NEW RESULTS ON THE STABILITY ANALYSIS OF THE LMF (LEAST MEAN FOURTH) ADAPTIVE ALGORITHM

Pedro I. Hübscher[1], *Vítor H. Nascimento*[2] *and José C. M. Bermudez*[3*]

[1] INPE - National Institute for Space Research - São José dos Campos - SP - Brazil
[2] USP - University of São Paulo - São Paulo - SP - Brazil
[3] UFSC - Federal University of Santa Catarina - Florianópolis - SC - Brazil
e-mails: pedro@lit.inpe.br, vitor@lps.usp.br, j.bermudez@ieee.org

ABSTRACT

This paper presents a new analysis for the convergence of the LMF (Least Mean Fourth) adaptive algorithm. The analysis improves previous results because it explicitly shows how the stability of the algorithm depends on the initial conditions of the weights, i.e., the analysis is also valid when the algorithm is initialized far from the optimum weight vector. Analytical expressions are derived relating the limiting values of the adaptation constant and the initial weight error vector. The analysis assumes a white zero-mean Gaussian reference signal and a white measurement noise with any even probability density function (p.d.f.) and finds conditions for convergence in the mean square sense.

1. INTRODUCTION

There are several approaches to analyze the convergence of adaptive algorithms: deterministic (worst-case) and stochastic (in the mean, in the mean-square, and almost-sure) [1]. Walach and Widrow [2] studied the convergence properties (in the mean-square sense) of the LMF algorithm. Their analysis was restricted to steady-state, and the stability limit was not expressed as a function of the initial conditions, even though the reported simulation results indicated this dependence. In [3], the ODE method is used to analyze general fixed-step adaptive algorithms (including LMF). However, no analytical expression is given for the LMF stability conditions. In [4], the authors comment on the dependence of LMF's stability on its initial conditions. An expression is provided for the maximum adaptation constant for convergence in the mean. However, the analysis in [4] assumes that both the input signal and the measurement noise to be Gaussian.

More recently, [5] has shown that the stability of the LMF algorithm depends on the initial conditions. However, such dependence was not explicitly determined.

This paper presents a new convergence analysis (in the mean-square sense) of the LMF algorithm, considering a white zero-mean Gaussian reference signal and a white zero-mean measurement noise with any even probability density function (p.d.f.). The dependence on the initial conditions is explicitly shown through analytical expressions. The algorithm is considered stable if the mean-square error (MSE) remains stable during the adaptation process, and converges to a steady-state value. Since we require mean-square stability, our conditions are more restrictive (and more useful in practice) than those presented in other works.

*This work has been supported in part by CNPq (Brazilian Ministry of Science and Technology) under Grant No. 352084/92-8. The work of V.H.N. was also supported in part by FAPESP – São Paulo State Research Council.

2. DEFINITION OF THE PROBLEM

Figure 1 shows a block diagram of the problem studied here. $W^0 = [w_1^0, w_2^0, ..., w_N^0]^T$ is the impulse response vector of a linear system, $W(n) = [w_1(n), w_2(n), ..., w_N(n)]^T$ is the adaptive weight vector, $x(n)$ is assumed stationary, white, 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, $o(n)$ is the adaptive filter output, and $e(n)$ is the error signal. $z(n)$ is the measurement noise, assumed stationary, white, zero-mean with variance σ_z^2 and independent of any other signal. Moreover, it is assumed that $z(n)$ can have any distribution with an even p.d.f.

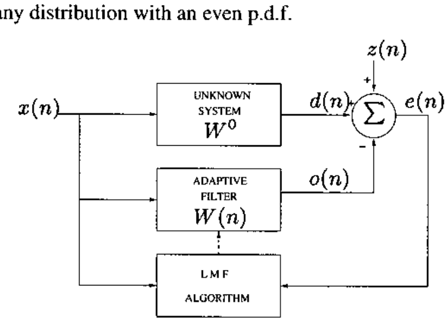

Fig. 1. LMF applied for System Identification.

3. TRACE OF THE WEIGHT ERROR AUTOCORRELATION MATRIX

Though the conditions for convergence in the mean can provide some insight, second moment stability is far more important in determining conditions for algorithm's convergence. Thus, we restrict this analysis to the study of conditions for second order moment convergence.

