A Robust Variable Step-Size LMS-Type Algorithm: Analysis and Simulations

Tyseer Aboulnasr, *Member, IEEE*, and K. Mayyas

***Abstract*—A number of time-varying step-size algorithms have been proposed to enhance the performance of the conventional LMS algorithm. Experimentation with these algorithms indicates that their performance is highly sensitive to the noise disturbance. This paper presents a robust variable step-size LMS-type algorithm providing fast convergence at early stages of adaptation while ensuring small final misadjustment. The performance of the algorithm is not affected by existing uncorrelated noise disturbances. An approximate analysis of convergence and steady-state performance for zero-mean stationary Gaussian inputs and for nonstationary optimal weight vector is provided. Simulation results comparing the proposed algorithm to current variable step-size algorithms clearly indicate its superior performance for cases of stationary environments. For nonstationary environments, our algorithm performs as well as other variable step-size algorithms in providing performance equivalent to that of the regular LMS algorithm.**

I. Introduction

SINCE its introduction, the LMS algorithm has been the focus of much study due to its simplicity and robustness, leading to its implementation in many applications. It is well known that the final excess mean square error (MSE) is directly proportional to the adaptation step size of the LMS while the convergence time increases as the step size decreases. This inherent limitation of the LMS necessitates a compromise between the opposing fundamental requirements of fast convergence rate and small misadjustment demanded in most adaptive filtering applications. As a result, researchers have constantly looked for alternative means to improve its performance. One popular approach is to employ a time-varying step size in the standard LMS weight update recursion [1]–[4]. This is based on using large step-size values when the algorithm is far from the optimal solution, thus speeding up the convergence rate. When the algorithm is near the optimum, small step-size values are used to achieve a low level of misadjustment, thus achieving better overall performance. This can be obtained by adjusting the step-size value in accordance with some criterion that can provide an approximate measure of the adaptation process state. Several criteria have been used: squared instantaneous error [1], sign changes of successive samples of the gradient [3], attempting to reduce the squared error at each instant [2], or cross correlation of input and error [4]. Experimental results show that the performance of existing variable step size (VSS) algorithms is quite sensitive to the noise disturbance [5], [6]. Their advantageous performance over the LMS algorithm is generally attained only in a high signal-to-noise environment. This is intuitively obvious by noting that the criteria controlling the step-size update of these algorithms are directly obtained from the instantaneous error that is contaminated by the disturbance noise. Since measurement noise is a reality in any practical system, the usefulness of any adaptive algorithm is judged by its performance *in the presence of this noise*.

Manuscript received January 20, 1995; revised August 26, 1996. The associate editor coordinating the review of this paper and approving it for publication was Dr. Jose Carlos M. Bermudez.

T. Aboulnasr is with the Department of Electrical Engineering, University of Ottawa, Ottawa, Ont., Canada K1N 6N5.

K. Mayyas is with the Department of Electrical Engineering, Jordan University of Science and Technology, Irbid, Jordan.

Publisher Item Identifier S 1053-587X(97)01890-4.

As an example of the sensitivity of the current algorithms to noise, the algorithm in [1] is studied in Section II in the presence of noise. This algorithm was shown to outperform the fixed step-size LMS algorithm and to exhibit favorable performance over other existing VSS algorithms [1]. However, it will be shown that its performance deteriorates in the presence of measurement noise. We then propose a new VSS LMS algorithm, where the step size of the algorithm is adjusted according to the square of the time-averaged estimate of the autocorrelation of $e(n)$ and $e(n-1)$. As a result, the algorithm can effectively adjust the step size as in [1] while maintaining the immunity against independent noise disturbance. It will be shown that the proposed algorithm allows more flexible control of misadjustment and convergence time without the need to compromise one for the other. The algorithm in [4] can be shown to have negative step size and diverge and, thus, will not be considered in this paper.

Convergence and steady-state analyses of the proposed algorithm are introduced in Section III. Section IV presents simulation examples and comparison with the standard LMS algorithm and other variable step-size algorithms as described in [1] and [2]. Conclusions are then given in Section V.

II. Algorithm Formulation

In [1], the adaptation step size is adjusted using the energy of the instantaneous error. The weight update recursion is given by

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu(n)e(n)\mathbf{X}(n) \tag{1}$$

and the step-size update expression is

$$\mu(n+1) = \alpha\mu(n) + \gamma e^2(n) \tag{2}$$

where $0 < \alpha < 1$, $\gamma > 0$, and $\mu(n+1)$ is set to $\mu_{\min}$ or $\mu_{\max}$ when it falls below or above these lower and upper bounds, respectively. The constant $\mu_{\max}$ is normally selected near the point of instability of the conventional LMS to provide the maximum possible convergence speed. The value of $\mu_{\min}$ is chosen as a compromise between the desired level of steady-state misadjustment and the required tracking capabilities of the algorithm. The parameter γ controls the convergence time as well as the level of misadjustment of the algorithm. The algorithm has preferable performance over the fixed step-size LMS: At early stages of adaptation, the error is large, causing the step size to increase, thus providing faster convergence speed. When the error decreases, the step size decreases, thus yielding smaller misadjustment near the optimum. However, using the instantaneous error *energy* as a measure to sense the state of the adaptation process does not perform as well as expected in the presence of measurement noise. This can be seen from (2). The output error of the identification system is

$$e(n) = d(n) - \mathbf{X^T}(n)\mathbf{W}(n) \tag{3}$$

where the desired signal $d(n)$ is given by

$$d(n) = \mathbf{X^T}(n)\mathbf{W}^*(n) + \xi(n) \tag{4}$$

$\xi(n)$ is a zero-mean independent disturbance, and $\mathbf{W}^*(n)$ is the time-varying optimal weight vector. Substituting (3) and (4) in the step-size recursion, we get

$$\begin{aligned}\mu(n+1) = &\alpha\mu(n) + \gamma\mathbf{V^T}(n)\mathbf{X}(n)\mathbf{X^T}(n)\mathbf{V}(n)\\ &+ \gamma\xi^2(n) - 2\gamma\xi(n)\mathbf{V^T}(n)\mathbf{X}(n)\end{aligned} \tag{5}$$

where $\mathbf{V}(n) = \mathbf{W}(n) - \mathbf{W}^*(n)$ is the weight error vector. The input signal autocorrelation matrix, which is defined as $\mathbf{R} = E\{\mathbf{X}(n)\mathbf{X^T}(n)\}$, can be expressed as $\mathbf{R} = \mathbf{Q\Lambda Q^T}$, where $\mathbf{\Lambda}$ is the matrix of eigenvalues, and $\mathbf{Q}$ is the modal matrix of $\mathbf{R}$. Using $\acute{\mathbf{V}}(n) = \mathbf{Q^T}\mathbf{V}(n)$ and $\acute{\mathbf{X}}(n) = \mathbf{Q^T}\mathbf{X}(n)$, then the statistical behavior of $\mu(n+1)$ is determined by taking the expected average of (5)

