DIXIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 20 au 24 MAI 1985

A NORMALIZED LMS ALGORITHM: MEAN AND SECOND MOMENT WEIGHT BEHAVIOR

N. J. Bershad[1]

University of California, Irvine, Electrical Engineering Department, Irvine, California 92717

RESUME

L'ALGORITHME LMS a filtre adaptif exige une connaissance a priori du niveau de pouvoir d'entrée pour choisir le gain d'algorithme, parametre μ , pour la stabilité et la focalisation. Puisque le niveau de pouvoir d'entreé est habitvellement une des invonnves statistiques, if est normalement evalvé d'apres les donneés antérievros ou de'but du processus d'adaptation. Il est ensuite supposé que l'evaluation est parfsite dans toute analyses ulte' Rieure du comportement de l' algorithme LMS.

Dans cet' article, les effets de l'estimation du niveou de pouvoir sont incorporés dans des données dépendantes μ qui apparaissent explicitment dans l'algorithme. On estime le comportement moyen et variable de l'algorithme normalisé en prenant en compte la dépendante statistique explicite de μ sur les donneés d'entreé.

Le comportement moyen de l'algorithme est prouvé converger vers le poids wiener. Les effets de fluctuation des poids de l'algorithme sont egalement etudiés. Une equation a constant coefficient matriciel est dérivee d'apres les fluctuations de poids du poids wiener. L'equation est re'solve poiur une matrice de covariance o donneés blanches et poru l'ALE ovec une frequence unique dons un état stable pour μ. Des expressions pouir l'erreur de movais ojustement sont oussi' présenteés. Il est prouvé que dans le cas de le matrice de covariance o donneés blonches la mise en moyenne d'un e chantillon d'a peu pres dux donneés produit des dégredations négligeobles comparé e l' algorithme LMS Dons l'application de l'ALE, les fluctuations de poids de l'etot-stable sont prouveés ehe dependantes modalement, etant plus grondes o lo frequence de l'énergie obsorbeé.

SUMMARY

The LMS adaptive filter algorithm requires a priori knowledge of the input power level to select the algorithm gain parameter μ for stability and convergence. Since the input power level is usually one of the statistical unknowns, it is normally estimated from the data prior to beginning the adaptation process. It is then assumed that the estimate is perfect in any subsequent analysis of the LMS algorithm behavior.

In this paper, the effects of the power level estimate are incorporated in a data dependent μ that appears explicitly within the algorithm. The mean and fluctuation behavior of the normalized LMS algorithm is evaluated, taking into account the explicit statistical dependence of μ upon the input data.

The mean behavior of the algorithm is shown to converge to the Wiener weight. The effects of fluctuations of the algorithm weights are also investigated. A constant coefficient matrix difference equation is derived for the covariance matrix of the weight fluctuations about the Wiener weight. The equation is solved for a white data covariance matrix and for the Adaptive Line Enhancer with a single frequency input in steady-state for small μ. Expressions for the misadjustment error are also presented. It is shown for the white data covariance matrix case that the averaging of about ten data samples causes negligible degradation as compared to the LMS algorithm. In the ALE application, the steady-state weight fluctuations are shown to be modal dependent, being largest at the frequency of the input.

I. INTRODUCTION

The time domain LMS adaptive filter algorithm [1] has found many applications where the statistics of the input processes are unknown or changing. These include noise cancelling [2], line enhancing [3-7] and adaptive array processing [8,9]. The algorithm uses a transversal filter structure driven by a primary input. The filter weights are up-dated iteratively based upon the differences between the filter output and a reference input, so as to minimize the mean square error of the difference. In all cases, the stability, convergence time and fluctuations of the adaptation process are governed by the product of the feedback coefficient, μ, and the eigenvalues of the data covariance matrix R_{XX} of the input to the adaptive filter.

Since the eigenvalues of the data covariance matrix are usually one of the statistical unknowns to be measured by the adaptive filter, μ is usually selected conservatively using the formula $0<\mu \, Tr[R_{XX}] < 2$. Now since $Tr[R_{XX}]$ is also unknown, it must be somehow estimated prior to adaptation. This usually involves estimating the input power level. Thus, in all practical applications, there is an implicit automatic gain control (AGC) on the input to the adaptive filter. The AGC insures that the μ-power product is maintained within acceptable design limits. For the properly selected μ, the algorithm convergence time, stability and mean square error performance depend upon the ratio of the largest and smallest eigenvalues of RXX. The smaller is the ratio, the better is the convergence and misadjustment noise properties of the algorithm [1].

