USC–SIPI REPORT #280

New Results on the Mean and Mean Square Convergence of LMS

by

Esmat Bekir and C.-C. Jay Kuo

April 1995

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

New Results on the Mean and Mean Square Convergence of LMS*

Esmat Bekir† and C.-C. Jay Kuo†

April 3, 1995, EDICS: SPL.SP.2.6

Abstract

New results on the mean and mean square convergence of the LMS algorithm are presented. First, a new approach is proposed for the derivation of the necessary and sufficient conditions for mean square convergence. Second, we derive an explicit expression on the bound of the step size μ and show an interesting relationship between the mean and mean square convergence.

I. Introduction

The LMS (Least-Mean-Square) algorithm has been extensively used in many adaptive signal processing applications due to its striking simplicity. One key issue on the application of LMS is its convergence property. Let us use $u(n)$, $y(n)$ and $d(n)$ to denote the filter input, output and desired output, respectively, and assume that $u(n)$ is a zero-mean sequence. Let $\mathbf{w}(n) = (w_0(n), w_1(n), \cdots w_{M-1}(n))^T$ be the tap-weight vector of size M and $\mathbf{u}(n) = (u(n), u(n-1), \cdots u(n-M+1))^T$ be the tap-input vector. With proper initialization (usually, $\mathbf{w}(1) = \mathbf{0}$), the LMS algorithm can written as [2]:

$$\begin{aligned} y(n) &= \mathbf{w}^H(n)\mathbf{u}(n), \\ e(n) &= d(n) - y(n), \\ \mathbf{w}(n+1) &= \mathbf{w}(n) + \mu\mathbf{u}(n)e^*(n), \end{aligned}$$

for $n = 1, 2, 3, \cdots$, and where μ is called the step size parameter. The LMS algorithm is said to be convergent in the mean, if the mean $E[\mathbf{w}(n)]$ of the tap-weight vector converges to the optimum Wiener filter solution. It is convergent in the mean square, if the mean-squared error $J(n) = E[|e(n)|^2]$ converges to a finite steady-state value $J(\infty)$.

The mean and mean square convergence properties of LMS are basically determined by the step size parameter μ and the eigenvalue distribution of the correlation matrix of the tap-input vector, i.e. $\mathbf{R} = E[\mathbf{u}(n)\mathbf{u}^H(n)]$. Conditions have been derived to guarantee the mean and mean square convergence of LMS

*This work was supported by the National Science Foundation Presidential Faculty Fellow Award ASC-9350309.

†The authors are with the Signal and Image Processing Institute and the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, California 90089-2564. E-mail: bekir@aludra.usc.edu and cckuo@sipi.usc.edu.

[2], [3]. Necessary and sufficient conditions for the mean square convergence are:

$$0 < \mu < \frac{2}{\lambda_{\max}}, \tag{1}$$

$$\sum_{i=1}^{M} \frac{\mu\lambda_i}{2 - \mu\lambda_i} < 1, \tag{2}$$

where λ_i, $1 \leq i \leq M$, are the eigenvalues of $\mathbf{R}$ and $\lambda_{\max}$ is the largest one[2] among them. Condition (1) alone is necessary and sufficient for the mean convergence.

New results on the mean and mean square convergence of LMS will be presented in this work. First, we propose a new approach to derive (1) and (2) for the mean square convergence in Section II. In contrast with the classical approach adopted by Ungerboeck [3], our derivation appears to be more direct and transparent, since it does not rely on the Perron-Frobenius Theorem [4]. Second, condition (2) is implicit so that it does not lend itself for easy understanding and practical design. We study a new condition which gives an explicit expression on the bound of step size μ in Section III. This new result provides an interesting link between the mean and mean square convergence.

II. Conditions for Mean Square Convergence

We can express $J(n)$ by [2]

$$J(n) = J_{\min} + J_{\text{ex}}(n),$$

where $J_{\min}$ is the minimum mean-squared error produced by the Wiener filter and

$$J_{\text{ex}}(n) = \boldsymbol{\lambda}^T \mathbf{x}(n), \quad \boldsymbol{\lambda}^T = (\lambda_1, \lambda_2, \cdots, \lambda_M), \tag{3}$$

and $\mathbf{x}(n)$ is the state vector satisfying the following recursion

$$\mathbf{x}(n+1) = \mathbf{B}\mathbf{x}(n) + \mu^2 J_{\min} \boldsymbol{\lambda}. \tag{4}$$

The $\mathbf{B}$ in (4) is a diagonal-plus-dyad matrix, i.e.

$$\mathbf{B} = \mathbf{D} + \mu^2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T, \tag{5}$$

where

$$\mathbf{D} = \text{diag}[d_1, d_2, \cdots d_M], \quad \text{and} \quad d_i = (1 - \mu\lambda_i)^2.$$

The convergence of $J(n)$ is equivalent to the convergence of iteration (4). This is true if and only if that the magnitude of all eigenvalues of $\mathbf{B}$ is less than unity. In the following, we use γ and $\mathbf{y}$ to denote an arbitrary eigenvalue-eigenvector pair for $\mathbf{B}$. Since $\mathbf{B}$ is symmetric and positive definite, γ is real and positive. Thus, only $\gamma < 1$ is required for (4) to converge.

