AN IMPROVED STOCHASTIC MODEL FOR THE LEAST MEAN FOURTH (LMF) ADAPTIVE ALGORITHM

Pedro I. Hübscher and José C. M. Bermudez

Electrical Engineering Department
Federal University of Santa Catarina
Florianópolis - SC - Brasil
e-mail: pedro@eel.ufsc.br, j.bermudez@ieee.org

ABSTRACT

This paper presents a new statistical analysis of the least mean fourth (LMF) adaptive algorithm behavior. Nonlinear recursive equations are derived which predict the behavior of the first and second order moments of the adaptive weights for Gaussian inputs. These recursions can be used to predict the mean square error (MSE) behavior. The new model improves the available models in that it predicts both the transient and steady-state behaviors for measurement noise having any zero-mean probability density function (pdf) and for any signal-to-noise ratio. This is important because the LMF algorithm is known to outperform the LMS algorithm for non-Gaussian noise distributions and for large signal-to-noise ratios. In addition, the new model explicitly shows the dependence of the algorithm's dynamics on the initial weight conditions. Computer simulations illustrate the accuracy of the new model in predicting the algorithm behavior.

1. INTRODUCTION

The LMS adaptive algorithm is the most employed adaptive algorithm in a variety of practical applications, maily due to its simplicity of implementation [1]. Analytical models for the LMS algorithm behavior under different input conditions are available. This facilitates the design and adds to its popularity. Adaptive algorithms based on higher order moments of the error signal have been shown to perform better than LMS in some important applications. However, the practical use of such algorithms has been restricted due to the lack of accurate analytical models to predict their behavior.

The LMF algorithm seeks to minimize the mean fourth error, which is a convex function (and thus unimodal) of the weight vector [3], [2]. It has been shown in [3] that the LMF algorithm outperforms the LMS algorithm when the additive noise in non-Gaussian. In such cases, the LMF algorithm can lead to much lower noise in the weights for the same speed of convergence. In [4], it was shown that the LMF algorithm can outperform LMS with gaussian noise for a sufficiently high signal-to-noise ratio (*SNR*).

Reference [3] presented a stochastic analysis of the LMF algorithm behavior about the optimum weigth error. Thus, the analytical model derived in [3] is accurate in steady-state. In the transient phase of adaptation, the model in [3] is accurate only for very small *SNR*, an uncommon practical situation. The simulation results in [3] also revealed that the algorithm's stability is dependent on the initial weight conditions. The analytical model derived in [3], however, cannot predict this dependence. [5] extended the analysis in [3] to include the transient behavior, but the analysis considered only the case of Gaussian noise. More recently, [6] presented an analysis based on energy arguments for a class of algorithms that use nonlinear error functions for weight adaptation. This class includes the LMF algorithm. The analysis is restricted to the development of a recursive equation for the quadratic norm of the weight error for Gaussian noise. No expression is derived for mean weight behavior or for MSE. [6] also shows that the stability of the LMF algorithm depends on the initial conditions, although this dependence is not explicitly determined.

This paper presents a new stochastic model for the behavior of the LMF adaptive algorithm. The new model assumes a zero-mean, wide-sense stationary Gaussian input and any zero-mean symmetric probability density function for the additive noise. Recursive equations are derived for the mean weight, for the weight error correlation matrix and for the MSE. The new model accurately predicts the algorithm behavior in both transient and steady-state for any *SNR*. It also clearly shows the dependence of the convergence properties on the initial conditions. Monte Carlo simulation results are presented to illustrate the accuracy of the proposed model.

2. MEAN WEIGHT BEHAVIOR

Figure 1 shows the block diagram of the problem studied. $W^0 = [w_0^0, w_1^0, ..., w_{N-1}^0]^T$ is the impulse response vector of a linear system. $W(n) = [w_0(n), w_1(n), ..., w_{N-1}(n)]^T$ is the adaptive weight vector. $x(n)$ is assumed stationary, zero-mean and Gaussian with variance σ_x^2. $X(n) = [x(n), x(n-1), ..., x(n-N+1)]^T$ is the observed data vector. $z(n)$ is the measurement noise, assumed stationary, white, zero-mean with variance σ_z^2 and uncorrelated with any other signal. $y(n)$ is the adaptive filter output and $e(n)$ is the error signal.