A NORMALIZED LMS ALGORITHM: MEAN AND SECOND MOMENT WEIGHT BEHAVIOR

The usual approach for analyzing the behavior of the LMS algorithm in a variety of configurations and input signal environments is to assume that the AGC makes a perfect estimate of the input power and ignore any AGC effects in the subsequent analysis. This model may be valid when the AGC averages a large number of input data samples (slow acting AGC). On the other hand, if the AGC is to respond quickly to changes in the input statistics (fast-acting AGC) the number of data samples may be small enough so that the AGC no longer makes an error free estimate of the input power. It would be quite valuable to know how few samples are really needed to accurately estimate the power level while not disturbing the basic behavior of the LMS algorithm.

The purpose of the paper is to analyze the mean and fluctuation behavior of the LMS algorithm when the AGC operation is explicitly contained within the LMS algorithm as a data-dependent μ. The most natural choice for the AGC algorithm is to average the square of the data in the taps and divide the filter input and reference by a number proportional to this quantity. Since this estimate operates on the same input data as the LMS algorithm, the data-dependent estimate of μ will be statistically related to other inputs to the algorithm. This dependence enormously complicates the analysis of the algorithm behavior. In this paper, the mean and fluctuation behavior of the weight vector is analyzed and the resultant mean square error performance is calculated.

II. ANALYSIS

A) The LMS Algorithm

The algorithm for changing the weights of the LMS adaptive filter is given by [1,2].

$$W(n+1) = W(n) + \mu\,[d(n) - X^T(n)W(n)]X(n)$$
$$= W(n) + \mu[d(n)X(n) - X(n)X^T(n)W(n)] \qquad (1)$$

where $W(n)$ = filter weight vector at time n,
$d(n)$ = desired scalar signal,
$X^T(n)$ = observed data vector at time n (vector of tap values of the transversal filter)
$= (x(n), x(n-1), x(n-2),\ldots,x(n-N+1))$,
T = transpose and N = number of filter taps.

Under the assumption that the data sequence X(n) is statistically independent over time [1-6], the present weight vector and the present data vector are statistically independent [3-7]. A difference equation for the mean weight vector behavior can be obtained by averaging Eq. (1) to yield

$$E[W(n+1)]=E[W(n)]+\mu[R_{dX}(n)-R_{XX}(n)E[W(n)]] \qquad (2)$$

where E[.] denotes statistical expectation

and
$$R_{dX}(n) = E[d(n)X(n)] \qquad (3)$$
$$R_{XX}(n) = E[X(n)X^T(n)]$$

For stationary input processes, when R_{dX} and R_{XX} are not functions of time, knowledge of the eigenvalues of R_{XX} enables one to determine both the transient and steady-state mean weight behavior of the algorithm [1,2]. Results for the covariance matrix of the weights are also available [7,9].

B) Normalized LMS Algorithm

The algorithm to be investigated in this paper replaces the fixed μ in Eq. (1) by a data dependent $\mu(n)$ given by

$$\mu(n) = \mu_0\,[\frac{1}{N}\sum_{i=0}^{N-1} x^2(n-i)]^{-1} = \frac{N\mu_0}{X^T(n)X(n)} \qquad (4)$$

Hence the weight update equation is

$$W(n+1)=[I - N\mu_0\frac{X(n)X^T(n)}{X^T(n)X(n)}]\;W(n)+N\mu_0\frac{d(n)X(n)}{X^T(n)X(n)} \qquad (5)$$

Eq. (4) implies a start-up time of N data samples which shall be ignored in the subsequent analysis.

The Normalized LMS algorithm is closely related to an algorithm presented in [10-Eq. (20)]. However, no analysis of the mean behavior of the weights is presented. Furthermore, only the white noise data case is considered.

1) Mean Behavior

Averaging both sides of Eq. (5), using the same assumptions as following Eq. (1) yields

$$E[W(n+1)]=\{I-N\mu_0 E\left[\frac{X(n)X^T(n)}{X^T(n)X(n)}\right]\}E[W(n)]+N\mu_0 E\left[\frac{d(n)X(n)}{X^T(n)X(n)}\right] \qquad (6)$$

In order to perform the data dependent expectations in Eq. (6), it will be further assumed that d(n) and X(n) are zero mean jointly gaussian with correlations given by Eq. (3). To further simplify the analysis of the mean behavior, consider deviations of the mean from the Wiener weight $W_o = R_{XX}^{-1}\,R_{dX}$.

