Analysis of the LMS Algorithm Family for Uncorrelated Gaussian Data

*Seung Hyon Nam and **Insung Lee

Abstract

In this paper, convergence properties of the LMS, LMF, and LVCMS algorithms are investigated under the assumption of the uncorrelated Gaussian input data. By treating these algorithms as special cases of more general algorithm family, unified results on these algorithms are obtained. First the upper bound on the step size parameter is obtained. Second, an expression for misadjustment is obtained. These theoretical results confirm earlier LMS works. Further, the results explain why the LMF and LVCMS algorithms are experiencing difficulties with plant noise having heavier tailed densities. Simulation results agree with theoretical expectation closely for various plant noise statistics.

I. Introduction

The least variance subject to a constraint on the mean square error (LVCMS) adaptive filtering algorithm was introduced by Gibson and Gray in [5]. The LVCMS algorithm is motivated by the steepest descent method like the LMS and LMF algorithms [3, 4]. In [5], the convergence in the mean coefficient error of the LVCMS algorithm was analyzed without assumptions on the density function of the input data as in [2, 3] for the LMS algorithm and in [4] for the LMF algorithm. The resulting upper bound on the step size parameter is quite loose and the actual step size should be chosen much smaller than the upper bound to ensure the convergence of the MSE in practice. In [7], however, the convergence in the covariance matrix of the coefficient error vector was studied for the complex and real LMS algorithms using the Gaussian assumption on the input data vector. The resulting upper bound on the step size is much tighter than the previous bound in [3]. Further, Feuer and Weinstein showed the same result using a different approach, and derived new expressions for the misadjustment and the rate of convergence [8].

In this study, the LMS, LMF, and LVCMS algorithms are considered as special cases of a more general adaptive algorithm [6], and the convergence analysis of the general algorithm is given using a Gaussian assumption on the input data vector. The tighter upper bounds so obtained explain earlier simulation results on the LMF and LVCMS algorithms in [5], which indicate that these algorithms are sensitive to noise with a heavier-tailed density. Furthermore, a new expression on the misadjustment is obtained. Simulations for the system identification problem are given to support the theoretical results.

*Division of Computer, Electronics, and Information Communications Engineering PaiChai University
** Department of Radio Engineering Chungbuk National University

Manuscript Received May 18, 1996.

II. The Generalized Error Criterion

The adaptive signal processing configuration of interest is depicted in Fig. 1. The input data vector at time k is given by $\mathbf{X}(k)=[x_1(k), x_2(k), \cdots, x_N(k)]^T$, and the $\mathbf{X}(k)$, $k=0, 1, 2, \cdots$, are assumed to be uncorrelated. The error signal at time k is given by

$$\begin{aligned}\epsilon(k) &= d(k)-\mathbf{W}^T(k)\mathbf{X}(k)\\ &= n(k)-(\mathbf{W}(k)-\mathbf{W}^*)^T\mathbf{X}(k)\\ &= n(k)-\mathbf{V}^T(k)\mathbf{X}(k)\end{aligned} \tag{1}$$

where $\mathbf{W}^*$ is the vector of optimal coefficients, $\mathbf{W}(k)$ is the coefficient vector at time k so that

$$\mathbf{V}(k)=\mathbf{W}(k)-\mathbf{W}^* \tag{2}$$

is the coefficient error vector at time k, and $n(k)$ is the noise such that

$$n(k)=d(k)-\mathbf{X}^T(k)\mathbf{W}^*.$$

The noise $n(k)$ is assumed to be white and to have a symmetrical probability density function with zero mean and finite higher order moments. Further, the noise $n(k)$ is assumed to be independent of the input data vector $\mathbf{X}(k)$.

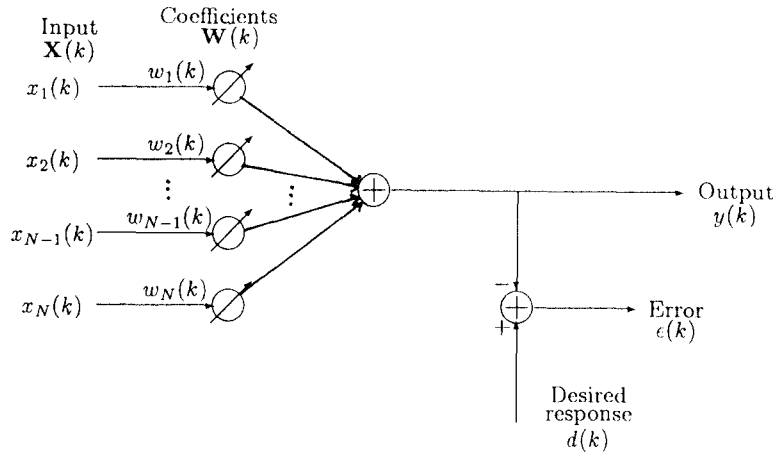

Figure 1. Adaptive Signal Processing Configuration.

