

Analysis of the Error Signal of the LMS Algorithm

Wilmar Hernandez, *Member, IEEE*, M. Elena Domínguez, and Gabriela Sansigre

Abstract—An analysis of the error signal of the Least-Mean-Square (LMS) algorithm is conducted from the robust control theory viewpoint. The difference equation that relates the input of the LMS algorithm and the error signal is presented. This equation is used to build the matrix S that maps the input vector to the error vector. It is shown that S has at least one singular value greater than 1. Therefore, the system may amplify noise at high frequencies. Nevertheless, the tap-weight vector may be chosen to prevent that noise amplification and improve the disturbance rejection performance of the LMS algorithm.

Index Terms—Error signal, LMS algorithm, singular values, tap-weight vector.

I. INTRODUCTION

THE LMS algorithm has the diversity to be used in a broad range of applications [1], [2]. However, testing the robustness of such an algorithm has not been an easy task and several researchers have tried it from the control theory point of view [2]–[9]. Also, in [10] the authors carried out a study of the stability of the algorithm.

In this paper, we formulate the LMS as usual [2], [7]. The entries of the algorithm are $\mathbf{u}(n)$ (the tap-reference vector at the time instant n), $d(n)$ (the primary input at the time instant n) and $\hat{\mathbf{w}}(0)$ (the initial value of the tap-weight vector of the transversal filter). After N iterations (starting from $n = 0$), we obtain the error vector $\mathbf{e} = [e(0) \dots e(N-1)]^T$ and the vector \mathbf{r} given by

$$\mathbf{r} = \begin{bmatrix} d(0) - \hat{\mathbf{w}}^H(0)\mathbf{u}(0) \\ \vdots \\ d(N-1) - \hat{\mathbf{w}}^H(0)\mathbf{u}(N-1) \end{bmatrix}. \quad (1)$$

Here we show that, under certain conditions,

$$G_{re} := \frac{\|\mathbf{e}\|}{\|\mathbf{r}\|} > 1. \quad (2)$$

By taking into consideration the one degree-of-freedom control configuration for the LMS algorithm shown in [3], from [11] it can be seen that G_{re} is equal to or less than the ∞ -norm of the sensitivity function [3]–[5]. Therefore, for the case under anal-

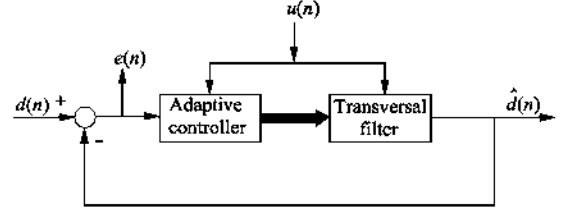


Fig. 1. General block diagram of LMS adaptive filter.

ysis, the magnitude of such a function may be greater than 1. What is more, as this function is the same as the transfer function from the disturbance on the plant output to the estimate of the important information [2] (i.e., the output signal in [3]), (2) means that the LMS algorithm may amplify measurement noise and disturbances on the plant output at high frequencies [11], [12].

Moreover, in this paper it is shown that in spite of the fact that the designer chooses the value of the step-size parameter for H^∞ optimality as given in [2], [8], [9], G_{re} may still be greater than 1. The methods developed in [3]–[7] are only valid for deterministic or stationary signals. They use modern control (stochastic control) or classical control techniques [13], [14], and they do not guarantee that $G_{re} \leq 1$.

Finally, we show that this problem may be solved by choosing a suitable initial value of the tap-weight vector of the transversal filter $\hat{\mathbf{w}}(0)$.

II. PROBLEM FORMULATION

We consider the closed-loop system shown in Fig. 1 (see [2, Fig. 5.1]).

At the time instant n , $d(n)$ and $u(n)$ are samples of the input signals—primary and reference signals, respectively—of the algorithm used to build the input vector \mathbf{r} defined in (1), where the tap-reference vector $\mathbf{u}(n)$ is given by

$$\mathbf{u}(n) = [u(n) \quad \dots \quad u(n-M+1)]^T \quad (3)$$

and M is the length of the transversal filter. Also, $\hat{d}(n)$ is an estimate of the important information, and $\hat{e}(n)$ is the control error.

