Local Stability of the Median LMS Filter

W. A. Sethares, *Member, IEEE,* and J. A. Bucklew, *Member, IEEE*

***Abstract*—Local stability properties of the recently proposed median LMS adaptive filter are investigated by relating the behavior of the algorithm to the behavior of an associated ordinary differential equation. With independent inputs, the differential equation and the algorithm are shown to be locally stable. On the other hand, several classes of (periodic, nonindependent) inputs are described which cause the differential equation and the algorithm to be unstable about its equilibrium, even in the no disturbance case. This will help delineate those applications for which the median LMS is an appropriate adaptive algorithm.**

I. Introduction

Conventional least mean squares (LMS) adaptive algorithms adjust a vector of adaptive filter weights $\hat{W}_k$ using an instantaneous approximation to the gradient of the error surface [13], resulting in the algorithm

$$\hat{W}_{k+1} = \hat{W}_k + \mu X_k e_k \tag{1}$$

where X_k is typically a regressor vector of past inputs

$$X_k = (x_k, x_{k-1}, \cdots, x_{k-n+1})^T \tag{2}$$

e_k is the scalar error between the desired signal and the filter output, and μ is the adaptive gain. Observe that this algorithm is vulnerable to impulsive disturbances in the input since a single noise impulse in the input x_k will directly corrupt the weight estimates in (1) for n successive time steps.

Recently, the median LMS (MLMS) algorithm has been proposed [5]

$$\hat{W}_{k+1} = \hat{W}_k + \mu \mathrm{med}_m(X_k e_k, X_{k-1}e_{k-1}, \cdots, X_{k-m+1}e_{k-m+1}) \tag{3}$$

(m is odd) to ameliorate this vulnerability. The function med_m in (3), operating on m n-tuples, should be interpreted as an element by element median of length m, that is

$$\mathrm{med}_m\left(\begin{pmatrix} a_{11}\\ a_{12}\\ \cdot\\ \cdot\\ \cdot\\ a_{1n}\end{pmatrix}, \begin{pmatrix} a_{21}\\ a_{22}\\ \cdot\\ \cdot\\ \cdot\\ a_{2n}\end{pmatrix}, \cdots, \begin{pmatrix} a_{m1}\\ a_{m2}\\ \cdot\\ \cdot\\ \cdot\\ a_{mn}\end{pmatrix}\right)$$

$$= \begin{pmatrix} \mathrm{med}(a_{11}, a_{21}, \cdots, a_{m1})\\ \mathrm{med}(a_{12}, a_{22}, \cdots, a_{m2})\\ \cdot\\ \cdot\\ \cdot\\ \mathrm{med}(a_{1n}, a_{2n}, \cdots, a_{mn})\end{pmatrix} \tag{4}$$

in which $\mathrm{med}(a_{1i}, a_{2i}, \cdots, a_{mi})$ is the median of the m scalars $a_{1i}, a_{2i}, \cdots, a_{mi}$. Intuitively, there is good reason to expect the MLMS algorithm (3) to outperform LMS (1) in an impulsive environment, since the presence of the median function tends to discard impulsive values. Indeed, extensive analysis and simulations were carried out in [15] which convincingly verify this intuition for certain classes of inputs and impulsive disturbances.

Recall that LMS is derived as an approximation to a gradient descent on a quadratic error surface. The MLMS algorithm, on the other hand, has no such interpretation, and there is the possibility that certain inputs might cause the algorithm to climb, rather than descend, the squared error surface. This was pointed out in [15], where a simple 3 periodic input sequence was shown to destabilize the algorithm. The current paper pursues the stability question for MLMS using the weak convergence approaches of [2], [7], [10] which relate the behavior of an adaptive algorithm to an associated ordinary differential equation (ODE). The (local) stability and instability properties of the ODE can be readily determined, and translated back to (local) stability and instability results for the algorithm. A key feature of many of the examples of divergence for MLMS is that the mean and the median of the input process have different signs. This suggests that applications which have inputs with symmetric densities (and others for which the means and medians have the same sign) are good candidate applications for the median LMS, while those which fail this property may encounter stability problems.