$$\begin{aligned}E\{\mu(n+1)\} = &\alpha E\{\mu(n)\} + \gamma(E\{\xi^2(n)\}\\ &+ E\{\acute{\mathbf{V}}^\mathbf{T}(n)\mathbf{\Lambda}\acute{\mathbf{V}}(n)\})\end{aligned} \tag{6}$$

where we have made use of the common independence assumption of $\acute{\mathbf{V}}(n)$ and $\acute{\mathbf{X}}(n)$ [7]. Clearly, the term $E\{\acute{\mathbf{V}}^\mathbf{T}(n)\mathbf{\Lambda}\acute{\mathbf{V}}(n)\}$ influences the proximity of the adaptive system to the optimal solution, and $\mu(n+1)$ is adjusted accordingly. However, due to the presence of $E\{\xi^2(n)\}$, the step-size update is not an accurate reflection of the state of adaptation before or after convergence. This reduces the efficiency of the algorithm significantly. More specifically, close to the optimum, $\mu(n)$ will still be large due to the presence of the noise term $E\{\xi^2(n)\}$. This results in large misadjustment due to the large fluctuations around the optimum. In this paper, a different approach is proposed to control step-size adaptation. The objective is to ensure large $\mu(n)$ when the algorithm is far from the optimum with $\mu(n)$ decreasing as we approach the optimum *even in the presence of this noise*. The proposed algorithm achieves this objective by using an estimate of the autocorrelation between $e(n)$ and $e(n-1)$ to control step-size update. The estimate is a time average of $e(n)e(n-1)$ that is described as

$$p(n) = \beta p(n-1) + (1-\beta)e(n)e(n-1). \tag{7}$$

The use of $p(n)$ in the update of $\mu(n)$ serves two objectives. First, the error autocorrelation is generally a good measure of the proximity to the optimum. Second, it rejects the effect of the uncorrelated noise sequence on the step-size update. In the early stages of adaptation, the error autocorrelation estimate $p^2(n)$ is large, resulting in a large $\mu(n)$. As we approach the optimum, the error autocorrelation approaches zero, resulting in a smaller step size. This provides the fast convergence due to large initial $\mu(n)$ while ensuring low misadjustment near optimum due to the small final $\mu(n)$ *even in the presence of* $\xi(n)$. Thus, the proposed step size update is given by

$$\mu(n+1) = \alpha\mu(n) + \gamma p(n)^2 \tag{8}$$

where limits on $\mu(n+1)$, α, and γ are the same as those of the VSS LMS algorithm in [1]. The positive constant β ($0 < \beta < 1$) is an exponential weighting parameter that governs the averaging time constant, i.e., the quality of the estimation. In stationary environments, previous samples contain information that is relevant to determining an accurate measure of adaptation state, i.e., the proximity of the adaptive filter coefficients to the optimal ones. Therefore, β should be $\approx$1. For nonstationary optimal coefficients, the time averaging window should be small enough to allow for forgetting of the deep past and adapting to the current statistics, i.e., $\beta < 1$. The step size in (8) can be rewritten as

$$\begin{aligned}&\mu(n+1) =\\ &\quad\alpha\mu(n) + \gamma[E\{\mathbf{V^T}(n)\mathbf{X}(n)\mathbf{X^T}(n-1)\mathbf{V}(n-1)\}]^2.\end{aligned} \tag{9}$$

Assuming perfect estimation of the autocorrelation of $e(n)$ and $e(n-1)$, we note that as a result of the averaging operation, the instantaneous behavior of the step size will be smoother. It is also clear from (9) that the update of $\mu(n)$ is dependent on how far we are from the optimum and is not affected by independent disturbance noise.

Finally, the proposed algorithm involves two additional update equations ((7) and (8)) compared with the standard LMS algorithm. Therefore, the added complexity is six multiplications per iteration. These multiplications can be reduced to shifts if the parameters α, β, γ are chosen as powers of 2. Compared with the VSS LMS algorithm in [1], the proposed algorithm adds a new equation (7) and a corresponding parameter β.

III. Performance Analysis of the Proposed Algorithm

We will now provide a performance analysis of the proposed algorithm when operating in stationary and nonstationary environments. The input signal is assumed to be a zero-mean, stationary Gaussian. The analysis is conducted assuming exact modeling of the unknown system, i.e., the number of coefficients of the optimal weight vector $\mathbf{W}^*(n)$ is the same as the number of the coefficients of the adaptive filter $\mathbf{W}(n)$. In addition, the measurement noise sequence, which is represented by $\xi(n)$ in (4), is zero-mean and white. Recall that the weight update recursion of the proposed algorithm is given in (1), where $e(n)$ is defined in (3) and (4), and $\mu(n)$ is defined in (7) and (8). The nonstationary environment is modeled by a time-varying optimal weight vector generated by a random walk model [8] as

$$\mathbf{W}^*(n) = \mathbf{W}^*(n-1) + \eta(n-1) \tag{10}$$

where $\eta(n)$ is a stationary noise process of zero-mean and correlation matrix $\sigma_n^2\mathbf{I}$. For a stationary system, $\sigma_n^2 = 0$, and $\mathbf{W}^*(n) = \mathbf{W}^*$. Substituting (3), (4), and (10) in (1) results in

$$\begin{aligned}\acute{\mathbf{V}}(n+1) = &[\mathbf{I} - \mu(n)\acute{\mathbf{X}}(n)\acute{\mathbf{X}}^{\mathbf{T}}(n)]\acute{\mathbf{V}}(n) \\ &+ \mu(n)\xi(n)\acute{\mathbf{X}}(n) - \acute{\eta}(n)\end{aligned} \tag{11}$$

where $\acute{\eta}(n) = \mathbf{Q}^{\mathbf{T}}\eta(n)$. Normally, γ is chosen to be a very small value; hence, $\mu(n)$ is slowly varying when compared with $e(n)$ and $X(n)$. This justifies the independence assumption of $\mu(n)$ and $\mu^2(n)$ with $e(n)$, $\mathbf{W}(n)$, and $\mathbf{X}(n)$. Accordingly, the following condition can be obtained from (11) to ensure convergence of the weight vector mean

$$0 < E\{\mu(n)\} < \frac{2}{\lambda_{\max}} \tag{12}$$

where $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{R}$. However, convergence of the mean weight vector cannot guarantee convergence of the mean square error. Therefore, we need to determine conditions that will ensure convergence in the mean square sense. To evaluate the performance of the system, approximate expressions for misadjustment are derived. This leads to tight conditions for the selection of the parameters α, β, and γ. Finally, we present some guidelines to selecting γ to guarantee MSE convergence while producing the desirable misadjustment level. The MSE is given by [9]

$$\begin{aligned}E\{e^2(n)\} &= \epsilon_{\min} + \epsilon_{ex}(n) \\ &= \epsilon_{\min} + E\{\acute{\mathbf{V}}^{\mathbf{T}}(n)\mathbf{\Lambda}\acute{\mathbf{V}}(n)\}\end{aligned} \tag{13}$$