First, by imposing $\gamma < 1$, we have

$$1 > \gamma > d_i = (1 - \mu\lambda_i)^2, \quad 1 \leq i \leq M, \tag{6}$$

where the last inequality is due to the matrix ordering property applied to (5) (see [1]). By expanding the quadratic term, the above inequality can be further simplified to

$$\mu < \frac{2}{\lambda_{\max}}. \tag{7}$$

which is condition (1). Next, by using the definition of $\mathbf{B}$, we have

$$\mathbf{D}\mathbf{y} + \mu^2(\boldsymbol{\lambda}^T\mathbf{y})\boldsymbol{\lambda} = \gamma\mathbf{y}. \tag{8}$$

which can be rearranged to

$$\mathbf{y} = \mu^2(\boldsymbol{\lambda}^T\mathbf{y})(\gamma\mathbf{I} - \mathbf{D})^{-1}\boldsymbol{\lambda}. \tag{9}$$

Premultiplying (9) by $\boldsymbol{\lambda}^T$ and using the definitions of $\boldsymbol{\lambda}$ and $\mathbf{D}$ gives

$$\mu^2 \sum_{i=1}^{M} \frac{\lambda_i}{\gamma - (1 - \mu\lambda_i)^2} = 1. \tag{10}$$

With (10), the constraint $\gamma < 1$ is equivalent to

$$1 = \mu^2 \sum_{i=1}^{M} \frac{\lambda_i}{\gamma - (1 - \mu\lambda_i)^2} > \mu^2 \sum_{i=1}^{M} \frac{\lambda_i}{1 - (1 - \mu\lambda_i)^2} = \sum_{i=1}^{M} \frac{\mu\lambda_i}{2 - \mu\lambda_i}, \tag{11}$$

which is condition (2).

Following the above discussion, an expression for the ratio of $J_{\text{ex}}(\infty)$ and $J_{\min}$, called the misadjustment $\mathcal{M}$, can also be easily obtained. Based on (3) and (5), we have

$$\mathcal{M} = \frac{J_{\text{ex}}(\infty)}{J_{\min}} = \mu^2\boldsymbol{\lambda}^T(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\lambda}. \tag{12}$$

Note that

$$(\mathbf{I} - \mathbf{B})^{-1} = (\tilde{\mathbf{D}} - \mu^2\boldsymbol{\lambda}\boldsymbol{\lambda}^T)^{-1},$$

where

$$\tilde{\mathbf{D}} = \mathbf{I} - \mathbf{D} = \text{diag}(\tilde{d}_1, \tilde{d}_2, \cdots \tilde{d}_M),$$

and where

$$\tilde{d}_i = 1 - (1 - \mu\lambda_i)^2 = 2\mu\lambda_i - \mu^2\lambda_i^2 > 0. \tag{13}$$

The last inequality is due to (6). By using the matrix inversion formula derived in Appendix, we have

$$(\mathbf{I} - \mathbf{B})^{-1} = \tilde{\mathbf{D}}^{-1} - \beta\tilde{\mathbf{D}}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^T\tilde{\mathbf{D}}^{-1}, \quad \beta = \frac{-\mu^2}{1 - \mu^2\boldsymbol{\lambda}^T\tilde{\mathbf{D}}^{-1}\boldsymbol{\lambda}}.$$

As shown in (12), we are interested in

$$\begin{aligned} \mu^2\boldsymbol{\lambda}^T(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\lambda} &= \mu^2\boldsymbol{\lambda}^T\tilde{\mathbf{D}}^{-1}\boldsymbol{\lambda} - \mu^2\beta(\boldsymbol{\lambda}^T\tilde{\mathbf{D}}^{-1}\boldsymbol{\lambda})^2 \\ &= \frac{\mu^2\boldsymbol{\lambda}^T\tilde{\mathbf{D}}^{-1}\boldsymbol{\lambda}}{1 - \mu^2\boldsymbol{\lambda}^T\tilde{\mathbf{D}}^{-1}\boldsymbol{\lambda}}. \end{aligned} \tag{14}$$

The common term in the numerator and denominator in (14) is

$$\mu^2 \boldsymbol{\lambda}^T \tilde{\mathbf{D}}^{-1} \boldsymbol{\lambda} = \sum_{i=1}^{M} \frac{\mu^2 \lambda_i^2}{2\mu\lambda_i - \mu^2 \lambda_i^2} = \sum_{i=1}^{M} \frac{\mu\lambda_i}{2 - \mu\lambda_i}. \tag{15}$$

The desired result for the misadjustment can be obtained by substituting (15) into (14), i.e.

$$\mathcal{M} = \frac{\sum_{i=1}^{M} \mu\lambda_i/(2 - \mu\lambda_i)}{1 - \sum_{i=1}^{M} \mu\lambda_i/(2 - \mu\lambda_i)}.$$

III. Explicit Condition on Step Size Selection

We can rewrite the left-hand-side of (2) as

$$\sum_{i=1}^{M} \frac{\mu\lambda_i}{2 - \mu\lambda_i} = -M + \sum_{i=1}^{M} \frac{1}{1 - \mu'\lambda_i}, \quad \mu' \equiv \frac{\mu}{2}. \tag{16}$$

