH IMPULSIVE DATA

We have shown that MLMS is effective in removing impulsive components of the coefficient updates. What happens if the data sequence itself is impulsive, or if it possesses a significant impulsive component? Our intuition says that in some sense, MLMS will tend to ignore

the impulsive components in the data. If that data is predominantly impulsive, then misbehaviors in the MLMS algorithm may possibly arise.

A. Impulsive Data Sequences

Consider the homogeneous problem (Case I of Section II), in which the data sequence is given by $x_n = i_n$, with i_n described by (2.5) and (2.6). In this situation, the sequence x_n is mostly zero, save for sparsely occurring impulses. The update term in the MLMS algorithms is

$$\mu \text{ med } \{x_n e_n\}_N = 0 \quad (5.1)$$

unless more than $M = (N + 1)/2$ impulses have occurred during the median window. Since we are implicitly assuming that the probability of impulse occurrence (c in (2.6)) is small, having M or more impulses in an N -length window is quite unlikely. Alternatively, we have the parameter errors propagating according to (3.13). In this case, β_M^2 equals zero for impulsive x_n provided the probabilities $p(C_{2:N})$, $p(C_{3:N})$, \dots , $p(C_{N:N})$ defined by (2.7) are negligible. Effectively, the MLMS update is always zero under these circumstances.

Therefore, we see that MLMS does not react to an impulsive data sequence. This observation is in keeping with the intuition that MLMS suppresses the effects of impulses. Hence, one must exercise caution in the use of MLMS in environments where the data sequence x_n has an impulsive character.

B. Partially Impulsive Data Sequences

Once more consider the homogeneous problem (Case I), but now suppose that the data sequence x_n is given by

$$x_n = x_n^w + i_n \quad (5.2)$$

with x_n^w a nonimpulsive i.i.d. random sequence, and with i_n described by (2.5) and (2.6). The model (5.2) for x_n recognizes an impulsive component in an otherwise nonimpulsive data sequence.

The parameter error equation here is derived as in Section III-B, resulting in

$$E\{v_{n+1}(i)\} = [1 - \mu E\{\text{med}\{x_{n-i}^2\}_N\}]E\{v_n(i)\}. \quad (5.3)$$

Equation (5.3) differs from (3.13) only in the nature of x_n . The presence of impulses in x_n is reflected by expanding the expected median as

$$E\{\text{med}\{x_{n-i}^2\}_N\} = \sum_{j=0}^N E\{\text{med}\{x_{n-i}^2\}_N | C_{j:N}\} P(C_{j:N}). \quad (5.4)$$

The expected value of the median, conditioned on $C_{0:N}$, is determined as

$$E\{\text{med}\{x_{n-i}^2\}_N | C_{0:N}\} = E\{\text{med}\{(x_{n-i}^w)^2\}_N\} = \beta_M^2. \quad (5.5)$$

Here β_i^2 denotes the expected value of the i th ordered value of $\{(x_n^w)^2\}_N$.

When conditioning on $C_{j:N}$ for $j > 0$, the occurrence of the impulses has the potential to change the median value. For instance, consider a ranking $(x_n^w)_{(k)}^2$ of $(x_n^w)^2$, and assume that the impulse amplitude is larger than the amplitude of any of the samples x_{n-i}^w . If event $C_{1:N}$ has occurred (one impulse arrived), but the impulse appears at a time such that the ranking k of the corresponding normal component $(x_n^w)^2$ satisfies $k > M$, then the median value is unchanged from the value which would occur were no impulse present. If $k \leq M$, then the median value is now shifted to what would have been the $(M+1)$ th order statistic of $(x_n^w)^2$. With each possible position for the impulse arrival having equal probability, we have

$$p(k > M) = \frac{N-1}{2} = \frac{N-1}{2N}$$

$$p(k \leq M) = \frac{N+1}{2} = \frac{N+1}{2N}. \quad (5.6)$$

Then

$$E\{\text{med}\{x_{n-i}^2\}_N | C_{1:N}\} = \frac{N-1}{2N} \beta_M^2 + \frac{N+1}{2N} \beta_{M+1}^2. \quad (5.7)$$