The LMF algorithm weight update equation is [3]

$$W(n+1) = W(n) + \mu e^3(n) X(n) \qquad (1)$$

where μ is the adaptation constant. Defining the weight error vector $V(n) = W(n) - W^0$ about optimal solution, (1) can be written as

$$V(n+1) = V(n) + \mu e^3(n)X(n) \qquad (2)$$

From Fig. 1, $e(n) = z(n) - X^T(n)V(n)$. Using this expression in (2), yields:

$$V(n+1) = V(n)$$
$$+ \mu \sum_{i=0}^{3} \binom{3}{i} (z^i(n)(-X^T(n)V(n))^{3-i})X(n) \qquad (3)$$

As $z(n)$ is zero-mean, i.i.d. and has zero odd moments (symmetric pdf), the expected value of (3) is given by

$$E[V(n+1)] = E[V(n)]$$
$$- 3\mu E[z^2(n)]E[X(n)X^T(n)V(n)] \qquad (4)$$
$$- \mu E[(X^T(n)V(n))^3 X(n)]$$

To proceed with the calculations, it is assumed that $X(n)$ and $V(n)$ can be considered independent for sufficiently small μ [1]. Thus $E[X(n)X^T(n)|V(n)] = E[X(n)X^T(n)] = R$ (correlation matrix of $X(n)$). The first expectation in (4) is then given by

$$E[X(n)X^T(n)V(n)] = RE[V(n)] \qquad (5)$$

The second expection, conditioned on $V(n)$, can be written as

$$E[(X^T(n)V(n))^3 X(n)|V(n)]$$
$$= E[(X^T(n)V(n))^2 X(n)X^T(n)V(n)|V(n)] \qquad (6)$$
$$= E[(X^T(n)V(n))^2 X(n)X^T(n)|V(n)]V(n)$$

Using the same methodology as in [7], (6) can be written as

$$E[(y_1)^2 X(n)X^T(n)|V(n)] =$$
$$E[X(n)X^T(n)|V(n)]E[(y_1)^2|V(n)] \qquad (7)$$
$$+ E[y_1 X(n)|V(n)]E[y_1 X^T(n)|V(n)]B(y_1|V(n))$$

where $B(y_1|V(n)) = \frac{1}{E[y_1^2|V(n)]}\left(\frac{E[y_1^4|V(n)]}{E[y_1^2|V(n)]} - E[y_1^2|V(n)]\right)$ and $y_1 = X^T(n)V(n)$. Note that for $x(n)$ zero-mean Gaussian, $y_1(n)$ is also zero-mean Gaussian when conditioned on $V(n)$. Using again the independence assumption, the terms in (7) are given by

$$E[y_1^2|V(n)] = V^T(n)RV(n)$$
$$E[y_1^4|V(n)] = 3E^2[y_1^2|V(n)]$$
$$B(y_1) = 2 \qquad (8)$$
$$E[y_1 X(n)|V(n)] = RV(n)$$

Using the above results in (6) leads to

$$E[(X^T(n)V(n))^3 X(n)|V(n)]$$
$$= V^T(n)RV(n)RV(n) + 2RV(n)V^T(n)RV(n) \qquad (9)$$

Averaging (9) over $V(n)$ requires extra approximations, since the pdf of $V(n)$ is not known. Assuming slow learning and N sufficiently large, the following approximations are used:

$i. \ E[V^T(n)RV(n)RV(n)]$
$$\approx E[V^T(n)RV(n)]RE[V(n)] = tr(RK(n))RE[V(n)]$$

$ii. \ E[RV(n)V^T(n)RV(n)]$
$$\approx RE[V(n)]E[V^T(n)RV(n)] = RE[V(n)]tr(RK(n))$$

where $K(n) = E[V(n)V^T(n)]$ is the correlation matrix of $V(n)$ and $tr(.)$ represents the trace of a matrix. Approximations (i) and (ii) are based on the fact that each component $v_i(n)$ of $V(n)$ contributes to only N of the N^2 terms in $V^T(n)RV(n)$. Thus, for large N, each $v_i(n)$ can be considered weakly correlated with $V^T(n)RV(n)$.

