A novel Approach for the Convergence Analysis of the Least-Mean Fourth Algorithm

Azzedine Zerguine
Electrical Engineering Department
King Fahd University of Petroleum & Minerals
KFUPM, Dhahran 31261, Saudi Arabia.
E-mail:azzedine@kfupm.edu.sa

Abstract

The convergence analysis of the least-mean fourth (LMF) algorithm is derived. A novel approach is used to study the convergence behavior of the algorithm. As a by-product of this novel approach, expressions for more general and new sufficient conditions for convergence and the excess steady-state error for the LMF algorithm are derived.

1 Introduction

The least-mean square (LMS) algorithm [1] is one of the most widely used adaptive schemes. It has several desirable features and some limitations. As such, several LMS-variants have been proposed that trade some of the LMS features for an enhanced performance in some of its limitations. Of particular importance is the class of least-mean square algorithms that employ an error nonlinearity $f(e_n)$ instead of the (linear) error term in LMS adaptation [2]-[4]. Examples include the sign-error algorithm [5], the least-mean fourth (LMF) algorithm and its family [6], and the least-mean mixed norm algorithm [7], all of which are intuitively motivated. Table 1 defines $f(e_n)$ for many famous algorithms. Also, mentioned in Table 1 is $f(e_n) = \alpha e_n + 2(1-\alpha)e_n^3$ which is the error nonlinearity used in the mixed LMS-LMF algorithm with α as the mixing parameter. This algorithm is found to result in better performance than either the LMS or the LMF algorithms in Gaussian and non-Gaussian environments.

The least mean-square algorithm and the least mean-fourth algorithm fall under the generalized minimization of the mean-pth-error function, that is $J_n = E[e_n^p]$, p being a positive integer, where $p = 2$ and $p = 4$ result, respectively, in the LMS and LMF algorithms [1].

While the LMS algorithm is very well established in adaptive filtering, the LMF algorithm has been proposed by [6] and has recently gained attention [7]-[9]. The two

Algorithm	$f(e_n)$
LMS	e_n
NLMS	$\frac{e_n}{\|\mathbf{x}_n\|^2}$
Sign-LMS (Sign-Error)	$sign[e_n]$
Sign-LMS (Sign-Regressor)	$sign[\mathbf{x}_n]$
Sign-LMS (Sign-Error, Sign-Regressor)	$sign[e_n]sign[\mathbf{x}_n]$
LMF	e_n^3
Mixed LMS-LMF	$\alpha e_n + 2(1-\alpha)e_n^3$

Table 1: Examples for $f(e_n)$.

algorithms have different convergence behavior and robustness to noise statistics (Gaussian versus non-Gaussian noise) [6]. For example, the LMF algorithm will clearly have a larger gradient driving it to converge faster when away from the optimum ($e_n^4 > e_n^2$ for $e_n^2 > 1$). However, the LMS will have more desirable characteristics in the neighborhood of the optimum.

The LMF algorithm is defined by the following cost function [6]:

$$J_n = E[e_n^4], \tag{1}$$

where the error $e_n = d_n + w_n - \mathbf{c}_n^T\mathbf{x}_n$, d_n is the desired value, $\mathbf{c}_n$ is the filter coefficient of the adaptive filter (with $\mathbf{c}_{opt}$ is its optimal value), $\mathbf{x}_n$ is the input vector and w_n is the additive noise.

In this work, the convergence and the steady-state analysis of the LMF algorithm are derived using a novel approach. Eventhough, some of the results are identical to those found in [6], expressions for more general and new sufficient conditions for convergence and the excess steady-state error for the LMF algorithm are obtained.