For $C_{j:N}$ with $j \geq 2$, similar considerations need to be made to calculate $E\{\text{med}\{x_{n-i}^2\}_N | C_{j:N}\}$. However, when the impulse arrival probability c is small, $p(C_{j:N})$ for $j \geq 2$ will also be small. Hence, we neglect in the computa-

tion of (5.4) terms corresponding to $j \geq 2$. We then have

$$E\{\text{med}\{x_{n-i}^2\}_N\} = p(C_{0:N})\beta_M^2 + p(C_{1:N}) \cdot \left[\frac{N-1}{2N} \beta_M^2 + \frac{N+1}{2N} \beta_{M+1}^2 \right]. \quad (5.8)$$

The above shows that for the data sequence x_n of (5.2), MLMS behaves as

$$E\{v_{n+1}(i)\} = [1 - \mu\gamma^2]E\{v_n(i)\}. \quad (5.9)$$

with γ^2 given by the right-hand side of (5.8).

Notice, however, that γ^2 in (5.9) is determined by a linear combination of statistics associated with the nonimpulsive component x_n^w of the data sequence. Hence, in its convergence characteristics, MLMS is largely ignoring the impulsive component of the data, and as long as β_M^2 or β_{M+1}^2 is nonzero, the error system (5.9) indicates that MLMS still converges in the mean.

It is enlightening to compare the convergence of MLMS under these conditions with the convergence of LMS. For LMS, the analysis of Case I in Section III-A applies directly, with (3.4) governing the evolution of the parameter estimates. The LMS convergence rate is determined by $\gamma_{\text{LMS}}^2 = \sigma_x^2 = \sigma_{x^w}^2 + \sigma_I^2$. When $\sigma_I^2 \gg \sigma_{x^w}^2$, with $\sigma_{x^w}^2$ having the same order of magnitude as β_M^2 and β_{M+1}^2 , γ_{MLMS}^2 given by (5.8) will be much smaller than γ_{LMS}^2 . Therefore, if we choose μ in each algorithm in order to equalize the convergence rates (i.e., we set $\mu\gamma^2$ equal to the same constant for each algorithm), μ_{LMS} will be much smaller than μ_{MLMS} . Though the overall convergence rate is the same, MLMS displays much smoother parameter convergence under these conditions. LMS will converge very slowly when no impulses are present, because the step size μ must be chosen small enough to reflect the impulse power. When an impulse arrives, the parameter estimates will "jump," since the update at these moments is much larger. On the other hand, MLMS displays smooth convergence, since the updates as well as the step size parameter μ are dependent only upon the nonimpulsive data component. This difference in the convergence character is reflected in the parameter estimate trajectories shown in Fig. 8. These simulations were conducted using the framework of the example in Section II-C, in the homogeneous case where x_n is distributed as i.i.d. standardized Gaussian variables. The step sizes chosen for this trial were $\mu_{\text{LMS}} = 0.05$, $\mu_{\text{MLMS}} = 0.4$. Once again the behavior observed in this trial is typical.

C. Misbehavior in a Deterministic Example

The analysis of Section V-A implies that at worst, MLMS will fail to update when the data sequence has an impulsive character. However, it is possible for more serious misbehaviors to arise, as is demonstrated in the following deterministic example. Let $L = 3$ (three parameters), $N = 3$ (median window length of three), and