Introducing the "ideal" value W^* and the parameter error term $W_k = W^* - \hat{W}_k$ shifts the equilibrium solution of (3) to the origin. The error e_k can be rewritten as $e_k = W_k^T X_k + D_k$, where D_k represents both disturbances to the algorithm and any unmodelled dynamics which cannot be matched by the adaptive filter. The algorithm is then described by

$$\begin{aligned} W_{k+1} = &W_k - \mu \mathrm{med}_m\{X_k(W_k^T X_k + D_k),\\ &X_{k-1}(W_{k-1}^T X_{k-1} + D_{k-1}),\\ &\cdots, X_{k-m+1}(W_{k-m+1}^T X_{k-m+1} + D_{k-m+1})\}. \end{aligned}$$

For simplicity, suppose that the median length is $m = 3$ (any

Manuscript received March 2, 1992; revised January 31, 1994. The associate editor coordinating the review of this paper and approving it for publication was Dr. James Zeidler.

The authors are with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706-1691 USA.

IEEE Log Number 9404795.

odd m will proceed identically). Define

$$Y_k = (X_k, X_{k-1}, X_{k-2}, D_{k-1}, D_{k-2})$$
$$U_{k+1} = D_k$$

and H in the obvious way. Then $\{W_k\}$ is of the form

$$W_{k+1} = W_k - \mu H(W_k, W_{k-1}, W_{k-2}, Y_k, U_{k+1}). \quad (5)$$

To initialize this difference equation, assume $W_0 = W_{-1} = W_{-2} = w_0$. The first step in the analysis (Section II) is to relate (5) to the limiting differential equation

$$W(t) = w_0 - \int_0^t \hat{H}(W(s), W(s), W(s))ds \quad (6)$$

where $\hat{H}$ is a "smoothed" version of the nonlinearity H. This ODE is then linearized about its equilibrium at $W = 0$ and its local stability properties can be determined as in Section III. It will be seen that stability and instability are primarily dependent on the properties of the input sequence X_k, and examples of both stability and instability are given. Conclusions are drawn in the final section.

II. Technical Setup

This section presents the theoretical results which relate the behavior of the adaptive algorithm (5) to the ODE (6). The adaptive update term $H(\cdot)$ in (5) has three arguments

- the parameter estimate error W_k
- a function Y_k of the inputs to the adaptive filter
- the present disturbance term U_{k+1}.

Assume that the $\{W_k\}$ are $\Re^d$ valued random variables, where d is the number of adaptive parameters and that $\{Y_k\}$ is a stationary, ergodic, random sequence with distribution ν_Y. The disturbance term is allowed to be a function of both the state W_k and the inputs Y_k, though the random component of U_k must be independent of W_k and Y_k. Thus, assume there exists i.i.d. random variables $\{\psi_k\}$, independent of $\{Y_k\}$, and a measurable function q such that $U_{k+1} = q(W_k, Y_k, \psi_k)$. For instance, the disturbance may be a moving average of either the state or input, hence it is not required that the disturbance itself be i.i.d.

Define

$$\begin{aligned} &P\big(U_{k+1} \in C | (W_l, Y_l, U_l)_{l=-\infty}^k\big) \\ &\quad = P\big(q(W_k, Y_k, \psi_k) \in C | (W_l, Y_l, U_l)_{l=-\infty}^k\big) \\ &\quad = \eta(W_k, Y_k, C). \end{aligned}$$

Also, define a smoothed version of the update term

$$\bar{H}(w, y) = \int H(w, y, u)\eta(w, y, du) \quad (7)$$

where $w \in \Re^d \times \Re^d \times \Re^d$. Assume that $\bar{H}$ is continuous (in both its arguments) and that the expectations of H and $\bar{H}$ are finite, that is, for some $K \in \Re^+$

$$\begin{aligned} &E\Big\{\sup\nolimits_{w:|w|\le K} |H(w, Y_k, q(w, Y_k, \psi_k))|\Big\} < \infty \\ &E\Big\{\sup\nolimits_{w:|w|\le K} |\bar{H}(w, Y_k)|\Big\} < \infty. \end{aligned} \quad (8)$$

Note that there are no assumptions on the autocorrelations of the inputs or disturbances. H is allowed to be discontinuous, provided that the expectation over η is smooth enough to make $\overline{H}$ continuous. Just as $\overline{H}$ averages H, the distribution of Y_k is used to average $\overline{H}$ over the inputs Y_k, and the doubly averaged quantity

$$\hat{H}(w) = \int \bar{H}(w, y)\nu_Y(dy) \quad (9)$$

is the key ingredient in the ODE and to the questions of stability.

