

Performance Analysis of the LMS Algorithm with a Tapped Delay Line (Two-Dimensional Case)

SHAUL FLORIAN AND ARIE FEUER

Abstract—This paper presents exact analytical results of the LMS performances on correlated data vectors. The case addressed here is of data vectors constructed from a tapped delay line. While the results are for the two-dimensional case, they do provide indications of the effect this structure of data vectors, so common to LMS applications, has on LMS performance. They show conclusively the possibility of bias in weight vector mean value and its source. There is also a clear increase in MSE values over the case with uncorrelated data and a decrease of the convergence region of adaptation coefficient. The results presented here strongly suggest that the correlation inherent to the above-mentioned data vectors could be a factor in some applications.

I. INTRODUCTION

THE least mean square (LMS) is one of the most commonly used algorithms in adaptive signal processing. As a result, many researchers have investigated its convergence and performance properties. However, to date, the analysis is still behind, and the algorithm “works” for cases which are not supported by analytical results.

The basic problem for which the LMS is designed can be described as follows. Sequences of n -dimensional vectors $\{X(k)\}$ and scalars $\{d(k)\}$ are measured. It is desired to find a weight vector W so that

$$y(k) = W^T X(k) \quad (1)$$

is as close to $d(k)$ as possible in the mean square error (MSE) sense. Namely, a W that will minimize

$$\begin{aligned} \epsilon(k) &= E\{e(k)^2\}^1 \\ &= E\{[d(k) - W^T X(k)]^2\}. \end{aligned} \quad (2)$$

Assuming that the measured data are stationary, the solution is well known and of the form

$$W^* = R^{-1}P \quad (3)$$

where

$$R = E\{X(k) X(k)^T\} \quad (4)$$

and

$$P = E\{d(k) X(k)\}. \quad (5)$$

This optimal solution can also be obtained recursively through the steepest descent (SD) algorithm, provided R and P are available. The LMS is applicable for the more common situation when R and P are not available. It can be viewed as the simplest approximation of the SD algorithm and has the form

$$W(k+1) = (I - 2\mu X(k) X(k)^T) W(k) + 2\mu d(k) X(k) \quad (6)$$

where $W(k)$ is a vector of weights updated every sampling interval and $\mu > 0$ is a constant parameter which determines the step size.

To be able to carry out a complete analysis, resulting in convergence conditions for both the $E\{W(k)\}$ and the MSE as well as quantitative performance investigation, the sequence $\{X(k)\}$ had to be assumed statistically independent in time (see, e.g., [1]–[4]). Unfortunately, in most applications of this algorithm this assumption does not hold, and then the results available are considerably more limited. References [5]–[8] provide a sample of the existing results when the above assumption does not hold. Generally, these results either depend on assuming a small step size μ or are in the form of bounds on the algorithm performance. We refer the reader to [8] for a more detailed overview of previous results.

In this paper, we try to isolate one main source of dependence in the sequence $\{X(k)\}$ in many applications. We consider the case where $X(k)$ is constructed from a tapped delay line (TDL), namely,

$$X(k) = [x(k), x(k-1), \dots, x(k-n+1)]^T \quad (7)$$

where $\{x(k)\}$ is a scalar i.i.d. sequence, Gaussian, with zero mean and variance σ^2 . We note that even with this independence assumption on $\{x(k)\}$ we have

$$E\{X(k) X(k-1)^T\} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ \sigma^2 & 0 & \dots & 0 & 0 \\ 0 & \sigma^2 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & \sigma^2 & 0 \end{bmatrix} \neq 0.$$

Hence, there is a clear correlation in $\{X(k)\}$ which is inherent in the structure (7), and the effects of this correlation are the subject of our investigation.

The analysis is still very difficult, and even the conver-

Manuscript received January 17, 1986; revised May 7, 1986.
S. Florian is with RAFAEL, P.O. Box 2250, Haifa, Israel.
A. Feuer is with the Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel.
IEEE Log Number 8610385.

¹ $E\{\bullet\}$ denotes the expected value of the term inside the braces.

gence of $E\{W(k)\}$ to the optimal value W^* is not guaranteed, as will be shown in the sequel. However, we limit our discussion to the two-dimensional case, through which we introduce an approach that may be applicable to higher dimensions. For this case, we derive the exact equations governing the behavior of the $E\{W(k)\}$ and the MSE. From these equations, we observe that while the transient behavior of $E\{W(k)\}$ is similar to the case when $\{X(k)\}$ are independent, there is here a possibility for bias in the steady state. The behavior of the MSE is different both in its transient and in its steady-state values. Some simulation experiments were performed to verify the analytical results, and a sample of their results is presented here.