where $\epsilon_{ex}(n)$ is the excess MSE, and $\epsilon_{\min} = E\{\xi^2(n)\}$ is the minimum value of the MSE. Equation (13) shows that the MSE is directly related to the diagonal elements of $E\{\acute{\mathbf{V}}(n)\acute{\mathbf{V}}^{\mathbf{T}}(n)\}$. Consequently, stability of the MSE is ensured by the stability of these elements. Postmultiplying both sides of (11) by $\acute{\mathbf{V}}^{\mathbf{T}}(n+1)$ and then taking the expected value yields [1]

$$\begin{aligned}&E\{\acute{\mathbf{V}}(n+1)\acute{\mathbf{V}}^{\mathbf{T}}(n+1)\} \\ &\quad= E\{\acute{\mathbf{V}}(n)\acute{\mathbf{V}}^{\mathbf{T}}(n)\} - E\{\mu(n)\}E\{\acute{\mathbf{V}}(n)\acute{\mathbf{V}}^{\mathbf{T}}(n)\}\mathbf{\Lambda} \\ &\qquad - E\{\mu(n)\}\mathbf{\Lambda}E\{\acute{\mathbf{V}}(n)\acute{\mathbf{V}}^{\mathbf{T}}(n)\} \\ &\qquad + 2E\{\mu^2(n)\}\mathbf{\Lambda}E\{\acute{\mathbf{V}}(n)\acute{\mathbf{V}}^{\mathbf{T}}(n)\}\mathbf{\Lambda} \\ &\qquad + E\{\mu^2(n)\}\mathbf{\Lambda}\,\mathrm{tr}\,(\mathbf{\Lambda}E\{\acute{\mathbf{V}}(n)\acute{\mathbf{V}}^{\mathbf{T}}(n)\}) \\ &\qquad + E\{\mu^2(n)\}\epsilon_{\min}\mathbf{\Lambda} + \sigma_n^2\mathbf{I}.\end{aligned} \tag{14}$$

Note that in (14), we have used the Gaussian factoring theorem to simplify the expression $E\{\acute{\mathbf{X}}(n)\acute{\mathbf{X}}^{\mathbf{T}}(n)$ $\acute{\mathbf{V}}(n)\acute{\mathbf{V}}^{\mathbf{T}}(n)\acute{\mathbf{X}}(n)\acute{\mathbf{X}}^{\mathbf{T}}(n)\}$ into a sum of second-order moments [8]. From (7), $p(n)$ can be obtained recursively as

$$p(n) = (1-\beta)\sum_{i=0}^{n-1}\beta^i e(n-i)e(n-i-1) \tag{15}$$

and

$$\begin{aligned}p(n)^2 = &(1-\beta)^2\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\beta^i\beta^j e(n-i) \\ &\cdot e(n-i-1)e(n-j)e(n-j-1).\end{aligned} \tag{16}$$

In the following analysis, we study the steady state performance of the proposed algorithm. Therefore, we assume in our analysis that the algorithm has converged. In this case, the samples of the error $e(n)$ can be assumed uncorrelated, i.e., $E\{e(n-i)e(n-j)\} = 0 \;\forall i \neq j$ for the case of sufficient modeling order. Using (8), (15), and (16), the mean and the mean-square behavior of the step-size $\mu(n)$, upon convergence, are

$$\begin{aligned}&E\{\mu(n+1)\} \\ &\quad= \alpha E\{\mu(n)\} + \gamma(1-\beta)^2\sum_{i=0}^{n-1}\beta^{2i}E\{e^2(n-i)\} \\ &\qquad \cdot E\{e^2(n-i-1)\}\end{aligned} \tag{17}$$

and

$$\begin{aligned}&E\{\mu^2(n+1)\} \\ &\quad= \alpha^2E\{\mu^2(n)\} + 2\alpha\gamma E\{\mu(n)\}(1-\beta)^2\sum_{i=0}^{n-1}\beta^{2i} \\ &\qquad \cdot E\{e^2(n-i)\}E\{e^2(n-i-1)\} + \gamma^2E\{p^4(n)\}.\end{aligned} \tag{18}$$

For small γ, the last term in (18) involving γ^2 is negligible compared with the other terms. Thus

$$\begin{aligned}E\{\mu^2(n+1)\} \approx &\alpha^2E\{\mu^2(n)\} + 2\alpha\gamma E\{\mu(n)\}(1-\beta)^2 \\ &\cdot\sum_{i=0}^{n-1}\beta^{2i}E\{e^2(n-i)\}E\{e^2(n-i-1)\}.\end{aligned} \tag{19}$$

Following the same argument in [1], a sufficient condition that ensures convergence of the MSE is

$$0 < \frac{E\{\mu^2(\infty)\}}{E\{\mu(\infty)\}} \leq \frac{2}{3\,\mathrm{tr}\,(\mathbf{R})} \tag{20}$$

where $E\{\mu(\infty)\}$, and $E\{\mu^2(\infty)\}$ are the steady-state values of $E\{\mu(n)\}$ and $E\{\mu^2(n)\}$. The misadjustment is defined as [9]

$$M = \frac{\epsilon_{ex}(\infty)}{\epsilon_{\min}}. \tag{21}$$

When the algorithm parameters are chosen such that (20) holds, we can use (13) and (14) in (21) to find (22), shown at the bottom of the next page. Recall that

$$E\{e^2(\infty)\} = \epsilon_{\min} + \epsilon_{ex}(\infty) \tag{23}$$

and then, from (17)

$$E\{\mu(\infty)\} = \frac{\gamma(1-\beta)(\epsilon_{\min} + \epsilon_{ex}(\infty))^2}{(1-\alpha)(1+\beta).} \tag{24}$$

Substituting (24) in (19) yields

$$E\{\mu^2(\infty)\} \approx \frac{2\gamma^2\alpha(1-\beta)^2(\epsilon_{\min} + \epsilon_{ex}(\infty))^4}{(1-\alpha^2)(1-\alpha)(1+\beta)^2}. \tag{25}$$

Assuming that $\epsilon_{ex}(\infty) \ll \epsilon_{\min}$ [1], [2], then, from (24) and (25), we have

$$\frac{E\{\mu^2(\infty)\}}{E\{\mu(\infty)\}} \approx y \tag{26}$$

where

$$y = \frac{2\gamma\alpha\epsilon_{\min}^2(1-\beta)}{(1-\alpha^2)(1+\beta)}. \tag{27}$$

Noting that $\gamma > 0$, $0 < \beta < 1$, and $0 < \alpha < 1$, then

$$y < \frac{2\gamma\epsilon_{\min}^2(1-\beta)}{1-\alpha^2}. \tag{28}$$

From (20), (26), and (28), the following condition is imposed on γ, α, and β to guarantee stability of the MSE

$$0 < \frac{\gamma\epsilon_{\min}^2(1-\beta)}{1-\alpha^2} \le \frac{1}{3\,\mathrm{tr}\,(\mathbf{R}).} \tag{29}$$