Using these approximations, the expected value of (9) becomes

$$E[(X^T(n)V(n))^3 X(n)] \approx 3tr(RK(n))RE[V(n)] \qquad (10)$$

Using (10) and (5) in (4) yields an expression for the mean weight behavior

$$E[V(n+1)] = E[V(n)]$$
$$- 3\mu\left(\sigma_z^2 + tr(RK(n))\right)RE[V(n)] \qquad (11)$$

Note that the model of [3, Eq.(21), for K=2] reduces to

$$E[V(n+1)] = (1 - 3\mu\sigma_z^2 R)E[V(n)] \qquad (12)$$

The two models (11) and (12) tend to coincide for very small *SNR*. For moderate and large *SNR*, the contribution of $K(n)$ in (11) will improve the model accuracy during the transient adaptation phase. Finally, (11) evidences the dependence of the LMF algorithm dynamic behavior on the initial condition $K(0) = V(0)V^T(0)$.

The next section presents the derivation of a recursive model for $K(n) = E[V(n)V^T(n)]$, which is required for (11). This result will also be used to determine the MSE behavior.

3. SECOND MOMENT ANALYSIS

Multiplying (3) by its transpose and taking the expected value yields

$$E[V(n+1)V^T(n+1)] = E[V(n)V^T(n)]$$
$$+ \mu E[e^3(n)(V(n)X^T(n) + X(n)V^T(n))] \qquad (13)$$
$$+ \mu^2 E[e^6(n)X(n)X^T(n)]$$

The same technique used in (3)-(9) can be used to determine the expected values in (13). Assuming independence of $X(n)$ and

$V(n)$, the expectation in the second line of (13) yields

$$
\begin{aligned}
&E[e^3(n)(V(n)X^T(n) + X(n)V^T(n))] \\
&= -3\sigma_z^2(RE[V(n)V^T(n)] + E[V(n)V^T(n)]R) \\
&\quad - 2RE[V(n)V^T(n)RV(n)V^T(n)] \\
&\quad - E[V^T(n)RV(n)RV(n)V^T(n)] \\
&\quad - 2E[V(n)V^T(n)RV(n)V^T(n)]R \\
&\quad - E[V(n)V^T(n)V^T(n)RV(n)]R
\end{aligned}
\tag{14}
$$

Using the same reasoning used to obtain (10), the moments of $V(n)$ with order higher than 2 in (14) can be approximated by:

$$
\begin{aligned}
i.\ &E[V(n)V^T(n)RV(n)V^T(n)] \\
&= E[V^T(n)RV(n)V(n)V^T(n)] \\
&\approx E[V^T(n)RV(n)]E[V(n)V^T(n)] = tr(RK(n))K(n)
\end{aligned}
$$

$$
\begin{aligned}
ii.\ &E[V^T(n)RV(n)RV(n)V^T(n)] \\
&\approx E[V^T(n)RV(n)]RE[V(n)V^T(n)] \\
&= tr(RK(n))RK(n)
\end{aligned}
$$

$$
\begin{aligned}
iii.\ &E[V(n)V^T(n)V^T(n)RV(n)] \\
&\approx E[V(n)V^T(n)]E[V^T(n)RV(n)] = tr(RK(n))K(n)
\end{aligned}
$$

Using these approximations in (14) yields

$$
\begin{aligned}
&E[e^3(n)(V(n)X^T(n) + X(n)V^T(n))] \\
&\approx -3\Big(\sigma_z^2 + tr(RK(n))\Big)\Big(RK(n) + K(n)R\Big)
\end{aligned}
\tag{15}
$$

In determining the last expected value in (13), the terms $E[(X^T(n)V(n))^{2k}X(n)X^T(n)]$ are neglected for $k > 1$. These terms represent moments of order higher than two which are multiplied by μ^2. Thus, their influence can be neglected. With this consideration,

$$
\begin{aligned}
&E[e^6(n)X(n)X^T(n)] \approx E[z^6(n)]R \\
&\quad + 15E[z^4(n)]\Big(tr(RK(n))R + 2RK(n)R\Big)
\end{aligned}
\tag{16}
$$