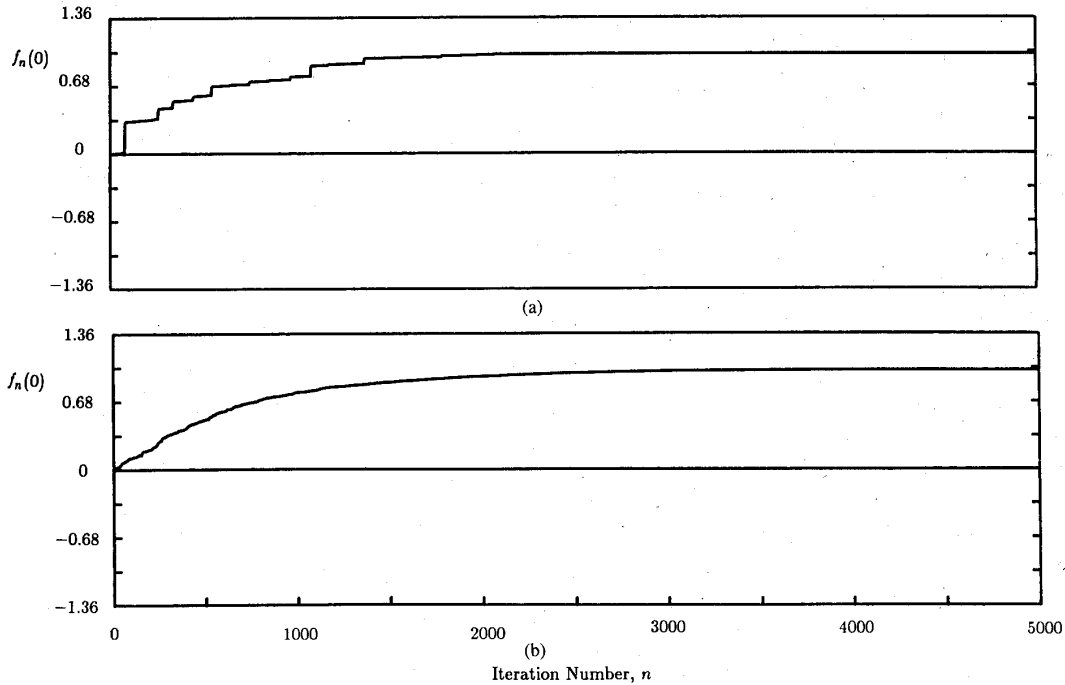


Fig. 8. $v_n(0)$ for Case I with $x_n = x_n^* + i_n$, x_n i.i.d. $N(0, 0.1)$. Parameters $L = 10$, $N = 3$. (a) LMS, $\mu_{LMS} = 0.05$, (b) MLMS, $\mu_{MLMS} = 0.4$.

suppose that the data sequence is

$$x_n = \begin{cases} 3; & n \bmod 3 = 0 \\ -1; & \text{otherwise.} \end{cases} \quad (5.10)$$

One can view the data sequence of (5.10) as a DC signal with a regular arrival of an impulse.

If the initial parameter error is $\mathbf{v}_0 = [1, 1, 1]^T$, then the data and error products $\mathbf{x}_n \mathbf{e}_n$ at $n = 0, -1, -2$ are given by

$$\begin{aligned} -\mathbf{x}_0 \mathbf{x}_0^T \mathbf{v}_0 &= \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}, & -\mathbf{x}_{-1} \mathbf{x}_{-1}^T \mathbf{v}_{-1} &= \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}, \\ -\mathbf{x}_{-2} \mathbf{x}_{-2}^T \mathbf{v}_{-2} &= \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} \end{aligned} \quad (5.11)$$

(assuming that $\mathbf{v}_{-2} \approx \mathbf{v}_{-1} \approx \mathbf{v}_0$). Therefore, the parameter errors update at $n = 1$ to

$$\begin{aligned} \mathbf{v}_1 &= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \mu \operatorname{med} \left\{ \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} \right\} \\ &= \begin{bmatrix} 1 + \mu \\ 1 + \mu \\ 1 + \mu \end{bmatrix}. \end{aligned} \quad (5.12)$$

This pattern then repeats giving a general expression for \mathbf{v}_n as

$$\mathbf{v}_n = (1 + \mu)^n [1 \ 1 \ 1]^T$$

and the parameter estimates diverge.

