Convergence and Performance Analysis of the Normalized LMS Algorithm with Uncorrelated Gaussian Data

MOSHE TARRAB, MEMBER, IEEE, AND ARIE FEUER, MEMBER, IEEE

Abstract —**The normalized least mean square (NLMS) algorithm can be viewed as a modification of the widely used LMS algorithm. It has an important advantage over the LMS—its convergence is independent of environmental changes. This fact is proven. In addition, we present a comprehensive study of the first- and second-order statistic behavior in the NLMS algorithm. We show that the NLMS algorithm exhibits significant improvement over the LMS algorithm in convergence rate, while its steady-state performance is considerably worse.**

I. Introduction

The LEAST mean square (LMS) algorithm as introduced by Widrow (see, e.g., [1]) has gained much popularity due to its simplicity and ease of implementation. One of its major advantages over stochastic approximation (SA) type algorithms (where the adaptation gains decay to zero) is the ability to "stay alive" and to respond to a changing environment (often the main reason for introducing the adaptive algorithm). However, as the analysis of the LMS algorithm has indicated (see [2]–[4]), the relationship between the fixed adaptation gain and the statistics of the environment determines the algorithm convergence and its performance. A gain choice which is good for certain environments may result in poor performance with a change in environment or even in divergence of the algorithm. As a result users of the LMS tend to be overly conservative in their choice of adaptation gain value, hence causing unnecessarily slow convergence. With this as a motivation, one would like to be able to choose the adaptation gain according to the incoming data. One possibility of accomplishing this is the normalized LMS (NLMS) as presented by Nagumo and Noda [5].

Bitmead and Anderson have investigated some convergence conditions for the NLMS [6], but the first attempt at a quantitative analysis of this algorithm was carried out by Bershard [7]. While some of the results in [7] are similar to some of our results, the conclusions drawn are considerably different. Bershard concentrates on one aspect only of the algorithm—its steady-state performance. As a result, he limits his discussion and conclusions (especially for the case with distinct eigenvalues) to the approximation of small gain. By doing so, we feel, he strips the NLMS algorithm of one of its main performance advantages over the LMS: the speed of convergence. The work reported here complements [7] and details a comprehensive study of the NLMS and its relation to the LMS.

In Section II we present a new motivation for the NLMS algorithm which provides, from the start, an intuitive explanation of why the NLMS is expected to converge faster than the LMS. In Section III we prove convergence conditions for the algorithm (both first- and second-order statistics), while Section IV deals with a quantitative comparison between the LMS and NLMS. Because of analytical difficulties, the comparison is done through two special cases. To support the analysis, simulation experiments were performed, and their results are reported in Section V. Section VI concludes the paper and summarizes the main observations of the study.

II. The Normalized Least Mean Square Algorithm

The normalized LMS algorithm has been presented in the literature (see, e.g., [5] and [7]), but we shall arrive at it from a new direction which, we believe, adds some insight into its properties. The basic problem under consideration can be described as follows. A sequence of n-dimensional vectors $\{X_k\}$ and a sequence of scalars $\{d_k\}$ are measured. It is desired to find a weight vector W, so that $y_k = W^T X_k$ is as close to d_k as possible in the mean square error (mse) sense. The solution W^*, often referred to as the Wiener solution, is well-known and has the form

$$W^* = R^{-1}P \tag{2.1}$$

where

$$R = E\{X_k X_k^T\}$$
$$P = E\{d_k X_k\}. \tag{2.2}$$

To get a constant W^*, it is assumed that X_k and d_k are jointly stationary.[1]

[1] Practically, it allows for either a slowly varying environment or sharp changes with long times in between.

Manuscript received January 27, 1986; revised September 11, 1986.

M. Tarrab was with the Department of Electrical Engineering, Technion, Haifa, Israel. He is now with the DSP Group, Motorola, Israel.
A Feuer is with the Department of Electrical Engineering, Technion, Haifa, 32000 Israel.

IEEE Log Number 8822311.

0018-9448/88/0700-0680$01.00 ©1988 IEEE

The vector W^* can also be computed recursively using the steepest descent (SD) algorithm as follows (see, e.g., [1]):

$$W_{k+1} = W_k + 2\mu[P - RW_k] = (I - 2\mu R)W_k + 2\mu P. \tag{2.3}$$

The convergence of the SD is guaranteed if

$$0 < \mu < \frac{1}{\lambda_{\max}} \tag{2.4}$$

and (see [8]) the fastest convergence is achieved with the choice

$$\mu^* = \frac{1}{\lambda_{\min} + \lambda_{\max}} \tag{2.5}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of R, respectively.

In the case where R and P are not available, they can be replaced by some estimates. Using $X_k X_k^T$ as an instantaneous estimate of R and $d_k X_k$ for P in (2.3), we get the well-known least mean square algorithm

$$W_{k+1} = \left(I - 2\mu X_k X_k^T\right)W_k + 2\mu d_k X_k. \tag{2.6}$$

Note that, to guarantee convergence of the LMS, considerably more stringent conditions on μ are required (see [3]).

Going back to the SD algorithm one can ask, at each step, what is the optimal step size to be taken. Then, instead of a constant μ, we get

$$\mu_k^* = \frac{a_k^T a_k}{2a_k^T R a_k} \tag{2.7}$$

where

$$a_k = RW_k - P. \tag{2.8}$$

Substituting (2.7) in (2.3) we get the normalized SD (NSD) algorithm

$$W_{k+1} = W_k - \frac{a_k^T a_k}{a_k^T R a_k} a_k. \tag{2.9}$$

First, we observe that if the initial values are such that $(W_0 - W^*)$ is an eigenvector of R then $W_1 = W^*$. Namely, in one step we get the right value. In case all eigenvalues of R are equal, namely, $R = \lambda I$, the above is true for all initial conditions. Actually, in this case $\mu_k^* = 1/2\lambda$, so the NSD becomes like the SD with the choice (2.5) for μ.

For the general case, when the ratio $\lambda_{\max}/\lambda_{\min}$ is increased, the SD performance is known to deteriorate and that is when the NSD performance becomes clearly superior. Fig. 1 demonstrates the effects of increase in $\lambda_{\max}/\lambda_{\min}$ on each algorithm in the case where the eigenvalues of R are $\lambda_1 > \lambda_2 = \lambda_3 = \cdots = \lambda_n$. The superiority of the NSD is clear for larger spread of the eigenvalues.

Let us again consider the case when R and P are not known, and we use again $X_k X_k^T$ and $d_k X_k$, respectively, as

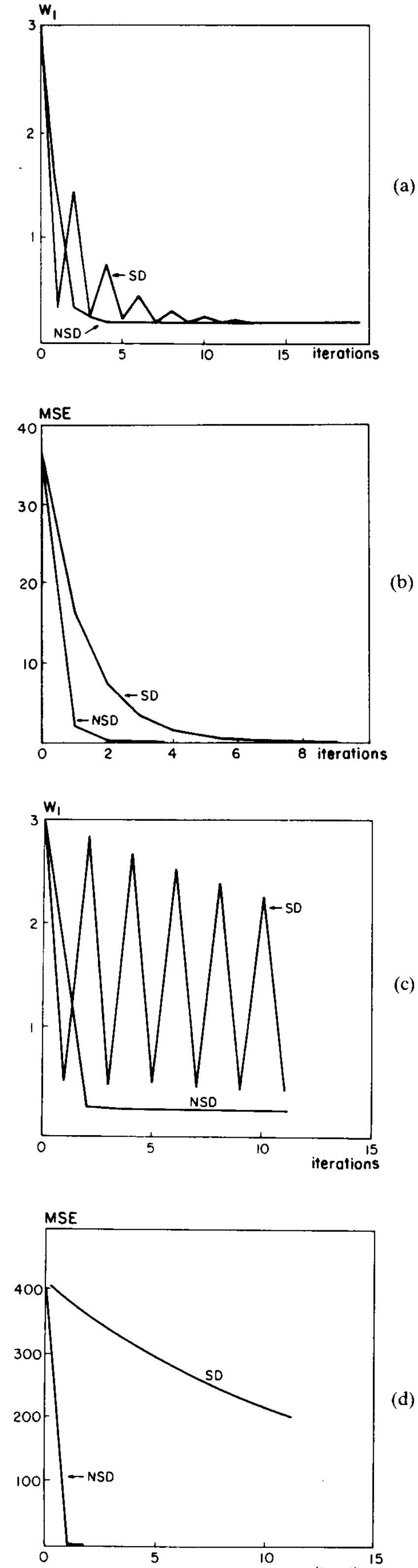

Fig. 1. Convergence comparison between SD and NSD. (a) W_1 behavior, $\lambda_{\max}/\lambda_{\min} = 5$. (b) mse behavior, $\lambda_{\max}/\lambda_{\min} = 5$. (c) W_1 behavior, $\lambda_{\max}/\lambda_{\min} = 65$. (d) mse behavior, $\lambda_{\max}/\lambda_{\min} = 65$.