Let $V(n) = W(n) - W_o$. Then Eq. (6) simplifies to

$$E[V(n+1)]=\left\{I-\mu_0 N\,E\left[\frac{X(n)X^T(n)}{X^T(n)X(n)}\right]\right\}E[V(n)] + \mu_0 N\left\{E\left[\frac{d(n)X(n)}{X^T(n)X(n)}\right] - E\left[\frac{X(n)X^T(n)}{X^T(n)X(n)}\right]W_o\right\} \qquad (7)$$

In [11], it is shown that

1) $$E\left[\frac{X(n)X^T(n)}{X^T(n)X(n)}\right]_{i,j} = F_{ij}(0) \qquad (8)$$

where $F_{ij}(\beta)$ satisfies the differential equation

$$\frac{dF_{ij}(\beta)}{d\beta} = \frac{-1}{|I + 2\beta R_{XX}|^{1/2}}\left[R_{XX}[I + 2\beta R_{XX}]^{-1}\right]_{i,j} \qquad (9)$$

with boundary condition $F_{ij}(\infty) = 0$.

2) $$E\left[\frac{d(n)X(n)}{X^T(n)X(n)}\right]_i = G_i(0) \qquad (10)$$

where $G_i(\beta)$ satisfies the differential equation

$$\frac{dG_i(\beta)}{d\beta} = \frac{-\{[I + 2\beta R_{XX}]^{-1}R_{dX}\}_i}{|I + 2\beta R_{XX}|^{1/2}} \qquad (11)$$

with boundary condition $G_i(\infty) = 0$. Thus, Eq. (7) simplifies to

$$E[V(n+1)] = \{I - \mu_0 N\ F(0)\}\ E[V(n)] + \mu_0 N\{G(0) - F(0)W_o\} . \qquad (12)$$

Consider the driver term in Eq. (12)

$$G(0)-F(0)W_o = -\int^{\beta} \frac{1}{|I+2\beta R_{XX}|^{1/2}} \Big\{ \Big[I + 2\beta R_{XX}\Big]^{-1} R_{dX} - R_{XX}\Big[I + 2\beta R_{XX}\Big]^{-1} W_o \Big\} d\beta . \qquad (13)$$

But $W_o = R_{XX}^{-1} R_{dX}$ and

$$[I + 2\beta R_{XX}]^{-1} - R_{XX}[I - 2\beta R_{XX}]^{-1} R_{XX}^{-1} = \underline{0} .$$

Hence, Eq. (12) simplifies to

$$E[V(n+1)] = \{I - \mu_0 NF(0)\}\ E[V(n)] \qquad (14)$$

with solution

$$E[V(n)] = [I - \mu_0 NF(0)]^n\ E[V(0)] . \qquad (15)$$

If the eigenvalues of $[I - \mu_0 NF(0)]$ are all less than unity, the mean weight converges to the Wiener weight since $\lim_{n\to\infty} E[V(n)] = \underline{0}$. Thus, Eq. (5) converges on the average to the Wiener weight as does Eq. (1). Hence, using the data dependent $\mu(n)$ as given by Eq. (4) yields an unbiased modification of the LMS algorithm.

2) Transient Mean Behavior

The transient behavior of Eq. (14) is determined by the eigenvalues of $[I - \mu_0 N\ F(0)]$. The eigenvalues of $[I - \mu_0 N\ F(0)]$, in turn, can be shown to depend on the eigenvalues of R_{XX}. Let $Y(n) = Q\ E[V(n)]$ where Q is the orthonormal matrix that diagonalizes R_{XX}, $QR_{XX}Q^{-1} = \mathrm{Diag}[\lambda_1, \lambda_2, \ldots, \lambda_N]$ where λ_i are the eigenvalues of R_{XX}. Pre-multiplying Eq. (14) by Q and using Y(n), Eq. (14) can be written as

$$Y(n+1) = [I - \mu_0 NH]\ Y(n) \qquad (16)$$

where H is a diagonal matrix with

$$H_{kk} = -\int^{\beta} \frac{1}{|I + 2\beta R_{XX}|^{1/2}} \frac{\lambda_k}{1 + 2\beta\lambda_k} d\beta \Big|_{\beta=0} \qquad (17)$$

Eq. (16) follows from Eq. (14) because Q also diagonalizes $\frac{dF}{d\beta}$. Hence the transient behavior of the normalized LMS algorithm depends on the quantities $1 - \mu_0 N\ H_{kk}$, k=1,...,N. H_{kk} is evaluated for three cases in [11] and shown in Table I. For comparison purposes, the modal response is also shown for the LMS algorithm with a perfect AGC, i.e., $Tr[R_{XX}]$ is known a priori. Thus, it is seen that for the equal eigenvalue case, the average behavior of the Normalized LMS algorithm is the same as for the LMS algorithm.