For white inputs, the second order moments of the weights are related to the MSE through [2]

$$\xi(n) = \sigma_z^2 + \sigma_x^2 E[V^T(n)V(n)], \tag{1}$$

where $V(n) = W(n) - W^0$ is the weight error vector. Hence, the MSE convergence can be studied through the convergence properties of $E[V^T(n)V(n)]$.

A recursive expression for the behavior of $E[V^T(n)V(n)]$ could be easily obtained by taking the trace of the recursion derived in [6] for the weight error correlation matrix $K(n) = E[V(n)V^T(n)]$ of the LMF algorithm. However, terms neglected in [6], which were not significant for the analysis made there, become important in an analysis[1] considering large values of μ and, consequently, for a stability analysis. Therefore, a recursive expression for $E[V^T(n)V(n)]$, must be determined starting from the LMF weight error updating equation [2]

$$V(n+1) = V(n) + \mu e^3(n)X(n). \tag{2}$$

Pre-multiplying (2) by its transpose, taking the expected value and using the statistical properties of $z(n)$ (odd moments equal to zero and independent of any other signal), leads to

$$\begin{aligned} &E[V^T(n+1)V(n+1)] = E[V^T(n)V(n)] \\ &- 2\mu E[(X^T(n)V(n))^4] - 6\mu E[z^2(n)]E[(X^T(n)V(n))^2] \\ &+ \mu^2 E[(X^T(n)V(n))^6 X^T(n)X(n)] \\ &+ 15\mu^2 E[z^2(n)]E[(X^T(n)V(n))^4 X^T(n)X(n)] \\ &+ 15\mu^2 E[z^4(n)]E[(X^T(n)V(n))^2 X^T(n)X(n)] \\ &+ \mu^2 E[z^6(n)]E[X^T(n)X(n)] \end{aligned} \tag{3}$$

In the following analysis, we assume that the effects of the statistical dependence of $X(n)$ and $V(n)$ can be neglected. The expected values of (3) are calculated as follows:

Expected Value 1: $E[(X^T(n)V(n))^4]$

As $X(n)$ is Gaussian and independent of $V(n)$, $X^T(n)V(n)$ is also Gaussian when conditioned on $V(n)$. Therefore, we can write

$$\begin{aligned} &E[(X^T(n)V(n))^{2k}|V(n)] \\ &= E[(X^T(n)V(n))^2|V(n)]^k \prod_{m=1}^{k}(2m-1) \end{aligned} \tag{4}$$

and

$$\begin{aligned} &E[(X^T(n)V(n))^4|V(n)] \\ &= 3E^2[(X^T(n)V(n))^2|V(n)] \\ &= 3\{E[V^T(n)X(n)X^T(n)V(n)|V(n)]\}^2 \\ &= 3\{V^T(n)RV(n)\}^2 = 3\{\sigma_x^2 V^T(n)V(n)\}^2 \\ &= 3\sigma_x^4 V^T(n)V(n)V^T(n)V(n) \end{aligned} \tag{5}$$

[1]The matrix $K(n)$ in [6] was derived neglecting the terms $E[(X^T(n)V(n))^{2k}X(n)X^T(n)]$ for $k > 1$, and considering small μ and large number of weights.

Averaging (5) over $V(n)$ requires extra approximations, since the p.d.f. of $V(n)$ is unknown. The following approximation is used.

$$E[V^T(n)V(n)V^T(n)V(n)] \approx E[V^T(n)V(n)]E[V^T(n)V(n)] \tag{6}$$

Approximation (6) assumes that the variance of $V^T(n)V(n)$ is much smaller than its mean value (this can be considered reasonable in the beginning of the adaptation process). In steady-state the higher-order moments of the weights can be neglected (since $V(n)$ should be small in steady-state). Extensive simulation results have shown that this approximation[2] leads to good accuracy in determining the stability conditions.

Using this (6), (5) becomes

$$\begin{aligned} E[(X^T(n)V(n))^4] &= 3\sigma_x^4 E[V^T(n)V(n)V^T(n)V(n)] \\ &\approx 3\sigma_x^4 E[V^T(n)V(n)]E[V^T(n)V(n)] \end{aligned} \tag{7}$$

Expected Value 2: $E[(X^T(n)V(n))^2]$

$$\begin{aligned} E[(X^T(n)V(n))^2|V(n)] &= E[V^T(n)X(n)X^T(n)V(n)|V(n)] \\ &= V^T(n)E[X(n)X^T(n)|V(n)]V(n) \\ &= V^T(n)RV(n) = \sigma_x^2 V^T(n)V(n) \end{aligned} \tag{8}$$

Averaging over $V(n)$, (8) gives

$$E[(X^T(n)V(n))^2] = \sigma_x^2 E[V^T(n)V(n)] \tag{9}$$

Next, we evaluate the expected values that are multiplied by μ^2 in (3). They are derived using the same methodology presented in [6] and [7], and also using approximations similar to (6).