The gain from the input \mathbf{r} to the error \mathbf{e} is defined by

$$G_{re} = \frac{\|\mathbf{e}\|}{\|\mathbf{r}\|}. \quad (4)$$

The *problem* is to design an LMS adaptive filter such that the gain G_{re} given by (4) meets the performance requirement that $G_{re} \leq 1$.

III. ERROR SIGNAL OF THE LMS ALGORITHM

For the analysis of the LMS algorithm, let us consider the input and output signals as they are defined in Section II and

shown in Fig. 1. Also, at the time instant n , the tap-weight vector of the transversal filter is given by

$$\hat{\mathbf{w}}(n) = [\hat{w}_0(n) \quad \hat{w}_1(n) \quad \cdots \quad \hat{w}_{M-1}(n)]^T \quad (5)$$

and the estimate of the important information is

$$\hat{d}(n) = \hat{\mathbf{w}}^H(n)\mathbf{u}(n). \quad (6)$$

According to [1], [2], the LMS algorithm can be summarized in the following two steps:

- 1) The algorithm should be initialized by using a prior knowledge of $\hat{\mathbf{w}}(0)$ if available or, if not, set $\hat{\mathbf{w}}(0) = \mathbf{0}$.
- 2) From $n = 0, \dots, N - 1$ compute the error signal and update the tap-weight vector of the transversal filter by using (7):

$$\begin{aligned} e(n) &= d(n) - \hat{d}(n) \\ \hat{\mathbf{w}}(n+1) &= \hat{\mathbf{w}}(n) + \mu e^*(n)\mathbf{u}(n) \end{aligned} \quad (7)$$

where μ is the step-size parameter.

Therefore, by iterating (7) $\hat{\mathbf{w}}(n)$ and $d(n)$ can be written as

$$\begin{aligned} \hat{\mathbf{w}}(n) &= \hat{\mathbf{w}}(0) + \mu \sum_{k=0}^{n-1} e^*(k)\mathbf{u}(k) \\ d(n) &= c(n) + \hat{\mathbf{w}}^H(0)\mathbf{u}(n) + \mu \sum_{k=0}^{n-1} e(k)\mathbf{u}^H(k)\mathbf{u}(n). \end{aligned} \quad (8)$$

Thus, if the LMS algorithm is run for N iterations, starting from $n = 0$, and we define $c(n) = \hat{\mathbf{w}}^H(0)\mathbf{u}(n)$, then (9) can be written in matrix form as follows

$$\mathbf{d} = \mathbf{c} + \mathbf{Q}\mathbf{e} \quad (10)$$

where \mathbf{e} is the error vector, $\mathbf{d} = [d(0) \cdots d(N-1)]^T$, $\mathbf{c} = [c(0) \cdots c(N-1)]^T$, $\mathbf{Q} = \mathbf{I} + \mu\mathbf{T}$ and

$$\mathbf{T} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \mathbf{u}^H(0)\mathbf{u}(1) & 0 & \cdots & 0 \\ \mathbf{u}^H(0)\mathbf{u}(2) & \mathbf{u}^H(1)\mathbf{u}(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}^H(0)\mathbf{u}(N-1) & \mathbf{u}^H(1)\mathbf{u}(N-1) & \cdots & 0 \end{bmatrix} \quad (11)$$

\mathbf{T} is a strictly lower triangular matrix and it does not vanish unless the reference input signal is a Delta signal. However, this case will not be taken into consideration in this paper. So, from now on, $\mathbf{T} \neq \mathbf{0}$.