II. CONVERGENCE OF WEIGHT MEAN VALUES

Recall that we are dealing with a two-dimensional case, namely,

$$X(k) = \begin{bmatrix} x(k) \\ x(k-1) \end{bmatrix}. \tag{8}$$

Denoting

$$V_i(k) = W_i(k) - W_i^*, \quad i = 1, 2 \tag{9}$$

and substituting both (8) and (9) in (6), we get

$$\begin{aligned} V_1(k+1) &= (1 - 2\mu x(k)^2) V_1(k) \\ &\quad - 2\mu x(k) x(k-1) V_2(k) + 2\mu e^*(k) x(k) \end{aligned} \tag{10}$$

$$\begin{aligned} V_2(k+1) &= -2\mu x(k) x(k-1) V_1(k) \\ &\quad + (1 - 2\mu x(k-1)^2) V_2(k) \\ &\quad + 2\mu e^*(k) x(k-1) \end{aligned} \tag{11}$$

where

$$e^*(k) = d(k) - X(k)^T W^*. \tag{12}$$

Since $\{x(k)\}$ is i.i.d., we observe that both $V_1(k)$ and $V_2(k)$ are independent of $x(k)$. It can also be readily verified that

$$E\{e^*(k) X(k)\} = 0; \tag{13}$$

hence, taking the expected value of both sides of (10) and (11) will result in

$$E\{V_1(k+1)\} = (1 - 2\mu\sigma^2) E\{V_1(k)\} \tag{14}$$

$$E\{V_2(k+1)\} = E\{V_2(k)\} - 2\mu E\{x(k-1)^2 V_2(k)\}. \tag{15}$$

Let us now define

$$V_3(k) = x(k-1)^2 V_2(k); \tag{16}$$

then from (11) we get

$$\begin{aligned} V_3(k+1) &= 2\mu x(k)^3 x(k-1) V_1(k) \\ &\quad + x(k)^2 (1 - 2\mu x(k-1)^2) V_2(k) \\ &\quad + 2\mu e^*(k) x(k)^2 x(k-1), \end{aligned} \tag{17}$$

or by taking the expected value on both sides, we have

$$\begin{aligned} E\{V_3(k+1)\} &= \sigma^2 E\{V_2(k)\} - 2\mu\sigma^2 E\{V_3(k)\} \\ &\quad + 2\mu E\{e^*(k) x(k)^2 x(k-1)\}. \end{aligned} \tag{18}$$

Combining (14), (15), and (18), we have

$$\begin{aligned} \begin{bmatrix} E\{V_1(k+1)\} \\ E\{V_2(k+1)\} \\ E\{V_3(k+1)\} \end{bmatrix} &= A \begin{bmatrix} E\{V_1(k)\} \\ E\{V_2(k)\} \\ E\{V_3(k)\} \end{bmatrix} \\ &\quad + 2\mu \begin{bmatrix} 0 \\ 0 \\ E\{e^*(k) x(k)^2 x(k-1)\} \end{bmatrix} \end{aligned} \tag{19}$$

where

$$A = \begin{bmatrix} 1 - 2\mu\sigma^2 & 0 & 0 \\ 0 & 1 & -2\mu \\ 0 & \sigma^2 & -2\mu\sigma^2 \end{bmatrix}.$$

This equation will converge if all eigenvalues of A are inside the unit circle. The eigenvalues of A are $0, 1 - 2\mu\sigma^2, 1 - 2\mu\sigma^2$, so the convergence is guaranteed if

$$\gamma = \mu\sigma^2 < 1. \tag{20}$$

This condition is identical to the condition derived when $\{X(k)\}$ are assumed independent. However, the steady-state values here will be

$$\begin{aligned} E\{V_1(\infty)\} &= 0 \\ E\{V_2(\infty)\} &= -\frac{2\mu}{\sigma^2} E\{e^*(k) x(k)^2 x(k-1)\}, \end{aligned} \tag{21}$$

so whether $E\{W_2(k)\}$ will converge to W_2^* depends on whether $E\{e^*(k) x(k)^2 x(k-1)\}$ is equal to zero or not. In many cases, it is equal to zero [e.g., $(d(k), x(k))$ jointly Gaussian], but we will show later a case where it is not zero.

III. CONVERGENCE OF WEIGHT VARIANCES AND MSE

As was argued in [2] and [4], convergence of the mean does not guarantee finite variances, and hence, is not sufficient to guarantee algorithm convergence. To investigate the behavior of weight variances, we square both sides of (10) and (11), take expected values, and make use of the independence between $x(k)$ and $V_1(k), V_2(k)$ to get

$$\begin{aligned} E\{V_1(k+1)^2\} &= (1 - 4\mu\sigma^2 + 12\mu^2\sigma^4) E\{V_1(k)^2\} \\ &\quad + 4\mu^2\sigma^2 E\{x(k-1)^2 V_2(k)^2\} \\ &\quad + 4\mu^2 E\{e^*(k)^2 x(k)^2\} \\ &\quad + 4\mu E\{(1 - 2\mu x(k)^2) x(k) e^*(k) V_1(k)\} \\ &\quad - 8\mu^2 E\{x(k)^2 x(k-1) e^*(k) V_2(k)\} \end{aligned} \tag{22}$$

$$\begin{aligned}
E\{V_2(k+1)^2\} &= 4\mu^2\sigma^2 E\{x(k-1)^2 V_1(k)^2\} \\
&+ E\{V_2(k)^2\} - 4\mu E\{x(k-1)^2 V_2(k)^2\} \\
&+ 4\mu^2 E\{x(k-1)^4 V_2(k)^2\} \\
&+ 4\mu^2 E\{x(k-1)^2 e^*(k)^2\} \\
&- 8\mu^2 E\{x(k) x(k-1)^2 e^*(k) V_1(k)\} \\
&+ 4\mu E\{(1-2\mu x(k-1)^2) \\
&\cdot x(k-1) e^*(k) V_2(k)\}. \quad (23)
\end{aligned}$$