The method of steepest descent uses gradients of a function of the error $\epsilon(k)$ to find the optimum coefficient vector $\mathbf{W}^*$. The function or error criterion characterizes the adaptive algorithm. In [6], a general error criterion

$$H_\theta(\mathbf{W}, a, b, c, d) = aE\{|\epsilon^2(k) - E(\epsilon^2(k))|^2\} + b[E\{\epsilon^2(k)\}]^2 + cE\{\epsilon^2(k)\} + d \quad (4)$$

is considered. This admits the LMS[2, 3], LMF[4], and LVCMS[5] criteria for appropriate selections of a, b, c, and d. That is,

$$H_\theta(\mathbf{W}, 0, 0, 1, 0) = E\{\epsilon^2(k)\}, \quad (5)$$

$$H_\theta(\mathbf{W}, 1, 1, 0, 0) = E\{\epsilon^4(k)\}, \quad (6)$$

and

$$H_\theta(\mathbf{W}, 1, 0, -\lambda, \lambda\sigma_v^2) = E\{|\epsilon^2(k) - E(\epsilon^2(k))|^2\} + \lambda\{\sigma_v^2 - E(\epsilon^2(k))\}. \quad (7)$$

Taking partial derivative of $H_\theta(\mathbf{W}, a, b, c, d)$ with respect to $\mathbf{W}$ yields

$$\frac{\partial H_\theta(\mathbf{W}, a, b, c, d)}{\partial \mathbf{W}} = -4aE\{\epsilon^3(k)\mathbf{X}(k)\} - 4(b-a)E\{\epsilon^3(k)\} E\{\epsilon(k)\mathbf{X}(k)\} + 4cE\{\epsilon(k)\mathbf{X}(k)\}. \quad (8)$$

Thus, the instantaneous gradient-based coefficient adaptation rule which approximately minimizes H_θ in (4) is given by

$$\mathbf{W}(k+1) = \mathbf{W}(k) + 4\mu b\epsilon^3(k)\mathbf{X}(k) + 2\mu c\epsilon(k)\mathbf{X}(k). \quad (9)$$

We should note that only b and c are included in the generalized algorithm (9) since the instantaneous estimate of the gradient is used. The generalized algorithm in (9) corresponds to the LMS algorithm if $(b, c) = (0, 1)$, the LMF algorithm if $(b, c) = (1, 0)$, and the LVCMS algorithm if $(b, c) = (1, -(2\sigma_v + \lambda))$. In the sequel, the general adaptation rule (9) will be used in the analysis, and the results could be applied to the LMS, LMF, and LVCMS algorithms.

III. Convergence Analysis of the Generalized Algorithm

We now assume that the input data $\mathbf{X}(k)$ is Gaussian, so that it is also independent. Therefore, we can apply the independence assumptions widely used in stochastic analyses of the LMS type algorithm [?].

Subtract $\mathbf{W}^*$ from both sides of (9) to obtain the recursion for $\mathbf{V}(k)$,

$$\mathbf{V}(k+1) = \mathbf{V}(k) + 4\mu b\epsilon^3(k)\mathbf{X}(k) + 2\mu c\epsilon(k)\mathbf{X}(k), \quad (10)$$

and then use (1) for $\epsilon(k)$ to obtain

$$\mathbf{V}(k+1) = \mathbf{V}(k) + 4\mu b\mathbf{X}(k)\sum_{i=0}^{3}\binom{3}{i} n^i(k)(-\mathbf{X}^T(k)\mathbf{V}(k))^{3-i} + 2\mu c(n(k) - \mathbf{X}^T(k)\mathbf{V}(k))\mathbf{X}(k). \quad (11)$$

Since $\mathbf{R}$ is symmetric, one can define the unitary matrix $\mathbf{U}$ as

$$E\{\mathbf{U}\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{U}^T\} = \Gamma = diag(\gamma_1, \gamma_2, \cdots, \gamma_N). \quad (12)$$

