A Stochastic Gradient Adaptive Filter with Gradient Adaptive Step Size

V. John Mathews, *Senior Member, IEEE,* and Zhenhua Xie, *Member, IEEE*

***Abstract*—This paper presents an adaptive step-size gradient adaptive filter. The step size of the adaptive filter is changed according to a gradient descent algorithm designed to reduce the squared estimation error during each iteration. An approximate analysis of the performance of the adaptive filter when its inputs are zero mean, white, and Gaussian and the set of optimal coefficients are time varying according to a random-walk model is presented in the paper. The algorithm has very good convergence speed and low steady-state misadjustment. Furthermore, the tracking performance of these algorithms in nonstationary environments is relatively insensitive to the choice of the parameters of the adaptive filter and is very close to the best possible performance of the least mean square (LMS) algorithm for a large range of values of the step size of the step-size adaptation algorithm. Several simulation examples demonstrating the good properties of the adaptive filter as well as verifying the analytical results are also presented in the paper.**

I. Introduction

STOCHASTIC gradient adaptive filters are extremely popular because of their inherent simplicity. However, they suffer from relatively slow and data-dependent convergence behavior. It is well known that the performance of stochastic gradient methods is adversely affected by high eigenvalue spreads of the autocorrelation matrix of the input vector.

Traditional approaches for improving the speed of convergence of the gradient adaptive filters have been to employ time-varying step-size sequences [4]–[6], [9], [10]. The idea is to somehow sense how far away the adaptive filter coefficients are from the optimal filter coefficients and use step sizes that are small when adaptive filter coefficients are close to the optimal values and use large step sizes otherwise. The approach is heuristically sound and has resulted in several *ad hoc* techniques, where the selection of the convergence parameter is based on the magnitude of the estimation error [6], polarity of the successive samples of the estimation error [4], measurement of the cross correlation of the estimation error with input data [5], [10], and so on. Experimentation with these techniques has shown that their performance is highly dependent on the selection of certain parameters in the algorithms and, furthermore, the optimal choice of these parameters is highly data dependent. This fact has severely limited the usefulness of such algorithms in practical applications. Mikhael *et al.* [9] have proposed methods for selecting the step sizes that would give the fastest speed of convergence among all gradient adaptive algorithms attempting to reduce the squared estimate error. Unfortunately, their choices of the step sizes will also result in fairly large values of steady-state excess mean-squared estimation error.

This paper presents a stochastic gradient adaptive filtering algorithm that overcomes many of the limitations of the methods discussed above. The idea is to change the time-varying convergence parameters in such a way that the change is proportional to the negative of the gradient of the squared estimation error with respect to the convergence parameter. The method was originally introduced by Shin and Lee [11]. Their analysis, which was based on several simplifying assumptions, indicated that the steady-state behavior of the adaptive filter depended on the initial choice of the step size. Specifically, their analysis predicted that the steady-state value of the step size is always larger than the initial value of the step size and is a function of the initial step size. This implies that the steady-state misadjustment will be large and will depend on the initial step size. These statements are contradictory to what has been observed in practice. Experiments have shown that the algorithms have very good convergence speeds as well as small misadjustments, irrespective of the initial step sizes. Our objectives in discussing the algorithm are essentially twofold: 1) Publicize this relatively unknown, but powerful adaptive filtering algorithm to the signal processing community, and 2) present an analysis of the adaptive filter that matches the observed behavior of the algorithm. Another algorithm that is conceptually similar to, but computationally more expensive than the one discussed here, was recently presented in [1].

The rest of the paper is organized as follows. In the next section, we will derive the adaptive step size, stochastic gradient adaptive filter. In Section III, we present a theoretical performance analysis of the algorithm. Simulation examples demonstrating the good properties of the adaptive filter and also demonstrating the validity of the analysis are presented in Section IV. Finally, concluding remarks are made in Section V.

Manuscript received September 25, 1990; revised July 13, 1992. The associate editor coordinating the review of this paper and approving it for publication was Dr. Erlandur Karlsson. This work was supported in part by NSF Grants MIP 8708970 and MIP 8922146, and a University of Utah Faculty Fellow Award. Parts of this paper were presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Albuquerque, NM, Apr. 3–6, 1990.

V. J. Mathews is with the Department of Electrical Engineering, University of Utah, Salt Lake City, UT 84112.

Z. Xie was with the Department of Electrical Engineering, University of Utah, Salt Lake City, UT 84112. He is now with the Technical Center, R. R. Donnelley & Sons Company, Lisle, IL 60532.