their estimates. Substitution in (2.8) and (2.9) will result in

$$W_{k+1}=\left(I-\beta\frac{X_kX_k^T}{X_k^TX_k}\right)W_k+\beta\frac{d_k}{X_k^TX_k}X_k \quad (2.10)$$

where the coefficient β was introduced for additional control. This is the normalized LMS algorithm in the exact form as it appears in [5] and [7].

In the following sections we investigate what is required to guarantee the convergence of both the weight vector mean and the MSE as well as some performance characteristics. The comparison between the SD and the NSD suggests that the convergence rates of the NLMS will be superior to those of the LMS. We will show that, contrary to the conclusion in [7], this in fact is the case. Before proceeding we make the following two additional assumptions on the data, similar to those made in [2] and [3] for the LMS analysis and in [7].

Assumption 1: The sequence $\{X_k\}$ is zero mean and jointly Gaussian with $\{d_k\}$.

Assumption 2: The sequence $\{X_k\}$ is uncorrelated in time, namely,

$$E\{X_kX_j^T\}=0 \text{ for } j\neq k.$$

III. Convergence of the Normalized LMS Algorithm

Since there exists an optimal solution to the problem posed, an immediate question is whether the algorithm converges, in some sense, to this optimal solution. We will show that the convergence of the weight vector mean to the optimal weight W^*, can indeed be guaranteed. However, as was pointed out in [2] and [3], since the weights are stochastic processes the above is not satisfactory. We must guarantee a finite variance for the weight vector as well as a finite mean square error—it will be shown that the two are strongly related.

As a first step we derive the equations governing the behavior of the weight vector mean and variance. To do that it will be convenient to carry out the following change of coordinates. Let Q be such that

$$Q^TQ=I$$

$$QRQ^T=\Lambda=\operatorname{diag}(\lambda_i). \quad (3.1)^2$$

Then define

$$\tilde{X}_k=QX_k \quad (3.2)$$

and

$$V_k=Q(W_k-W^*). \quad (3.3)$$

Next, premultiply (2.10) by Q and substitute (3.2) and (3.3) to get

$$V_{k+1}=\left[I-\beta\frac{\tilde{X}_k\tilde{X}_k^T}{\tilde{X}_k^T\tilde{X}_k}\right]V_k+\frac{\beta}{\tilde{X}_k^T\tilde{X}_k}e_k^*\tilde{X}_k \quad (3.4)$$

where $e_k^*=d_k-X_k^TW^*$.

$$V_{k+1}V_{k+1}^T=V_kV_k^T-\beta\left[\frac{\tilde{X}_k\tilde{X}_k^T}{\tilde{X}_k^T\tilde{X}_k}V_kV_k^T+V_kV_k^T\frac{\tilde{X}_k\tilde{X}_k^T}{\tilde{X}_k^T\tilde{X}_k}\right]+\beta^2\frac{\tilde{X}_k\tilde{X}_k^T}{\tilde{X}_k^T\tilde{X}_k}V_kV_k^T\frac{\tilde{X}_k\tilde{X}_k^T}{\tilde{X}_k^T\tilde{X}_k}+\frac{\beta e_k^*}{\tilde{X}_k^T\tilde{X}_k}\left[V_k\tilde{X}_k^T+\tilde{X}_kV_k^T-\frac{\tilde{X}_k\tilde{X}_k^T}{\tilde{X}_k^T\tilde{X}_k}V_k\tilde{X}_k^T-\tilde{X}_kV_k^T\frac{\tilde{X}_k\tilde{X}_k^T}{\tilde{X}_k^T\tilde{X}_k}\right]+\frac{\beta^2(e_k^*)^2}{(\tilde{X}_k^T\tilde{X}_k)^2}\tilde{X}_k\tilde{X}_k^T. \quad (3.5)$$

We now make the following observations based on Assumptions 1 and 2 and (3.4):

Observation 1: Since

$$E\{e_k^*X_k\}=E\{d_kX_k\}-E\{X_kX_k^T\}W^*=P-RR^{-1}P=0$$

and e_k^*, X_k are jointly Gaussian, they are independent as well.

Observation 2: Since X_j and X_k are uncorrelated for $j\neq k$ and Gaussian, they are independent. Hence, from (3.4) and Observation 1, V_k and $\tilde{X}_k$ are independent.

Observation 3: X_k being Gaussian with zero mean and (3.1) with (3.2) imply that the expected value of any function of $\tilde{X}_k$, which is odd with respect to at least one entry[3] of $\tilde{X}_k$, is zero.

Next, take the expected value of both sides of (3.4) and (3.5) and use Observations 1–3 to get

$$E\{V_{k+1}\}=[I-\beta B]E\{V_k\} \quad (3.6)$$

and

$$C_{k+1}=C_k-\beta(BC_k+C_kB)+\beta^2D_k+\beta^2\epsilon^*H \quad (3.7)$$

where $\epsilon^*=E\{(e_k^*)^2\}$ is minimal mse,

$$B=E\left\{\frac{\tilde{X}_k\tilde{X}_k^T}{\tilde{X}_k^T\tilde{X}_k}\right\} \quad (3.8)$$

$$C_k=E\{V_kV_k^T\} \quad (3.9)$$

$$D_k=E\left\{\frac{\tilde{X}_k\tilde{X}_k^T}{\tilde{X}_k^T\tilde{X}_k}C_k\frac{\tilde{X}_k\tilde{X}_k^T}{\tilde{X}_k^T\tilde{X}_k}\right\} \quad (3.10)$$

$$H=E\left\{\frac{\tilde{X}_k\tilde{X}_k^T}{(\tilde{X}_k^T\tilde{X}_k)^2}\right\}. \quad (3.11)$$

Based on Observation 3, we add the following.

Observation 4: B and H are diagonal matrices and D_k has the property that only the diagonal entries of C_k

[2]A Q satisfying (3.1) exists since R is symmetric.

[3]This means that, if $g(x_1,x_2,\cdots,x_n)$ is the given function, then $g(x_1,x_2,\cdots,x_i,\cdots,x_n)=-g(x_1,x_2,\cdots,x_i,\cdots,x_n)$.

appear in the diagonal entries of D_k. Specifically,

$$(D_k)_{i,i} = \sum_{j=1}^{n} G_{ij}(C_k)_{j,j} \tag{3.12}$$

and also for $i \neq j$

$$(D_k)_{i,j} = 2G_{i,j}(C_k)_{i,j} \tag{3.13}$$

where

$$G_{i,j} = E\left\{ \frac{(\tilde{X}_k)_i^2 (\tilde{X}_k)_j^2}{(\tilde{X}_k^T \tilde{X}_k)^2} \right\}.$$

Since both (3.6) and (3.7) are linear time-invariant difference equations the asymptotic stability of their autonomous part and the boundedness of the forcing term (in (3.7)) will guarantee their convergence. We start with the latter.