Though for more random data we expect MLMS to “ignore” the impulsive data, we see here that explicit behavior of MLMS can be much more complicated. In some sense, MLMS removes the effect of the impulse in the data vector \mathbf{x}_n , but the impulse presence in the error term \mathbf{e}_n still influences the update direction. However, the type of behavior evidenced in this example is atypical of MLMS, and to a large extent is a consequence of the repetition of a particularly unfortunate pattern in the data. It is worth noting that this cannot be considered as a counterexample to the developments of Section III because the signal considered violates all of the assumptions used in that analysis.

VI. CONCLUSIONS

We have addressed in this paper two important facets of MLMS algorithm behavior: 1) convergence in the mean, and 2) a cost comparison with LMS. Our analysis demonstrates first that with nonimpulsive, zero mean, i.i.d. inputs, MLMS converges exponentially in the mean. This exponential convergence occurs regardless of the presence of impulsive interference, either in the measured data or in the measured desired signal. Furthermore, unlike for LMS, the convergence rate for MLMS depends predominantly upon statistics of the actual data, with little if any dependence on the character of the impulsive noise.

To attain a cost comparison between LMS and MLMS, we have assessed the average deviation of the parameter estimates from their optimal values caused by the arrival of a single noise impulse, either in the desired signal or in the data. In both cases, MLMS displays a significant performance improvement over LMS. Our work therefore provides heretofore unavailable analytic support for the previously observed benefits of using the MLMS algorithm for adaptive filters operating in impulsive environments.

The convergence analysis demonstrates that when excited by nonimpulsive i.i.d. data, the introduction of the median operation into LMS to create MLMS does not severely alter the attractive convergence properties of LMS. As noted in Section V, however, impulsive data sequences may cause possible misbehaviors in MLMS convergence. Further work is needed to understand the roots of these misbehaviors, and then to characterize the class of input signals for which one has satisfactory convergence of the MLMS algorithm. For instance, theoretical support is currently lacking for the exponential convergence of MLMS when the assumption of independence of successive data points is relaxed. Nonetheless, this work lays the foundation for understanding MLMS convergence behavior by providing analytic results for MLMS behavior with i.i.d. inputs.