Since \mathbf{Q} is invertible, the error vector can be given by

$$\mathbf{e} = \mathbf{Q}^{-1}(\mathbf{d} - \mathbf{c}). \quad (12)$$

Finally, defining $\mathbf{S} = \mathbf{Q}^{-1}$ and recalling that $\mathbf{r} = \mathbf{d} - \mathbf{c}$ we obtain

$$\mathbf{e} = \mathbf{S}\mathbf{r}. \quad (13)$$

IV. ANALYSIS OF THE ERROR SIGNAL OF THE LMS ALGORITHM

Before presenting the fundamental theorems of this section, recall that the norm of a matrix \mathbf{A} is defined as

$$\|\mathbf{A}\| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \quad (14)$$

and this supremum is achieved [15]. Also, $\|\mathbf{A}\| = \sigma_{\max}(\mathbf{A})$ (largest singular value of \mathbf{A}). If \mathbf{A} is invertible, then

$$\|\mathbf{A}^{-1}\| = \sigma_{\max}(\mathbf{A}^{-1}) = \frac{1}{\sigma_{\min}(\mathbf{A})} \quad (15)$$

where $\sigma_{\min}(\mathbf{A})$ is the least singular value of \mathbf{A} .

Having made the above statements, at this point it is important to highlight that the performance behavior of the LMS algorithm is related to the value of the quotient given by (4): $G_{re} = \|\mathbf{e}\|/\|\mathbf{r}\|$ and the aim of this section is to analyze the gain G_{re} of the LMS algorithm.

Theorem 1: Let $\mathbf{L} \neq \mathbf{0}$ be a square strictly lower triangular matrix (i.e., $\ell_{i,j} = 0$, $i \leq j$) and \mathbf{I} the identity matrix; define $\mathbf{A} := \mathbf{I} + \mathbf{L}$ then

$$0 < \sigma_{\min}(\mathbf{A}) < 1 < \sigma_{\max}(\mathbf{A}). \quad (16)$$

Proof: Let $\mathbf{L}_j \neq \mathbf{0}$ be the j th column of \mathbf{L} and write it as

$$\mathbf{L}_j = \begin{bmatrix} \mathbf{0} \\ \mathbf{b}_j \end{bmatrix}. \quad (17)$$

If \mathbf{I}_j is the unit vector with 1 in the j th position, we have

$$\begin{aligned} \sigma_{\max}(\mathbf{A}) &= \|\mathbf{A}\| \geq \|\mathbf{A}\mathbf{I}_j\| = \|\mathbf{I}_j + \mathbf{L}_j\| \\ &= \sqrt{1 + \|\mathbf{b}_j\|^2} > 1. \end{aligned} \quad (18)$$

The other inequality follows easily taking into account that the norm of \mathbf{A}^{-1} is greater than 1 because \mathbf{A}^{-1} has the same structure as \mathbf{A} , i.e., lower triangular with 1's on the diagonal. Then $\|\mathbf{A}^{-1}\| > 1$ and

$$\sigma_{\min}(\mathbf{A}) = \frac{1}{\|\mathbf{A}^{-1}\|} < 1. \quad (19)$$

Theorem 2: For any choice of $\hat{\mathbf{w}}(0)$, $\mu \neq 0$ and \mathbf{u} (with at least two nonzero components), there exists a vector \mathbf{d}_0 such that the corresponding \mathbf{r}_0 and the error \mathbf{e}_0 verify

$$\frac{\|\mathbf{e}_0\|}{\|\mathbf{r}_0\|} > 1. \quad (20)$$

Proof: The algorithm admits the matrix expression $\mathbf{e} = \mathbf{S}\mathbf{r}$ of (13). Taking norms, we consider the relative error

$$\frac{\|\mathbf{e}\|}{\|\mathbf{r}\|} = \frac{\|\mathbf{S}\mathbf{r}\|}{\|\mathbf{r}\|} \leq \|\mathbf{S}\| \quad (21)$$

and, in accordance with (14), the spectral norm of \mathbf{S} is achieved for a vector \mathbf{r}_0 and following Theorem 1, it is greater than 1:

$$\frac{\|\mathbf{S}\mathbf{r}_0\|}{\|\mathbf{r}_0\|} = \|\mathbf{S}\| = \sigma_{\max}(\mathbf{S}) > 1. \quad (22)$$

Once \mathbf{r}_0 is given, for the reference signal $\mathbf{d}_0 = \mathbf{r}_0 + \mathbf{c}$, the corresponding gain [see (2)] verifies $G_{r_0, e_0} > 1$. ■