Then

$$\mathbf{UV}(k+1)\mathbf{V}^T(k+1)\mathbf{U}^T = \mathbf{UV}(k)\mathbf{V}^T(k)\mathbf{U}^T$$

$$+16\mu^2 b^2 \left\{ \sum_{i=0}^{3} \binom{3}{i} n^i(k)(-\mathbf{X}^T(k)\mathbf{V}(k))^{3-i} \right\}^2 \mathbf{UX}(k)\mathbf{X}^T(k)\mathbf{U}^T$$

$$+4\mu^2 c^2 [n(k)-\mathbf{X}^T(k)\mathbf{V}(k)]^2 \mathbf{UX}(k)\mathbf{X}^T(k)\mathbf{U}^T$$

$$+4\mu b \left\{ \sum_{i=0}^{3} \binom{3}{i} n^i(k)(-\mathbf{X}^T(k)\mathbf{V}(k))^{3-i} \right\}$$

$$\{\mathbf{UX}(k)\mathbf{V}^T(k)\mathbf{U}^T + \mathbf{UV}(k)\mathbf{X}^T(k)\mathbf{U}^T\}$$

$$+4\mu c[n(k)-\mathbf{X}^T(k)\mathbf{V}(k)]\{\mathbf{UX}(k)\mathbf{V}^T(k)\mathbf{U}^T + \mathbf{UV}(k)\mathbf{X}^T(k)\mathbf{U}^T\}$$

$$+16\mu^2 bc \left\{ \sum_{i=0}^{3} \binom{3}{i} n^i(k)(-\mathbf{X}^T(k)\mathbf{V}(k))^{3-i} \right\}$$

$$[n(k)-\mathbf{X}^T(k)\mathbf{V}(k)]\mathbf{UX}(k)\mathbf{X}^T(k)\mathbf{U}^T. \tag{13}$$

Taking the expectation of both sides of (13) and assuming that the noise has a symmetric density function with zero mean and finite higher order moments, one finds

$$\begin{aligned} E\{\mathbf{UV}(k+1)\mathbf{V}^T(k+1)\mathbf{U}^T\} &= E\{\mathbf{UV}(k)\mathbf{V}^T(k)\mathbf{U}^T\} \\ &-2\mu\bar{a}E\{\mathbf{X}^T(k)\mathbf{V}(k)[\mathbf{UX}(k)\mathbf{V}^T(k)\mathbf{U}^T + \mathbf{UV}(k)\mathbf{X}^T(k)\mathbf{U}^T]\} \\ &+4\mu^2\bar{b}E\{(\mathbf{X}^T(k)\mathbf{V}(k))^2\mathbf{UX}(k)\mathbf{X}^T(k)\mathbf{U}^T\} \\ &+4\mu^2\bar{c}E\{\mathbf{UX}(k)\mathbf{X}^T(k)\mathbf{U}^T\} \\ &+16\mu^2\bar{d}E\{(\mathbf{X}^T(k)\mathbf{V}(k))^4\mathbf{UX}(k)\mathbf{X}^T(k)\mathbf{U}^T\} \\ &+16\mu^2 b^2 E\{(\mathbf{X}^T(k)\mathbf{V}(k))^6\mathbf{UX}(k)\mathbf{X}^T(k)\mathbf{U}^T\} \\ &-4\mu b E\{(\mathbf{X}^T(k)\mathbf{V}(k))^3[\mathbf{UX}(k)\mathbf{V}^T(k)\mathbf{U}^T + \mathbf{UV}(k)\mathbf{X}^T(k)\mathbf{U}^T]\} \end{aligned} \tag{14}$$

where

$$\bar{a} = 6bE(n^2(k)) + c \tag{15}$$

$$\bar{b} = 60b^2E(n^4(k)) + 24bcE(n^2(k)) + c^2 \tag{16}$$

$$\bar{c} = 4b^2E(n^6(k)) + 4bcE(n^4(k)) + c^2E(n^2(k)) \tag{17}$$

$$\bar{d} = 15b^2E(n^2(k)) + bc. \tag{18}$$

From the independence theory (see [12]), the noise $n(k)$ is independent of $\mathbf{X}(k)$ and $\mathbf{V}(k)$. Further, if we assume that $\mathbf{V}(k)$ is small enough, then we can neglect the terms of power greater than 2 in (14). A more general proof that includes all the higher order terms is omitted here for space. But the analysis shows that the higher order terms can be considered as an additional constant term to the recursion for $\mathbf{C}(k)$ below. The constant term does not affect the upper bound on μ.

Define

$$\mathbf{C}(k) = E\{\mathbf{UV}(k)\mathbf{V}^T(k)\mathbf{U}^T\}. \tag{19}$$

It can be shown that

$$E\{\mathbf{X}^T(k)\mathbf{V}(k)[\mathbf{UX}(k)\mathbf{V}^T(k)\mathbf{U}^T + \mathbf{UV}(k)\mathbf{X}^T(k)\mathbf{U}^T]\} = \boldsymbol{\Gamma}\mathbf{C}(k) + \mathbf{C}(k)\boldsymbol{\Gamma}. \tag{20}$$

and using the fourth moment expansion for Gaussian random variables, that

$$E\{(\mathbf{X}^T(k)\mathbf{V}(k))^2\mathbf{UV}(k)\mathbf{V}^T(k)\mathbf{U}^T\} = 2\boldsymbol{\Gamma}\mathbf{C}(k)\boldsymbol{\Gamma} + tr[\boldsymbol{\Gamma}\mathbf{C}(k)]\boldsymbol{\Gamma}. \tag{21}$$