Clearly,

$$|H_{i,i}| \leq \text{tr}[H] = E\left\{ \frac{1}{\tilde{X}_k^T \tilde{X}_k} \right\} = E\left\{ \frac{1}{\tilde{X}_k^T \Lambda^{1/2} \Lambda^{-1} \Lambda^{1/2} \tilde{X}_k} \right\}$$
$$\leq \frac{1}{\lambda_{\min}} E\left\{ \frac{1}{\tilde{X}_k^T \Lambda^{-1} \tilde{X}_k} \right\}.$$

The entries of the vector $\Lambda^{-1/2}\tilde{X}_k$ [4] are independently identically distributed (i.i.d.) Gaussian with zero mean and variance 1. Hence, from (A3) in Appendix I we get

$$|H_{i,i}| \leq \frac{1}{(n-2)\lambda_{\min}}. \tag{3.14}$$

Now we prove the following:

Proposition 3.1: Consider (3.6) and (3.7). Then, if

$$0 < \beta < 2, \tag{3.15}$$

both equations are asymptotically stable. (See also [6].)

Proof: Consider first (3.6). Let α and $\boldsymbol{a}$ be an eigenvalue and the corresponding eigenvector of $[I - \beta B]$. Note that both are real since the matrix is symmetric. Then

$$a^T[I - \beta B]a = \alpha a^T a$$

or

$$\beta a^T B a = (1-\alpha) a^T a. \tag{3.16}$$

Also, from (3.8) and the Cauchy–Schwartz inequality,

$$0 < a^T B a \leq a^T a. \tag{3.17}$$

Combining (3.15)–(3.17) we can conclude that

$$0 < 1 - \alpha < \beta < 2$$

so

$$|\alpha| < 1.$$

Hence, since all eigenvales of $[I - \beta B]$ are within the unit circle, (3.6) is asymptotically stable.

[4]Since Λ is a diagonal matrix with positive values λ_i on its diagonal, $\Lambda^{-1/2} = \text{diag}\{\lambda_1^{-1/2}, \lambda_2^{-1/2}, \cdots, \lambda_n^{-1/2}\}$.

Now we consider (3.7). We treat separately the off-diagonal and diagonal terms of C_k. For the off-diagonal terms we have, because $(C_k)_{i,j} = (C_k)_{j,i}$ and using Observation 4,

$$(C_{k+1})_{i,j} = \gamma_{i,j}(C_k)_{i,j} \tag{3.18}$$

where

$$\gamma_{i,j} = 1 - \beta(B_{i,i} + B_{j,j}) + 2\beta^2 G_{i,j}, \qquad i \neq j.$$

Since

$$2(\tilde{X}_k)_i(\tilde{X}_k)_j \leq (\tilde{X}_k)_i^2 + (\tilde{X}_k)_j^2, \tag{3.19}$$

we have, with (3.15),

$$1 - \gamma_{i,j} = \beta[(B_{i,i} + B_{j,j}) - 2\beta G_{i,j}]$$
$$\geq \beta E\left\{ \frac{(\tilde{X}_k)_i^2 + (\tilde{X}_k)_j^2}{\tilde{X}_k^T \tilde{X}_k} \left[1 - \frac{\beta}{2} \cdot \frac{(\tilde{X}_k)_i^2 + (\tilde{X}_k)_j^2}{\tilde{X}_k^T \tilde{X}_k} \right] \right\}$$
$$> 0.$$

On the other hand,

$$1 - \gamma_{i,j} < \beta(B_{i,i} + B_{j,j}) = \beta E\left\{ \frac{(\tilde{X}_k)_i^2 + (\tilde{X}_k)_j^2}{\tilde{X}_k^T \tilde{X}_k} \right\} < \beta < 2.$$

Hence

$$|\gamma_{i,j}| < 1$$

and (3.18) is asymptotically stable.

For the diagonal entries of C_k we can define

$$\sigma_k = [(C_k)_{1,1}, (C_k)_{2,2}, \cdots, (C_k)_{n,n}]^T \tag{3.20}$$

and use (3.7) to write

$$\sigma_{k+1} = F\sigma_k + \beta^2 \epsilon^* \boldsymbol{h} \tag{3.21}$$

where

$$\boldsymbol{h} = [H_{1,1}, H_{2,2}, \cdots, H_{n,n}]^T$$
$$F = \text{diag}\{(1 - 2\beta B_{i,i})\} + \beta^2 G. \tag{3.22}$$

For stability analysis we consider the autonomous part, namely, the equation

$$\tilde{\sigma}_{k+1} = F\tilde{\sigma}_k \tag{3.23}$$

where

$$\tilde{\sigma}_0 = \sigma_0. \tag{3.24}$$

For the above equation we note the following:

$$(\tilde{\sigma}_0)_i \geq 0. \tag{3.25}$$

Since by the Cauchy–Schwarz inequality

$$G_{i,i} \geq B_{i,i}^2$$

we have from (3.21) that if $(\tilde{\sigma}_k)_i \geq 0$, then

$$(\tilde{\sigma}_{k+1})_i = (F\sigma_k)_i = (1 - 2\beta B_{i,i})(\tilde{\sigma}_k)_i + \beta^2 \sum_{j=1}^{n} G_{i,j}(\tilde{\sigma}_k)_j$$
$$= (1 - 2\beta B_{i,i} + \beta^2 G_{i,i})(\tilde{\sigma}_k)_j + \beta^2 \sum_{j \neq i}^{n} G_{i,j}(\tilde{\sigma}_k)_j$$
$$> [1 - 2\beta B_{i,i} + \beta^2 (B_{i,i})^2](\tilde{\sigma}_k)_j \geq 0.$$

Hence, by induction

$$(\tilde{\sigma}_k)_i \geq 0, \qquad i = 1, 2, \cdots, n \tag{3.26}$$

for all k.

Now if we denote

$$v_k = \sum_{i=1}^{n} (\tilde{\sigma}_k)_i,$$

we have from (3.22) and (3.23)

$$v_{k+1} = v_k - \beta \sum_{i=1}^{n} \left[2B_{i,i}(\tilde{\sigma}_k)_i - \beta \sum_{j=1}^{n} G_{i,j}(\tilde{\sigma}_k)_j \right].$$

Since $\sum_{j=1}^{n} G_{i,j} = B_{i,i}$ and by the Cauchy–Schwarz inequality

$$E\{(\tilde{X}_k)_i^2\} = E\left\{ \frac{(\tilde{X}_k)_i}{(\tilde{X}_k^T \tilde{X}_k)^{1/2}} \cdot (\tilde{X}_k)_i (\tilde{X}_k^T \tilde{X}_k)^{1/2} \right\}$$
$$\leq \left[B_{i,i} \cdot E\{(\tilde{X}_k)_i^2 \cdot \tilde{X}_k^T \tilde{X}_k\} \right]^{1/2}$$

or

$$B_{i,i} \geq \frac{\lambda_i}{2\lambda_i + \sum_{j=1}^{n} \lambda_j},$$

we get

$$v_{k+1} \leq \left[1 - \beta(2-\beta) \frac{\lambda_{\min}}{2\lambda_{\max} + \operatorname{tr}\Lambda} \right] v_k.$$

By (3.15),

$$0 < \beta(2-\beta) \frac{\lambda_{\min}}{2\lambda_{\max} + \operatorname{tr}\Lambda} < 1$$

and by (3.26) $v_k \geq 0$ for all k, so the above implies that

$$\lim_{k \to \infty} v_k = 0$$

and from (3.26) with the definition of v_k we conclude that

$$\lim_{k \to \infty} \tilde{\sigma}_k = 0.$$

so (3.23) and hence (3.21) are asymptotically stable.

IV. Performance Comparison Between the NLMS and LMS Algorithms

Since exact quantitative analysis of (3.6) and (3.7) for the general case seems very difficult if possible at all, we will in the sequel concentrate on two special cases: 1) all eigenvalues of R are equal and 2) one eigenvalue of R is larger than the others which are equal.