Substituting (26) in (22) yields the following expression for algorithm misadjustment:

$$M \approx \frac{\displaystyle\sum_{j=1}^{N} \frac{y\lambda_j}{2-2y\lambda_j} + \sum_{j=1}^{N} \frac{\sigma_n^2}{E\{\mu(\infty)\}(2-2y\lambda_j)\epsilon_{\min}}}{1 - \displaystyle\sum_{j=1}^{N} \frac{y\lambda_j}{2-2y\lambda_j}.} \tag{30}$$

In the event of small values of misadjustment so that $\sum_j y\lambda_j \ll 1$

$$M \approx \frac{y}{2}\,\mathrm{tr}\,(\mathbf{R}) + \frac{N\sigma_n^2}{2E\{\mu(\infty)\}\epsilon_{\min}} \tag{31}$$

where

$$E\{\mu(\infty)\} \approx \frac{\gamma(1-\beta)}{(1-\alpha)(1+\beta)}\,\epsilon_{\min}^2. \tag{32}$$

In a stationary environment, $\sigma_n^2 = 0$, and the misadjustment is given by

$$M \approx \frac{y}{2}\,\mathrm{tr}\,(\mathbf{R}). \tag{33}$$

Practically, α, γ, and β are selected to produce the same MSE attained by the fixed step-size LMS (FSS) while improving the convergence speed. Accordingly, α, γ, and β could be selected to satisfy $y \le \mu_{\mathrm{FSS}}$, where μ_{FSS} is the adaptation step size of the FSS algorithm. In a stationary environment, a large number of samples would be used in the estimation of $E\{e(n)e(n-1)\}$, i.e., the exponential weighting parameter β should be close to one. From (32) and (33), this value for β results in a reduced misadjustment. Thus, for the same level of misadjustment, a larger γ can be used while maintaining the stability of the algorithm (see (29)). A larger γ results in a larger step size in the initial stages of adaptation, i.e., faster convergence. Thus, the parameters β and γ provide the algorithm with an extra degree of freedom that facilitates better simultaneous control of both convergence speed and final excess MSE. In [1], the level of misadjustment of the VSS algorithm can be lowered by reducing the value of γ. However, this is achieved at the expense of slower convergence speed. When operating in a nonstationary environment, the choice of β becomes crucial. To provide good tracking capabilities, β should be small to cope with the time-varying statistics of the environment. Although this will decrease the second term in the misadjustment expression in (31), it will at the same time increase the first term. Therefore, the selection of β should achieve a compromise between tracking speed and excess MSE. Since the first term in (31) is directly proportional to γ and the second term is inversely proportional to γ, we can optimize the choice of γ for given α and β to obtain the minimum misadjustment (when both terms are equal). From (27), (31), and (32), the optimal value of γ is

$$\gamma^* = \sqrt{\frac{N\sigma_n^2}{q\epsilon_{\min}\,\mathrm{tr}\,(\mathbf{R})}} \tag{34}$$

where

$$q = \frac{2\alpha\epsilon_{\min}^4(1-\beta)^2}{(1-\alpha^2)(1-\alpha)(1+\beta)^2}. \tag{35}$$

IV. Simulation

Here, the proposed modified variable step-size LMS (MVSS) algorithm is implemented for stationary and non-stationary environments in a system identification setup. The performance of the algorithm is compared with the variable step-size LMS (VSS) algorithm [1], the stochastic gradient algorithm with gradient adaptive step size (SGA-GAS) [2], and the fixed step-size LMS (FSS) algorithm [9]. Parameters of these algorithms are selected to produce a comparable level of misadjustment. Moreover, our choice of these parameters is also guided by the recommended values in their corresponding publication.

In all simulations presented here, except for the final example, the desired signal $d(n)$ is corrupted by zero-mean, uncorrelated Gaussian noise of $\epsilon_{\min}$ variance. Results are obtained by averaging over 200 independent runs.

A. Example 1: White Input, Low SNR

The unknown moving average system has four time-invariant coefficients, and the FIR adaptive filter is of equal order. Both are excited by a zero-mean, white Gaussian signal

$$M = \frac{\displaystyle\sum_{j=1}^{N} \frac{E\{\mu^2(\infty)\}\lambda_j}{2E\{\mu(\infty)\} - 2E\{\mu^2(\infty)\}\lambda_j} + \sum_{j=1}^{N} \frac{\sigma_n^2}{(2E\{\mu(\infty)\} - 2E\{\mu^2(\infty)\}\lambda_j)\epsilon_{\min}}}{1 - \displaystyle\sum_{j=1}^{N} \frac{E\{\mu^2(\infty)\}\lambda_j}{2E\{\mu(\infty)\} - 2E\{\mu^2(\infty)\}\lambda_j}} \tag{22}$$

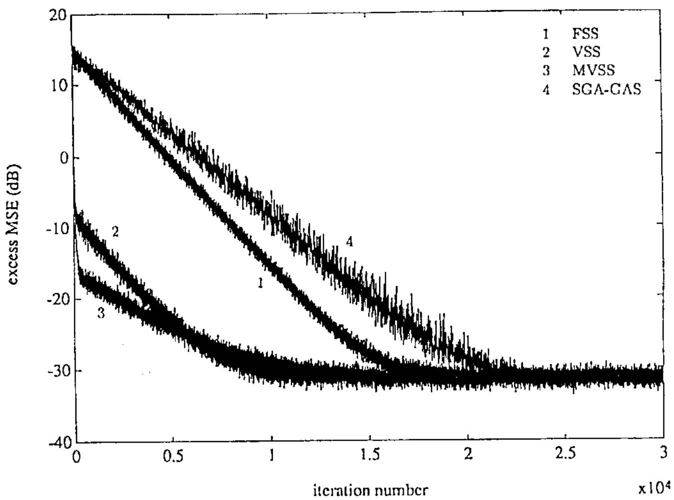

Fig. 1. Comparison of excess MSE of various adaptive algorithms for white input case SNR = 0 dB.

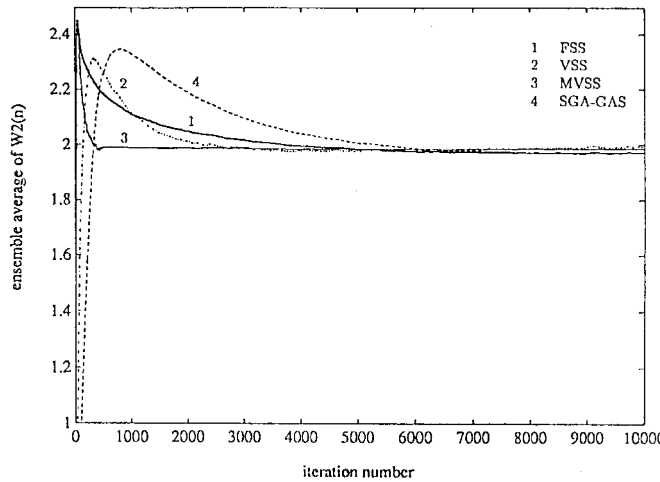

Fig. 2. Ensemble average of the second component of $\mathbf{W}(n)$ for white input case SNR = 0 dB.