Therefore, combining (12), (20), and (21), the recursion (14) is simplified to

$$\begin{aligned} \mathbf{C}(k+1) &= \mathbf{C}(k) - 2\mu\bar{a}\{\boldsymbol{\Gamma}\mathbf{C}(k) + \mathbf{C}(k)\boldsymbol{\Gamma}\} \\ &+4\mu^2 b\{2\boldsymbol{\Gamma}\mathbf{C}(k)\boldsymbol{\Gamma} + tr[\boldsymbol{\Gamma}\mathbf{C}(k)]\boldsymbol{\Gamma}\} + 4\mu^2\bar{c}\boldsymbol{\Gamma}. \end{aligned} \tag{22}$$

Following Feuer and Weinstein's LMS work [8], decompose the recursion for $\mathbf{C}(k)$ to obtain

$$C_{ii}(k+1) = \rho_{ii}C_{ii}(k) + 4\mu^2\bar{b}\gamma_i \sum_{p=1}^{N} \gamma_p C_{pp}(k) + 4\mu^2 c\gamma_i, \tag{23}$$

and

$$C_{ij}(k+1) = \rho_{ij}C_{ij}(k) \qquad \text{for } i \neq j \tag{24}$$

where

$$\rho_{ij} = 1 - 2\mu\bar{a}(\gamma_i + \gamma_j) + 8\mu^2 b\gamma_i\gamma_j. \tag{25}$$

Since the matrix $\mathbf{C}(k)$ is symmetric and positive definite, it is diagonally dominant [11]; that is, $C_{ij}^2(k) \leq C_{ii}(k)C_{jj}(k)$. Therefore, the convergence of the diagonal elements of $C(k)$ ensures the convergence of the off-diagonal elements.

Define

$$\bar{\mathbf{C}}(k) = [C_{11}(k)\, C_{22}(k) \cdots C_{NN}(k)]^T \tag{26}$$

and

$$\mathbf{F} = \mathrm{diag}(\rho_1\, \rho_2 \cdots \rho_N) + 4\mu^2\bar{b}\mathbf{L}\mathbf{L}^T \tag{27}$$

where $\rho_i = \rho_{ii}$ and

$$\mathbf{L} = [\gamma_1\, \gamma_2 \cdots \gamma_N]^T. \tag{28}$$

Then the diagonal term (23) can be rewritten as

$$\bar{\mathbf{C}}(k+1) = \mathbf{F}\bar{\mathbf{C}}(k) + 4\mu^2\bar{c}\mathbf{L}. \tag{29}$$

Hence, $\bar{\mathbf{C}}(k)$ converges if and only if all of the eigenvalues of **F** lie inside the unit circle.

Following Foley and Boland's LMS work [9], we use the property of nonnegative matrices stated by Gantmacher [10] as follows. (For an alternative approach, see [8].)

Theorem [9]: A necessary and sufficient condition that the real number q be greater than the dominant eigenvalue of the nonnegative matrix **F** is that all the leading principal minors of the characteristic matrix

$$\mathbf{F}_q = q\mathbf{I} - \mathbf{F}$$

be positive. □

From (25) and (27), we can see that, if

$$\bar{a}^2 < 3\bar{b} \tag{30}$$

all the elements of **F** are nonnegative. Further, **F** is irreducible for $\bar{b} \neq 0$ since **R** is assumed to be positive definite. Hence, there exists a dominant real eigenvalue by the Perron-Frobenius theorem [10], [11], and the theorem given above can be applied.

With $q = 1$, we investigate the principal minors of

$$\mathbf{F}_1 = 4\mu\bar{a}\,\mathbf{\Gamma} - 8\mu\bar{b}\,\mathbf{\Gamma}^2 - 4\mu^2\bar{b}\,\mathbf{L}\mathbf{L}^T. \tag{31}$$

First, consider the principal minor of order N. Let

$$\mathbf{D} = diag(d_1, d_2, \cdots, d_N) = 4\mu\bar{a}\,\mathbf{\Gamma} - 8\mu\bar{b}\,\mathbf{\Gamma}, \tag{32}$$

so that the principal minor of order N is

$$\begin{aligned}\Delta_N &= det(\mathbf{D} - 4\mu^2\bar{b}\,\mathbf{L}\mathbf{L}^T)\\ &= \prod_{i=1}^{N} d_i - 4\mu^2\bar{b}(\gamma_1^2 d_2 d_3 \cdots d_N) - 4\mu^2\bar{b}(d_1\gamma_2^2 d_3 \cdots d_N)\\ &\quad - 4\mu^2\bar{b}(d_1 d_2 \cdots d_{N-1}\gamma_N^2) + det(4\mu^2\bar{b}\,\mathbf{L}\mathbf{L}^T).\end{aligned} \tag{33}$$