However, before we turn to these special cases, the following comments are in order. Noting that relationships discussed in Section II between the LMS and SD on one hand and the NLMS and NSD on the other, one would expect the NLMS to have better convergence rate than the LMS. Contrary to the conclusions of [7], both our analysis and simulations indicate that this indeed is the case.

In our comparison of NLMS and LMS performances, since convergence conditions were established, we study convergence rate of weight mean values and mse behavior —both its convergence rate and steady-state value. Let us recall from [2], [3], or [4] that

$$\epsilon_k = E\{(e_k)^2\} = \epsilon^* + \operatorname{tr}[\Lambda C_k]$$
$$= \epsilon^* + \sum_{i=1}^{n} \lambda_i (\sigma_k)_i. \tag{4.1}$$

This and the structure of (3.7) enables us to lump together all $(\sigma_k)_i$ which correspond to equal eigenvalues of R. The measures we use for algorithm performance are the misadjustment (see [1])

$$M_s = \frac{\epsilon_\infty - \epsilon^*}{\epsilon^*} \tag{4.2}$$

for steady-state performance and eigenvalues of the matrices involved for convergence rates.

A. Case 1: $\lambda_i = \lambda$, $i = 1, 2, \cdots, n$

For this case $R = \lambda I$ hence $\tilde{X}_k = X_k$, and the following result from [9] can be applied: $(X_k^T X_k)$ is statistically independent of any function of X_k, $h(X_k)$, which has the property $h(\alpha X_k) = h(X_k)$ for all $\alpha \neq 0$. Using this, we have

$$E\{(X_k)_i^2\} = E\left\{ \frac{(X_k)_i^2}{X_k^T X_k} \cdot X_k^T X_k \right\}$$
$$= B_{i,i} E\{X_k^T X_k\}$$

$$E\{(X_k)_i^2 (X_k)_j^2\} = E\left\{ \frac{(X_k)_i^2 (X_k)_j^2}{(X_k^T X_k^T)^2} \cdot (X_k^T X_k)^2 \right\}$$
$$= G_{i,j} E\{(X_k^T X_k)^2\}.$$

Hence

$$B_{i,i} = \frac{1}{n}, \qquad i = 1, 2, \cdots, n \tag{4.3}$$

$$G_{i,i} = \frac{3}{n(n+2)}, \qquad i = 1, 2, \cdots, n \tag{4.4}$$

and

$$G_{i,j} = \frac{1}{n(n+2)}, \qquad i \neq j. \tag{4.5}$$

Also, using the same technique as in Appendix I one can conclude

$$H_{i,i} = E\left\{ \frac{(X_k)_i^2}{(X_k^T X_k)^2} \right\}$$
$$= \frac{1}{n(n-2)\lambda}. \tag{4.6}$$

Substituting (4.3) through (4.6) in (3.6) and (3.21), we get

$$E\{V_{k+1}\} = \left(1-\frac{\beta}{n}\right)E\{V_k\} \tag{4.7}$$

and for $v_k = \Sigma_{i=1}^n(\sigma_k)_i$

$$v_{k+1} = \left(1-\frac{\beta(2-\beta)}{n}\right)v_k + \frac{\beta^2\epsilon^*}{(n-2)\lambda}. \tag{4.8}$$ [5]

From (4.8) the fastest convergence for v_k (and equivalently for $\epsilon_k = \epsilon^* + \lambda v_k$) is achieved for

$$\beta^* = 1 \tag{4.9}$$

then ϵ_k will converge as $(1-(1/n))^k$. Also from (4.8), clearly

$$\epsilon_\infty = \epsilon^* + \frac{n\beta}{(n-2)(2-\beta)}\epsilon^*$$

and

$$M_s = \frac{n\beta}{(n-2)(2-\beta)}.$$

For the choice (4.9)

$$M_s^* = \frac{n}{n-2}. \tag{4.10}$$

As a basis for comparison we take the value of μ resulting in the fastest mse convergence in the LMS (see [2] and [3])

$$\mu^* = \frac{1}{2(n+2)\lambda}. \tag{4.11}$$

With this choice the mse will converge (for the LMS) as $[1-1/(n+1)]^k$, slower than for the NLMS, while the misadjustment becomes

$$\tilde{M}_s = \frac{n}{n+2} \tag{4.12}$$

consistently smaller than with the NLMS. From the equations of the weight-mean values we observe that with the NLMS ($\beta^*=1$) the convergence is as $(1-(1/n))^k$ while with the LMS ($\mu^* = 1/2(n+2)\lambda$) as $[1-1/(n+2)]^k$. To conclude, for this case, while not significantly different, the NLMS converges faster with worse steady-state performance ($M_s > \tilde{M}_s$).

Remark: An interesting relationship can be observed between the choice $\beta^*=1$ for the NLMS and $\mu^* = 1/2(n+2)\lambda$ for the LMS. With $\beta^*=1$ we have in the NLMS

$$\mu_k = \frac{1}{2X_k^TX_k}$$

which is a stationary stochastic process. It can be shown that

$$\text{mode}(\mu_k) = \mu^*$$

where by mode(μ_k) we denote the value of μ at which the probability density function of μ_k achieves its maximum.

B. Case 2: $\lambda_1 > \lambda_2 = \lambda_3 = \cdots = \lambda_n = \lambda$

The difficulty in analyzing the NLMS lies in computing B, G, and H (see (3.8)–(3.13)). In Appendix II we demonstrate how, for the case we have, they can be computed. The results are

$$H_{1,1} = \frac{1}{(\lambda_1\lambda)^{1/2}n(n-2)}F\left(\frac{3}{2},\frac{n-2}{2};\frac{n+2}{2};\left(1-\frac{\lambda}{\lambda_1}\right)\right) \triangleq \frac{1}{(\lambda_1\lambda)^{1/2}n(n-2)}f_1 \tag{4.13}$$

$$H_{i,i} = \frac{1}{(\lambda_1\lambda)^{1/2}n(n-2)}F\left(\frac{1}{2},\frac{n-2}{2};\frac{n+2}{2};\left(1-\frac{\lambda}{\lambda_1}\right)\right) \triangleq \frac{1}{(\lambda_1\lambda)^{1/2}n(n-2)}f_2, \quad i \geq 2 \tag{4.14}$$

$$G_{1,1} = \frac{3(\lambda/\lambda_1)^{1/2}}{n(n+2)}F\left(\frac{5}{2},\frac{n}{2};\frac{n+4}{2};\left(1-\frac{\lambda}{\lambda_1}\right)\right) \triangleq \frac{3(\lambda/\lambda_1)^{1/2}}{n(n+2)}f_3 \tag{4.15}$$

$$G_{i,i} = \frac{3(\lambda/\lambda_1)^{1/2}}{n(n+2)}F\left(\frac{1}{2},\frac{n}{2};\frac{n+4}{2};\left(1-\frac{\lambda}{\lambda_1}\right)\right) \triangleq \frac{3(\lambda/\lambda_1)^{1/2}}{n(n+2)}f_4, \quad i \geq 2 \tag{4.16}$$

$$G_{i,j} = \frac{(\lambda/\lambda_1)^{1/2}}{n(n+2)}F\left(\frac{1}{2},\frac{n}{2};\frac{n+4}{2};\left(1-\frac{\lambda}{\lambda_1}\right)\right) \triangleq \frac{(\lambda/\lambda_1)^{1/2}}{n(n+2)}f_4, \quad i,j \geq 2 \tag{4.17}$$

$$G_{1,i} = \frac{(\lambda/\lambda_1)^{1/2}}{n(n+2)}F\left(\frac{3}{2},\frac{n}{2};\frac{n+4}{2};\left(1-\frac{\lambda}{\lambda_1}\right)\right) \triangleq \frac{(\lambda/\lambda_1)^{1/2}}{n(n+2)}f_5, \quad i \geq 2 \tag{4.18}$$

$$B_{1,1} = \frac{(\lambda/\lambda_1)^{1/2}}{n}F\left(\frac{3}{2},\frac{n}{2};\frac{n+2}{2};\left(1-\frac{\lambda}{\lambda_1}\right)\right) \triangleq \frac{(\lambda/\lambda_1)^{1/2}}{n}f_6 \tag{4.19}$$

$$B_{i,i} = \frac{(\lambda/\lambda_1)^{1/2}}{n}F\left(\frac{1}{2},\frac{n}{2};\frac{n+2}{2};\left(1-\frac{\lambda}{\lambda_1}\right)\right) \triangleq \frac{(\lambda/\lambda_1)^{1/2}}{n}f_7, \quad i \geq 2 \tag{4.20}$$

where $F(\alpha, \beta; \gamma; z)$ is a hypergeometric sequence defined

[5] These same equations have been independently derived in [7]. However, the approach used here is simpler and more straightforward.