IEEE Log Number 9208247.

II. The Gradient Adaptive Step-Size Algorithm

Consider the problem of estimating the desired response signal $d(n)$ as a linear combination of the elements of $X(n)$, the N-dimensional input vector sequence to the adaptive filter. The popular least mean square (LMS) adaptive filter updates the filter coefficients in the following manner:

$$e(n) = d(n) - X^T(n)\, H(n) \tag{1}$$

and

$$H(n+1) = H(n) + \mu X(n) e(n). \tag{2}$$

Here, $(\cdot)^T$ denotes the matrix transpose of $(\cdot)$, $H(n)$ is the coefficient vector at time n, and μ is the step-size parameter that controls the speed of convergence as well as the steady-state and/or tracking behavior of the adaptive filter. The selection of μ is very critical for the LMS algorithm. A small μ (small compared to the reciprocal of the input signal strength) will ensure small misadjustments in steady state, but the algorithm will converge slowly and may not track the nonstationary behavior of the operating environment very well. On the other hand, a large μ will in general provide faster convergence and better tracking capabilities at the cost of higher misadjustments. Any selection must be a compromise between the two scenarios described above.

The adaptive step-size algorithm that will be introduced now is designed to eliminate the "guesswork" involved in selection of the step-size parameter, and at the same time satisfy the following requirements: 1) The speed of convergence should be fast; 2) when operating in stationary environments, the steady-state misadjustment values should be very small; and 3) when operating in nonstationary environments, the algorithm should be able to sense the rate at which the optimal coefficients are changing and select step-sizes that can result in estimates that are close to the best possible in the mean-squared-error sense. Our approach to achieving the above goals is to adapt the step-size sequence using a gradient descent algorithm so as to reduce the squared-estimation error at each time. This approach will result in the following algorithm [11]:

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

$$\mu(n) = \mu(n-1) - \frac{\rho}{2} \frac{\partial}{\partial \mu(n-1)} e^2(n) \tag{4a}$$

$$= \mu(n-1) - \frac{\rho}{2} \frac{\partial^T e^2(n)}{\partial H(n)} \cdot \frac{\partial H(n)}{\partial \mu(n-1)} \tag{4b}$$

$$= \mu(n-1) + \rho e(n) e(n-1) X^T(n-1) X(n) \tag{5}$$

and

$$H(n+1) = H(n) - \frac{\mu(n)}{2} \frac{\partial e^2(n)}{\partial H(n)} \tag{6}$$

$$= H(n) + \mu(n) e(n) X(n). \tag{7}$$

In the above equations, ρ is a small positive constant that controls the adaptive behavior of the step-size sequence $\mu(n)$.

Remark 1: The increase in computational complexity of the above algorithm over the LMS algorithm is minimal. If the algorithm is used in single-channel applications so that the input vector is formed using the N most recent samples of the input signal, i.e.,

$$X(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^T \tag{8}$$

the additional computations correspond to four multiplications and three additions per sample. In addition, the algorithm requires approximately N more memory locations than the LMS algorithm.

Remark 2: Deriving conditions on ρ so that convergence of the adaptive system can be guaranteed appears to be a difficult task. However, we can guarantee (mean-squared) convergence of the adaptive filter by restricting $\mu(n)$ to be such that it always stays within the range that would ensure convergence. A sufficient, but not necessary, condition on $\mu(n)$ (assuming stationarity of the input process) to ensure mean-squared convergence of the adaptive filter is [3]

$$0 < \mu(n) < \frac{2}{3 \operatorname{tr}\{R\}} \tag{9}$$

where $\operatorname{tr}\{(\cdot)\}$ denotes trace of the matrix $(\cdot)$ and R is the autocorrelation matrix of the input vector given by

$$R = E\{X(n) X^T(n)\}. \tag{10}$$

If $\mu(n)$ falls outside the range in (9), we can bring it inside the range by setting it to the closest of 0 or $2/3 \operatorname{tr}\{R\}$. (Note that [3] uses 2μ as the step size and, consequently, the upper bound in (9) is different from that in [3] by a factor of 2.)