3) Derivation of Matrix Difference Equation for the Second Moment of the Weight Vector

The second moment behavior of the weights can be evaluated about a number of possible fixed vectors. The most natural candidates are the zero vector, the mean weight vector, E[W(n)], and the Wiener weight vector W_o. We chose to use W_o. Our choice is influenced by two factors - 1) the misadjustment error is a simple function of the second moment about W_o, 2) in the real LMS algorithm, the matrix difference equation for the second moment about W_o is diagonalized by the matrix that diagonalizes the data covariance matrix [9]. The latter factor leads to significant analytic simplifications for solving the matrix difference equation. It will be shown that this property also holds for the Normalized LMS algorithm.

Consider the error between the weight and the Wiener weight $V(n) = W(n) - W_o$ as given above Eq. (7),

$$V(n+1) = \{I - N \frac{\mu_0 X(n)X^T(n)}{X^T(n)X(n)}\}\ V(n) + N \frac{\mu_0}{X^T(n)X(n)} [d(n)X(n) - X(n)X^T(n)W_o]$$

$$\text{Let} \quad K_{VV}(n) = E[V(n)V^T(n)] \qquad (18)$$

Then, post-multiplying Eq. (18) by its transpose and averaging yields

$$K_{VV}(n+1) = K_{VV}(n) - N\mu_0 K_{VV}(n) \times$$

(1) (2)

$$E[\frac{X(n)X^T(n)}{X^T(n)X(n)}] - N\mu_0\ E[\frac{X(n)X^T(n)}{X^T(n)X(n)}]\ K_{VV}(n)$$

(3)

$$+ \mu_0^2 N^2 E[\frac{X(n)X^T(n)\ K_{VV}(n)X(n)X^T(n)}{[X^T(n)X(n)]^2}]$$

(4)

$$+ N\mu_0\ E[\frac{d(n)X(n) - X(n)X^T(n)W_0}{X^T(n)X(n)}]\quad E[V(n)]^T$$

$$+ N\mu_0 E[V/n]E\ \frac{d(n)X^T(n) - X(n)X^T(n)W_0{}^T}{X^T(n)X(n)}$$

$$- N^2\mu_0^2 \times$$

(5) (10)

$$E[\frac{d(n)X(n)V^T(n)X(n)X^T(n) - X(n)X^T(n)W_oV^T(n)X^T(n)}{[X^T(n)X(n)]^2}]$$

$$- N^2\mu_0^2 \times$$

$$E[\frac{d(n)X(n)V^T(n)X(n)X^T(n) - X(n)X^T(n)W_oV^T(n)X(n)X^T(n)_T}{[X^T(n)X(n)]^2}]$$

$$+ N^2\mu_0^2 \times$$

$$E\,\frac{\overset{⑥}{d^2(n)X(n)X^T(n)}-\overset{⑦}{X(n)X^T(n)W_o d(n)X^T(n)}-\overset{⑧}{d(n)X(n)W_o^T X(n)X^T(n)} + \overset{⑨}{X(n)X^T(n)W_oW_o^T X(n)X^T(n)}}{[X^T(n)X(n)]^2} \tag{19}$$

From [11], it follows that

$$① = ② = F(0)$$
$$④ = 0$$

③ and ⑤ thru ⑩ are new expectations which must be evaluated.

These unknown expectations are evaluated in [12], yielding

$$K_{VV}(n+1)=K_{VV}(n)-N\mu_0 K_{VV}(n)F(0)-N\mu_0 F(0)K_{VV}(n) \tag{20}$$

$$+ N^2\mu_0^2 \iint \frac{1}{|I+2\beta R_{XX}|^{1/2}}[2\,B(\beta)\,K_{VV}B(\beta)$$

$$+ Tr[B(\beta)K_{VV}(n)]B(\beta)]\; d\beta_1 d\beta_2\Big|_{\beta=0}$$

$$+ N^2\mu^2{}_0\,\xi_0 \iint \frac{R_{XX}[I+2\beta R_{XX}]^{-1}}{|I+2\beta\; R_{XX}|^{1/2}}\; d\beta_1 d\beta_2\Big|_{\beta=0}$$

where

$$F(0) = -\int \frac{1}{|I+2\beta R_{XX}|^{1/2}} B(\beta)\; d\beta\Big|_{\beta=0} \tag{21}$$

and

$$B(\beta) = R_{XX}[I + 2\beta R_{XX}]^{-1} \tag{22}$$

Diagonalization of Eq. (20)