Let $W_\mu(t) = W_{[t/\mu]}$ define a time scaled continuous version of the discrete W_k process where $[z]$ represents the integer part of z. For $K \in \Re^+$, let $\tau_\mu^K = \inf\{t{:}|W_\mu(t)| \ge K\}$ be the first time that $W_\mu(t)$ reaches the value K, and let $W_\mu^{\tau_\mu^K}(\cdot) = W_\mu(\cdot \wedge \tau_\mu^K)$ define the corresponding "stopped" process, that is, the process that is equal to $W_\mu(t)$ from time zero to the stopping time τ_μ^K, and is then held constant for all $t > \tau_\mu^K$. The terminology $a \wedge b$ means the minimum of a and b.

If the solution (denote it $W(t)$) of the ODE (6) is unique, then for every $\epsilon > 0, 0 < T < \infty$

$$\lim_{\mu\to 0} P\left(\sup_{0\le t<\tau_\mu^K \wedge T} |W_\mu^{\tau_\mu^K} - W(t)| > \epsilon\right) = 0. \quad (10)$$

A version of this result was presented in [3] and it is proven in detail in [4]. It is closely related to prior results due to Benveniste, Kushner, Ljung and their coworkers, and the books [1], [7] provide an excellent introduction to this area.

Equation (10) says that for small μ and finite observation times, the algorithm generically follows the trajectory of the ODE solution. Thus the behavior of the algorithm can be predicted by investigating the behavior of the ODE. In particular, if the ODE is locally stable at the origin (its equilibrium point), then the algorithm will be well behaved near the origin also, for a finite observation time and small enough μ. There is a complicated interconnection between the length of time over which the algorithm remains near the ODE and the stepsize. Any model of this type will eventually make large excursions from a given operating point. In [12], we argued that the time intervals between these large excursions could be approximated by a suitable compound Poisson process.

Suppose for instance that the ODE solution does not blow up in finite time. Then we may dispense with the stopping times and it is true that

$$\lim_{\mu\to 0} P\left(\sup_{0\le t<T} |W_\mu(t) - W(t)| > \epsilon\right) = 0.$$

However, typically $P(\sup_{0\le t<\infty} |W_\mu(t) - W(t)| > \epsilon) = 1$, for all $\mu > 0$. One must be careful. The step size must be made smaller as the observation window becomes larger. It is possible to be more precise here via related Central Limit Theorem type results. It can be shown that a scaled version of $W_\mu(t) - W(t)$ is asymptotically a Gaussian random process with a certain covariance structure (see e.g., [1], [3], [4], [7]).

III. Median LMS Algorithm

Application of the methodology of Section II to the median LMS filtering problem requires derivation of the functional forms of $\bar{H}$ and $\hat{H}$. The local stability of the algorithm can then be readily determined for particular inputs by checking the stability of the linearized system with transition matrix $\frac{\partial}{\partial w}\hat{H}(w,w,w)$. The first step is to smooth H over the present disturbance. For notational simplicity, let X_3, w_3, D_3 represent the present terms, while X_2, w_2, D_2 are delayed one timestep, and X_1, w_1, D_1 are delayed two timesteps. Denote $\underline{w} = (w_3, w_2, w_1)$ and $\underline{D} = (D_3, D_2, D_1)$. Define

$$g_i(\underline{w}, \underline{D}) = \operatorname{med}\{x_{3i}(w_3^T X_3 + D_3), x_{2i}(w_2^T X_2 + D_2), x_{1i}(w_1^T X_1 + D_1)\} \qquad i = 1, \cdots, n.$$

where x_{ij} is the jth element of the vector X_i. Then

$$\bar{H}(w_3, w_2, w_1, y) = E_{D_3}\begin{Bmatrix} g_1(\underline{w}, \underline{D}) \\ g_2(\underline{w}, \underline{D}) \\ \vdots \\ g_n(\underline{w}, \underline{D}) \end{Bmatrix}.$$