But $det(\mathbf{L}\mathbf{L}^T) = 0$, therefore,

$$\Delta_N = det\,\mathbf{D} - 4\mu^2\bar{b}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})\,det\,\mathbf{D}. \tag{34}$$

Then using the definition of **D** and **L** in (32) and (28), respectively, we obtain

$$\Delta_N = (4\mu)^N \left\{ \prod_{i=1}^{N} \gamma_i(\bar{a} - 2\mu\bar{b}\gamma_i) \right\} \left\{ 1 - \sum_{i=1}^{N} \frac{\mu\bar{b}\gamma_i}{\bar{a} - 2\mu\bar{b}\gamma_i} \right\}. \tag{35}$$

Since $\mu > 0$, the necessary conditions for $\Delta_N > 0$ are given by

$$\mu < \frac{a}{2\bar{b}\gamma_i} \tag{36}$$

and

$$\sum_{i=1}^{N} \frac{\mu\bar{b}\gamma_i}{\bar{a} - 2\mu\bar{b}\gamma_i} < 1. \tag{37}$$

Also we can see that the principal minor of order less than N will give the conditions that is identical to (37) except the upper limit on summation.

To find an explicit bound on μ from (37), we follow Feuer and Weinstein's LMS work [8]. The new bound on μ is then given by

$$\mu \leq \frac{N}{(\frac{\bar{b}}{\bar{a}})\left\{ 3\sum_{i=1}^{N}\gamma_i + \sqrt{9(N-1)^2(\sum_{i=1}^{N}\gamma_i)^2 - 8(N-1)\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}\gamma_i\gamma_j} \right\}}, \tag{38}$$

which can be simplified to

$$\mu \leq \frac{\bar{a}}{3\bar{b}\sum_{i=1}^{N}\gamma_i} = \frac{\bar{a}}{3\bar{b}\,tr(\mathbf{R})} \tag{39}$$

where $tr(\mathbf{R})$ is the trace of the autocorrelation matrix for the input $\mathbf{X}(k)$.

Now, we need to check if (30) is satisfied. For noise with zero mean, symmetric probability density,

$$E(n^4(k)) = g\,[E(n^2(k))]^2 \tag{40}$$

holds for some g. For example, $g = 9/5$, 3, 6 for uniform, Gaussian, and Laplacian densities, respectively. Then we may rewrite (16) as

$$b = a^2 + 12\,a\,E(n^2(k)) + (60g - 108)b^2\,[E(n^2(k))]^2 \tag{41}$$

For the LMS, LMF, LVCMS algorithms, $\bar{a} \geq 0$ with the choices of b, c mentioned earlier since $\lambda \leq 0$. Hence, (30) holds for uniform, Gaussian, and Laplacian densities.

If we substitute the choices of b, c, we may have

$$\mu \leq \frac{1}{\bar{a}\,tr(\mathbf{R})} \tag{42}$$

as an alternative expression for the previous upper bounds [2, 4, 5] obtained without

Gaussian assumption on input data. Comparing (39) and (42), we can see that the new bound is tighter than the previou one if the condition (30) holds.

Table I. Upper bounds on the step size parameter with and without Gaussian assumption.

algorithm	without Gaussian assumption	with Gaussian assumption
LMS	$\frac{1}{tr(\mathbf{R})}$	$\frac{1}{3tr(\mathbf{R})}$
LMF	$\frac{1}{6E(n^2(k))tr(\mathbf{R})}$	$\frac{1}{30E(n^4(k))tr(\mathbf{R})}$
LVCMS	$\frac{1}{\bar{a}tr(\mathbf{R})}$	$\frac{\bar{a}}{3\bar{b}tr(\mathbf{R})}$

$\bar{a} = 6E(n^2(k)) - (2\sigma_0^2 + \lambda)$

$\bar{b} = 60E(n^4(k)) - 24(2\sigma_0^2 + \lambda)E(n^2(k)) + (2\sigma_0^2 + \lambda)^2$

The upper bounds, with and without the Gaussian assumption, on μ for the LMS, LMF, and LVCMS algorithms are compared in Table I. For the LMS algorithm, (39) corresponds to the earlier result by Feure and Weinstein in [8]. Both the new and previous bounds for the LMS algorithm depend only on the trace of the autocorrelation matrix of the input data vector. On the other hand, the new bounds for the LMF and LVCMS algorithms depend on the second and fourth order moments of the noise whereas the previous bounds depend only on the second moment. This explains the reason why it is difficult to get convergence in the error covariance of the coefficients in the LMF and LVCMS algorithms when the noise has a heavier-tailed density.