Comparing this to similar developments in [2] or [4], we note a considerable complication due to the correlations present here. To proceed, we define

$$V_3(k) = x(k-1) V_1(k)$$

$$V_4(k) = x(k-1) V_2(k)$$

and use (10) and (11) to get

$$\begin{aligned}
V_3(k+1) &= x(k)(1-2\mu x(k)^2) V_1(k) - 2\mu x(k)^2 V_4(k) \\
&+ 2\mu e^*(k) x(k)^2 \quad (24)
\end{aligned}$$

$$\begin{aligned}
V_4(k+1) &= -2\mu x(k)^2 V_3(k) + x(k)(1-2\mu x(k-1)^2) \\
&\cdot V_2(k) + 2\mu e^*(k) x(k) x(k-1). \quad (25)
\end{aligned}$$

Now, squaring (17), (24), and (25), taking the expected values, and using the independence of $x(k)$, we get, together with (22) and (23),

$$U(k+1) = FU(k) + 4\mu^2 g + 4\mu b(k) - 8\mu^2 c(k) \quad (26)$$

where

$$\begin{aligned}
U(k) &= [E\{V_1(k)^2\}, E\{V_2(k)^2\}, E\{V_3(k)^2\}, E\{V_4(k)^2\}, E\{V_5(k)^2\}]^T \\
F &= \begin{bmatrix} 1 - 4\mu\sigma^2 + 12\mu^2\sigma^4 & | & 0 & | & 0 & | & 4\mu^2\sigma^2 & | & 0 \\ 0 & | & 1 & | & 4\mu^2\sigma^2 & | & -4\mu & | & 4\mu^2 \\ \sigma^2(1 - 12\mu\sigma^2 + 60\mu^2\sigma^4) & | & 0 & | & 0 & | & 12\mu^2\sigma^4 & | & 0 \\ 0 & | & \sigma^2 & | & 12\mu^2\sigma^4 & | & -4\mu\sigma^2 & | & 4\mu^2\sigma^2 \\ 0 & | & 3\sigma^4 & | & 60\mu^2\sigma^6 & | & -12\mu\sigma^4 & | & 12\mu^2\sigma^4 \end{bmatrix} \quad (27)
\end{aligned}$$

$$\begin{aligned}
g &= \begin{bmatrix} E\{x(k)^2 e^*(k)^2\} \\ E\{x(k-1)^2 e^*(k)^2\} \\ E\{x(k)^4 e^*(k)^2\} \\ E\{x(k)^2 x(k-1)^2 e^*(k)^2\} \\ E\{x(k)^4 x(k-1)^2 e^*(k)^2\} \end{bmatrix} \quad (28)
\end{aligned}$$

$$\begin{aligned}
b(k) &= \begin{bmatrix} E\{x(k) e^*(k) V_1(k)\} \\ E\{x(k-1) e^*(k) V_2(k)\} \\ E\{x(k)^3 e^*(k) V_1(k)\} \\ E\{x(k)^2 x(k-1) e^*(k) V_2(k)\} \\ E\{x(k)^4 x(k-1) e^*(k) V_2(k)\} \end{bmatrix} \quad (29)
\end{aligned}$$

and

$$\begin{aligned}
c(k) &= \begin{bmatrix} E\{x(k)^3 e^*(k) V_1(k)\} + E\{x(k)^2 x(k-1) e^*(k) V_2(k)\} \\ E\{x(k) x(k-1)^2 e^*(k) V_1(k)\} + E\{x(k-1)^3 e^*(k) V_2(k)\} \\ E\{x(k)^5 e^*(k) V_1(k)\} + E\{x(k)^4 x(k-1) e^*(k) V_2(k)\} \\ E\{x(k)^3 x(k-1)^2 e^*(k) V_1(k)\} + E\{x(k)^2 x(k-1)^3 e^*(k) V_2(k)\} \\ E\{x(k)^5 x(k-1)^2 e^*(k) V_1(k)\} + E\{x(k)^4 x(k-1)^3 e^*(k) V_2(k)\} \end{bmatrix} \quad (30)
\end{aligned}$$

Assuming that g , $b(k)$, and $c(k)$ are bounded, the boundedness of $U(k)$ will depend on the eigenvalues of F ; they must all be inside the unit circle to guarantee stability.

The characteristic equation of F is

$$\det(\alpha I - F) = \alpha[\alpha^4 - 2(1 - 4\gamma + 12\gamma^2)\alpha^3 + (1 - 4\gamma)(1 - 4\gamma + 24\gamma^2)\alpha^2 + 192\gamma^4(1 - 12\gamma^2)\alpha - 64\gamma^4(1 - 24\gamma^2 + 144\gamma^4)] = 0$$

where $\gamma = \mu\sigma^2$.

Using root locus² methods (with γ as a parameter), the condition for all roots to be inside the unit circle was found to be

$$0 < \mu\sigma^2 < 0.20116. \quad (31)$$

Hence, (31) is a necessary condition for the convergence of $U(k)$. For the case where $\{X(k)\}$ are independent, the equivalent condition is (see [4])

$$0 < \mu\sigma^2 < 0.25, \quad (32)$$

so, as a result of the dependence in $\{X(k)\}$, the stability region got smaller.