A careful study of the various terms in Eq. (20) suggests pre- and post-multiplication by Q, Q^{-1} respectively.

$$Q\,K_{VV}(n+1)Q^{-1} = Q\,K_{VV}(n)Q^{-1}$$

$$+ \mu_0 N \int \frac{[Q\,K_{VV}(n)Q^{-1}]\Lambda(I+2\beta\Lambda)^{-1}}{|I\;+\;2\beta\;R_{XX}|^{1/2}}\; d\beta\Big|_{\beta=0}$$

$$+ \mu_0 N \int \frac{\Lambda(I+2\beta\Lambda)^{-1}(Q\,K_{VV}(n)Q^{-1})}{|I\;+\;2\beta\;R_{XX}|^{1/2}}\; d\beta\Big|_{\beta=0}$$

$$+ N^2\mu_0^2 \iint \frac{1}{|1+2\beta\;R_{XX}|^{1/2}}$$

$$2\Lambda(I+2\beta\Lambda^{-1})\;(QK_{VV}(n)Q^{-1})\Lambda(I+2\beta\Lambda)^{-1}$$

$$+ Tr[\Lambda(I+2\beta\Lambda)^{-1}[Q\;K_{VV}(n)Q^{-1}]\;\Lambda(I+2\beta\Lambda)^{-1}]d\beta_1 d\beta_2\Big|_{\beta=0} \tag{21}$$

The result

$$B(\beta) = Q^{-1}\Lambda(I+2\beta\Lambda)^{-1}Q \tag{22}$$

has been used to simplify the terms containing $B(\beta)$.

Note that every matrix on the right hand side of Eq. (21) is a diagonal except $QK_{VV}(n)Q^{-1}$. Hence, the following theorem holds for the second moments of the Normalized LMS algorithm about the Wiener weight.

Theorem: If Q diagonalizes R_{XX}, then Q diagonalizes $K_{VV}(n)$ if it diagonalizes $K_{VV}(0)$.

This theorem is similar to an analogous theorem in [7] for the complex LMS algorithm. The proof of the theorem easily follows by induction as in [7]. However, since $K_W(0)$ is the dyad $W_0W_0^T$, the initial conditions introduce off-diagonal terms. However, from the form of Eq. (21), it can be shown that the diagonal and off-diagonal terms are uncoupled. This same phenomena was observed in [9]. Thus, even though Q does not diagonalize $K_{VV}(0)$, the solution of Eq. (21) for the diagonal entries can proceed without knowledge of the off-diagonal terms.

The mean square error at any stage of the weight adaptation is given by the expression

$$\begin{aligned}\xi(n) &= \xi_o + Tr[R_{XX}K_{VV}(n)]\\ &= \xi_o + Tr[\Lambda\;QK_{VV}(n)Q^{-1}]\\ &= \xi_o + \sum_i \lambda_i[Q\;K_{VV}(n)Q^{-1}]_{ii}\end{aligned} \tag{23}$$

ξ_o is the mean square error obtained using W_o. Hence, the diagonal entries in Eq. (21), are all that is needed to describe the mean square error behavior as a function of n. Let

$$\gamma_i(n) = [Q\;K_{VV}(n)Q^{-1}]_{ii} \tag{24}$$

Thus, the diagonal entries of Eq. (21) can be written as

$$\gamma_i(n+1) = \gamma_i(n)+2\mu_o N \int \frac{\lambda_i d\beta}{(1+2\beta\lambda_i)\prod_i(1+2\beta\lambda_i)^{1/2}}\Bigg|_{\beta=0} \gamma_i(n)$$

$$+ N^2\mu_0^2 \iint \frac{d\beta_1 d\beta_2}{\prod_i(1+2\beta\lambda_i)^{1/2}}\;\frac{2\lambda_i^2}{(1+2\beta\lambda_i)^2}\;\gamma_i(n)$$

$$+ \sum_{j=1}^{N} \frac{\lambda_j}{1+2\beta\lambda_j}\,\gamma_j(n)\;\frac{\lambda_i}{1+2\beta\lambda_i}\Bigg|_{\beta=0}$$

$$+ N^2\mu_0^2\,\xi_o \iint \frac{\lambda_i\;d\beta}{\prod_i(1+2\beta\lambda_i)^{1/2}(1+2\beta\lambda_i)}\Bigg|_{\beta=0} \qquad i=1,2,\ldots,N \tag{25}$$