IV. Steady State Performance of The Generalized Algorithm

In [2], the misadjustment of the LMS algorithm was found using the assumption that the coefficient vector $\mathbf{W}(k)$ of the adaptive filter is very close the true coefficient vector $\mathbf{W}^*$. On the other hand, Horowitz and Senne [7] (and Feuer and Weinstein [8] later) obtained the misadjustement for the LMS algorithm without using the assumption when the input data is Gaussian. In this study, however, we assumed that $\mathbf{W}(k) \approx \mathbf{W}^*$ in obtaining the recursion (22) due to the terms of higher power (greater than 2) in $\mathbf{V}(k)$. As pointed out earlier, this assumption is fictitious for the LMS algorithm as demonstrated in [7, 8].

Define

$$J_{mse}(k) = E(\epsilon^2(k)) = E\{[d(k) - \mathbf{W}^T(k)\mathbf{X}(k)]^2\} \tag{43}$$

as the MSE at time k. Then the misadjustment is given by

$$M = \frac{J_{mse}(\infty) - J_{mse}^*(k)}{J_{mse}^*(k)} \tag{44}$$

where J_{mse}^* is the minimum MSE and $J_{mse}(\infty)$ is the steady state MSE. Then

$$\begin{aligned} J_{mse}(k) &= E\{\epsilon^2(k)\} = E\{[d(k) - \mathbf{X}^T(k)\mathbf{W}(k)]^2\} \\ &= E\{\mathbf{X}^T(k)\mathbf{W}^* + n(k) - \mathbf{X}^T(k)\mathbf{W}(k)]^2\} \\ &= E\{[n(k) - \mathbf{X}^T(k)\mathbf{V}(k)]^2\} \\ &= E\{n^2(k)\} - 2E\{n(k)\mathbf{X}^T(k)\mathbf{V}(k)\} \\ &\quad + E\{\mathbf{V}^T(k)\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{V}(k)\}. \end{aligned} \tag{45}$$

However, $n(k)$ has zero mean and is independent of $\mathbf{X}(k)$ and $\mathbf{V}(k)$. Furthermore, $\mathbf{X}(k)$ and $\mathbf{V}(k)$ are independent since $\mathbf{X}(k)$ is Gaussian and is uncorrelated in time. Thus, (45) becomes

$$\begin{aligned} J_{mse}(k) &= J_{mse}^* + tr[E\{\mathbf{V}(k)\mathbf{V}^T(k)\}E\{\mathbf{X}(k)\mathbf{X}^T(k)\}] \\ &= J_{mse}^* + \mathbf{L}^T\bar{\mathbf{C}}(k) \end{aligned} \tag{46}$$

where $J_{mse}^* = E\{n^2(k)\}$ and $\mathbf{L}$ and $\bar{\mathbf{C}}(k)$ are defined in (28) and (26), respectively.

As $\bar{\mathbf{C}}(k)$ converges to a steady state value, the MSE becomes

$$J_{mse}(\infty) = J_{mse}^* + \mathbf{L}^T\bar{\mathbf{C}}(\infty). \tag{47}$$

From (29)

$$\bar{\mathbf{C}}(\infty) = 4\mu^2\bar{c}(\mathbf{I} - \mathbf{F})^{-1}\mathbf{L}, \tag{48}$$

and hence, substituting (47) and (48) into (44) yields

$$M = \frac{4\mu^2\bar{c}\,\mathbf{L}^T(\mathbf{I} - \mathbf{F})^{-1}\mathbf{L}}{J_{mse}^*}. \tag{49}$$

Substituting (27) and (28) into (49) and using the matrix inversion lemma [12] to invert the matrix will yield

$$M = \frac{\bar{c}}{J_{mse}^*} \cdot \frac{\sum_{i=1}^{N} \frac{\mu\gamma_i}{a - 2\mu b\gamma_i}}{1 - \bar{b}\sum_{i=1}^{N} \frac{\mu\gamma_i}{\bar{a} - 2\mu\bar{b}\gamma_i}}. \tag{50}$$

As mentioned in Section III, (22) should contain a constant term associated with μ^2 which comes from higher order terms. Although the constant term does not affect the upper bound on the step size but affect the misadjustment value for the LMF and LVCMS algorithms. However, the effect of the constant term will be reduced as the step size μ becomes smaller since it is of second order in μ.