As was noted in [2], the MSE is strongly related to the variances of the weight vector in the independent case. Here we get

$$e(k) = d(k) - X(k)^T W(k) = e^*(k) - x(k) V_1(k) - x(k-1) V_2(k)$$

and

$$\begin{aligned} \epsilon(k) &= E\{e^*(k)^2\} + E\{x(k)^2 V_1(k)^2\} \\ &+ E\{x(k-1)^2 V_2(k)^2\} - 2[E\{x(k) e^*(k) V_1(k)\} \\ &+ E\{x(k-1) e^*(k) V_2(k)\}], \end{aligned}$$

or using the notation of (26)–(29),

$$\epsilon(k) = \epsilon^* + \sigma^2 U_1(k) + U_4(k) - 2(b_1(k) + b_2(k)) \quad (33)$$

where $\epsilon^* = E\{(e^*(k))^2\}$ is the optimal MSE, so here the MSE depends on more than just the variances. However, the convergence of $U(k)$ will guarantee the MSE convergence as well.

At this point, to further illustrate the effects of the way $X(k)$ is constructed here, we have chosen two special cases, which will be discussed in the next section.

IV. STEADY-STATE MSE FOR TWO SPECIAL CASES

Case 1: An Estimation Problem

Let

$$d(k) = a_1 x(k) + a_2 x(k-1) + n(k) \quad (34)$$

where a_1, a_2 are fixed scalars and $\{n(k)\}$ is an i.i.d. se-

quence with zero mean and variance σ_n^2 , independent of $\{x(k)\}$ [recall that we assume $x(k)$ to be Gaussian $(0, \sigma^2)$].

With the above, clearly

$$W^* = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

and

$$e^*(k) = n(k).$$

This and (11) imply that $e^*(k)$ and $V_2(k)$ are independent. Using all this and (28)–(30), we get

$$g = \sigma^2 \epsilon^* \begin{bmatrix} 1 \\ 1 \\ 3\sigma^2 \\ \sigma^2 \\ 3\sigma^4 \end{bmatrix} \quad (35)$$

and

$$b(k) = c(k) = 0.$$

Then from (33) we get

$$\epsilon(\infty) = \epsilon^* + \sigma^2 U_1(\infty) + U_3(\infty) \quad (36)$$

where from (26)

$$U(\infty) = 4\mu^2(I - F)^{-1}g. \quad (37)$$

Substituting (27), (35), and (37) in (36) will result in

$$\epsilon(\infty) = \frac{(1 - 2\gamma)^2 \sigma_n^2}{1 - 6\gamma + 8\gamma^2 - 48\gamma^4 - 576\gamma^6}, \quad (38)$$

while for the independent $\{X_k\}$ one would get

$$\tilde{\epsilon}(\infty) = \frac{(1 - 2\gamma)\sigma_n^2}{1 - 4\gamma}. \quad (39)$$

It can be observed that (39) approximates (38) for sufficiently small γ [neglecting the γ^4 and γ^6 terms in (38) will bring (38) to (39)].

In Fig. 1, we compare the results through the commonly used measure

$$M_s = \frac{\epsilon(\infty) - \epsilon^*}{\epsilon^*},$$

often referred to as misadjustment. In addition, we picked an approximation presented in [7], which for the case we deal with will result in

$$\hat{\epsilon}(\infty) = (1 + 2\gamma)\sigma_n^2, \quad (40)$$

and the corresponding misadjustment is presented in Fig. 1. We observe that all three are close for small γ . The approximation (40) becomes quite poor for large γ . The comparison of (38) and (39) clearly indicates that the inherent dependence of $X(k)$ in (7) results, for a fixed γ , in deteriorated steady-state performance (larger MSE in steady state).

²One could also use Jury's stability test to get the same result.

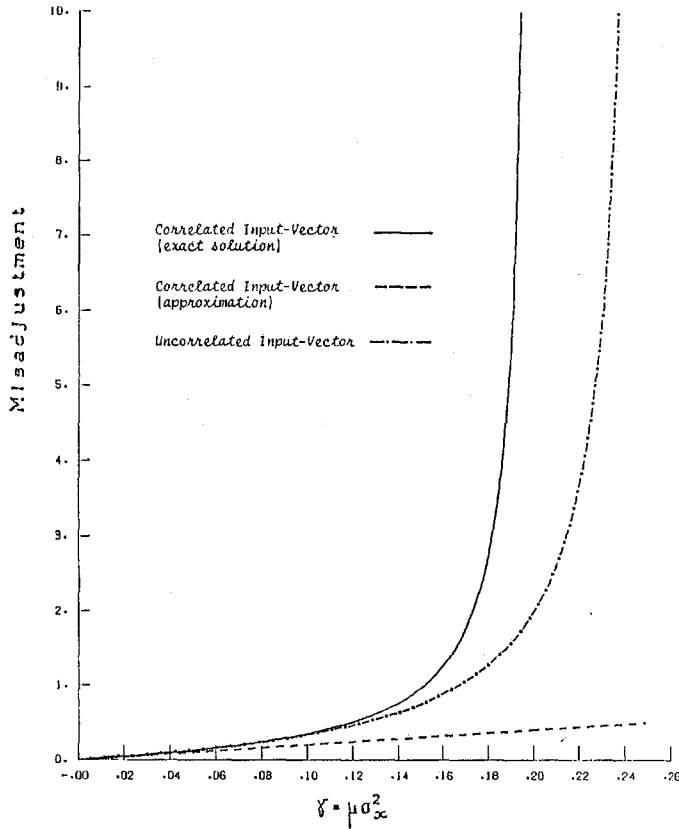


Fig. 1. Comparison of misadjustment for Case 1.