of unity variance. $\epsilon_{\min}$ is equal to 1. The proposed MVSS is implemented with the parameters $\alpha = 0.97$ (which is found to be a good choice in stationary and nonstationary environments) and $\beta = 0.99$. In accordance with (33), γ is set to 10^{-3} to produce a steady-state excess MSE of about -34 dB. Note that these parameter values satisfy the condition in (29) for MSE convergence. The VSS algorithm is used with $\alpha_{vss} = 0.97$, and as required by (43) in [1], $\gamma_{vss} = 1 \times 10^{-5}$ to obtain the same level of the steady state excess MSE. For the algorithm in [2], we chose $\rho = 2 \times 10^{-8}$ to obtain a comparable misadjustment value with the other algorithms as required by (45) in [2]. For all previous algorithms, we used $\mu_{\max} = 0.1$, and $\mu_{\min} = 1 \times 10^{-5}$ [1]. In addition, we used $\mu = 3.5 \times 10^{-4}$ for the FSS [9]. Fig. 1 shows that the MVSS algorithm provides the fastest speed of convergence among all other algorithms while retaining the same small level of misadjustment. This is confirmed by the plot of the mean behavior of one of the weights in Fig. 2, where the actual coefficient value to be identified is 2.0.

B. Example 2: Correlated Input, Low SNR

Both the unknown system and the adaptive filters are of order four and excited by a correlated signal $x(n)$ generated by [1]

$$x(n) = 0.9x(n-1) + a(n) \tag{36}$$

where $a(n)$ is a zero-mean, uncorrelated Gaussian noise of unity variance. This input results in flattened elliptical contours, which usually cause difficulties in the convergence of gradient algorithms. $\epsilon_{\min}$, α, and β are chosen as in Example 1, whereas γ is chosen as 8×10^{-4} to obtain a final steady-state excess MSE around -28 dB. Note that for this example, $\mathrm{tr}(\mathbf{R}) = 21.0526$. The VSS algorithm is used with $\alpha_{vss} = 0.97$, and to obtain the same level of misadjustment, γ_{vss} is set to 8×10^{-6}. Since there is no theoretical analysis of misadjustment for SGA-GAS for correlated input signals, we found experimentally that $\rho = 1 \times 10^{-7}$ provides the desired misadjustment. For all previous algorithms, we used $\mu_{\max} = 0.008$, and $\mu_{\min} = 1 \times 10^{-4}$. The FSS algorithm is used with $\mu_{FSS} = 3 \times 10^{-4}$. Fig. 3 shows that for correlated input signals, the MVSS is superior in its convergence rate to the VSS, SGA-GAS, and the FSS algorithms while providing the same steady-state MSE. This can be seen from the step-size evolution in Fig. 4. The step size of MVSS remains near the $\mu_{\max}$ value until the algorithm is fairly close to steady state, where it automatically decreases to its minimum value to produce low misadjustment. On the other hand, the SGA-GAS step size stays high even after algorithm convergence and decreases slowly to its steady-state value after 50 000 iterations. Therefore, the excess MSE approaches the same level as the other algorithms only after 50 000 samples. The VSS algorithm, however, is still showing dramatic improvement over the FSS algorithm in this environment.

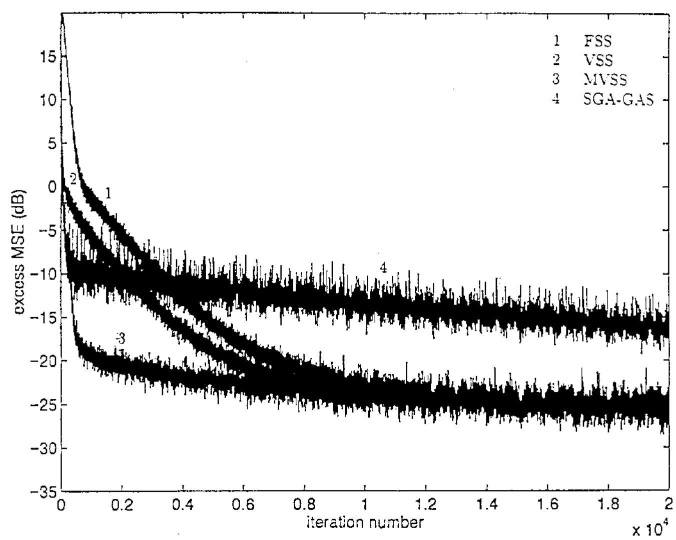

Fig. 3. Comparison of excess MSE of various adaptive algorithms for correlated input case SNR $\simeq$ 7 dB.

Figs. 5 and 6 show the same example but with an abrupt change in the coefficient values (all multiplied by -1 at itera-

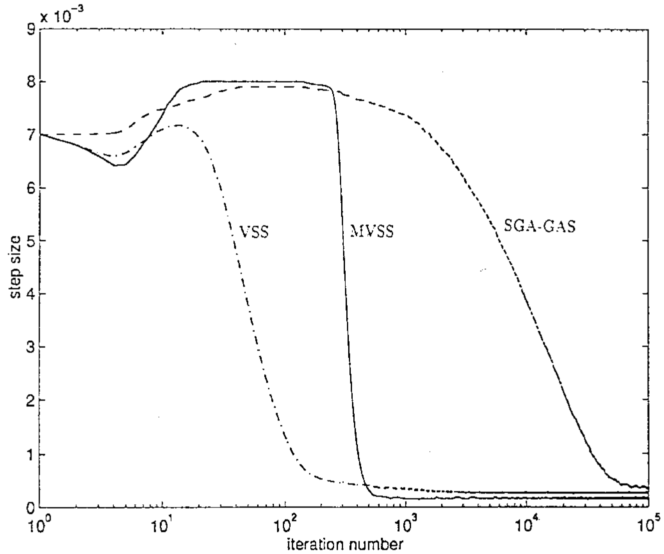

Fig. 4. Comparison of mean step size evolution of the VSS, SGA-GAS, and the MVSS algorithms for correlated input case SNR $\simeq$ 7 dB.

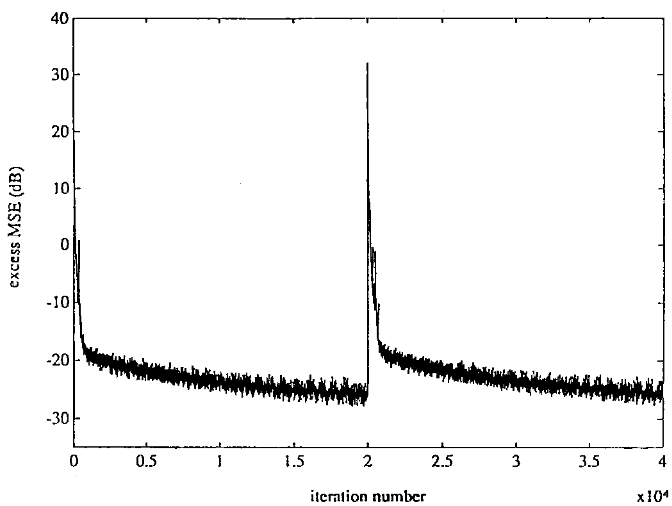

Fig. 5. Excess MSE of the MVSS algorithm for an abrupt change in the unknown system parameters.

tion 20 000). The step size of the MVSS algorithm immediately increases to $\mu_{\max}$ to provide the fastest speed to track changes in the system.