Remark 1:

- Note that being \mathbf{T} a nilpotent matrix of order N , one has

$$\mathbf{S} = (\mathbf{I} + \mu\mathbf{T})^{-1} = \mathbf{I} - \mu\mathbf{T} + \dots + (-1)^{N-1}\mu^{N-1}\mathbf{T}^{N-1}.$$

It is obvious that the j th column of \mathbf{S} can be written as

$$\mathbf{S}_j = \begin{bmatrix} \mathbf{0} \\ 1 \\ \mu \mathbf{v}_j \end{bmatrix} \quad (23)$$

for some vector \mathbf{v}_j depending (polynomially) itself on μ .

Hence, for any value of $\mu \neq 0$,

$$\|\mathbf{S}\| \geq \|\mathbf{S}_j\| = (1 + |\mu|^2 \|\mathbf{v}_j\|^2)^{1/2} > 1. \quad (24)$$

- Considering the LMS algorithm as a system that transforms the excitation \mathbf{r} into the error \mathbf{e} , it can be said that the gain of such a system may be greater than 1 regardless the choice of $\hat{\mathbf{w}}(0)$ and $\mu \neq 0$.

Corollary 1: Given the matrix \mathbf{S} , for each value of $\alpha \in [\sigma_{\min}(\mathbf{S}), \sigma_{\max}(\mathbf{S})]$ there exists a vector \mathbf{x} such that

$$\frac{\|\mathbf{S}\mathbf{x}\|}{\|\mathbf{x}\|} = \alpha. \quad (25)$$

In particular, for $\alpha \leq 1$, if $\mathbf{x} = \mathbf{d} - \mathbf{c}$ for some $\hat{\mathbf{w}}(0)$ (recall that the components of \mathbf{c} are $c(n) = \hat{\mathbf{w}}^H(0)\mathbf{u}(n)$), then the corresponding error $\mathbf{e} = \mathbf{S}(\mathbf{d} - \mathbf{c})$ will be bounded in norm by $\|\mathbf{d} - \mathbf{c}\|$. The aim of next section is to discuss the best choice of $\hat{\mathbf{w}}(0)$ and when it is feasible.

V. BEST CHOICE OF $\hat{\mathbf{w}}(0)$

As we have shown, given any \mathbf{d} , \mathbf{u} , μ and $\hat{\mathbf{w}}(0)$, the error norm $\|\mathbf{e}\|$ is *not always* guaranteed to be bounded by $\|\mathbf{r}\|$: according to Corollary 1, for some choices of $\hat{\mathbf{w}}(0)$ we may obtain $\|\mathbf{e}\| > \|\mathbf{r}\|$, and for other choices of $\hat{\mathbf{w}}(0)$ we may obtain $\|\mathbf{e}\| \leq \|\mathbf{r}\|$. Hence, the key question arises: how can we choose $\hat{\mathbf{w}}(0)$ such that the associate vector \mathbf{c} guarantees that $\|\mathbf{e}\| \leq \|\mathbf{d} - \mathbf{c}\|$?

To this aim, let us compute the singular value decomposition of \mathbf{S} : $\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ where \mathbf{U} , \mathbf{V} are unitary matrices, and $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_N)$ is the diagonal matrix which contains the singular values of \mathbf{S} , in decreasing order. As we know, some of them are greater than 1, and some others are less than 1. Let K denote the number of singular values greater than 1:

$$\sigma_1 \geq \dots \geq \sigma_K > 1 \geq \sigma_{K+1} \geq \dots > \dots \geq \sigma_N.$$