For the LMS algorithm($\bar{a} = \bar{b} = 1$), the misadjustment in (50) corresponds to the earlier result by Feuer and Weinstein in [8]. For the special case of $\gamma_i = \gamma_j$, $i \neq j$, (50) simplifies to

Table II. Misadjustment values with and without Gaussian assumption.

algorithm	without Gaussian assumption	with Gaussian assumption
LMS	$\mu tr(\mathbf{R})$	$\frac{\mu tr(\mathbf{R})}{1-\mu(1+2/N)tr(\mathbf{R})}$
LMF	$\frac{2\mu E(n^6(k))tr(\mathbf{R})}{3[E(n^2(k))]^2}$	$\frac{2\mu E(n^6(k))tr(\mathbf{R})}{3E(n^2(k))[E(n^2(k))-10\mu E(n^4(k))(1+2/N)tr(\mathbf{R})}$
LVCMS	$\frac{\mu\bar{c}tr(\mathbf{R})}{\bar{a}E(n^2(k))}$	$\frac{\mu\bar{c}tr(\mathbf{R})}{E(n^2(k))[\bar{a}-\mu\bar{b}(1+2/N)tr(\mathbf{R})]}$

$a = 6E(n^2(k)) - (2\sigma_0^2 + \lambda)$
$\bar{b} = 60E(n^4(k)) - 24(2\sigma_0^2 + \lambda)E(n^2(k)) + (2\sigma_0^2 + \lambda)^2$
$c = 4E(n^6(k)) - 4(2\sigma_0^2 + \lambda)E(n^4(k)) + (2\sigma_0^2 + \lambda)^2 E(n^2(k))$

$$M = \frac{c}{J_{mse}^*} \cdot \frac{\mu tr\mathbf{R}}{\bar{a} - \mu\bar{b}(1+2/N)tr\mathbf{R}} \quad (51)$$

The misadjustment expressions with and without the Gaussian assumption for the LMS, LMF, and LVCMS algorithms are compared in Table II. Clearly, the misadjustment values with the Gaussian assumption are larger than those without the assumption in all three algorithms.

V. Simulation Results

Simulations were performed on the system identification problem studied in [4, 5]. The transfer function of the system is given by

$$P(z) = 0.1 + 0.2z^{-1} + 0.3z^{-2} + 0.4z^{-3} + 0.5z^{-4} + 0.4z^{-5} + 0.3z^{-6} + 0.2z^{-7} + 0.1z^{-8}.$$

The value of the step size μ for each algorithm is chosen to give the same rate of convergence. The time constant is fixed at 555 samples. The input signal $\mathbf{X}(k)$ is white and Gaussian with zero mean and unit variance. The plant noise $n(k)$ is white with zero mean and variance 100. The noise is independent of the input signal. The adaptive filter is initialized by adding to each of the coefficients a zero mean, Gaussian random variable with standard deviation 0.75.

Tables III-V show the upper bounds on μ calculated with and without the Gaussian assumption for this system identification example. It is clear that the new bounds are much tighter than the original bounds for all three algorithm and for all noise densities.

The mean squared coefficient error for the LMS, LMF, and LVCMS algorithms are obtained by averaging 40 independent runs. Curves look very similar to those shown in [5] and are omitted here. Theoretical and actual misadjustments for three algorithms are shown in Tables VI-VII.

The actual misadjustment values are obtained by averaging the last 125 data samples out of 2000 samples. Note that the misadjustment values computed from the new expressions are slightly closer to the actual values than those from the original expressions for all cases.

Table III. Upper bounds on the step size parameter with and without Gaussian assumption for the system identification problem: uniform noise density.

algorithm	without Gaussian assumption	with Gaussian assumption
LMS	0.11	3.70×10^{-2}
LMF	1.85×10^{-4}	2.06×10^{-5}
LVCMS	2.78×10^{-4}	2.31×10^{-5}

Table IV. Upper bounds on the step size parameter with and without Gaussian assumption for the system identification problem: Gaussian noise density.

algorithm	without Gaussian assumption	with Gaussian assumption
LMS	0.11	3.70×10^{-2}
LMF	1.85×10^{-4}	1.23×10^{-4}
LVCMS	8.55×10^{-5}	1.21×10^{-5}

Table V. Upper bounds on the step size parameter with and without Gaussian assumption for the system identification problem: Laplace noise density.

algorithm	without Gaussian assumption	with Gaussian assumption
LMS	0.11	3.70×10^{-2}
LMF	1.85×10^{-4}	6.17×10^{-6}
LVCMS	7.94×10^{-5}	8.42×10^{-6}

Table VI. Misadjustment values with and without Gaussian assumption for the system identification problem: uniform noise density.

algorithm	theory		simulation
	without Gaussian assumption	with Gaussian assumption	
LMS	8.11×10^{-3}	8.19×10^{-3}	8.51×10^{-3}
LMF	3.47×10^{-3}	3.49×10^{-3}	4.02×10^{-3}
LVCMS	2.55×10^{-3}	2.65×10^{-3}	3.05×10^{-3}

Table Ⅵ. Misadjustment values with and without Gaussian assumption for the system identification problem: Gaussian noise density.