C. Example 3: High SNR

Examples 1 and 2 examined the performance of the various algorithms under low signal-to-noise ratio (SNR) conditions. These two examples are now repeated for higher SNR using $\epsilon_{\min} = 0.001$.

Example 1 is repeated with the parameters $\mu_{\max} = 0.1$, $\mu_{\min} = 5 \times 10^{-4}$, $\alpha = 0.97$, $\beta = 0.99$, $\gamma = 1$, $\alpha_{vss} = 0.97$, $\gamma_{vss} = 0.02$, $\rho = 0.001$, and $\mu_{\text{FSS}} = 5 \times 10^{-4}$. Those parameters are chosen to achieve about -60 dB steady-state excess MSE for all algorithms. Fig. 7 shows the result of comparison the various algorithms. It should be noted that the SGA-GAS algorithm takes about 70 000 iterations in this example to converge to the same steady-state excess MSE of the other algorithms.

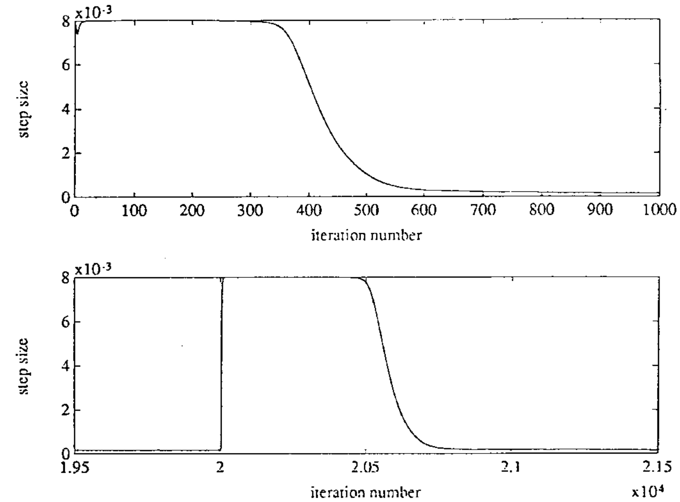

Fig. 6. Mean step size evolution of the MVSS algorithm for an abrupt change in the unknown system parameters.

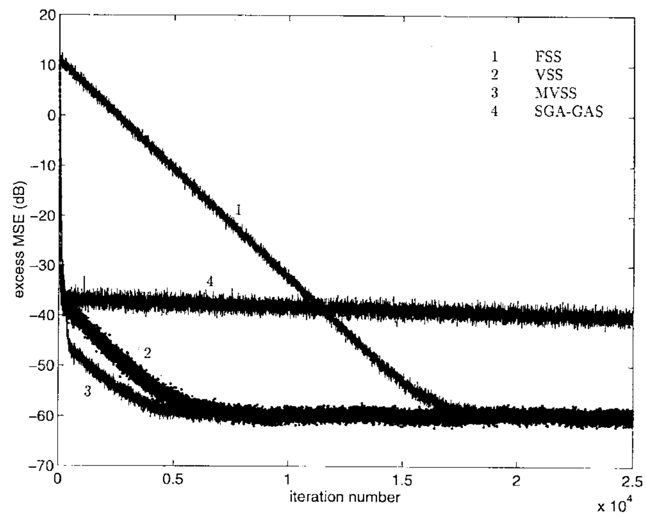

Fig. 7. Comparison of excess MSE of various adaptive algorithms for the white input case when SNR = 30 dB.

Example 2 is repeated using the parameters $\mu_{\max} = 0.008$, $\mu_{\min} = 1 \times 10^{-3}$, $\alpha = 0.97$, $\beta = 0.99$, $\gamma = 0.5$, $\alpha_{vss} = 0.97$, $\gamma_{vss} = 0.01$, $\rho = 0.001$, and $\mu_{\text{FSS}} = 1 \times 10^{-3}$. The parameters are selected to achieve a steady-state excess MSE of about -50 dB for all algorithms. Comparison of the various algorithms is shown in Fig. 8. Note that in Fig. 8, the VSS and SGA-GAS algorithms exhibit similar behavior in this case. Comparing Figs. 1 and 3 and Figs. 7 and 8, respectively, we can see that the performance of the VSS and SGA-GAS deteriorates significantly as the noise level increases from $\epsilon_{\min} = 0.001$ in Figs. 7 and 8 to $\epsilon_{\min} = 1$ in Figs. 1 and 3. On the other hand, the MVSS algorithm maintains the same performance for high and low levels of uncorrelated noise.

D. Example 4: Time Varying Optimal Weight Vector

The unknown system model here is similar to that used in the first example, except that the optimal weight vector

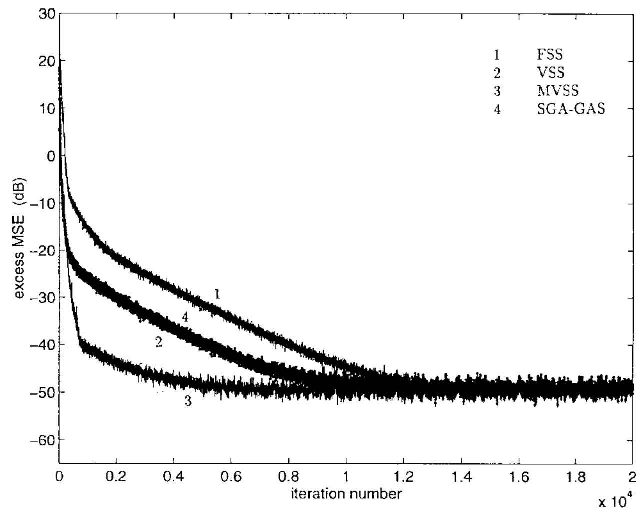

Fig. 8. Comparison of excess MSE of various adaptive algorithms for correlated input case SNR $\simeq$ 37 dB.