With this notation we state that

$$\|\mathbf{e}\| = \|\mathbf{S}\mathbf{r}\| = \|\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H\mathbf{r}\| = \|\mathbf{\Sigma}\mathbf{V}^H\mathbf{r}\| = \|\mathbf{\Sigma}\mathbf{y}\| \quad (26)$$

where we have defined $\mathbf{y} = \mathbf{V}^H\mathbf{r}$, and we have used the fact that the unitary matrices \mathbf{U} , \mathbf{V} preserve the Euclidean norm. By the same reason, $\|\mathbf{y}\| = \|\mathbf{V}^H\mathbf{r}\| = \|\mathbf{r}\|$. Hence,

$$\|\mathbf{e}\| \leq \|\mathbf{r}\| \iff \|\mathbf{\Sigma}\mathbf{y}\| \leq \|\mathbf{y}\| \quad (27)$$

and this property is fulfilled if and only if

$$\sum_{j=1}^N \sigma_j^2 |y_j|^2 \leq \sum_{j=1}^N |y_j|^2. \quad (28)$$

Unfortunately, not every vector \mathbf{y} satisfies this condition, because not every σ_j is less than or equal to 1.

But not everything is lost: if we impose the K first components of \mathbf{y} to be null, then condition (28) is satisfied:

$$\sum_{j=K+1}^N \sigma_j^2 |y_j|^2 \leq \sum_{j=K+1}^N |y_j|^2 \quad (29)$$

since $\sigma_j \leq 1$ for $K+1 \leq j \leq N$. In other words, it suffices to choose \mathbf{c} such that $\mathbf{y} = \mathbf{V}^H\mathbf{r}$ verifies $y_1 = \dots = y_K = 0$; in this case, $\|\mathbf{e}\| \leq \|\mathbf{r}\|$ will be guaranteed.

This sufficient condition is equivalent to the matrix equation

$$\mathbf{V}_K^H \mathbf{r} = \mathbf{0} \quad (30)$$

where \mathbf{V}_K^H is the submatrix of \mathbf{V}^H containing its first K rows.

As $\mathbf{r} = \mathbf{d} - \mathbf{c}$, it suffices to solve the linear system

$$\mathbf{V}_K^H \mathbf{c} = \mathbf{V}_K^H \mathbf{d}. \quad (31)$$

But \mathbf{c} cannot be any arbitrary vector: recall that $c(n) = \hat{\mathbf{w}}^H(0)\mathbf{u}(n) = \mathbf{u}^T(n)\hat{\mathbf{w}}^*(0)$. Matricially,

$$\begin{aligned} \mathbf{c} &= \begin{bmatrix} \mathbf{u}^T(0) \\ \mathbf{u}^T(1) \\ \vdots \\ \mathbf{u}^T(N-1) \end{bmatrix} \hat{\mathbf{w}}^*(0) = \\ &= \begin{bmatrix} u(0) & 0 & \dots & 0 \\ u(1) & u(0) & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ u(N-1) & \dots & \dots & u(N-M) \end{bmatrix} \hat{\mathbf{w}}^*(0) \\ &= \mathbf{M}\hat{\mathbf{w}}^*(0) \end{aligned} \quad (32)$$

where \mathbf{M} is an N -by- M lower triangular Toeplitz matrix whose j th row contains vector $\mathbf{u}^T(j)$; besides, its columns contain the signal \mathbf{u} , windowed by a shifted window of width N .

By introducing (32) in (31), we have proven the following result:

Theorem 3: If $\hat{\mathbf{w}}(0)$ verifies

$$\mathbf{V}_K^H \mathbf{M}\hat{\mathbf{w}}^*(0) = \mathbf{V}_K^H \mathbf{d} \quad (33)$$

then $\hat{\mathbf{w}}(0)$ is the best choice for initial vector of the LMS algorithm, since it guarantees that $\|\mathbf{e}\| \leq \|\mathbf{r}\|$.

If the system (33) is consistent but has infinite solutions, we consider the minimal-norm one. If (33) is inconsistent, we compute the minimal-norm least-squares solution.

Alternatively, for the inconsistent cases, we also propose to reduce the linear system (33) by only considering its first M equations:

$$\mathbf{V}_M^H \mathbf{M}\hat{\mathbf{w}}^*(0) = \mathbf{V}_M^H \mathbf{d} \quad (34)$$

where matrix \mathbf{V}_M^H is used instead of \mathbf{V}_K^H . This way, the new coefficient matrix $\mathbf{V}_M^H \mathbf{M}$ is M -by- M ; if it is invertible, we compute the unique solution $\hat{\mathbf{w}}(0)$ of the system (34). Although it is not the best choice, because it is not a solution of (33), we have used it satisfactorily in the Matlab simulations of the next section.