Case 2: $d(k) = x(k)^2 x(k - 1)$

This case is constructed only for the purpose of demonstrating a bias in the convergence of the mean.

Here

$$W^* = \begin{bmatrix} 0 \\ \sigma^2 \end{bmatrix} \tag{41}$$

and

$$e^*(k) = [x(k)^2 - \sigma^2] x(k - 1). \tag{42}$$

Then

$$E\{x(k)^2 e^*(k) x(k - 1)\} = 2\sigma^6 \neq 0$$

and [see (21)]

$$E\{W(\infty)\} = \begin{bmatrix} 0 \\ \sigma^2 - 4\mu\sigma^4 \end{bmatrix} \neq W^*, \tag{43}$$

so the mean does not converge to the optimal solution.

For the MSE, we get [see (28)–(30)]

$$g = 2\sigma^8 \begin{bmatrix} 5 \\ 3 \\ 39\sigma^2 \\ 15\sigma^2 \\ 117\sigma^4 \end{bmatrix} \tag{44}$$

$$b(k) = 2\sigma^4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 6\sigma^2 \end{bmatrix} E\{V_5(k)\} \tag{45}$$

and

$$c(k) = 2\sigma^4 \begin{bmatrix} E\{V_5(k)\} \\ 0 \\ 6\sigma^2 E\{V_5(k)\} \\ E\{V_6(k)\} \\ 6\sigma^2 E\{V_6(k)\} \end{bmatrix} \tag{46}$$

where we made use of the independence between $x(k)$ and $V_1(k)$, $V_2(k)$ and $x(k - 1)$, $V_5(k)$ is defined in (16) and

$$V_6(k) = x(k - 1)^4 V_2(k). \tag{47}$$

To find the steady-state MSE, we use (33), and because of (45), (36) also applies here. However, to find $U(\infty)$, we need to find $E\{V_5(\infty)\}$ and $E\{V_6(\infty)\}$. From (19), it can easily be verified that

$$E\{V_5(\infty)\} = 0. \tag{48}$$

For $V_6(k)$, we use (11) and (42) to get

$$V_6(k + 1) = -2\mu x(k)^5 V_3(k) + x(k)^4 [V_2(k) - 2\mu V_5(k)] + 2\mu [x(k)^6 - x(k)^4 \sigma^2] x(k - 1)^2;$$

taking expected values on both sides and comparing to (18), we observe that

$$E\{V_6(k + 1)\} = 3\sigma^2 E\{V_5(k + 1)\} + 12\mu\sigma^8.$$

Hence, from (48) we have

$$E\{V_6(\infty)\} = 12\mu\sigma^8 \tag{49}$$

and

$$b(\infty) = 0$$

$$c(\infty) = 24\mu\sigma^{12} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 6\sigma^2 \end{bmatrix}$$

Now, since from (26) and $b(\infty) = 0$

$$U(\infty) = 4\mu^2(I - F)^{-1}[g - 2c(\infty)],$$

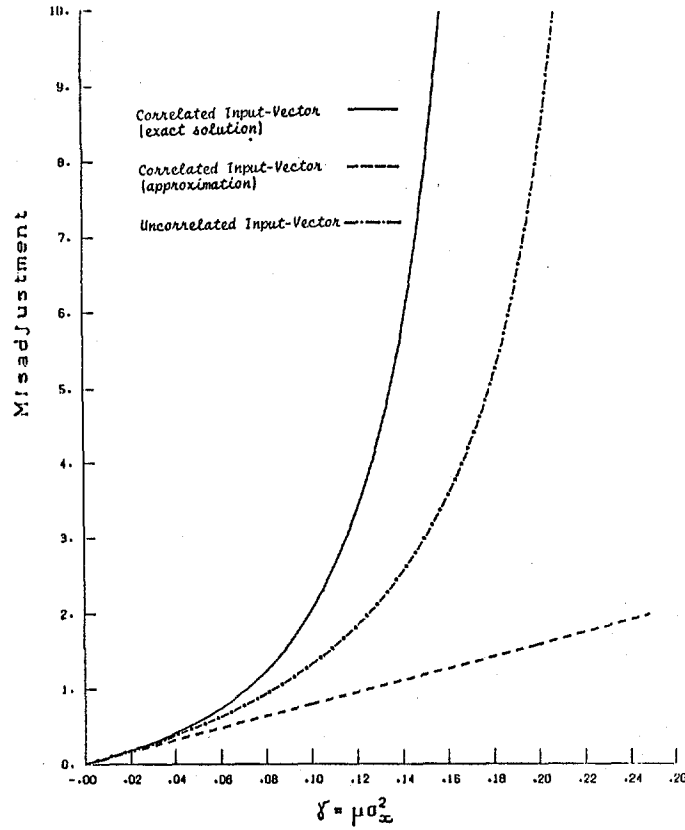


Fig. 2. Comparison of misadjustment for Case 2.