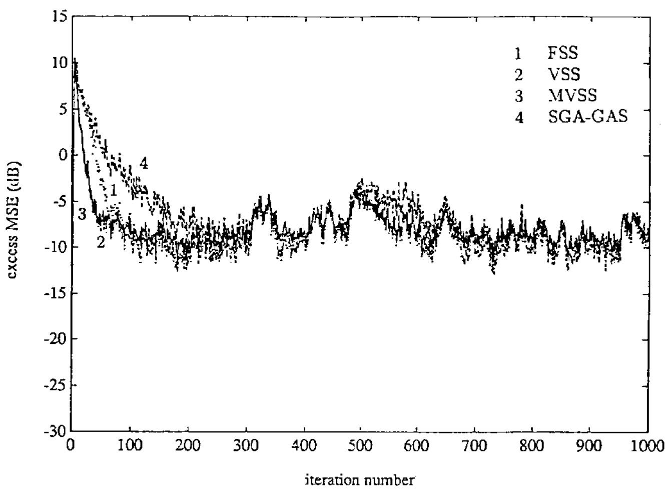

Fig. 9. Comparison of excess MSE of various adaptive algorithms in nonstationary environment.

is nonstationary and generated according to the random walk model in (10). Fig. 9 compares the four algorithms for $\sigma_n^2 = 0.001$. Optimal parameters for a given level of nonstationarity were calculated to achieve minimum misadjustment. The MVSS algorithm is used with $\beta = 0.6$, $\alpha = 0.97$, and $\gamma^* = 3.8 \times 10^{-3}$. For the VSS algorithm, $\alpha_{vss} = 0.97$ and $\gamma_{vss} = 7.65 \times 10^{-4}$, as given in [1]. For this level of nonstationarity, the optimal step size for the FSS algorithm is found to be $\mu^*_{\text{FSS}} = 0.0316$ [9]. Since the excess MSE of SGA-GAS is relatively insensitive to the selection of ρ in nonstationary environments [2], we used $\rho = 1 \times 10^{-4}$ chosen experimentally to obtain the best performance of the algorithm. Fig. 9 illustrates the ability of the MVSS to operate as well as the FSS and the VSS in a nonstationary environment. In Table I, we compare the MVSS with the other algorithms for different levels of nonstationarity. Table I also compares theoretical results of misadjustment obtained from (31) with results of simulation. When $\sigma_n^2 = 0.01$, the MVSS is used with $\alpha = 0.97$ and $\beta = 0.4$, and we found that $\gamma^* = 6.9 \times 10^{-3}$. The VSS and the SGA-GAS algorithms were used with

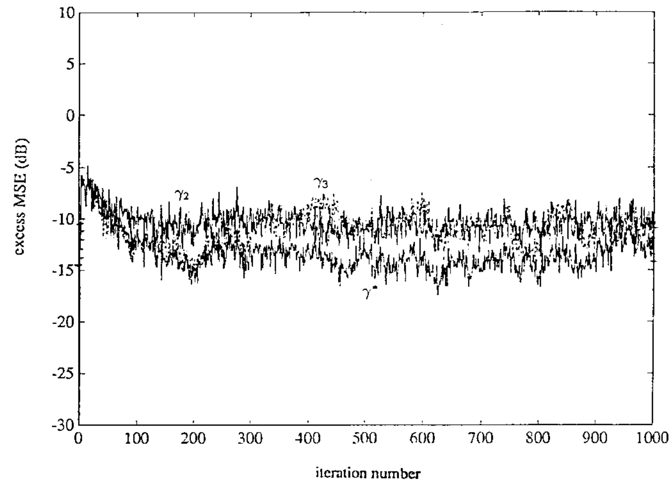

Fig. 10. Comparison of excess MSE of the MVSS algorithm in nonstationary environment.

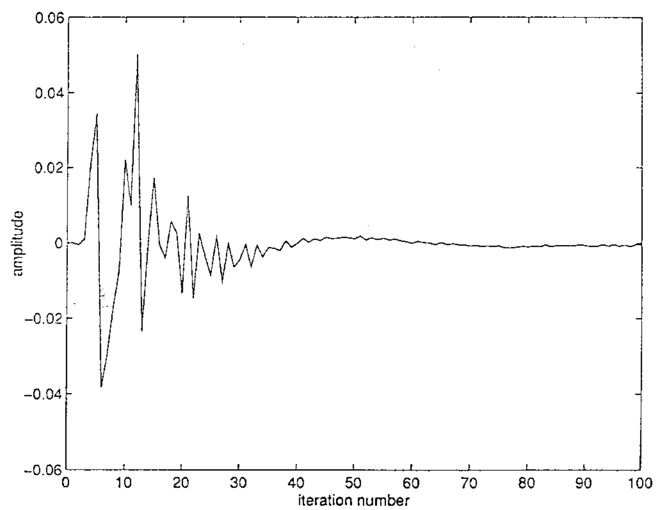

Fig. 11. Measured impulse response of a real hybrid. The sampling frequency is 8 kHz.

$\alpha_{vss} = 0.97$, $\gamma_{vss} = 1 \times 10^{-3}$, and $\rho = 1 \times 10^{-4}$, which were chosen experimentally to provide the minimum level of misadjustment. The FSS algorithm is used with the optimum step size $\mu^*_{\text{FSS}} = 0.1$. For $\sigma_n^2 = 0.0001$, the MVSS is used with $\alpha = 0.97$, $\beta = 0.8$, and $\gamma^* = 2.7 \times 10^{-3}$. The VSS and SGA-GAS were used with $\alpha_{vss} = 0.97$, $\gamma_{vss} = 1 \times 10^{-4}$, and $\rho = 1 \times 10^{-4}$. The FSS is used with $\mu^* = 0.01$. It is obvious from Table I that the MVSS algorithm is as good as the FSS and the VSS algorithms even when operating under varying level of nonstationarities. Furthermore, it can be seen that the misadjustment predicted in (31) agrees well with the simulation results of misadjustment. Fig. 10 displays the excess MSE of the MVSS algorithm with $\alpha = 0.97$, $\beta = 0.8$, $\gamma_1 = \gamma^* = 2.7 \times 10^{-3}$, $\gamma_2 = 1 \times 10^{-4}$, and $\gamma_3 = 8 \times 10^{-3}$ when $\sigma_n^2 = 0.0001$. Note that $\gamma_2 < \gamma^* < \gamma_3$. In Fig. 10, plots for γ_2, γ_3 almost overlap. It is clear from Fig. 10 that the optimal value $\gamma^* = \gamma_1$ obtained from (34) indeed provides the minimum possible misadjustment for a given level of nonstationary.

It should be mentioned that theoretical and simulation results presented for the proposed algorithm assumed uncorre-

TABLE I
COMPARISON OF THEORETICAL AND EXPERIMENTAL MISADJUSTMENT OF VARIOUS VSS LMS ALGORITHMS

	Calculated Misadjustment		Measured Misadjustment			
σ_n^2	MVSS (31)	FSS	MVSS	SGA-GAS	FSS	VSS
0.01	0.416	0.4	0.57	0.59	0.56	0.55
0.001	0.129	0.126	0.157	0.19	0.136	0.137
0.0001	0.04	0.04	0.044	0.065	0.039	0.049