Eq. (25) is set of N simultaneous coupled equations in $\gamma_i(n)$, $i=1,2,\ldots,N$. Eq. (25) can be written as a vector difference equation in the vector $\gamma^T(n) = (\gamma_1(n), \gamma_2(n),\ldots,\gamma_N(n))$ and matrices obtained from Eq. (25). Writing Eq. (25) in vector form is of no use in obtaining an explicit solution. However, the vector form displays matrices whose eigenvalues determine the behavior of $\gamma(n)$ with n. The dimensionality of Eq. (25) is the same as the the number of distinct eigenvalues of RXX' not the dimensionality of R_{XX}. Hence, the easiest case to

A NORMALIZED LMS ALGORITHM:

MEAN AND SECOND MOMENT WEIGHT BEHAVIOR

analyze is when $R_{XX} = \sigma^2 I$. There is only one distinct eigenvalue, $\gamma_i = \sigma^2$, all i. The next simplest case is when there are two distinct eigenvalues, say $\gamma_1 = \gamma_2 = \gamma_j = a$, $\gamma_{j+1} = \gamma_{j+2}$, ..., $\gamma_N = b$. The ALE model investigated in [12] is a special case of two distinct eigenvalues.

III. APPLICATIONS AND DISCUSSION

A) Mean Weight Behavior

For an arbitrary set of eigenvalues for R_{XX}, one could, in theory, evaluate H_{kk} and study the modal responses. Fortunately, however, there are two very interesting cases that can be studied using the results in Table I, cases (a) and (b). Case (a) corresponds to a white data sequence input. For case (a), it can be seen from the table that the mean modal response is the same for the normalized LMS and for the LMS with a perfect AGC.

Case (b) is also very interesting because it corresponds to the Adaptive Line Enhancer (ALE) with a single frequency input [4-7]. Case (b) is useful since it corresponds to the other extreme distribution of the eigenvalues of R_{XX}, the greatest ratio of the largest to smallest eigenvalues. For the ALE, with a single frequency input with power $2\sigma_s^2$ in an additive white noise background of power σ_n^2,

$$R_{XX} = \sigma_n^2 I + \sigma_s^2 (dd^\dagger + d^* d^T) \qquad (26)$$

where $d^T = e^{j\omega_0(\Delta t)}, e^{j2\omega_0(\Delta t)} \ldots\ldots e^{jN\omega_0 \Delta t}$

Δt = tap spacing, ω_0 = input sine wave frequency,

$\dagger$ = conjugate transpose.

The eigenvalues of R_{XX} are given by

$$\lambda_{1,2} = \sigma_n^2 + N\sigma_s^2 \left(1 \pm \frac{\sin N\theta}{N \sin\theta}\right) \qquad \theta = \omega_0(\Delta t)$$

$$\lambda_3 = \lambda_4 = \lambda_N = \sigma_n^2 .$$

Hence, for N large or $\theta = \pi m$, m=1,2,...

$$\lambda_1 = \lambda_2 = \sigma_n^2 + N\sigma_s^2 . \qquad (27)$$

Hence, for these parameter values, the ALE corresponds to case (b). Furthermore, the interesting physical problem is usually when $\sigma_n^2 \geq \sigma_s^2$ but $N\sigma_s^2 >> \sigma_n^2$. From the third column of Table I, the noise eigenvector response decay factor is $1 - \mu_0 \frac{\sigma_n^2}{\sigma_s^2}$ and the signal eigenvalue response decay factor is $1 - \mu_0 \frac{N}{2}$. Thus for $\sigma_n^2 \simeq \sigma_s^2$, it is seen that the signal modal response is much faster than the noise modal response. This same phenomena occurs for the LMS algorithm with a perfect AGC. Using Table I, the modal response to the signal is $1 - \mu_0/2\ N \frac{\sigma_s^2}{\sigma_n^2 + \sigma_s^2}$ and the modal response to the noise is $1 - \mu_0 \frac{\sigma_n^2}{\sigma_n^2 + \sigma_s^2}$. Hence, the ratio of the modal decay factors is the same for both algorithms.