Expected Value 3: $E[(X^T(n)V(n))^6 X^T(n)X(n)]$

$$\begin{aligned} &E[(X^T(n)V(n))^6 X^T(n)X(n)] \\ &= tr\Big\{E[X^T(n)V(n))^6 X(n)X^T(n)]\Big\} \\ &= (15N + 90)\sigma_x^8 E^3[V^T(n)V(n)] \end{aligned} \tag{10}$$

Expected Value 4: $E[(X^T(n)V(n))^4 X^T(n)X(n)]$

$$\begin{aligned} &E[(X^T(n)V(n))^4 X^T(n)X(n)] \\ &= (3N + 12)\sigma_x^6 E^2[V^T(n)V(n)] \end{aligned} \tag{11}$$

Expected Value 5: $E[(X^T(n)V(n))^2 X^T(n)X(n)]$

$$E[(X^T(n)V(n))^2 X^T(n)X(n)] = (N+2)\sigma_x^4 E[V^T(n)V(n)] \tag{12}$$

Expected Value 6: $E[X^T(n)X(n)]$

$$E[X^T(n)X(n)|V(n)] = E[X^T(n)X(n)] = \sigma_x^2 N \tag{13}$$

[2]To be exact, one should use $E[V^T(n)V(n)V^T(n)V(n)] = E^2[V^T(n)V(n)] + \sigma^2_{V^T(n)V(n)}$, where $\sigma^2_{V^T(n)V(n)}$ is the variance of $V^T(n)V(n)$, which cannot be calculated as $V(n)$ has unknown p.d.f.

Using the expected values 1-6 in equation (3), results an expression for $E[V^T(n+1)V(n+1)]$.

$$y(n+1) = (1-a)y(n) - by^2(n) + cy^3(n) + d \qquad (14)$$

where:
$y(n) = E[V^T(n)V(n)]$;
$a = A_1\mu - A_2\mu^2$;
$b = B_1\mu - B_2\mu^2$;
$c = C\mu^2$;
$d = D\mu^2$;
$A_1 = 6\sigma_z^2\sigma_x^2$;
$A_2 = 15E[z^4(n)]\sigma_x^4(N+2)$;
$B_1 = 6\sigma_x^4$;
$B_2 = 15\sigma_z^2\sigma_x^6(3N+12)$;
$C = \sigma_x^8(15N+90)$;
$D = E[z^6(n)]\sigma_x^2 N$.

4. STABILITY ANALYSIS

Expression (14) is a nonlinear difference equation. Its convergence depends in general on the initial condition $y(0) = V^T(0)V(0)$, the squared Euclidean norm of the initial weight error vector.

To proceed with the determination of the stability conditions, we need to find the equilibrium points of (14). Writing $y(n+1) = y(n) = y_\infty$, we obtain

$$y(n+1) = (1-a)y(n) - by^2(n) + cy^3(n) + d = y(n) = y_\infty \qquad (15)$$

and

$$y_\infty^3 - \frac{b}{c}y_\infty^2 - \frac{a}{c}y_\infty + \frac{d}{c} = 0 \qquad (16)$$

Equation (16) has three roots, which represent the equilibrium points. These roots can be expressed in analytical form as follows.

$$\begin{aligned} y_{1\infty} &= (s_1+s_2) + \frac{b}{3c} \\ y_{2\infty} &= -\frac{1}{2}(s_1+s_2) + \frac{b}{3c} + \frac{j\sqrt{3}}{2}(s_1-s_2) \\ y_{3\infty} &= -\frac{1}{2}(s_1+s_2) + \frac{b}{3c} - \frac{j\sqrt{3}}{2}(s_1-s_2) \end{aligned} \qquad (17)$$

where
$s_1 = \left(r + \sqrt{q^3+r^2}\right)^{\frac{1}{3}}$, $s_2 = \left(r - \sqrt{q^3+r^2}\right)^{\frac{1}{3}}$;
$q = -\frac{a}{3c} - \frac{b^2}{9c^2}$, $r = \frac{1}{6}\left(\frac{ab}{c^2} - \frac{3d}{c}\right) + \frac{b^3}{27c^3}$.