algorithm	theory		simulation
	without Gaussian assumption	with Gaussian assumption	
LMS	8.11×10^{-3}	8.19×10^{-3}	8.98×10^{-3}
LMF	1.35×10^{-2}	1.36×10^{-2}	1.54×10^{-2}
LVCMS	9.26×10^{-3}	9.48×10^{-3}	1.02×10^{-2}

Table Ⅶ. Misadjustment values with and without Gaussian assumption for the system identification problem: Laplacian noise density.

algorithm	theory		simulation
	without Gaussian assumption	with Gaussian assumption	
LMS	8.11×10^{-3}	8.19×10^{-3}	8.20×10^{-3}
LMF	8.11×10^{-2}	8.13×10^{-2}	9.14×10^{-2}
LVCMS	2.55×10^{-2}	2.63×10^{-2}	2.83×10^{-2}

Ⅵ. Conclusions

In this study, convergence properties of the LMS, LMF, and LVCMS algorithms under the uncorrelated Gaussian input data are investigated. This is carried out by employing the generalized error criterion that admits all three algorithms. The analysis provides the unified results on the upper bound for the step size and the misadjustment expressions for all three algorithms under the uncorrelated Gaussian assumption. The results also explains convergence characteristics of the LMF and LVCMS algorithms envountered in the previous simulation studies. A new results also provide better accuracies to simulation results.

References

1. B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *1960 WESCON Conv. Rec.*, pt. 4, pp. 96-104.
2. B. Widrow, P. Mantey, L. Griffiths, and B. Goode, "Adaptive antenna systems," *Proc. IEEE*, vol. 55, pp. 2143-2159, Dec. 1967.
3. 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.
4. E. Walach and B. Widrow, "The least mean fourth algorithm and its family," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 275-283, Mar. 1984.
5. J. D. Gibson and S. D. Gray, "MVSE adaptive filtering subject to a constraint on MSE," *IEEE Trans. Circuits and Systems*, vol. CAS-35, no. 5, pp. 603-608, May 1988.
6. J. D. Gibson and S. D. Gray, "MVSE adaptive filtering subject to a constraint on MSE," TCSL Res. Rep. no. 86-05-01, Dept. Elec. Eng. Texas A&M Univ., Aug. 5, 1986.
7. L. I. Horowitz and K. D. Senne, "Performance adavantage of complex LMS for controlling narrow-bank adaptive arrays," *IEEE Trans. Circuits and Systems*, vol. CAS-28, no. 6, pp. 562-576, June 1981.
8. A. Feuer and E. Weinstein, "Convergence analysis of LMS Filters with uncorrelated Gaussian data," *IEEE Trans. Acoustic, Speech, Signal Processing*, vol. ASSP-33, no. 1, pp. 222-230, Feb. 1985.
9. J. B. Foley and F. M. Boland, "A note on the convergence analysis of LMS adaptive filters with Gaussian data," *IEEE Trans. Acoustic, Speech, Signal Processing*, vol. ASSP-36, no. 7, pp. 1087-1089, July 1988.
10. F. R. Gantmatcher, *Theory of Matrices*, vol. 2, New York: Chelsea, 1964.
11. P. Lancaster and M. Tismenetsky, *Theory of Matrices with application*, Academic Press, New York, 1985.
12. S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, N. J., 1986.

▲Seung Hyon Nam

Seung Hyon Nam received the B.S. degree in electronic engineering from Sogang University, Korea, in 1980, the M.S. degree in electrical and computer engineering from the University of Alabama, Huntsville, U.S. A., in 1987, and the Ph.D. degree in electrical engineering from Texas A&M University, U.S.A., in 1992.

From 1980 to 1985, he was with the Agency for Defense Development (ADD), Korea, as an engineer. Since 1993, he has been with PaiChai University, Korea, where he is currently an Assistant Professor of the Division of Computer, Electronics, and Information Communications Engineering.

His current research interests include audio and speech coding, adaptive filter theory and applications, and digital signal processing.

▲Insung Lee

Insung Lee received the B.S and M.S degrees in electronic engineering from Yonsei University, Korea, in 1983 and 1985, respectively, and the Ph.D degree in electrical engineering from Texas A&M University, U.S. A., in 1992.

From 1986 to 1987, he was a research engineer at the Korea Telecom Research Center, Korea. From 1989 to 1992, he was a graduate research assistant

in the Department of Electrical Engineering, Texas A&M University, U.S.A. From 1993 to 1995, he was with the Signal Processing Section of the Mobile Communication Division at the Electronics and Telecommunications Research Institute (ETRI), Korea, as a senior member of technical staff. Since 1995, he has been with the Department of Radio Engineering, Chungbuk National University, as an Assistant Professor.

His current research interests include speech and image compression, mobile communication, adaptive filters.