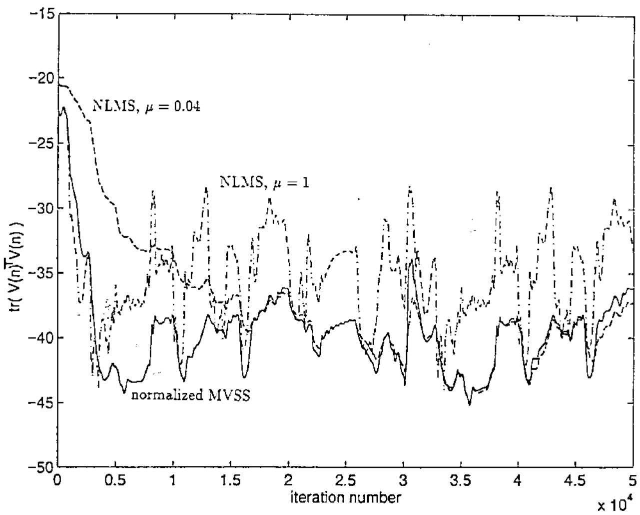

Fig. 12. Comparison of performance of the normalized MVSS and the NLMS algorithms.

lated noise sequence. However, in some applications, the noise can be correlated. For example, in acoustic echo cancellation, the acoustic noise can be a correlated one. Noise resulting from undermodeling of the unknown system is also a correlated one. In this situation, the term $E\{\xi(n)\xi(n-1)\}$ will not vanish, and it will appear in step-size equation (9). Consequently, the improvement provided by the MVSS algorithm over the VSS algorithm will depend on the relative ratio between $E\{\xi(n)\xi(n-1)\}$ and $E\{\xi^2(n)\}$. If the correlation between the noise samples is known to decrease as the time difference between them increases, we could, in this case, use the measure $e(n)e(n-D)$ such that $D \ll N$. In the worst case, when $E\{\xi(n)\xi(n-1)\} = E\{\xi^2(n)\}$, the MVSS algorithm will exhibit similar performance to the VSS algorithm.

E. Example 5: Nonstationary Input Signal

The MVSS is applied here to the cancellation of echo produced by a real hybrid of length 100. The input signal is a real speech of a male voice obtained at a sampling rate of 8 kHz. The impulse response of a real hybrid was measured at a sampling rate of 8 kHz and is shown in Fig. 11. In order to cope with the changes in the speech signal energy, the step size of the MVSS algorithm is normalized so that its recursion becomes

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \frac{\mu(n)}{\delta + \mathbf{X}^{\mathbf{T}}(n)\mathbf{X}(n)} e(n)\mathbf{X}(n) \quad (37)$$

where δ is a small positive number. Note that the asymptotic convergence of (37) is assured by [8]

$$0 < E\{\mu(n)\} < 2. \quad (38)$$

It is shown in [10] that the fastest speed of convergence of the normalized LMS (NLMS) algorithm is obtained when $\mu_{NLMS} = 1$. However, this value results in a large misadjustment value. Accordingly, $\mu_{\max}$ for the normalized MVSS is set to 1, and $\mu_{\min}$ is chosen to attain the required level of misadjustment. In this example, the adaptive filter has 30 coefficients and results in 20-dB steady-state echo return loss enhancement (ERLE), which is defined as

$$\text{ERLE} = 10 \log_{10} \left(\frac{E\{d^2(n)\}}{E\{e^2(n)\}} \right). \quad (39)$$

Note that the undermodeling will result in some correlation in the noise samples. For the normalized MVSS, we used $\alpha = 0.97$, $\beta = 0.95$, $\gamma = 1 \times 10^{-5}$, and $\mu_{\min} = 0.04$. The performance of the algorithm is compared with the NLMS algorithm with two step sizes: $\mu_{\text{NLMS}} = 1$ and $\mu_{\text{NLMS}} = 0.04$. In Fig. 12, we plot $\text{tr}(\mathbf{V}^{\mathbf{T}}(n)\mathbf{V}(n))$, which is the squared norm of the weight error vector. This measure is roughly proportional to algorithm misadjustment and is therefore more sensitive to and illustrative of the impact of the step-size value on the algorithm performance compared with the ERLE. Fig. 12 shows that for a speech input, the normalized MVSS behaves better than the NLMS algorithm in the sense that it can compromise simultaneously between the two states of the NLMS of fast convergence (attained when $\mu_{\text{NLMS}} = 1$) and low error in the adaptive filter coefficients (attained when $\mu_{\text{NLMS}} = 0.04$).

V. CONCLUSION

A new LMS-type algorithm employing a time-varying step size in the standard LMS weight update equation was introduced. The step size of the algorithm is adjusted according to the square of a time-averaging estimate of the autocorrelation of $e(n)$ and $e(n-1)$. As a result, the algorithm can effectively sense the proximity to the optimum independent of uncorrelated measurement noise. Approximate theoretical expressions for the steady-state behavior of the algorithm were derived. The performance of the algorithm was compared with that of the standard LMS [7] as well as other variable step-size LMS algorithms [1], [2] through simulations. Results show that the algorithm has a significant convergence rate improvement over those algorithms in a stationary environment for the same excess MSE in both high and low SNR environments. The performance in nonstationary cases is comparable with the standard LMS algorithm.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their constructive suggestions that helped in improving this manuscript.

References

[1] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Processing,* vol. 40, pp. 1633–1642, July 1992.

[2] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Processing,* vol. 41, pp. 2075–2087, June 1993.

[3] R. Harris *et al.,* "A variable step size (VSS) algorithm," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-34, pp. 499–510, June 1986.

[4] T. J. Shan and T. Kailaith, "Adaptive algorithms with an automatic gain control feature," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 35, pp. 122–127, Jan. 1988.

[5] J. Soo and K. K. Pang, "A multi step size (MSS) frequency domain adaptive filter," *IEEE Trans. Signal Processing,* vol. 39, pp. 115–121, Jan. 1991.

[6] J. B. Evans, P. Xue, and B. Liu, "Analysis and implementation of variable step size adaptive algorithms," *IEEE Trans. Signal Processing,* vol. 41, pp. 2517–2535, Aug. 1993.

[7] L. L. Horowitz and K. D. 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.

[8] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1991.

[9] B. Widrow *et al.,* "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE,* vol. 64, pp. 1151–1162, Aug. 1976.

[10] W. B. Mikhael *et al.,* "Adaptive filters with individual adaptation of parameters," *IEEE Trans. Circuits Syst.,* vol. CAS-33, pp. 677–685, July 1986.

Tyseer Aboulnasr (M'87) received the B.Eng. degree from Cairo University, Cairo, Egypt, and the M.Sc. and Ph.D. degrees from Queen's University, Kingston, Ont., Canada, all in electrical engineering.

She held an NSERC postdoctorate scholarship from 1985 to 1986. She later joined the University of Ottawa, Ottawa, Ont., Canada, where she is currently an Associate Professor and the Associate Dean of Academic Affairs. Her interests are in the area of adaptive filtering, algorithms and applications, filterbanks and applications, and applications of signal processing in communications.

K. Mayyas was born in Irbid, Jordan, in 1967. He received the B.S.E. and M.Sc. degrees from the Jordan University of Science and Technology, Irbid, in 1990 and 1992 and the Ph.D. degree from the University of Ottawa, Ottawa, Ont., Canada, in 1995, all in electrical engineering.

He is currently an Assistant Professor with the Department of Electrical Engineering, Jordan University of Science and Technology. His research interests are in various aspects of adaptive filters and their applications, including the development of efficient adaptive algorithms.