To streamline the following discussion, assume that the disturbance sequence $\{D_k\}$ is i.i.d. and symmetric with distribution function $F(\cdot)$ and density $f(\cdot)$. Then

$$\bar{H}(\underline{w}, y) = (\bar{h}_1(\underline{w}, y), \bar{h}_2(\underline{w}, y), \cdots, \bar{h}_n(\underline{w}, y))^T$$

where

$$\begin{aligned}\bar{h}_j(\underline{w}, y) =& F(z_j^{\min}) \min(u_{1j}, u_{2j}) 1_{x_{3j}>0} \\ &+ (1 - F(z_j^{\min})) \min(u_{1j}, u_{2j}) 1_{x_{3j}<0} \\ &+ (1 - F(z_j^{\max})) \max(u_{1j}, u_{2j}) 1_{x_{3j}>0} \\ &+ F(z_j^{\max}) \max(u_{1j}, u_{2j}) 1_{x_{3j}<0} \\ &+ x_{3j} 1_{x_{3j}>0} \int_{z_j^{\min}}^{z_j^{\max}} z dF(z) \\ &+ x_{3j} 1_{x_{3j}>0} (w_3^T X_3) [F(z_j^{\max}) - F(z_j^{\min})] \\ &+ x_{3j} 1_{x_{3j}<0} \int_{z_j^{\max}}^{z_j^{\min}} z dF(z) \\ &+ x_{3j} 1_{x_{3j}<0} (w_3^T X_3) [F(z_j^{\min}) - F(z_j^{\max})]\end{aligned}$$

and where

$$u_{ij} = x_{ij}(w_i^T X_i + D_i),$$
$$z_j^{\min} = \frac{\min(u_{1j}, u_{2j}) - x_{3j} w_3^T X_3}{x_{3j}}$$

and

$$z_j^{\max} = \frac{\max(u_{1j}, u_{2j}) - x_{3j} w_3^T X_3}{x_{3j}}.$$

Note that

$$F(z_j^{\min}) = P(u_{3j} = \min \text{of}\{u_{1j}, u_{2j}, u_{3j}\})$$

and

$$1 - F(z_j^{\max}) = P(u_{3j} = \max \text{of}\{u_{1j}, u_{2j}, u_{3j}\}).$$

Note also that $\bar{H}$ is continuous in all of its arguments. The next step is to smooth the nonlinearity over the inputs by taking the expectation with respect to the Y variables, yielding

$$\hat{H} = E_Y(\bar{H}). \qquad (11)$$

It is easy to see from (6) that an equilibrium for the differential equation occurs at $w = 0$. Local stability of the ODE about this equilibrium can then be determined by linearization, which requires calculation of

$$M = \frac{\partial}{\partial w}\hat{H}(w, w, w) = (\hat{h}_{jr}(w))$$

where

$$\begin{aligned}\hat{h}_{jr}(w)|_{w=0} &= E_Y\Bigg\{\left[F\left(\frac{x_{2j}D_2}{x_{3j}}\right)x_{2j}x_{2r} - F\left(\frac{x_{1j}D_1}{x_{3j}}\right)x_{1j}x_{1r}\right] \\ &+ x_{3j}x_{3r}\left[F\left(\frac{x_{1j}D_1}{x_{3j}}\right) - F\left(\frac{x_{2j}D_2}{x_{3j}}\right)\right] \\ &\times (1_{x_{3j}>0} - 1_{x_{3j}<0})(1_{D_1 x_{1j} > D_2 x_{2j}} - 1_{D_1 x_{1j} < D_2 x_{2j}}) \\ &+ \big[(x_{1j}x_{1r}1_{x_{3j}>0} + x_{2j}x_{2r}1_{x_{3j}<0})1_{D_1 x_{1j} > D_2 x_{2j}} \\ &+ (x_{1j}x_{1r}1_{x_{3j}<0} + x_{2j}x_{2r}1_{x_{3j}>0})1_{D_1 x_{1j} < D_2 x_{2j}}\big]\Bigg\}\end{aligned}$$

where all terms involving the density $f(\cdot)$ cancel because of the assumed symmetry. The question of stability becomes essentially a study of the positive definiteness of the above matrix. The ODE (and hence the algorithm) will be locally stable if M is positive definite, and will be unstable if M has any eigenvalues with negative real parts. Note that the above theoretical development makes no assumptions about the independence of the inputs and the disturbances. To calculate the required smoothed update terms $\bar{H}$ and $\hat{H}$ in closed form, however, it is simplest to suppose the i.i.d. case in which the input vectors X_k are independent. Strictly speaking, this cannot occur due to the nature of X_k as a regressor vector (2), but it is a common assumption in adaptive filtering [11], [13] and tends to give results which closely approximate the behavior of the algorithms.

Example: Suppose $\{X_i, X_k, X_l\}$ are zero mean independent random variables for all $i \neq k \neq l$. Then it is easy to verify that $\hat{h}_{jr}(w) = 0$ if $j \neq r$ and $\hat{h}_{jr}(w) > 0$ if $j = r$. Hence $\hat{h}_{jr}(w)|_{w=0}$ is positive definite and the median LMS algorithm is stable.