by

$$F(\alpha,\beta;\gamma;z)=1+\frac{\alpha\cdot\beta}{\gamma\cdot 1}z+\frac{\alpha(\alpha+1)\cdot\beta(\beta+1)}{\gamma(\gamma+1)\cdot 1\cdot 2}z^2+\frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{\gamma(\gamma+1)(\gamma+2)1\cdot 2\cdot 3}z^3+\cdots=\frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)}\sum_{M=0}^{\infty}\frac{\Gamma(\alpha+m)\Gamma(\beta+m)}{\Gamma(\gamma+m)m!}z^m \tag{4.21}$$

and $\Gamma(\cdot)$ is the Gamma function.

Using some algebraic properties of the hypergeometric sequence described in [10] the following can readily be derived

$$f_1=\frac{n+1}{n+2}f_4+\frac{1}{n+2}\left[1+(n+3)\left(1-\frac{\lambda}{\lambda_1}\right)\right]f_5 \tag{4.22}$$

$$f_2=\frac{n+1}{n+2}f_4+\frac{\lambda}{(n+2)\lambda_1}f_5 \tag{4.23}$$

$$f_3=\frac{(n+1)\lambda_1}{3\lambda}f_4-\frac{(n-3)\lambda+\lambda_1}{3\lambda}f_5 \tag{4.24}$$

$$f_6=\frac{(n-1)\lambda_1}{(n-2)\lambda}f_4+\frac{1}{n+2}\left(2-\frac{\lambda_1}{\lambda}\right)f_5 \tag{4.25}$$

$$f_7=\frac{n+1}{n+2}f_4+\frac{1}{n+2}f_5 \tag{4.26}$$

where we have expressed all f_i through f_4 and f_5.

Turning now to (3.21) we note that by making use of the special case we consider here the behavior of σ_k and ϵ_k can be studied through a two-dimensional equation,

$$\begin{bmatrix}(\sigma_{k+1})_i\\ \sum_{i=2}^{n}(\sigma_{k+1})_i\end{bmatrix}=\begin{bmatrix}F_1 & F_2\\ F_3 & F_4\end{bmatrix}\begin{bmatrix}(\sigma_k)_1\\ \sum_{i=2}^{n}(\sigma_k)_i\end{bmatrix}+\beta^2\epsilon^*\begin{bmatrix}H_1\\ H_2\end{bmatrix} \tag{4.27}$$

where

$$F_1=F_{1,1} \tag{4.28}$$

$$F_2=F_{1,2}=F_{1,3}=\cdots=F_{1,n} \tag{4.29}$$

$$F_3=\sum_{i=2}^{n}F_{i,1}=(n-1)F_2 \tag{4.30}$$

$$F_4=F_{i,i}+(n-2)F_{i,j},\qquad i,j\geq 2,\ i\neq j \tag{4.31}$$

$$H_1=H_{1,1} \tag{4.32}$$

$$H_2=\sum_{i=2}^{n}H_{i,i}=(n-1)H_{2,2} \tag{4.33}$$

Substituting (4.13)–(4.21) and (3.22) into (4.28)–(4.33) will result in

$$F_1=1-\frac{\beta(\lambda/\lambda_1)^{1/2}}{n}\left(2f_6-\frac{3\beta}{n+2}f_3\right) \tag{4.34}$$

$$F_2=\frac{\beta^2(\lambda/\lambda_1)^{1/2}}{n(n+2)}f_5 \tag{4.35}$$

$$F_3=(n-1)F_2 \tag{4.36}$$

$$F_4=1-\frac{\beta(\lambda/\lambda_1)^{1/2}}{n}\left(2f_7-\frac{\beta(n+1)}{n+2}f_4\right) \tag{4.37}$$

$$H_1=\frac{1}{(\lambda\lambda_1)^{1/2}n(n-2)}f_1 \tag{4.38}$$

$$H_2=\frac{n-1}{(\lambda\lambda_1)^{1/2}n(n-2)}f_2. \tag{4.39}$$

From (4.1), (4.2), and (4.27) it follows that

$$M_s=\frac{\beta^2[\lambda_1(H_1(1-F_4)+H_2F_2)+\lambda(H_1F_2+H_2(1-F_1))]}{(1-F_1)(1-F_4)-F_2F_3}. \tag{4.40}$$

The value of M_s has been computed as a function of β with λ_1/λ as a parameter and fixed n (Fig. 2(a)) and with n as a parameter with a fixed λ_1/λ (Fig. 2(b)). We observe from the figure that M_s is monotonically increasing with β, equal to zero at $\beta=0$ and goes to infinity at

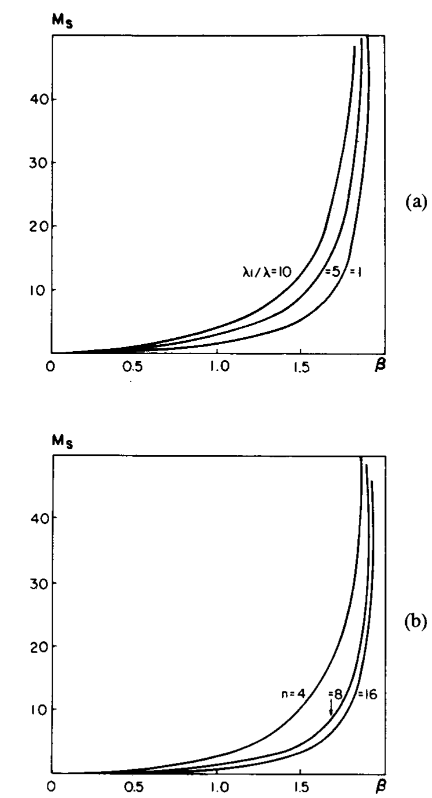

Fig. 2. Misadjustment dependence on β. (a) $n=4$. (b) $\lambda_1/\lambda=5$.

$\beta = 2$ (this last fact supports the necessity of $\beta > 2$ as a condition for convergence).

Since equation (4.27) governs the behavior of the mse we used the largest eigenvalue of the matrix (in absolute value)

$$\begin{bmatrix} F_1 & F_2 \\ F_3 & F_4 \end{bmatrix}$$

from this equation as a measure for speed of convergence. The values of this eigenvalue were computed as a function of β and are presented in Fig. 3. We note in this figure that even for values $\lambda_1/\lambda > 1$ optimal convergence rate (minimum of the computed eigenvalue) occurs around

$$\beta^* = 1. \tag{4.41}$$

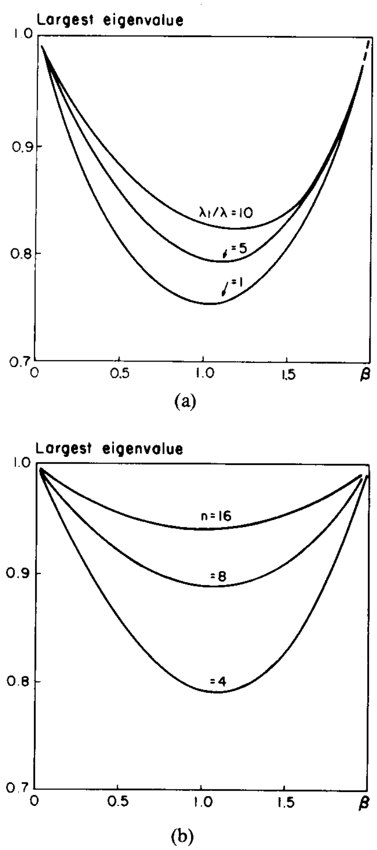

Fig. 3. Convergence rate (largest eigenvalue) dependence on β. (a) $n = 4$. (b) $\lambda_1/\lambda = 5$.