VI. MATLAB SIMULATIONS

In order to show the importance of choosing an appropriate value of $\hat{\mathbf{w}}(0)$ to meet the requirement that

$$G_{rc} = \frac{\|\mathbf{e}\|}{\|\mathbf{r}\|} \leq 1 \quad (35)$$

the results of several Matlab simulations for different choices of μ and $\hat{\mathbf{w}}(0)$ are presented in this section. In these simulations the LMS adaptive filter shown in Fig. 1 is used as an adaptive noise canceller [1], [2].

Here *primary signal* \mathbf{d} consists of a *relevant signal* that is a sinusoid of amplitude 2 V at 20 Hz corrupted by additive noise. Thus, \mathbf{d} is given by

$$\mathbf{d} = 2 \sin(2\pi 20t) + n_1,$$

where the interval of time of analysis is 0.1 seconds, the sampling frequency is 10 kHz, $t = 0 : 0.0001 : 0.1$. Also, let us assume both that part of the noise corrupts the primary signal without passing through any transfer function, and that part of the noise reaches the primary signal via a transfer function [7] that is a second order low-pass Butterworth filter with cutoff frequency equal to 100 Hz:

$$n_1 = \text{lsim}(\text{D1}, \text{N1}, \text{randn}(\text{size}(t)), t)' + 0.5 \text{randn}(\text{size}(t))$$

where $[\text{D1}, \text{N1}] = \text{butter}(2, 2\pi 100, 's')$.

In addition, assume both that part of the noise goes directly to the *reference signal* \mathbf{u} without passing through any transfer function, and that part of the noise reaches the *reference signal* via a transfer function [7] that is a second order low-pass Butterworth filter with cutoff frequency equal to 50 Hz,

$$\mathbf{u} = \text{lsim}(\text{D2}, \text{N2}, \text{randn}(\text{size}(t)), t)' + 0.5 \text{randn}(\text{size}(t))$$

and $[\text{D2}, \text{N2}] = \text{butter}(2, 2\pi 50, 's')$.

In these simulations, the length of the filter is $M = 20$ and, for any value of the parameter a of Table I, the values of μ are given by

$$\mu = \frac{a}{\max_{0 \leq n \leq N-1} \|\mathbf{u}(n)\|^2} - 0.0001.$$

In Table I, for each value of a a value of μ is obtained. Then, the gain G_{rc} is displayed for three different choices of the initial value of the tap-weight vector:

- $\hat{\mathbf{w}}(0)_0 = \mathbf{0}$;
- $\hat{\mathbf{w}}(0)_1$ is the conjugate of the minimal-norm least-squares solution of (33);
- $\hat{\mathbf{w}}(0)_2$ is the conjugate of the solution of (34).

Note that, in this example, when choosing $\hat{\mathbf{w}}(0)_0 = \mathbf{0}$ the energy gain G_{rc} is *always* greater than 1. However, for $\hat{\mathbf{w}}(0)_1$ G_{rc} is less than 1 for *small enough* values of μ ($0.1 \leq a \leq 2.5$) and for $\hat{\mathbf{w}}(0)_2$ G_{rc} is less than 1 for almost all the values of μ ($0.1 \leq a \leq 5.0$).

In order to show how different tap-weight vectors affect the convergence rate of the LMS algorithm, additional simulations