2 Convergence Analysis of the LMF Algorithm

Throughout our ensuing convergence analysis, the following commonly-used assumptions [1], [6] are made:

A.1 *The noise sequence* $\{w_n\}$ *is statistically independent of the input signal sequence* $\{x_n\}$ *and both sequences have zero mean.*

A.2 *The noise* w_n *has zero odd moments.*

A.3 *The weight error vector,* $\mathbf{v}_n \triangleq \mathbf{c}_n - \mathbf{c}_{opt}$*, is independent of the input* $\mathbf{x}_n$*.*

The proposed algorithm for recursively adjusting the coefficients of the system is expressed in the following form:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + 2\mu e_n^3 \mathbf{x}_n, \tag{2}$$

where μ is the step size.
To study the convergence of the algorithm in the mean-square, let $\mathbf{K}_n = E[\mathbf{v}_n \mathbf{v}_n^T]$ be the weight-error correlation matrix, where

$$\begin{aligned}\mathbf{v}_{n+1}\mathbf{v}_{n+1}^T \;=\; & \{\mathbf{v}_n\mathbf{v}_n^T - 6\mu w_n^2 \mathbf{x}_n\mathbf{x}_n^T\mathbf{v}_n\mathbf{v}_n^T\} \\ & \times\{\mathbf{I} - 6\mu w_n^2 \mathbf{x}_n\mathbf{x}_n^T\} \\ & +2\mu\{\mathbf{v}_n - 6\mu w_n^2 \mathbf{x}_n\mathbf{x}_n^T\mathbf{v}_n\}\mathbf{x}_n w_n^3 \\ & +2\mu w_n^3 \mathbf{x}_n\mathbf{v}_n^T\{\mathbf{I} - 6\mu w_n^2\mathbf{x}_n\mathbf{x}_n^T\} \\ & +4\mu^2 w_n^6 \mathbf{x}_n\mathbf{x}_n^T. \end{aligned} \tag{3}$$

Therefore, it can be shown that the weight-error correlation matrix is governed by the following recursion:

$$\begin{aligned}\mathbf{K}_{n+1} \;=\; & \mathbf{K}_n - 6\mu\sigma_w^2[\mathbf{R}\mathbf{K}_n + \mathbf{K}_n\mathbf{R}] \\ & +36\mu^2\chi_w^4[2\mathbf{R}\mathbf{K}_n\mathbf{R} + \mathbf{R}tr\{\mathbf{R}\mathbf{K}_n\}] \\ & +4\mu^2\phi_w^6\mathbf{R}, \end{aligned} \tag{4}$$

where σ_w^2, χ_w^4 and ϕ_w^6 are the noise power, the fourth- and the sixth-order moments of the noise, respectively, $\mathbf{R} = E[\mathbf{x}_n\mathbf{x}_n^T]$ is the autocorrelation matrix of the input signal, and $tr\{\}$ denotes trace operation.

It is assumed that the input autocorrelation matrix, $\mathbf{R}$, is positive definite with eigenvalues, $\lambda_i' s$ (λ_i is the i^{th} eigenvalue). Hence, it can be factorized as $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, where $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues, $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_N)$, and $\mathbf{Q}$ is the orthonormal matrix whose i^{th} column is the eigenvector of $\mathbf{R}$ associated with the i^{th} eigenvalue, that is, $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$. This results in $\mathbf{G}_n = \mathbf{Q}^T\mathbf{K}_n\mathbf{Q}$. Hence Equation (4) will take the following form:

$$\begin{aligned}\mathbf{G}_{n+1} \;=\; & \mathbf{G}_n - 6\mu\sigma_w^2[\mathbf{\Lambda}\mathbf{G}_n + \mathbf{G}_n\mathbf{\Lambda}] \\ & +36\mu^2\chi_w^4[2\mathbf{\Lambda}\mathbf{G}_n\mathbf{\Lambda} + \mathbf{\Lambda}tr\{\mathbf{\Lambda}\mathbf{G}_n\}] \\ & +4\mu^2\phi_w^6\mathbf{\Lambda}. \end{aligned} \tag{5}$$

Let $\mathbf{h}_n$ be a vector whose entries are the diagonal elements of $\mathbf{G}_n$, that is $\mathbf{h}_n^i = \mathbf{G}_n^{i,i}, i = 1, 2, \ldots N$, and let $\mathbf{\Gamma} = [\lambda_1, \lambda_2, \ldots, \lambda_N]^T$. Consequently, Equation (5) is transformed into the following form:

$$\mathbf{h}_{n+1} = \mathbf{A}\mathbf{h}_n + 4\mu^2\phi_w^6\mathbf{\Gamma} \tag{6}$$

with

$$\mathbf{A} = diag(\rho_1, \rho_2, \ldots, \rho_N) + 36\mu^2\chi_w^4\mathbf{\Gamma}\mathbf{\Gamma}^T, \tag{7}$$

and

$$\rho_i = 1 - 12\mu\sigma_w^2\lambda_i + 72\mu^2\chi_w^4\lambda_i^2, \qquad i = 1, \ldots N. \tag{8}$$