Depending on the values of q and r three cases can occur:

- Case 1: $q^3 + r^2 < 0$ (only real roots)

  In this case, (16) has three real negative roots (out of the region of interest, because $y(n)$ is a norm and cannot be negative) or two real positive roots and one real negative root (the negative root has again no physical sense). Figure 2 illustrates the case of one negative and two positive roots, represented by y_{neg}, y_c and $y(0)_{max}$. Root y_c corresponds to the stable equilibrium point and also represents the steady-state point $E[V^T(\infty)V(\infty)]$. Root $y(0)_{max}$ corresponds to an unstable point, which gives the maximum value of $y(0)$ that guarantees the stability of (14) for a specific value of μ. The smaller the value of μ, the larger the value for $y(0)_{max}$. As $\mu \to 0$, $y(0)_{max} \to \infty$. Root y_{neg} is always negative, because $y(n+1)$ is a third-degree polynomial of $y(n)$, with $d > 0$, and $y(n+1) \to -\infty$ if $y(n) \to -\infty$.

- Case 2: $q^3 + r^2 = 0$ (only real roots, and two of them are equal and nonzero)

  Equation (16) has two real and equal roots (y_c and $y(0)_{max}$). The curve $f(y(n)) = y(n+1)$ is tangent to the line $y(n+1) = y(n)$ at the point $y_c = y(0)_{max}$ (Figure 3). There is one real negative root, represented by y_{neg}.

  This case allows us to find the maximum value of μ. Writing $q^3 + r^2$ as a function of a, b, c and d, yields

$$\begin{aligned} q^3 + r^2 &= \left\{-\frac{a}{3c} - \frac{b^2}{9c^2}\right\}^3 + \left\{\frac{1}{6}\left(\frac{ab}{c^2} - \frac{3d}{c}\right) + \frac{b^3}{27c^3}\right\}^2 \\ &= -\frac{1}{729c^6}\left(3ac + b^2\right)^3 \\ &\quad + \frac{1}{729c^6}\left(\frac{9}{2}abc - \frac{27}{2}dc^2 + b^3\right)^2 \end{aligned} \qquad (18)$$

  As (18) is equal to zero for case 2, we conclude that

$$4\left(3ac + b^2\right)^3 = \left(9abc - 27dc^2 + 2b^3\right)^2 \qquad (19)$$

  Writing (19) in polynomial form, and substituting the variables a, b, c and d as functions of A_1, A_2, B_1, B_2, C, D and μ, results

$$P_4\mu^4 + P_3\mu^3 + P_2\mu^2 + P_1\mu + P_0 = 0 \qquad (20)$$

  where:
  $P_4 = -4A_2^3C + A_2^2B_2^2 + 18A_2B_2CD - 4B_2^3D - 27C^2D^2$;
  $P_3 = 12A_1A_2^2C - 2(A_1A_2B_2^2 + A_2^2B_1B_2) - 18(A_1B_2 + A_2B_1)CD + 12B_1B_2^2D$;
  $P_2 = -12A_1^2A_2C + A_1^2B_2^2 + A_2^2B_1^2 + 4A_1A_2B_1B_2 + 18A_1B_1CD - 12B_1^2B_2D$;
  $P_1 = 4A_1^3C - 2(A_1A_2B_1^2 + A_1^2B_1B_2) + 4B_1^3D$;
  $P_0 = A_1^2B_1^2$.

  The smallest positive and real root μ_0 of (20) gives the maximum value of μ that guarantees stability.

- Case 3: $q^3 + r^2 > 0$ (two complex roots)

  In this case there are one real root and two complex roots. There is no point in common between $f(y(n)) = y(n+1)$ and the line $y(n+1) = y(n)$ for $y(n) > 0$. Therefore, there is no value of $y(0) \geq 0$ which can guarantee the stability of (14). This case, represented in Figure 4, occurs when the value of μ is greater than the limit value given by μ_0.