In Fig. 4 we present NLMS performance as a function of the ratio λ_1/λ with β chosen as (4.41). Note that both transient and steady-state performance deteriorate with the increase of λ_1/λ.

It has been shown in [11] that, for the LMS to get close to optimal convergence rate, one can choose

$$\mu^* = \frac{1}{2[3\lambda_1 + (n-1)\lambda]}. \tag{4.42}$$

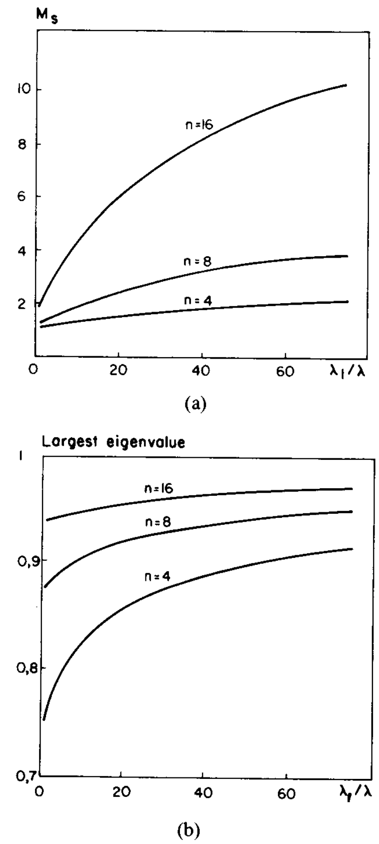

Fig. 4. NLMS performance dependence on eigenvalue spread λ_1/λ. (a) Misadjustment. (b) Convergence rate (largest eigenvalue).

The choices (4.41) and (4.42) will provide the basis for our comparison of LMS and NLMS performances. In Fig. 5 we have convergence rate comparisons of NLMS and LMS, weight mean value in Fig. 5(a) and (b) and mse in Fig. 5(c) and (d). We observe quite a significant superiority of NLMS over LMS as could be expected from the discussion in Section II. On the other hand, LMS steady-state behavior is considerably better than NLMS behavior, as can be observed in Fig. 6.

V. Simulation Results

To verify the analytical results presented in the previous sections we have run a number of simulation experiments on the computer. The case used was of the type analyzed in Section IV-B, similar to the one used in [3], where

$$X_k = \begin{bmatrix} s_k + n_{1,k} \\ s_k + n_{2,k} \\ s_k + n_{3,k} \\ s_k + n_{4,k} \end{bmatrix}$$

and

$$d_k = s_k.$$

$s_k, n_{i,k}$ are sample functions from mutually uncorrelated zero-mean white Gaussian sequences. The $n_{i,k}$ have unit variance; s_k has a variance σ^2.

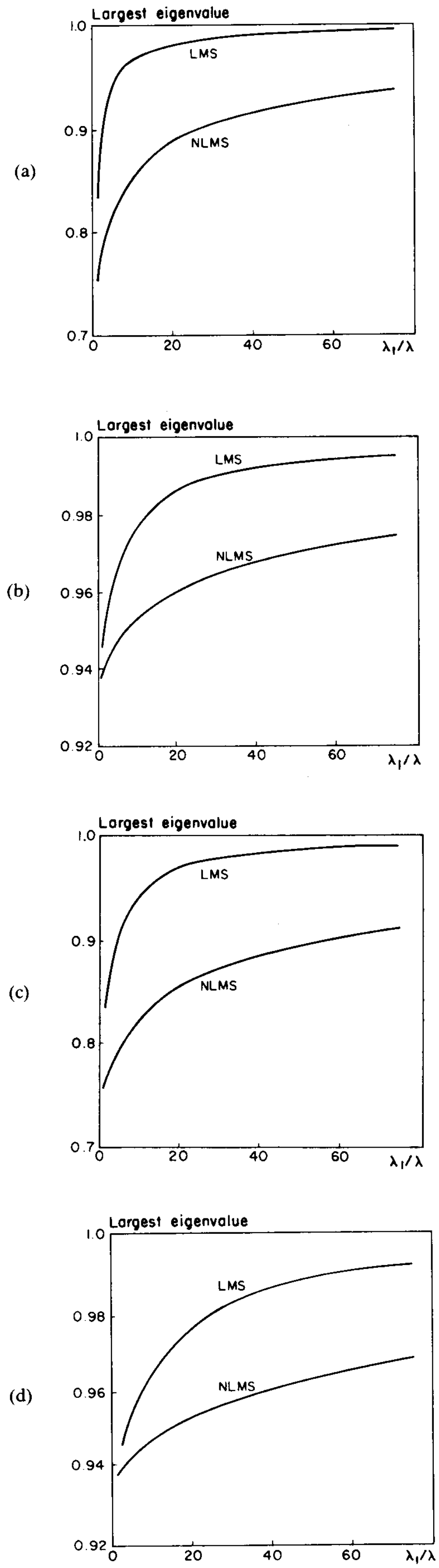

Fig. 5. Convergence rate comparison between LMS and NLMS. (a) $E\{W_1\}$, $n=4$. (b) $E\{W_1\}$, $n=16$. (c) mse, $n=4$. (d) mse, $n=16$.

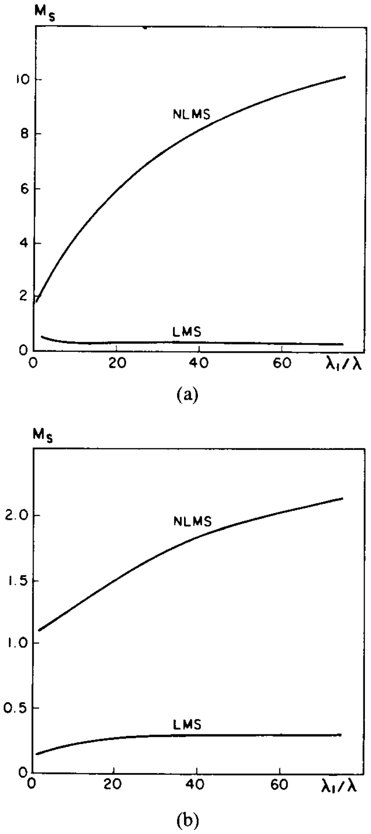

Fig. 6. Comparison of steady-state performance (misadjustment) between LMS and NLMS. (a) $n=4$. (b) $n=16$.

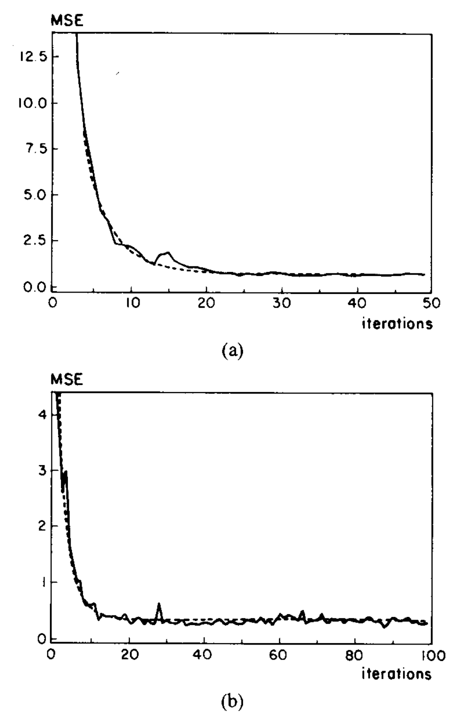

Fig. 7. Analytical results fit (dotted line) with simulation results—average of 100 runs. (a) $\sigma^2=1$, $n=4$. (b) $\sigma^2=0.2$, $n=4$.