TABLE I
GAIN G_{rc} FOR DIFFERENT CHOICES OF μ AND $\hat{\mathbf{w}}(0)$

a	μ	$\hat{\mathbf{w}}(0)_0$	$\hat{\mathbf{w}}(0)_1$	$\hat{\mathbf{w}}(0)_2$
0.1	$9.20 \cdot 10^{-3}$	1.01	0.48	0.47
0.5	$3.77 \cdot 10^{-2}$	1.11	0.46	0.22
1.0	$9.12 \cdot 10^{-2}$	1.27	0.28	0.16
2.0	$1.61 \cdot 10^{-1}$	1.49	0.49	0.13
2.5	$2.00 \cdot 10^{-1}$	1.90	0.68	0.13
3.0	$2.76 \cdot 10^{-1}$	3.38	1.48	0.11
3.5	$3.01 \cdot 10^{-1}$	12.41	4.37	0.11
4.0	$4.19 \cdot 10^{-1}$	85.88	27.92	$9.82 \cdot 10^{-2}$
4.5	$4.50 \cdot 10^{-1}$	$1.15 \cdot 10^6$	$3.38 \cdot 10^5$	$8.37 \cdot 10^{-2}$
5.0	$5.00 \cdot 10^{-1}$	$6.76 \cdot 10^{11}$	$2.14 \cdot 10^{11}$	$9.76 \cdot 10^{-2}$
5.5	$6.00 \cdot 10^{-1}$	$1.06 \cdot 10^{16}$	$9.29 \cdot 10^{15}$	4.61

were carried out. As a result, the fastest convergence rate was achieved with $\hat{\mathbf{w}}(0)_1$ and the second fastest with $\hat{\mathbf{w}}(0)_2$. When $a = 0.1$, for $\hat{\mathbf{w}}(0)_0$ the LMS algorithm converged in 1057 iterations, for $\hat{\mathbf{w}}(0)_1$ it converged in 162 iterations, and for $\hat{\mathbf{w}}(0)_2$ it converged in 719 iterations. Finally, when $a = 3.5$, for $\hat{\mathbf{w}}(0)_0$ the LMS algorithm did not converge, for $\hat{\mathbf{w}}(0)_1$ it converged in 68 iterations, and for $\hat{\mathbf{w}}(0)_2$ it converged in 140 iterations.

VII. CONCLUSION

In this paper, it has been shown that for the LMS algorithm the gain from the input signal to the error signal may be greater than 1 regardless of the value of $\mu \neq 0$. Also, in order to diminish such a gain and guarantee it to be less than or equal to 1, a novel approach has been presented. Finally, it has also been shown that, for reasonable values of μ , choosing the right initial value of the tap-weight vector $\hat{\mathbf{w}}(0)$ is the key issue to avoid amplification of the linear system from the input to the error signal.

REFERENCES

- [1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [2] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002, ch. 5, pp. 231–312.
- [3] C. P. Kwong, "Control-theoretic design of the LMS and the sign algorithms in nonstationary environments," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, pp. 253–259, Feb. 1990.
- [4] C. P. Kwong, "Further results on the control-theoretic analysis of the LMS algorithm," *IEEE Trans. Signal Process.*, vol. 41, pp. 943–946, Feb. 1993.
- [5] C. P. Kwong, "Robust design of the LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, pp. 2613–2616, Feb. 1992.
- [6] H. S. Dabis and T. J. Moir, "Least mean squares as a control system," *Int. J. Control*, vol. 54, pp. 321–335, 1991.
- [7] T. J. Moir, "On feed-through terms in the LMS algorithm," *Res. Lett. Inf. Math. Sci.*, vol. 6, pp. 71–82, 2004.
- [8] B. Hassibi, A. H. Sayed, and T. Kailath, "LMS is H^∞ optimal," in *Proc. IEEE Conf. Decision Contr.*, San Antonio, Texas, Dec. 1993, vol. 1, pp. 74–79.
- [9] B. Hassibi, A. H. Sayed, and T. Kailath, " H^∞ optimality of the LMS algorithm," *IEEE Trans. Signal Process.*, vol. 44, pp. 267–280, Feb. 1996.
- [10] L. Guo, L. Ljung, and G. -J. Wang, "Necessary and sufficient conditions for stability of LMS," *IEEE Trans. Automat. Contr.*, vol. 42, pp. 761–770, Jun. 1997.
- [11] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback Control Theory*. New York: McMillan, 1992.
- [12] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Upper Saddle River, NJ: Prentice-Hall, 1996.
- [13] B. D. O. Anderson and J. B. Moore, *Optimal Control: Linear Quadratic Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [14] B. C. Kuo, *Automatic Control Systems*, 7th ed. Upper Saddle River, NJ: Prentice-Hall, 1995.
- [15] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.