The convergence of (6) depends on $\mathbf{A}$. This will converge if and only if the eigenvalues of $\mathbf{A}$, i.e. the solutions of (9), lie within the unit circle:

$$det\,[\mathbf{A} - \gamma\mathbf{I}] = 0, \tag{9}$$

where $det\,[\mathbf{Z}]$ is the determinant of matrix $\mathbf{Z}$.

The determinant of $[\mathbf{A} - \gamma\mathbf{I}]$ can be shown to have the following form [10]:

$$\begin{aligned} det\,[\mathbf{A} - \gamma\mathbf{I}] \;=\; & \left[\Pi_{i=1}^N(\rho_i - \gamma)\right] \\ & \times\left[1 + 36\mu^2\chi_w^4\sum_{i=1}^N \frac{\lambda_i^2}{\rho_i - \gamma}\right]. \end{aligned} \tag{10}$$

Following the approach of [11], it can be shown that necessary and sufficient conditions for the roots of $det\,[\mathbf{A} - \gamma\mathbf{I}]$ to be inside the unit circle are:

$$\rho_i < 1, \qquad i = 1, 2, \ldots, N \tag{11}$$

and

$$1 + 36\mu^2\chi_w^4\sum_{i=1}^N \frac{\lambda_i^2}{\rho_i - 1} > 0. \tag{12}$$

Consequently, inequality (11) yields the following condition:

$$\mu\lambda_i[6\mu\chi_w^4\lambda_i - \sigma_w^2] < 0, \qquad i = 1, 2, \ldots, N. \tag{13}$$

Since the step-size parameter μ is a positive quantity and λ_i's are positive values (since $\mathbf{R}$ is positive definite) then inequality (13) leads to the following range for the step-size μ:

$$0 < \mu < \frac{\sigma_w^2}{6\chi_w^4\lambda_i}, \qquad i = 1, 2, \ldots, N. \tag{14}$$

Also, inequality (12) leads to a second condition on μ for convergence in the mean-square sense:

$$\sum_{i=1}^N \frac{3\mu\chi_w^4\lambda_i}{\sigma_w^2 - 6\mu\chi_w^4\lambda_i} < 1. \tag{15}$$

It is of important practical interest to translate conditions (14) and (15) into direct bounds on the step-size parameter μ. First, observe that the left-hand side of (15) is a strictly monotonically-increasing function of the step-size μ and is equal to zero for $\mu = 0$. Hence, if we let μ_i, $i = 1, 2, \ldots, N$, denote the solutions of the following equation:

$$\sum_{i=1}^{N} \frac{3\mu\chi_w^4\lambda_i}{\sigma_w^2 - 6\mu\chi_w^4\lambda_i} = 1, \tag{16}$$

assuming that the λ_i's are arranged in an increasing order, i.e. $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_N$, we have then:

$$0 < \mu_1 < \frac{\sigma_w^2}{6\chi_w^4\lambda_N} < \mu_2 < \frac{\sigma_w^2}{6\chi_w^4\lambda_{N-1}} < \ldots < \mu_N < \frac{\sigma_w^2}{6\chi_w^4\lambda_1}. \tag{17}$$

Therefore, for the conditions in (14) and (15) to hold, μ should be bounded by:

$$0 < \mu < \mu_1. \tag{18}$$

A closed-form expression for μ_1 cannot be found. However, following the analytical considerations outlined below, a tight lower bound on μ_1 is obtained.