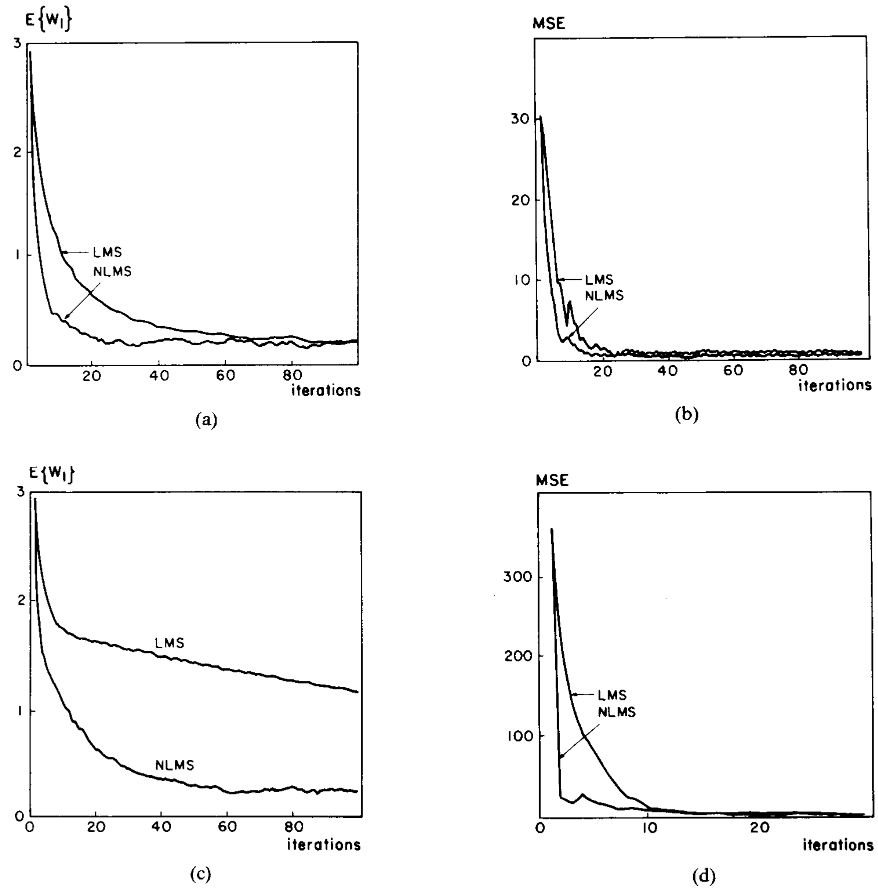

Fig. 8. Comparison of simulation results for LMS and NLMS—average of 100 runs. (Note different scales in (b) and (d).) (a) $E\{W_1\}$, $\sigma=1$. (b) mse, $\sigma=1$. (c) $E\{W_1\}$, $\sigma=4$. (d) mse, $\sigma=4$.

In Fig. 7 we compare the results of the analytical equations with the average of 100 stochastic runs and observe how close they are to each other. In Fig. 8 we present the results of using LMS and NLMS algorithms and clearly observe their support of our analysis of the previous section—the NLMS superiority in speed of convergence is very clear.

VI. Conclusion

By introducing the normalized LMS algorithm as an approximation of the NSD algorithm—a relationship similar to the one between the LMS and the SD—we feel that additional insight has been gained. After presenting the NLMS algorithm, convergence conditions for both first- and second-order statistics were given and their validity proven. We have limited our discussions to stationary Gaussian data without correlation in time.

To gain some quantitative appreciation for the NLMS, two special cases were chosen for which exact expressions were derived for both transient and steady-state performance. These results were then compared to the LMS performance with the following conclusions. Significant improvement of speed of convergence of the NLMS over the LMS, contrary to some previously published results in the literature, can be consistently observed at the expense of worse steady-state performance. A set of simulated stochastic runs on the computer has been carried out, and its results verify the analysis of earlier sections.

Appendix I

Given an i.i.d. sequence $\{y_i\}_1^n$, of Gaussian random variables with zero mean and variance one, we compute

$$I_1 = E\left\{\frac{1}{\sum_{i=1}^{n} y_i^2}\right\}.$$

To do that we make use of the generalized spherical coordinates through the following transformation:

$$y_1 = r\left(\prod_{k=1}^{i-1} \sin\phi_k\right)\cos\phi_i, \qquad 1 \le i \le n-2$$

$$y_{n-i} = r\left(\prod_{k=1}^{n-2} \sin\phi_k\right)\cos\theta$$

$$y_n = r\left(\prod_{k=1}^{n-2} \sin\phi_k\right)\sin\theta$$

where

$$0 \le \phi_k \le \pi$$
$$0 \le \theta \le 2\pi$$
$$0 \le r < \infty.$$

Clearly,

$$\sum_{i=1}^{n} y_i^2 = r^2,$$

so

$$I_1 = E\left\{\frac{1}{\sum_{i=1}^{n} y_i^2}\right\} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{\sum_{i=1}^{n} y_i^2}\frac{1}{(2\pi)^{n/2}} \cdot \exp\left(-1/2\sum_{i=1}^{n} y_i^2\right) dy_1\cdots dy_n$$

$$= \frac{1}{(2\pi)^{n/2}}\int_0^\infty dr\int_0^{2\pi} d\theta \int_0^\pi d\phi_1\cdots d\phi_{n-2} r^{n-3} \cdot \prod_{k=1}^{n-2}\sin^{n-1-k}\phi_k e^{-r^2/2}$$

$$= \frac{2\pi}{(2\pi)^{n/2}}\int_0^\infty\left[r^{n-3}e^{-r^2/2}\,dr\int_0^\pi \sin^{n-2}\phi_1\,d\phi_1\cdots \cdot\int_0^\pi \sin\phi_{n-2}\,d\phi_{n-2}.\right.$$

Substituting

$$v = r^2/2$$
$$r = (2v)^{1/2}$$
$$dr = dv/(2v)^{1/2},$$

we get

$$\int_0^\infty r^{n-3}e^{-r^2/2}\,dr = 2\frac{n-4}{2}\int_0^\infty v\frac{n-2}{2}-1e^{-v}\,dv$$

$$= 2\frac{n-4}{2}\Gamma\left(\frac{n-2}{2}\right), \qquad n>2. \tag{I.1}$$

We also have

$$\int_0^\pi \sin^k\phi\,d\phi = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k+2}{2}\right)}\pi^{1/2}, \tag{I.2}$$

hence substituting (I.1) and (I.2) we get

$$I_1 = \frac{2\pi}{(2\pi)^{n/2}}\cdot 2\frac{n-4}{2} \cdot\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}\cdots\frac{\Gamma(1)}{\Gamma\left(\frac{3}{2}\right)}\Pi^{(n-2)/2}$$

$$= \frac{\Gamma\left(\frac{n-2}{2}\right)}{2\Gamma\left(\frac{n}{2}\right)}, \qquad n>2,$$

or, since,

$$\Gamma(v+1) = v\Gamma(v),$$

$$I_1 = \frac{1}{n-2}, \qquad n>2. \tag{I.3}$$

Appendix II

To establish the expected values of the various functions of X_k, we first consider the following integral which will be central in the calculations to follow:

$$\int_0^\pi \frac{\sin^m x\cos^n x}{(a-b\cos^2 x)^k}\,dx. \tag{II.1}$$

Note first that

$$\int_0^\pi f(\sin x,\cos x)\,dx = \int_0^{\pi/2}[f(\sin x,\cos x)+f(\sin x,-\cos x)]\,dx. \tag{II.2}$$