A. No Disturbance Case

Though the median LMS is designed to reject impulsive disturbances, it is instructive to consider problems when such disturbances are not present.

For this setting, the local stability conditions can be expressed nicely. With $\{U_k\} = 0$, $\hat{H} = \bar{H}$. The expression of interest for $\hat{H}(w) = (\hat{h}_1, \hat{h}_2, \cdots, \hat{h}_n)^T$ is given by

$$\hat{h}_j(w) = E_{X_3, X_2, X_1}\{\operatorname{med}(x_{3j}(w^T X_3), x_{2j}(w^T X_2), x_{1j}(w^T X_1))\}.$$

Again, local stability at the origin can be assured when the matrix

$$\frac{\partial}{\partial w}\hat{H}(w,w,w)$$

has all of its eigenvalues positive. This leads immediately to several corollaries. We suppose that $x_{ji} = x_{j-1,i-1}\forall j, i = 2,\cdots,n$.

Corollary 1: In the scalar no disturbance case, the median LMS algorithm is locally stable at the origin.

Proof of Corollary: If the input is identically zero, then the algorithm remains where it is. Let us suppose then that the input is not identically zero. Note that $\hat{H}(w) = E\{\text{med}(X_3 w^T X_3, X_2 w^T X_2, X_1 w^T X_1)\} = wE\{\text{med}(X_3^2, X_2^2, X_1^2)\}$, since in the scalar case $X_i w^T X_i = wX_i^T X_i = wX_i^2$. Hence $\frac{\partial}{\partial w}\hat{H}(w) = E\{\text{med}(X_3^2,\ X_2^2, X_1^2)\} > 0$.

Corollary 2: Let $s_1, s_2, \cdots, s_n$ denote the elements of one period of an n-periodic input sequence. Let $m_1, m_2, \cdots, m_n$ denote one period of the n-periodic sequence obtained by passing the input sequence through the median filter. Then, if

$$\sum_i s_i > 0(< 0) \text{ and } \sum_i m_i < 0(> 0), \tag{12}$$

that sequence will cause the algorithm to not be locally stable at zero.

Proof of Corollary: All we need to do is show to instability in a particular w "direction." Choose $(w_1, w_2, \cdots, w_n) = (\Delta w, \Delta w, \cdots, \Delta w)$, i.e., all components of the w vector are equal. Denote $\sum_i s_i = \bar{s}, \sum_i m_i = \bar{m}$. Then it is straightforward to verify that $\hat{h}_j(w) = \Delta w \bar{s}\bar{m}, j = 1, \cdots, n$.

Example: For $n = 3$, and median window size of 3, consider any three periodic sequence whose median value is negative and whose average value is positive, e.g. $\{3, -1, -1\}$. By the previous corollary, the median LMS algorithm is not stable for that input.

Corollary 3: Let n be an even integer. Let $s_1,\ s_2, \cdots, s_n$ denote the elements of one period of an n-periodic input sequence. Define a new sequence $\hat{s}_i = (-1)^{i+1}s_i$. If the $\hat{s}$ sequence is not stable as in (12) of corollary 2, then the original sequence s causes the algorithm to not be locally stable at the origin.

Proof of Corollary: The proof follows as in the proof of the previous corollary except we choose the w direction to be $w = (\Delta w, -\Delta w, \Delta w,\ -\Delta w,\ \cdots, -\Delta w, \Delta w)$.

Example: For $n = 4$, and median window size of 3, consider $s = [5, 1, -2, 1]$. Hence $\hat{s} = [5, -1, -2, -1]$. This has a positive mean value and negative mean median value. Hence s causes instability of the median LMS algorithm.

It is possible that for a given fixed value of μ, the algorithm will not be driven into instability by one of the above "unstable sequences" since these sequences were derived by linearizing a nonlinear differential equation. For sufficiently small μ, however, the unstable sequences drive the algorithm into instability, and these the divergent behaviors are readily observable in simulations for a variety of stepsizes.

IV. Discussion and Conclusion

This paper examines the behavior of the median LMS adaptive algorithm using a weak convergence approach in which the evolution of the parameter estimates (for small step sizes) is related to the behavior of an associated ODE. Depending on the characteristics of the input sequence, the ODE can be either locally stable (indicating probable success of the adaptive scheme) or locally unstable (indicating that the algorithm will not remain in close proximity to the desired equilibrium). The ODE associated with the MLMS algorithm is fairly complicated. The stability results were derived in the simple i.i.d. setting, while the instability results were found by investigating certain directions where symmetry arguments could be exploited.