we get eventually

$$\epsilon(\infty) = \frac{1 + 2\gamma - 8\gamma^2 + 384\gamma^3 - 864\gamma^4 + 2880\gamma^5 - 2304\gamma^6}{1 - 6\gamma + 8\gamma^2 - 48\gamma^4 - 576\gamma^6} 2\sigma^6 \tag{50}$$

or

$$M_s = \frac{1 - 2\gamma + 48\gamma^2 - 102\gamma^3 + 360\gamma^4 - 216\gamma^5}{1 - 6\gamma + 8\gamma^2 - 48\gamma^4 - 576\gamma^6} 8\gamma. \tag{51}$$

Again, for comparison, if we work under the independence assumption, we get

$$\tilde{M}_s = \frac{8\gamma}{1 - 4\gamma}, \tag{52}$$

which can be reached from (51) when, for small γ , one neglects all powers of γ higher than 2.

The approximation in [7] for this case will be

$$\hat{M}_s = 8\gamma. \tag{53}$$

Equations (51)–(53) are presented in Fig. 2. We can observe here considerably larger differences between the three curves at smaller γ .

V. SIMULATION RESULTS

In our simulation experiments, we aimed to verify and support the analytical results and conclusions of previous sections. The various experiments are described in the following table.

	Case 1	Case 2
TDL-LMS	$X(k) = \begin{bmatrix} x(k) \\ x(k-1) \end{bmatrix}$ $d(k) = x(k) + x(k-1) + n(k)$	$X(k) = \begin{bmatrix} x(k) \\ x(k-1) \end{bmatrix}$ $d(k) = x(k)^2 x(k-1)$
Unc.-LMS	$X(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}$ $x_1(k), x_2(k) \text{ uncorrelated}$ $d(k) = x_1(k) + x_2(k) + n(k)$	$X(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}$ $x_1(k), x_2(k) \text{ uncorrelated}$ $d(k) = x_1(k)^2 x_2(k)$
TDL-MOD2-LMS	Same as TDL-LMS except the weights are updated only every two samples.	

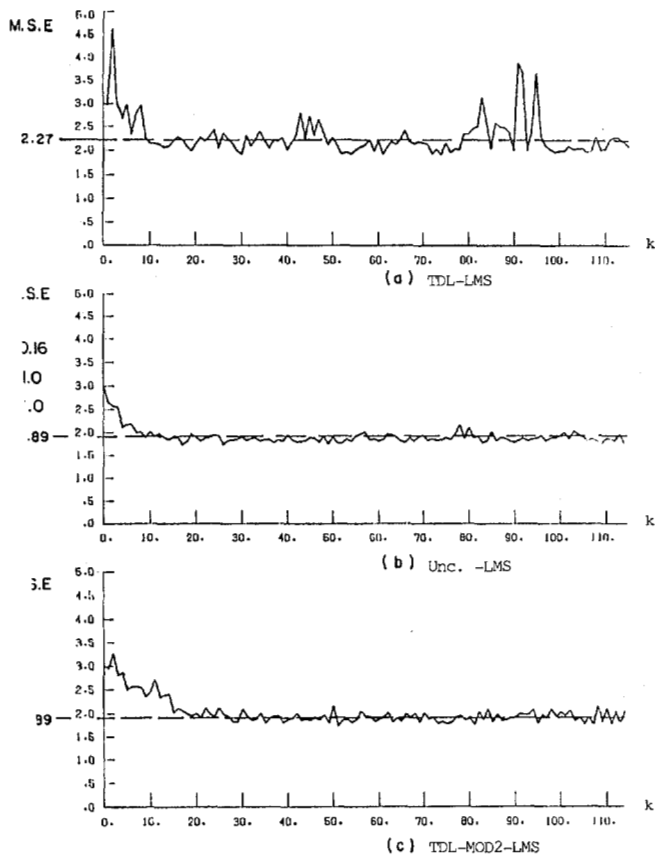


Fig. 3. MSE behavior for the three processings in Case 1.

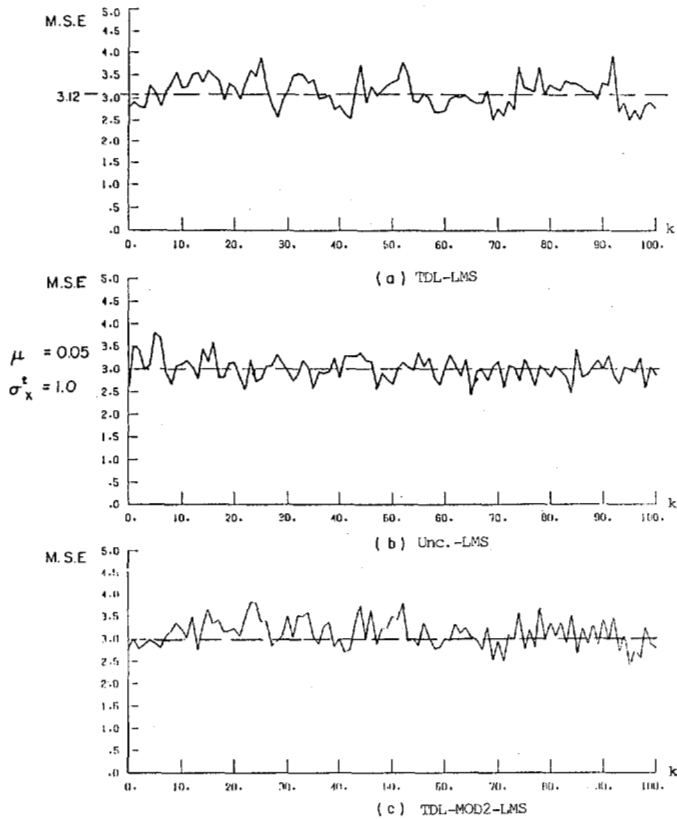


Fig. 4. MSE behavior for the three processings in Case 2, $\mu = 0.05$.