These results allow the explicit determination of the stability conditions for the LMF algorithm when applied to the system in Fig. 1 with a known W^o. Given the system parameters, the maximum value of μ (μ_o) can be determined from (20). Then, for any $\mu < \mu_o$, $y(0)_{max}$ can be determined from the solutions of (16). This procedure requires prior knowledge of the system to be identified in a design situation. This is a property of the algorithm, not a flaw in the analysis. Given some prior estimate of W^o, the analysis can then be used to study the robustness of the algorithm in solving the practical problem for a given error in the estimate.

The theoretical results have been extensively tested. Predictions of μ_o matched the simulation results within $\pm 10\%$. Predictions of $y(0)_{max}$ were, on average, 130% above the values determined by simulation. These errors are expected due to the complexity of the problem and the simplifications used. However, the predictions can serve as good guidance in a conservative design.

5. CONCLUSION

This paper presented a new convergence analysis for the LMF adaptive algorithm. The analysis improves previous results in that the dependence of the stability on the initial conditions is explicitly shown. The results reveal a relationship between the initial conditions and the step size in determining convergence. The smaller the value of μ, the larger the allowable values for the initial weight error vector. Simulations have shown that the theoretical predictions can be useful for design purposes.

6. REFERENCES

[1] N. Kalouptsidis and S. Theodoridis, *Adaptive System Identification and Signal Processing Algorithms*, Prentice-Hall, 1993.

[2] E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family", *IEEE Transactions on Information Theory*, 30(2), pp. 275-283, 1984.

[3] A. Sethares and S. Rajesh, "Asymptotic analysis of stochastic gradient-based adaptive filtering algorithms with general cost functions", *IEEE Transactions on Signal Processing*, Vol. 44, No. 9, Sept., 1996.

[4] 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.

[5] T. Al-Naffouri and A. Sayed, "Transient analysis of adaptive filters", *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-2001*, Vol. 6, pp. 3869-3872, 2001.

[6] Pedro I. Hübscher and José C. M. Bermudez, "An improved stochastic model for the least mean fourth (LMF) adaptive algorithm", *IEEE International Symposium on Circuits and Systems, ISCAS-2002*, Vol. 1, pp. 25-28, 26-27/May, 2002.

[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*, Vol. 47, No. 3, pp. 675-689, Mar., 1999.

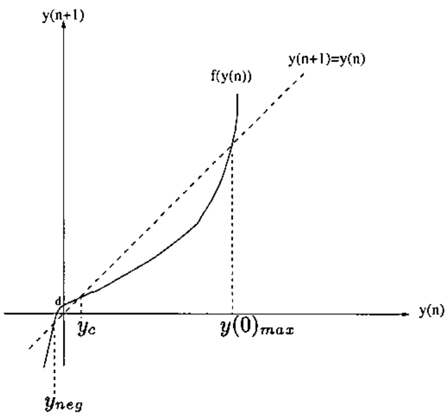

Fig. 2. Equilibrium points: case 1 ($y_c << y(0)_{max}$).

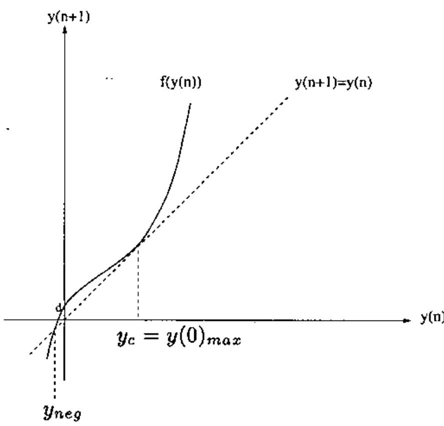

Fig. 3. Equilibrium points: case 2 ($y_c = y(0)_{max}$).

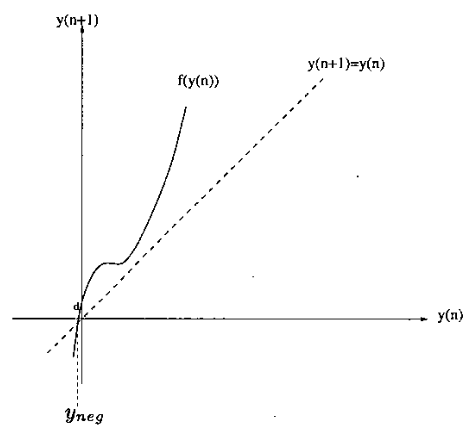

Fig. 4. Equilibrium points: case 3 (y_c and $y(0)_{max}$ complex).