Now from [10, p. 389] for $m>-1$, $n>-1$ and $a>|b|\ge 0$

$$\int_0^{\pi/2}\frac{\sin^m x\cos^n x}{(a-b\cos^2 x)^k}\,dx = \frac{1}{a^k}B\left(\frac{n+1}{2},\frac{n+1}{2}\right)F\left(\frac{n+1}{2},k;\frac{m+n+2}{2};\frac{b}{a}\right) \tag{II.3}$$

where $B(\cdot,\cdot)$ is a beta function for which

$$B\left(\frac{n+1}{2},\frac{m+1}{2}\right) = \frac{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{n+m+2}{2}\right)}, \tag{II.4}$$

$\Gamma(\cdot)$ being the gamma function, and $F(\alpha,\beta\colon\gamma,z)$ is a hypergeometric sequence defined in [10, p. 1039] as

$$F(\alpha,\beta;\gamma;z) = 1+\frac{\alpha\cdot\beta}{\gamma\cdot 1}z+\frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)1\cdot 2}z^2 + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{\gamma(\gamma+1)(\gamma+2)1\cdot 2\cdot 3}z^3+\cdots. \tag{II.5}$$

Now let $\{y_i\}_1^n$ be an in dependent sequence with $y_i \sim N(0,\lambda_i)$ and let $g(y_1,y_2,\cdots,y_n)$ be any function of y_i, then

$$E\{g(x_1,\cdots,x_n)\} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(y_1,\cdots,y_n)\frac{1}{(2\pi)^{n/2}\left(\prod_{i=1}^{n}\lambda_i\right)^{1/2}} \cdot\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{y_i^2}{\lambda_i}\right\}dy_1\cdots dy_n. \tag{II.6}$$

Now use the same coordinate transformation as in Appendix I, namely,

$$y_i = r\left(\prod_{k=1}^{i-1}\sin\phi_k\right)\cos\phi_i, \qquad i=1,2,\cdots,n-2$$

$$y_{n-i} = r\left(\prod_{k=1}^{n-2}\sin\phi_k\right)\cos\theta$$

$$y_n = r\left(\prod_{k=1}^{n-2}\sin\phi_k\right)\sin\theta. \tag{II.7}$$

Since we deal with the special case where $\lambda_i = \lambda \ i = 2,\cdots,n$ we observe from (B7) that

$$\sum_{i=n}^{j} y_i^2 = r^2\left(\prod_{k=1}^{j-1} \sin\phi_k\right)^2$$

so

$$\sum_{i=1}^{n} \frac{y_i^2}{\lambda_i} = r^2\left[\frac{1}{\lambda_1}\cos^2\phi_1 + \frac{1}{\lambda}\sin^2\phi_1\right] \tag{II.8}$$

and substituted in (II.6) we get

$$E\{g(y_1,\cdots,y_n)\} = \int_0^\infty dr \int_0^{2\pi} d\theta \int_0^\pi d\phi_1 \cdots \int_0^\pi d\phi_{n-2} \cdot \hat{g}(r,\theta,\phi_1,\cdots,\phi_{n-2}) \cdot r^{n-1} \prod_{k=1}^{n-2} \sin^{n-1-k}\phi_k \cdot \frac{1}{(2\pi)^{1/2}\lambda_1^{1/2}\lambda^{(n-1)/2}} \exp\left\{-\frac{r^2}{2}\left[\frac{1}{\lambda_1}\cos^2\phi_1 + \frac{1}{\lambda}\sin^2\phi_1\right]\right\} \tag{II.9}$$

where $\hat{g}(\cdot)$ is the expression resulting from substituting (II.7) in $g(\cdot)$.

From here the computation of the expected values is quite straightforward. Let

$$g(y_1,\cdots,y_n) = \frac{y_1^2}{\left(\sum_{i=1}^{n} y_i^2\right)^2}.$$

Then

$$I_1 = E\{g(y_1,\cdots,y_n)\} = \int_0^\infty dr \int_0^{2\pi} d\theta \int_0^\pi d\phi_1 \cdots \int_0^\pi d\phi_{n-2} \cdot \frac{r^{n-3}\prod_{k=1}^{n-2} \sin^{n-1-k}\phi_k \cdot \cos^2\phi_1}{(2\pi)^{n/2}\lambda_1^{1/2}\lambda^{(n-1)/2}} \cdot \exp\left\{-\frac{r^2}{2}\left[\frac{1}{\lambda_1}\cos^2\phi_1 + \frac{1}{\lambda}\sin^2\phi_1\right]\right\}.$$

We will integrate first with respect to r, using $v = cr^2$ we get

$$\int_0^\infty r^{n-3}e^{-cr^2}\,dr = \frac{1}{2}c^{-(n-2)/2}\int_0^\infty v^{[(n-2)/2]-1}e^{-v}\,dv = \frac{1}{2}c^{-(n-2/2)}\Gamma\left(\frac{n-2}{2}\right), \qquad n<2$$

where

$$c = \frac{1}{2}\left[\frac{1}{\lambda_1}\cos^2\phi_1 + \frac{1}{\lambda}\sin\phi_1\right].$$

Hence

$$I_1 = \frac{\frac{1}{2}\Gamma\left(\frac{n-2}{2}\right)}{(2\pi)^{n/2}\lambda_1^{1/2}\lambda^{(n-1)/2}}\int_0^{2\pi} d\theta \int_0^\pi d\phi_1 \sin^{n-2}\phi_1\cos^2\phi_1 c^{-(n-2/2)} \cdot \int_0^\pi d\phi_2 \sin^{n-3}\phi_2 \int_0^\pi d\phi_3 \sin^{n-4}\phi_3 \cdots \int_0^\pi d\phi_{n-2}\sin\phi_{n-2},$$

or using (I.2)

$$I_1 = \frac{\lambda_1^{(n-3)/2}\Gamma\left(\frac{n-2}{2}\right)}{2\pi^{1/2}\lambda^{1/2}\Gamma\left(\frac{n-1}{2}\right)}\int_0^\pi d\phi_1 \frac{\sin^{N-2}\phi_1\cos^2\phi_1}{[\lambda\cos^2\phi_1 + \lambda_1\sin^2\phi_1]^{(n-2)/2}}.$$

Substituting (II.1)–(II.5), we get

$$I_1 = E\left\{\frac{y_1^2}{\sum_{i=1}^{n} y_i^2}\right\} = \frac{1}{(\lambda_1\lambda)^{1/2}\cdot n(n-2)} \cdot F\left(\frac{3}{2},\frac{n-2}{2};\frac{n+2}{2};\left(1-\frac{\lambda}{\lambda_1}\right)\right).$$

Similarly, we can derive all the other expected values.

References

[1] B. Widrow, "Adaptive filters," in *Aspects of Network and System Theory*, R. E. Kalman and N. DeClaris, Eds. New York: Holt, Rinehart and Winston, 1970, pp. 563–587.

[2] L. I. Horowitz and K. D. Senne, "Performance advantage of complex LMS for controlling narrow-bank adaptive arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 722–736, June 1981.

[3] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 222–230, Feb. 1985.

[4] W. A. Gardner, "Learning characteristics of stochastic gradient descent algorithms: A general study, analysis and critique," *Signal Processing*, vol. 6, pp. 113–133, 1984.

[5] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 283–287, June 1967.

[6] R. R. Bitmead and B. D. O. Anderson, "Performance of adaptive estimation algorithms in dependent random environment," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 788–793, Aug. 1980.

[7] N. J. Bershard, "Analysis of the normalized LMS algorithm with gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 793–806, Aug. 1986.

[8] M. L. Honig and D. G. Messerschmitt, *Adaptive Filters: Structures, Algorithms and Applications*. Boston, MA: Kluwer Academic, 1984.

[9] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. I London: Charles Griffin, 1977.

[10] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*. New York: Academic, 1965.

[11] M. Tarrab, "Convergence and performance analysis of the normalized LMS algorithm" (in Hebrew), M.Sc. thesis, Technion, Haifa, Israel, 1985.