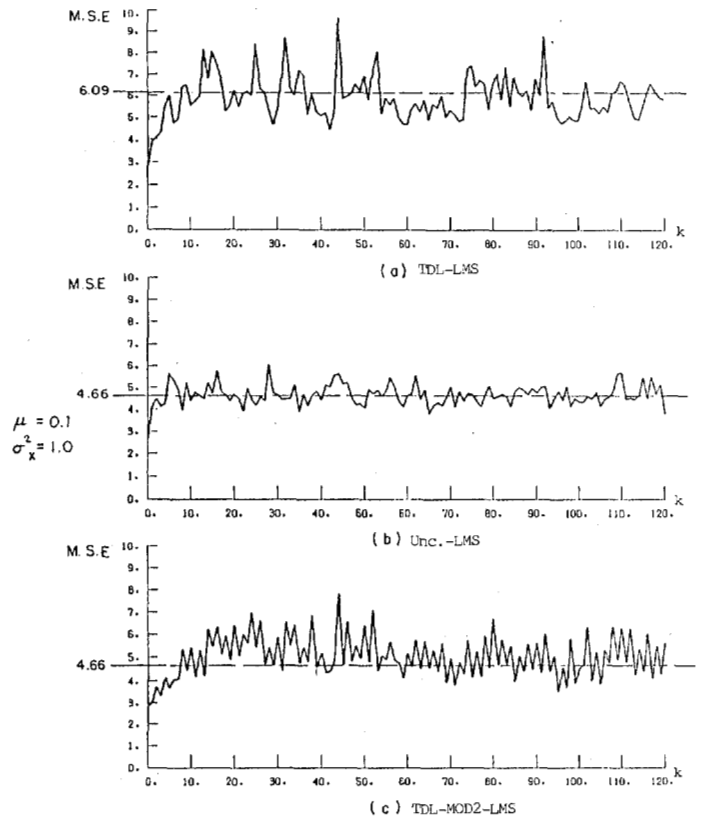


Fig. 5. MSE behavior for the three processings in Case 2, $\mu = 0.1$.

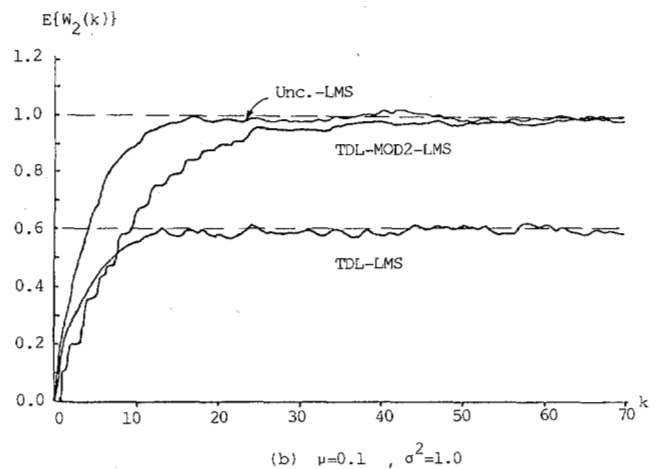
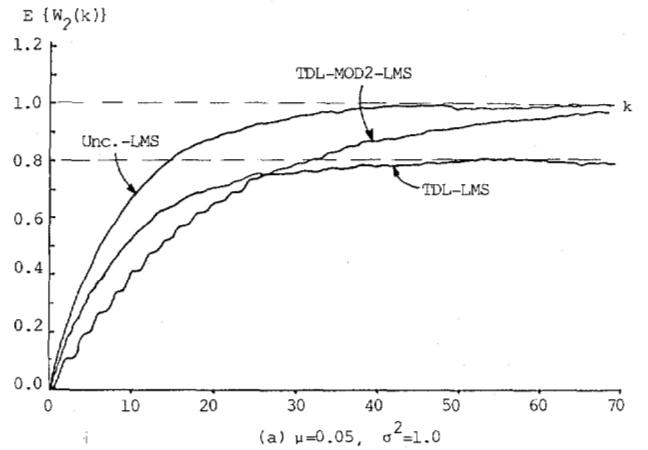


Fig. 6. $E\{W_2(k)\}$ behavior for the three processings in Case 2.