Thus, for Case (a) and (b), it is seen that the mean behavior of the Normalized LMS and LMS with perfect AGC are very similar. These two cases represent the two extremes of eigenvalue selection. Hence, it would be expected that the algorithms do not differ significantly in mean behavior for Case (c). One is led to the conclusion that the mathematical model of the mean behavior of the LMS algorithm with a perfect AGC is an accurate predictor of the behavior of the LMS algorithm with an imperfect AGC.

B) Second Moment Behavior

When $R_{XX} = \sigma^2 I$, Eq (25) reduces to a scalar difference equation since $\gamma_i(n) = \gamma_j(n) \triangleq \alpha(n)$ all i,j.

$$\alpha(n+1) = \alpha(n) + 2\mu_0 N\sigma^2 \int \frac{d\beta}{(1+2\beta\sigma^2)^{N/2+1}} \Bigg|_{\beta=0} \alpha(n) + (2+N)N^2\mu_0^2\sigma^4 \iint \frac{d\beta_1 d\beta_2}{(1+2\beta\sigma^2)^{N/2+2}} \Bigg|_{\beta=0} \alpha(n) + N^2\mu_0^2\,\xi_0\sigma^2 \iint \frac{d\beta_1 d\beta_2}{(1+2\beta\sigma^2)^{N/2+1}} \qquad (28)$$

Carrying out the integration, subject to the boundary conditions that the integrals disappear at $\beta = \infty$, and letting $\beta = 0$, yields

$$\alpha(n+1) = (1 - 2\mu_0 + N\mu_0^2)\alpha(n) + \left(\frac{\mu_0^2}{1 - \frac{2}{N}}\right)\frac{\xi_0}{\sigma^2} \qquad (29)$$

The solution to Eq. (29) is

$$\alpha(n) = (1-2\mu_0 + N\mu_0^2)\alpha(0) + \frac{\mu_0^2}{\left(1 - \frac{2}{N}\right)}\frac{\xi_0}{\sigma^2}\sum_{j=0}^{n-1}(1-2\mu_0 + N\mu_0^2)^{n-j} \qquad (30)$$

In steady-state,

$$\lim_{n\to\infty} \alpha(n) = \frac{\frac{\mu_0}{2}\left(\frac{1}{1-\frac{2}{N}}\right)\frac{\xi_0}{\sigma^2}}{1 - \frac{N}{2}\mu_0} \qquad (31)$$

The comparable equations for the LMS algorithm can be obtained from [9]. For the white noise case, the second moment about the Wiener weight satisfies the difference equation

$$\beta(n+1) = [1-2\mu\sigma^2 + \mu^2\sigma^4(N+2)]\beta(n) + \mu^2\sigma^2\xi_0 \qquad (32)$$

Eq. (32) has steady-state solution

$$\lim_{n\to\infty} \beta(n) = \frac{\frac{\mu}{2}\xi_0}{1 - \frac{\mu\sigma^2}{2}(N+2)} \qquad (33)$$

A NORMALIZED LMS ALGORITHM:

MEAN AND SECOND MOMENT WEIGHT BEHAVIOR

The two algorithms can be compared by constraining the transient behavior to be the same and comparing steady-state mean square errors. Thus

$$1 - 2\mu\sigma^2 + \mu^2\sigma^4(N+2) = 1 - 2\mu_0 + N\mu_0^2 \qquad (34)$$

Solving for the smaller root,

$$\mu_0 = \mu\sigma^2 - (\mu\sigma^2)^2 \left(\frac{N+2}{2}\right) \qquad \mu\sigma^2 \text{ for } \mu\sigma^2\left(\frac{N+2}{2}\right) << 1 \quad (35)$$

Thus, on comparing Eqs. (31) and (33), it is seen that the increase in the misadjustment error (second term of Eq. (23)) for the Normalized LMS is $(1-\frac{2}{N})^{-1}$. Hence, for N as small as 10, the misadjustment error increases only by 25%. Also, from Eq. (31), N > 2 for the misadjustment error to be bounded.

This analysis has been extended to the ALE in [12].