Let $\mu_1 \le \mu_2 \le \mu_3 \le \ldots \le \mu_N$ be the soultion of (16). Our objective is to set a lower bound on μ_1. To do that, let us rewrite (16) in the following form [11]:

$$\begin{aligned}\Pi_{i=1}^{N}\left(\frac{1}{\mu} - \frac{1}{\mu_i}\right) &= \left(\frac{1}{\mu}\right)^N - b_1\left(\frac{1}{\mu}\right)^{N-1} \\ &\quad + b_2\left(\frac{1}{\mu}\right)^{N-2} + \ldots + (-1)^N b_N \\ &= 0,\end{aligned} \tag{19}$$

where by comparison of similar terms in (16) and (19), one finds:

$$\begin{cases} b_1 = \sum_{i=1}^{N} \frac{1}{\mu_i} = \frac{9\chi_w^4}{\sigma_w^2}\sum_{i=1}^{N}\lambda_i \\ b_2 = \sum_{i\ne j}^{N}\sum^{N} \frac{1}{\mu_i}\frac{1}{\mu_j} = 72\left(\frac{\chi_w^4}{\sigma_w^2}\right)^2 \sum_{i\ne j}^{N}\sum^{N}\lambda_i\lambda_j. \end{cases} \tag{20}$$

A theorem established in [12] asserts that the smallest root μ_1 is lower bounded by:

$$\mu_1 \ge \frac{N}{S_1 + \sqrt{(N-1)(NS_2 - S_1^2)}}, \tag{21}$$

where

$$S_1 = \sum_{i=1}^{N}\frac{1}{\mu_i} \tag{22}$$

and

$$S_2 = \sum_{i=1}^{N}\left(\frac{1}{\mu_i}\right)^2 = \left(\sum_{i=1}^{N}\frac{1}{\mu_i}\right)^2 - \sum_{i\ne j}^{N}\sum^{N}\frac{1}{\mu_i}\frac{1}{\mu_j}. \tag{23}$$

Consequently, using (20), the values of S_1 and S_2 in terms of b_1 and b_2:

$$\begin{cases} S_1 = b_1 \\ S_2 = b_1^2 - 2b_2. \end{cases} \tag{24}$$

Substituting (24) in (21) yields in the following:

$$\mu_1 \ge \frac{N}{b_1 + \sqrt{b_1(N-1)^2 - 2b_2N(N-1)}} = \mu^*. \tag{25}$$

Thus, to ensure convergence in the mean square, μ should be bounded by

$$0 < \mu < \mu^*, \tag{26}$$

and to make the above range more practical, we note that

$$\mu \ge \frac{1}{b_1}. \tag{27}$$

Then, to ensure convergence in the mean square, μ should be bounded by:

$$0 < \mu < \frac{\sigma_w^2}{9\chi_w^4\sum_{i=1}^{N}\lambda_i}, \tag{28}$$

which is identical to that found in [6], even though it has resulted from an analysis which is totally different from that of [6].

3 The Excess Steady-State MSE

The excess steady-state MSE of the LMF algorithm can be calculated by evaluating the misadjustment factor. The quantity $E[(\mathbf{x}_n^T\mathbf{v}_n)^2]$ represents the excess MSE, i.e.:

$$\begin{aligned}\zeta_{excess} &= E[(\mathbf{x}_n^T\mathbf{v}_n)^2] \\ &= \sum_{i=1}^{N}\lambda_i g_n^i,\end{aligned} \tag{29}$$

where g_n^i is obtained directly from (5):

$$\begin{aligned}g_{n+1}^i &= g_n^i - 12\mu\sigma_w^2\lambda_i g_n^i \\ &\quad + 36\mu^2\chi_w^4\left[2\lambda_i^2 g_n^i + \lambda_i\sum_{i=1}^{N}\lambda_i g_n^i\right] \\ &\quad + 4\mu^2\phi_w^6\lambda_i.\end{aligned} \tag{30}$$

Consequently, the excess steady-state MSE for the LMF algorithm is given by:

$$\zeta_{excess} = \frac{\sum_{i=1}^{N} \frac{\mu\lambda_i \phi_w^6}{3\sigma_w^2 - 9\mu\lambda_i\chi_w^4}}{1 - 9\sum_{i=1}^{N} \frac{\mu\lambda_i\chi_w^4}{3\sigma_w^2 - 9\mu\lambda_i\chi_w^4}}. \quad (31)$$