We know from (13) that

$$1 - \mu'\lambda_i > 0, \quad 1 \le i \le M. \tag{17}$$

By using (17) and the Hölders inequality

$$(a_1 + a_2 + \cdots + a_M)(\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_M}) \ge M^2, \quad a_1, a_2, \cdots, a_M > 0,$$

we have

$$\sum_{i=1}^{M} \frac{1}{1 - \mu'\lambda_i} \ge \frac{M^2}{\sum_{i=1}^{M} 1 - \mu'\lambda_i} = \frac{M^2}{M - \mu' \sum_{i=1}^{M} \lambda_i} = \frac{M}{1 - \mu'\lambda_{\text{ave}}}, \tag{18}$$

where $\lambda_{\text{ave}} = \sum_{i=1}^{M} \lambda_i / M$. By combining results in (2), (16) and (18), it is straightforward to show that

$$\mu = 2\mu' < \frac{2}{(M+1)\lambda_{\text{ave}}}, \tag{19}$$

which is the desired explicit condition on step size μ. Therefore, conditions (19) and (1) can be integrated into one as

$$0 < \mu < \min\left(\frac{2}{\lambda_{\max}}, \frac{2}{(M+1)\lambda_{\text{ave}}}\right) \tag{20}$$

It is shown in [2] that it is sufficient with (1) for the mean converge while both (1) and (2) have to be satisfied for the mean square convergence. However, based on (20), we see that if

$$\frac{2}{\lambda_{\max}} \le \frac{2}{(M+1)\lambda_{\text{ave}}}, \tag{21}$$

the convergence in the mean in fact implies the convergence in the mean square. Furthermore, Equation (21) can be rearranged as

$$M \le \frac{\lambda_{\max} - \lambda_{\text{ave}}}{\lambda_{\text{ave}}}, \tag{22}$$

which has some interesting consequences.

To provide insights into (22), let us consider two cases. For the first case, we fix the right-hand-side of (22). For an adaptive filter with a shorter impulse response (or smaller M), the constraint $\mu < 2/\lambda_{\max}$ is sufficient for both mean and mean square convergences. Note that the step size μ is independent of the filter length M in this case. On the other hand, if the filter length is long enough, we have to impose the constraint $\mu < 2/[(M+1)\lambda_{\text{ave}}]$. The longer the filter length is, the smaller the step size to guarantee the mean square convergence.

For the second case, the value of M is fixed. When the eigenvalues of $\mathbf{R}$ are clustered (e.g. low SNR), the right-hand-side of (22) tends to be small and it is likely that $M > (\lambda_{\max} - \lambda_{\text{ave}})/\lambda_{\text{ave}}$. Convergence in the mean does not guarantee convergence in the mean square since data are too noisy. On the other hand, when the eigenvalues of $\mathbf{R}$ are dispersed (e.g. high SNR), Equation (22) can be more easily satisfied. Then, convergence in the mean does guarantee convergence in the mean square, and we only have to require $\mu < 2/\lambda_{\max}$ in selecting the step size μ.

IV. Conclusion

We provided a simple approach to derive the necessary and sufficient conditions for the step size selection of the mean square convergence of LMS. Then, we converted one explicit condition to an explicit one, and explained its physical meaning.

Appendix

In this appendix, we derive an inversion formula for a diagonal-plus-dyad matrix. Let $\mathbf{V}$ be a square matrix of size $M \times M$ and $\mathbf{a}$ denote a vector of length M. If

$$\mathbf{V} = \mathbf{I} + \alpha \mathbf{a}\mathbf{a}^H,$$

where $\mathbf{I}$ is the unity matrix, it is easy to show that

$$\mathbf{V}^{-1} = \mathbf{I} - \beta \mathbf{a}\mathbf{a}^H, \quad \text{where} \quad \beta = \frac{\alpha}{1 + \alpha \mathbf{a}^H \mathbf{a}}.$$

The above result can be generalized to the following case. If

$$\mathbf{V} = \mathbf{D} + \alpha \mathbf{a}\mathbf{a}^H,$$

where $\mathbf{D}$ a nonsingular diagonal matrix, then we have

$$\mathbf{V}^{-1} = \mathbf{D}^{-1} - \beta \mathbf{D}^{-1}\mathbf{a}\mathbf{a}^H \mathbf{D}^{-1}, \quad \text{where} \quad \beta = \frac{\alpha}{1 + \alpha \mathbf{a}^H \mathbf{D}^{-1} \mathbf{a}}.$$

References

[1] R. Bellman, *Introduction to Matrix Analysis*, MaGraw Hill, 1970.

[2] S. Haykin, *Adaptive Filter Theory*, 2nd Edition, Prentice Hall, 1991.

[3] G. Ungerboeck, "Theory on the speed of convergence in adaptive equalizers for digital communication," *IBM J. Res. Dev.*, Vol. 16, pp. 546-555.

[4] R. S. Varga, *Matrix Iterative Analysis*, Prentice Hall, 1962.