A common theme throughout the corollaries of Section III is that misbehavior tends to be associated with inputs that have a mean value opposite in sign from its median value. Why is this?

Consider the special case of an n-periodic input sequence in an n-dimensional problem. Speaking loosely, the direction $1^T = (1, 1, \cdots, 1)^T$ acts like an "eigenvector" of $\text{med}(XX^TW)$, that is

$$\text{med}(XX^T 1) = \text{med}(X)X^T 1 = c1.$$

If c is positive, then the algorithm tends to be stable (since the transition matrix $(I - \mu cI)$ is a contraction). If c is negative, the transition matrix is an expansion, and the algorithm tends to be destabilized. Since $X^T 1$ is precisely the mean of the input sequence x_i, c is the product of the mean and the median of the input sequence.

Applications in which the mean and the median of the input have the same sign (any symmetric density, for instance) are good candidate settings for application of the median LMS algorithm. Conversely, situations in which the mean and median of the input differ in sign are likely to be poor settings for the median LMS algorithm.

We reiterate that this is only a rule of thumb, since it is possible to concoct examples which violate it. Nevertheless, the heuristic arguments combined with the concrete analysis of the previous sections, suggests that this is a good general guideline to help delineate feasible application settings from those for which it may be dangerous (from a stability point of view) to use the median LMS.

The instability examples do not imply that the median LMS algorithm is useless. Indeed, several studies have shown that the algorithm is able to improve resistance to impulsive noises dramatically. Thus, this paper should be read as a warning to the potential user of MLMS: some *a priori* knowledge of the input statistics are needed.

References

[1] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations.* New York: Springer-Verlag, 1990.

[2] A. Benveniste, M. Goursat, and G. Ruget, "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communication algorithms," *IEEE Trans. Automat. Contr.*, vol. 25, no. 6, pp. 1042–1058, Dec. 1980.

[3] J. A. Bucklew, T. Kurtz, and W. A. Sethares, "Results on local stability and tracking behavior of fixed step size recursive algorithms," in *Proc. ICASSP 1992* (San Francisco, CA), Mar. 23–26.
[4] J. A. Bucklew, T. Kurtz, and W. A. Sethares, "Weak convergence and local stability of fixed step size recursive algorithms," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 966–978, May 1993.
[5] T. I. Haweel and P. M. Clarkson, "A class of order statistic LMS algorithms," *IEEE Trans. Signal Processing*, vol. 40, no. 1, pp. 44–53, Jan. 1992.
[6] T. G. Kurtz and P. Protter, "Weak limit theorems for stochastic integrals and stochastic differential equations," *Annals of Probability*, vol. 19, pp. 1035–1070, 1991.
[7] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes, M.I.T. Press Series in Signal Processing, Optimization, and Control.* Cambridge, MA: M.I.T. Press, 1984.
[8] H. J. Kushner and H. Huang, "Asymptotic properties of stochastic approximations with constant coefficients," *SIAM J. Contr., Optimization* vol. 19, no. 1, pp. 87–105, Jan. 1981.
[9] H. J. Kushner and A. Schwartz, "An invariant measure approach to the convergence of stochastic approximations with state dependent noise," *SIAM J. Contr., Optimization*, vol. 22, no. 1, pp. 13–27, Jan. 1984.
[10] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Automat. Contr.* vol. 22, no. 4, pp. 551–575, Aug. 1977.
[11] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, no. 5, May–June 1979.
[12] W. A. Sethares and J. A. Bucklew, "Excursions of adaptive algorithms via the Poisson clumping heuristic," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 40, no. 6, pp. 1443–1451, June 1992.
[13] B. Widrow, 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, no. 8, pp. 1151–1162, Aug. 1976.
[14] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
[15] G. A. Williamson, P. M. Clarkson, and W. A. Sethares, "Performance characteristics of the median LMS adaptive filter," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 667–680, Feb. 1993.

W. A. Sethares (S'86–M'87), for a photograph and biography please see page 2627 of the October 1994 issue of this TRANSACTIONS.

J. A. Bucklew (S'75–M'79), for a photograph and biography please see page 2627 of the October 1994 issue of this TRANSACTIONS.