REFERENCES

- [1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [2] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *Rec. IRE WESCON Conv.*, pt. 4, 1960, pp. 96-140.
- [3] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.
- [4] A. C. Bovik, T. S. Huang, and D. C. Munson, "A generalization of median filtering using linear combinations of order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1342-1349, 1983.
- [5] Y. H. Lee and S. A. Kassam, "Generalized median filtering and related nonlinear filtering techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 672-683, 1985.
- [6] N. S. Jayant, "Average and median-based smoothing techniques for improving digital speech quality in the presence of transmission errors," *IEEE Trans. Commun.*, vol. COM-24, pp. 1043-1045, 1976.
- [7] L. R. Rabiner, M. R. Sambur, and C. E. Schmidt, "Application of a nonlinear smoothing algorithm to speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 552-557, 1975.
- [8] A. Kunda, S. K. Mitra, and R. R. Vaidyanathan, "Application of two-dimensional generalized mean filtering for removal of impulse noises from images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 600-609, 1984.
- [9] P. M. Clarkson and T. I. Haweel, "A median LMS algorithm," *Electron. Lett.*, vol. 25, pp. 520-522, 1989.
- [10] N. C. Gallagher and G. L. Wise, "A theoretical analysis of the properties of median filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 1136-1141, 1981.
- [11] T. A. Nodes and N. C. Gallagher, "The output distribution of median type filters," *IEEE Trans. Commun.*, vol. COM-32, pp. 532-541, 1984.
- [12] H. G. Longbotham and A. C. Bovik, "Theory of order statistic filters and their relationship to linear FIR filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 275-287, 1989.
- [13] R. R. Bitmead and B. D. O. Anderson, "Performance of adaptive estimation algorithms in dependent random environments," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 788-794, 1980.
- [14] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1984.
- [15] T. I. Haweel and P. M. Clarkson, "Median adaptive filtering for spectral line enhancement," *J. Acoust. Soc. Amer.*, vol. 85(S1), p. S2, 1989.
- [16] T. I. Haweel and P. M. Clarkson, "Analysis and generalization of a median adaptive filter," in *Proc. IEEE ICASSP*, 1990, pp. 1269-1272.
- [17] R. T. Hoctor and S. A. Kassam, "An algorithm and a pipelined architecture for order statistic determination and L -filtering," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 344-351, 1989.
- [18] B. Widrow, and J. M. McCool, M. G. Larimore and C. R. Johnson Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151-1162, 1976.
- [19] J. R. Treichler, "Transient and convergent behavior of the adaptive line enhancer," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 53-64, 1979.
- [20] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, pp. 963-993, 1979.
- [21] H. A. David, *Order Statistics*. New York: Wiley, 1981.
- [22] B. D. O. Anderson et al., *Stability of Adaptive Systems: Passivity and Averaging Analysis*. Cambridge, MA: M.I.T. Press, 1986.
- [23] W. A. Sethares, B. D. O. Anderson, and C. R. Johnson, Jr., "Adaptive algorithms with filtered regressor and filtered error," *Math. Contr. Signal Processing*, vol. 2, pp. 381-403, 1989.
- [24] M. D. Springer, *The Algebra of Random Variables*. New York: Wiley, 1979.

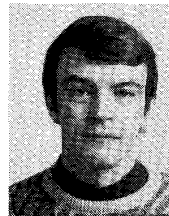


Geoffrey A. Williamson (S'82-M'83) was born in Kansas City, MO, on October 12, 1961. He received the B.S. (with distinction), M.S., and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY.

In 1989, he joined the Department of Electrical and Computer Engineering at the Illinois Institute of Technology, Chicago, IL, as an Assistant Professor. His current research interests are in the areas of system identification and adaptive parameter estimation, as applied in control and signal

processing problems.

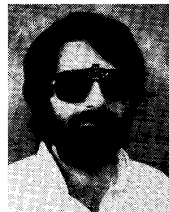
Dr. Williamson is a member of Eta Kappa Nu and Tau Beta Pi. He held a GenRad Foundation Fellowship, as part of the American Electronics Association's Faculty Development Program, from 1983 to 1987.



Peter M. Clarkson (M'86-SM'91) was born in Altrincham, England, in 1958. He obtained the B.Sc. degree, with first class honors, from Portsmouth Polytechnic, England, in 1980, and the Ph.D. degree from the University of Southampton, England, in 1984.

From 1980 to 1987 he was associated with the Institute of Sound and Vibration Research at the University of Southampton; from 1980 to 1984 as a Research Assistant, from 1984 to 1985 as a post-doctoral fellow of the U.K. Science and Engineering Research Council, and from 1985 to 1987 as a Research Lecturer funded by the U.K. Admiralty Research Establishment. Since 1987 he has been with the Department of Electrical and Computer Engineering, at the Illinois Institute of Technology, Chicago, where he currently is an Associate Professor.

Dr. Clarkson's research interests are in digital signal processing, particularly adaptive signal processing and its application to speech and other acoustic signals.



William A. Sethares (S'84-M'86) received the B. A. degree in mathematics from Brandeis University, Waltham, MA, and the M.S. and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY.

He worked at the Raytheon Company as a Systems Engineer and is currently on the faculty of the Department of Electrical and Computer Engineering at the University of Wisconsin, Madison. His research interests include adaptive systems in signal processing and communications and

control.