Finally, sufficient conditions for the convergence of the LMF algorithm are obtained as:

$$\begin{cases} 0 < \mu < \frac{\sigma_w^2}{6\chi_w^4\lambda_{max}} \\ \sum_{i=1}^{N} \frac{3\mu\lambda_i\chi_w^4}{\sigma_w^2 - 6\mu\lambda_i\chi_w^4} < 1. \end{cases} \quad (32)$$

Remarks:

1. Note that for small values of the step size μ, expression (31) can be approximated by:

$$\zeta_{excess} \simeq \frac{1}{3}\mu\frac{\phi_w^6}{\sigma_w^2}\sum_{i=1}^{N}\lambda_i. \quad (33)$$

2. Remark that (33), as obtained in [6], is a special case (for small μ) of our own expression of (31). Therefore, our analysis resulted in a more general expression for the excess steady-state MSE for the LMF algorithm than has been found in [6].

3. Moreover, our analysis resulted in a new and more general sufficient conditions for the LMF algorithm as given by (32).

4. For the case when the step-size parameter μ is small compared to $\frac{\sigma_w^2}{6\lambda_{max}\chi_w^4}$, condition $\sum_{i=1}^{N} \frac{3\mu\lambda_i\chi_w^4}{\sigma_w^2 - 6\mu\lambda_i\chi_w^4} < 1$ in (32) may be simplified as follows:

$$0 < \mu < \frac{\sigma_w^2}{3\chi_w^4 \sum_{i=1}^{N}\lambda_i}. \quad (34)$$

5. If one compares (28) and (34), (28) results in a lower range than that of (34). This should be expected since (34) is an approximation.

4 Conclusions

The analysis presented here resulted in a more general expression for the excess steady-state MSE as well as new sufficient conditions for convergence of the LMF algorithm.

Acknowledgment The author acknowledges the support of KFUPM.

References

[1] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1991.

[2] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust. Speech and Signal Process.*, 30(4):578–586, Aug. 1982.

[3] S. C. Douglas and T. H.-Y. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Trans. Signal Process.*, 42(6):1352–1365, June 1994.

[4] T. Y. Al-Naffouri, A. Zerguine, and M. Bettayeb, "A unifying view of error nonlinearities in lms adaptation," *Proceedings of the 1998 IEEE ICASSP*, 3:1697–1700, 1998.

[5] H. Sari, "Performance evaluation of three adaptive equalization algorithms," *in Proc. ICASSP'82*, pages 1385–1388, 1982.

[6] E. Walach and B. Widrow, "The Least Mean Fourth (LMF) Adaptive Algorithm and its Family," *IEEE Trans. Inf. Theory, vol.* 30, pp. 275-283, Feb. 1984.

[7] O. Tanrikulu and J. A. Chambers, " Convergence and steady-state properties of the least mean mixed-norm (lmmn) adaptive filtering," *IEE Proc.-VIS Image Signal Process.*, 143(3):137–142, June 1996.

[8] A. Zerguine, M. Bettayeb, and C. F. N. Cowan, "Hybrid lms-lmf scheme for echo cancellation," *IEE Proc.-VIS Image Signal Process.*, 146(4):173–180, August, 1999.

[9] A. Zerguine, C. F. N. Cowan, and M. Bettayeb, "Adaptive Echo Cancellation using Least Mean Mixed-Norm Algorithm," *IEEE Trans. Signal Processing, vol.* 45, pp. 1340-1343, May 1997.

[10] J. B. Foley and F. M. Boland, "A Note on the Convergence Analysis of LMS Adaptive Filters with Gaussian Data," *IEEE Trans. Signal Processing, vol.* 36, pp. 1087-1089, July 1988.

[11] A. Feuer and E. Weinstein, "Convergence Analysis of LMS Filters with Uncorrelated Gaussian Data," *IEEE Trans. Acoust., Speech, Signal Processing, vol.* 33, pp. 222-229, Feb. 1985.

[12] D. L. Jagerman, "Nonstationary Blocking in Telephone Traffic," *Bell Syst. Tech. Journal, vol.* 54, pp. 625-661, March 1975.