Using (16) and (15) in (13), a recursive expression is obtained for the correlation matriz $K(n) = E[V(n)V^T(n)]$:

$$
\begin{aligned}
K(n+1) &= K(n) \\
&\quad - 3\mu\Big(\sigma_z^2 + tr(RK(n))\Big)\Big(RK(n) + K(n)R\Big) \\
&\quad + \mu^2\Big(15E[z^4(n)](tr(RK(n))R + 2RK(n)R)\Big) \\
&\quad + \mu^2 E[z^6(n)]R
\end{aligned}
\tag{17}
$$

Eq. (17) can now be used in (11) to determine the mean weight behavior. Some interesting comments are in order regarding (17). First, note that the term controlling the dynamics of $K(n)$ is itself a function of $K(n)$. This shows that the convergence depends on the initial conditions. Second, (17) shows that, contrary to what happens with the LMS algorithm, the transient behavior of the

LMF algorithm behavior is affected by both the noise power an by the noise pdf. $E[z^2(n)] = \sigma_z^2$ and $E[z^4(n)]$ affect the transient behavior (stability and convergence speed). $E[z^6(n)]$ affects the steady-state behavior.

Using the results just derived, it is also possible determine the MSE for the LMF algorithm, since [1]

$$
\xi(n) = \sigma_z^2 + tr(RK(n))
\tag{18}
$$

The new analytical model for the LMF algorithm behavior is then composed by expressions (11), (17) and (18).

4. EXAMPLES

The model derived in this paper has been tested in several simulations using different noise distributions. This section presents a simulation example to illustrate the accuracy of the model. Consider the system in Fig. 1 with $W^0 = [0.3030; 0.8081; 0.5051]^T$, $W^{0T}W^0 = 1$. $z(n)$ is uniformly distributed with $\sigma_z^2 = 0.1$. The input signal is an autoregressive process defined by $x(n) = ax(n-1) + y(n)$ $(0 < a < 1)$. The variance σ_y^2 of $y(n)$ was chosen equal to 1. The weights were initialized at $W(0) = 0$. Fig. 2 shows the mean behavior of the first weight derived from (11) and (17), as well as from the model in [3] (Eq. (12)), for two values of μ: $\mu_1 = 0.0035 \approx \frac{\mu_{max}}{2}$ and $\mu_2 = 0.00035 \approx \frac{\mu_{max}}{20}$, for $a = 0.6$, yielding an eigenvalue spread of R equal to 6.55. Simulations are averaged over 50 runs. The other weights have similar behavior. The value for μ_{max} was obtained by simulation. Fig. 3 shows the MSE derived from (17), (18), by simulation (50 runs) and also from the model in [3], for the same values of μ_1 and μ_2. Fig. 3 also shows the mean square error during transient, averaged over 200 runs, for the same values of μ_1 and μ_2. Fig. 4 is equivalent to Fig. 3, but for $a = 0.875$, yielding an eigenvalue spread of R equal to 30.60. The MSE is shown again for two values of μ: $\mu_3 = 0.0013 \approx \frac{\mu_{max}}{2}$ and $\mu_4 = 0.00013 \approx \frac{\mu_{max}}{20}$. The value for μ_{max} was again determined by simulation. Note that the new model accurately predicts the algorithm behavior, even for μ as large as $\mu_{max}/2$.

5. CONCLUSION

This paper presented a new statistical analysis of the LMF algorithm. The new analysis improves previous results in that higher order moments of the weight error vector are not neglected, but approximated using assumptions valid for a large number of filter weights. Recursive equations were derived for the first and second order moments of the adaptive weights for a stationary Gaussian reference signal. The model accurately predicts the algorithm behavior in both transient and steady-state phases. It is valid for any zero-mean noise probability density function and for any signal-to-noise ratio. Computer simulations illustrate the accuracy of the new model in predicting the algorithm behavior for different values of step size.

Though the analysis was based on assumptions of slow learning and large number of weights, the results show excellent match between theory and simulations even for reasonable large μ and small number of taps. Nevertheless, a small μ is usually preferred

in practical applications, where only one realization of the process is available and the fluctuations about the mean weight values should be as small as possible.

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 3th edition, Englewood Cliffs, NJ:Prentice-Hall, 1996.

[2] A. Gersho, "Some Aspects of Linear Estimation with Non-Mean-Square Error Criteria," in *Asilomar Circuits and Systems Conference*, 1969.

[3] B. Widrow and E. Walach "The Least Mean Fourth (LMF) Adaptive Algorithm and Its Family," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 275-283, Aug. 1984.