Remark 3: There exist some signals such that $X^T(n-1) X(n) = 0$ for all n. An example[1] is that where $\{x(n)\}$ is given by the sequence $\{\cdots, 1, -1, +1, -1, +1, -1, +1, -1, 1, -1, \cdots\}$. When N is even, $X^T(n-1)X(n) = 0$ for all n. In such cases, the step-size adaptation algorithm will not update the step size. Consequently, it would be advisable to initialize the step size to some small positive value, so that the filter adapts to the environment always. However, we believe that occurrences of such signals as in the above example will be extremely rare in practice and we will assume that the step size will adapt to the environment. For the random signals considered for the analysis in the next section, the adaptation of the step sizes will occur with probability one.

Remark 4: It is possible to adapt the step size corresponding to each coefficient individually. Such an approach will result in the following algorithm [8]:

$$e(n) = d(n) - H^T(n) X(n) \tag{11}$$

$$\mu_i(n) = \mu_i(n-1) + \rho e(n) e(n-1)\, x_i(n) x_i(n-1); \quad i = 0, 1, \cdots, N-1 \tag{12}$$

[1]The authors thank one of the anonymous reviewers for pointing out this example.

and

$$h_i(n+1) = h_i(n) - \frac{\mu_i(n)}{2}\frac{\partial e^2(n)}{\partial h_i(n)}$$
$$= h_i(n) + \mu_i(n)e(n)x_i(n);$$
$$i = 0, 1, \cdots, N-1. \quad (13)$$

Here, $h_i(n)$ and $x_i(n)$ are the ith elements of $H(n)$ and $X(n)$, respectively. The computational complexity of this algorithm is $3N + 4$ multiplications per sample. We have found that this algorithm performs slightly better than the one in (3)–(7) for multichannel and nonlinear adaptive filtering applications. Both methods seem to perform about the same in single-channel applications. The rest of the paper deals only with the algorithm that utilizes a single step size.

III. Performance Analysis

For the performance analysis, we will assume that the adaptive filter structure is that of an N-point FIR filter, and the input vector $X(n)$ is obtained as a vector formed by the most recent N samples of the input sequence $x(n)$, i.e.,

$$X(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^T. \quad (14)$$

Let $H_{\text{opt}}(n)$ denote the optimal coefficient vector (in the minimum mean-squared estimation error sense) for estimating the desired response signal $d(n)$ using $X(n)$. We will assume that $H_{\text{opt}}(n)$ is time varying, and that the time variations are caused by a random disturbance of the optimal coefficient process. Thus, the behavior of the optimal coefficient process can be modeled as

$$H_{\text{opt}}(n) = H_{\text{opt}}(n-1) + C(n-1) \quad (15)$$

where $C(n-1)$ is the disturbance process that is a zero-mean and white vector process with covariance matrix $\sigma_c^2 I$. In order to make the analysis tractable, we will make use of the following assumptions and approximations.

i) $X(n)$, $d(n)$ are jointly Gaussian and zero-mean random processes. $X(n)$ is a stationary process. Moreover, $\{X(n), d(n)\}$ is uncorrelated with $\{X(k), d(k)\}$ if $n \neq k$. This is the commonly employed indepedence assumption and is seldom true in practice. However, analyses employing this assumption have produced reliable design rules in the past.

ii) The autocorrelation matrix R of the input vector $X(n)$ is a diagonal matrix and is given by

$$R = \sigma_x^2 I. \quad (16)$$

While this is a fairly restrictive assumption, it considerably simplifies the analysis. Furthermore, the white data model is a valid representation in many practical systems such as digital data transmission systems and analog systems that are sampled at the Nyquist rate and adapted using discrete-time algorithms.

iii) Let

$$d(n) = X^T(n)H_{\text{opt}}(n) + \zeta(n) \quad (17)$$

where $\zeta(n)$ corresponds to the optimal estimation error process. We will also assume that the triplet $\{X(n), C(n), \zeta(n)\}$ are statistically independent random processes.

iv) We will use the approximation that the convergence sequence parameter $\mu(n)$ is statistically independent of $X(n)$, $H(n)$, and $e(n)$. While this is never true, experiments have indicated that the approximations that $\mu(n)$ and $\mu^2(n)$ are uncorrelated with $X(n)$, $H(n)$, and $e(n)$ are reasonably accurate for small values of ρ and relatively white input signals (This is the only place where we need the whiteness assumption. Otherwise, the analysis can be easily extended to the colored input signal case.) Note that the condition under which the above approximation is accurate is when the statistical fluctuations of $\mu(n)$ are small when compared with that of $X(n)$ and $e(n)$. This condition is, in general, satisfied for small values of ρ.