REFERENCES

1. B. Widrow, "Adaptive Filters," in Aspects of Network and System Theory, R. Kalman and N. DeClaris, Eds. New York: Holt. Rinehart, and Winston, 1971, pp. 563-587.
2. B. Widrow et. al., "Adaptive Noise Cancelling Principles and Applications," Proc. IEEE, Vol. 63, pp. 1692-1716, Dec. 1975.
3. J. R. Treichler, "Transient and Convergent Behavior of the Adaptive Line Enhancer", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-27, pp. 53-69, Feb. 1979.
4. J. R. Zeidler, E. H. Sartorious, D. M. Chabries, and H. T. Wexler, "Adaptive Enhancement of Multiple Sinusoids in Uncorrelated Noise", IEEE Trans. Acoust., Speech and Signal Processing, Vol. ASSP-26, pp. 240-254, June 1978.
5. N. J. Bershad, P. L. Feintuch, F. A. Reed, and B. Fisher, "Tracking Characteristics of the LMS Adaptive Line Enhancer-Response to a Linear Chirp Signal in Noise", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-28, pp. 504-516, Oct. 1980.
6. J. T. Rickard and J. R. Zeidler, "Second-order Output Statistics of the Adaptive Line Enhancer", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27, pp, 31-39, Feb. 1979.
7. B. Fisher and N. J. Bershad, "The Complex LMS Adaptive Algorithm-Transient Weight Mean and Covariance with Applications to the ALE", IEEE Trans. Acoust., Speech and Signal Processing, Vol. ASSP-31, No. 1, pp. 34-44, Feb. 1983.
8. B. Widrow, P. Mantey, L. Griffiths, and B. Goode, "Adaptive Antenna Systems', Proc. IEEE, Vol. 55, pp. 3143-2159, Dec. 1967.
9. L. L. Horowitz and K. D. Senne, Performance Advantage of Complex LMS for Controlling Narrowband Adaptive Arrays", IEEE Trans. Circuits Syst., Vol. CAS-28, June 1981.
10. T.A.C.M. Claasen and W.F.G. Mecklenbrauker. "Comparison of the Convergence of Two Algorithms for Adaptive FIR Digital Filters", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-29, No. 3, June 1981, pp. 670-678.
11. N.J. Bershad "A Normalized LMS Algorithm: Part I - Mean Weight Behavior", submitted to IEEE Trans on Acoustics, Speech and Signal Processing. October 1985.
12. N.J. Bershad "A Normalized LMS Algorithm: Part II - Weight Second Moment Behavior", submitted to IEEE Trans. on Acoustics, Speech and Signal Processing, October 1985.

Normalized LMS	H_{kk}	Modal Response
a) $\lambda_1 = \lambda_2, \ldots\ldots, \lambda_N$	$\frac{1}{N}$	$1 - \mu_0$ all modes
b) $\lambda_3 = \lambda_2 = a$	$\frac{a}{2(a-b)} - \frac{ab(\frac{N}{2}-1)}{2(a-b)^2}\left(\frac{a}{a-b}\right)^{\frac{N}{2}-2}\left[\sum_{\frac{N}{2}-1}^{\infty}\left(\frac{a-b}{a}\right)^{v}\frac{1}{v}\right]$	a>>b $\quad 1 - \mu_0 \frac{N}{2}\left(\frac{a}{a-b}\right)$
$\lambda_3 = \lambda_4, \ldots, \lambda_N = b$	$\frac{1}{2}\left(\frac{a}{a-b}\right)^{N/2-1} \quad \frac{b}{a-b}\sum_{\frac{N}{2}}^{\infty}\left(\frac{a-b}{a}\right)^{v}\frac{1}{v}$	$1 - \mu_0 \frac{N}{2}\left(\frac{b}{a-b}\right)$
c) $\lambda_1 = \lambda_2 \neq \lambda_3 = \lambda_4 \neq \ldots \lambda_{N-1} = \lambda_N$	$C_k + \frac{A_{0k}}{2}$	$1 - \mu_0 N H_{kk}$
LMS with perfect AGC $\quad \mu = \frac{\mu_0 N}{E[X^T X]} = \frac{\mu_0 N}{Tr\ R_{XX}}$		$1 - \mu_0 N \frac{\lambda_i}{\sum_{i=1}^{N}\lambda_i}$
a) $\lambda_1 = \lambda_2, \ldots = \lambda_N$		$1 - \mu_0$
b) $\lambda_1 = \lambda_2 = a$		$1 - \frac{\mu_0 N a}{2a + (N-2)b}$
$\lambda_3 = \lambda_4 = \lambda_N = b$		$1 - \frac{\mu_0 N b}{2a + (N-2)b}$

Table I Comparison of NLMS with LMS with Perfect AGC