[4] S. V. Narasimhan, "Application of the Least Mean Fourth (LMF) Adaptive Algorithm to Signals Associated with Gaussian Noise," *International Journal of Electronics*, vol. 63, pp. 895-913, 1987.

[5] S. H. Cho, S. D. Kim and K. Y. Jeon, "Statistical Convergence of the Adaptive Least Mean Fourth Algorithm," *Proceedings of the ICSP'96*, pp.610-613.

[6] T.Y. Al-Naffouri and A. Sayed, "Transient Analysis of Adaptive Filters," *International Conference of Acoustics, Speech and Signal Processing - ICASSP*, Salt Lake City, USA, May 2001.

[7] N. J. Bershad, P. Celka and J. M. Vesin, " Stochastic Analyses Gradient Adaptive Identification of Nonlinear Systems with Memory for Gaussian Data and Noisy Input and Output Measurements," *IEEE Transactions on Signal Processing*, v. 47, n. 3, pp. 675-689, Mar. 1999.
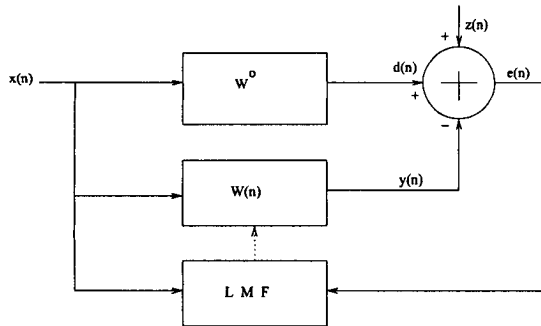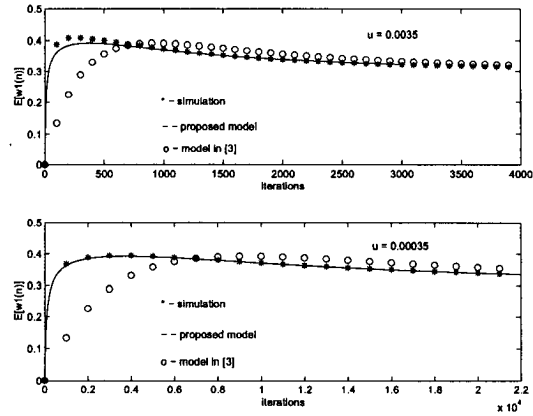
Fig. 1. Block Diagram.
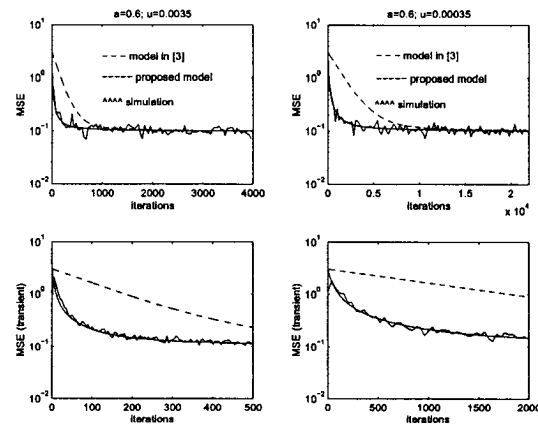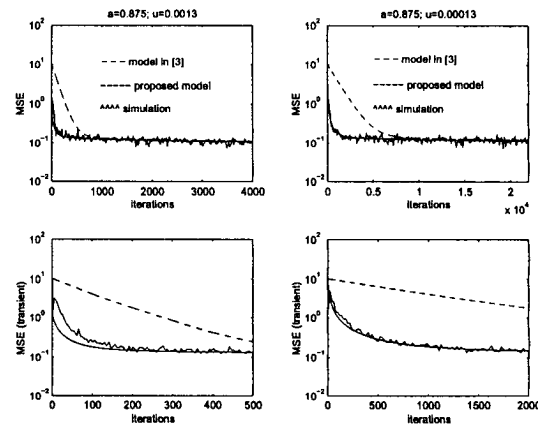

Fig. 2. Mean Value of First Weight (a = 0.6).

Fig. 3. Mean Square Error (a = 0.6).

Fig. 4. Mean Square Error (a = 0.875).

I - 28