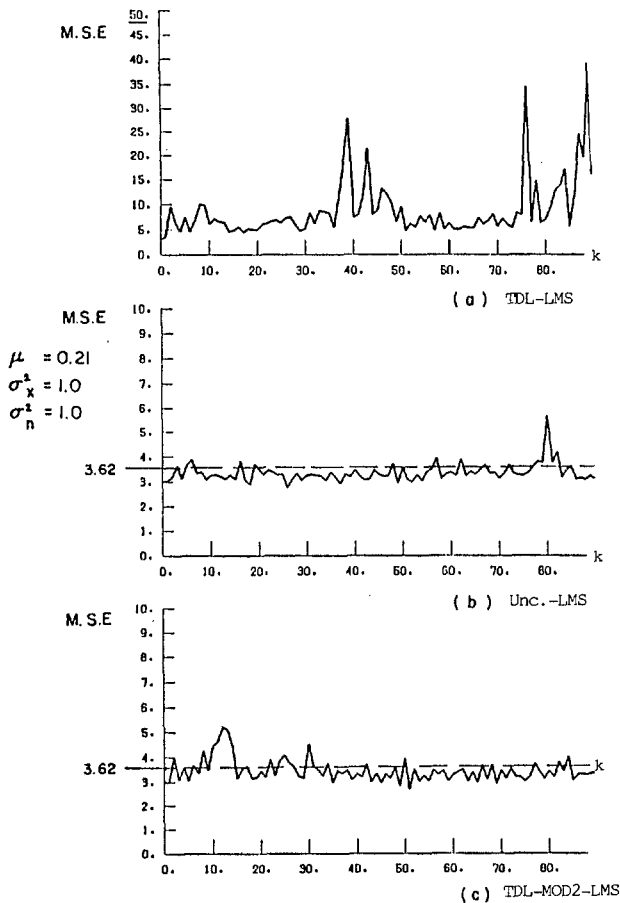


Fig. 7. Verification of stability bound (Case 1).

The experiments on Case 1 showed identical weight mean behavior in the TDL-LMS and the Unc.-LMS and twice slower convergence with the TDL-MOD2-LMS. In Fig. 3 we present the MSE behavior for this case. We observe that, as expected, there is a difference in steady-state values for the TDL-LMS and the Unc.-LMS, with the first being higher. The steady-state value for the TDL-MOD2-LMS is equal to the one for the Unc.-LMS, but its convergence is slower.

The same behavior of the MSE can be observed in the experiments on Case 2 for different choices of μ in Figs. 4 and 5. But here we can observe the bias in the weight mean as shown in Fig. 6. However, one should note that if the weight update is done every two samples, namely, in the TDL-MOD2-LMS, the bias disappears (with the tradeoff in slower convergence).

Finally, in Fig. 7 we demonstrate that a choice $\gamma = 0.21$ causes the TDL-LMS to diverge, while the other two still converge.

VI. CONCLUSION

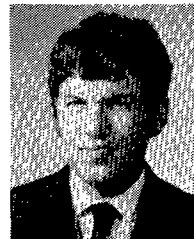
In this paper, we have presented exact analytical results for the LMS algorithm applied to special types of correlated data vectors—constructed from a TDL. This is a very common case in practical applications, but there are no

exact results for it in the literature. For the two-dimensional case, we have shown that the correlation does not affect the rate of weight mean convergence, but may cause (maybe for pathological cases only) bias in the steady-state values. For the MSE, we have shown that the correlation in the data results in higher MSE in the steady state, while, again, the transient behavior is hardly affected.

In conclusion, analyzing the LMS with correlated data under the assumption that this correlation does not exist, as was done in a number of results in the literature, may give a misleadingly optimistic performance picture that does not fit the reality.

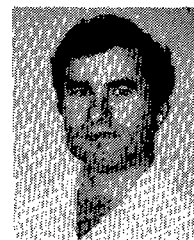
REFERENCES

- [1] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodin, "Adaptive noise cancellation: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692-1716, 1975.
- [2] L. L. Horowitz and K. S. Senne, "Performance advantage of complex LMS for controlling narrow-band adaptive arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 722-736, June 1981.
- [3] W. A. Gardner, "Learning characteristics of stochastic-gradient-descent algorithm: A general study, analysis and critique," *Signal Processing*, No. 6. Amsterdam, The Netherlands: North-Holland, 1984, pp. 113-133.
- [4] 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.
- [5] T. P. Daniell, "Adaptive estimation with mutually collated training sequences," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-6, pp. 12-19, Jan. 1970.
- [6] D. C. Farden, "Tracking properties of adaptive signal processing algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 439-446, June 1981.
- [7] S. K. Jones, R. K. Cavin, and W. M. Reed, "Analysis of error-gradient adaptive linear estimators for a class of stationary dependent processes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 318-329, Mar. 1982.
- [8] O. M. Macchi and E. Eweda, "Second-order convergence analysis of stochastic adaptive linear filtering," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 76-85, Jan. 1983.



Shaul Florian received the B.Sc. and M.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, in 1977 and 1986, respectively.

Since 1977 he has been employed by RAFAEL (Armament Development Authority) doing research on optimal waveform design and implementation for EW systems. His current research interests are in adaptive algorithms and optimal signal processing.



Arie Feuer received the B.Sc. and M.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, in 1967 and 1973, respectively, and the Ph.D. degree from Yale University, New Haven, CT, in 1978.

From 1967 to 1970 he was employed by Technomatic—Israel, working on factory automation. From 1978 through 1983 he was with Bell Laboratories, Holmdel, NJ, studying telephone network performance. Since 1983 he has been on the Faculty of the Department of Electrical Engineering, Technion—Israel Institute of Technology. His research interests are in adaptive control and adaptive signal processing.