v) We will use the approximation that the statistical expectation of $e^2(n)X(n)X^T(n)$ conditioned on the coefficient vector $H(n)$ is the same as the unconditional expectation, i.e.,

$$E\{e^2(n)X(n)X^T(n)|H(n)\} \approx E\{e^2(n)X(n)X^T(n)\}. \quad (18)$$

This approximation has been successfully employed for performance analysis of adaptive filters equipped with the sign algorithm [7] and for analyzing the behavior of some blind equalization algorithms [13]. Even though it is possible to rigorously justify this approximation only for small step sizes (i.e., for slowly varying coefficient values), it has been our experience that it works reasonably well in large step-size situations also. For the gradient adaptive step-size algorithm, $\mu(n)$ can become fairly large during the early stages of adaptation. Experimental results presented later in this paper show a reasonably good match between analytical and empirical results even during such times.

A. Mean Behavior of the Weight Vector

Let

$$V(n) = H(n) - H_{\text{opt}}(n) \quad (19)$$

denote the coefficient misalignment vector at time n. Then,

$$e(n) = \zeta(n) - V^T(n)X(n). \quad (20)$$

Substituting (17), (19), and (20) into (7), we can easily show that

$$V(n+1) = (I - \mu(n)X(n)X^T(n))V(n) + \mu(n)X(n)\zeta(n) - C(n). \quad (21)$$

It is straightforward using the independence assumption and the uncorrelatedness of $\mu(n)$ with $X(n)$ and $e(n)$ to show that

$$E\{V(n+1)\} = (1 - E\{\mu(n)\}\sigma_x^2)E\{V(n)\}. \quad (22)$$

B. Mean-Squared Behavior of the Weight Vector

Let

$$K(n) = E\{V(n)V^T(n)\} \quad (23)$$

denote a second moment matrix of the misalignment vector. Multiplying both sides of (21) with their respective transposes, we get the following equation:

$$\begin{aligned} V(n+1)V^T(n+1) &= (I - \mu(n)X(n)X^T(n))V(n)V^T(n)(I - \mu(n)X(n)X^T(n)) \\ &\quad + \mu^2(n)\zeta^2(n)X(n)X^T(n) + C(n)C^T(n) \\ &\quad + g(\mu(n), X(n), V(n), \zeta(n), C(n)) \end{aligned} \tag{24}$$

where $g(\mu(n), X(n), V(n), \zeta(n), C(n))$ corresponds to the sum of the six terms that are explicitly not listed in the expansion. Under our assumptions and approximations, the mean value of these six terms are all zero matrices. Combining usual analysis techniques for Gaussian input signals [3] with the approximation that $\mu(n)$ and $\mu^2(n)$ are uncorrelated with the data while taking the statistical expectation of (24) will result in the following evolution equation for the second moment matrix of the coefficient misalignment vector:

$$\begin{aligned} K(n+1) &= K(n) - 2E\{\mu(n)\}\sigma_x^2 K(n) \\ &\quad + E\{\mu^2(n)\}(2\sigma_x^4 K(n) + \sigma_x^2\sigma_e^2(n)I) + \sigma_c^2 I \end{aligned} \tag{25}$$

where

$$\sigma_e^2(n) = \xi_{\min} + \sigma_x^2 \operatorname{tr}\{K(n)\} \tag{26}$$

and

$$\xi_{\min} = E\{\zeta^2(n)\} \tag{27}$$

is the minimum value of the mean-squared estimation error.

The mean and mean-squared behavior of the step-size sequence $\mu(n)$ can be shown to follow the following nonlinear difference equations:

$$\begin{aligned} E\{\mu(n)\} &= E\{\mu(n-1)\}(1 - \rho\{N\sigma_e^2(n-1)\sigma_x^4 \\ &\quad + 2\sigma_x^6 \operatorname{tr}\{K(n-1)\}\}) + \rho\sigma_x^4 \operatorname{tr}\{K(n-1)\} \end{aligned} \tag{28}$$

and

$$\begin{aligned} E\{\mu^2(n)\} &= E\{\mu^2(n-1)\}(1 - 2\rho(N\sigma_e^2(n-1)\sigma_x^4 \\ &\quad + 2\sigma_x^6 \operatorname{tr}\{K(n-1)\})) \\ &\quad + 2\rho E\{\mu(n-1)\}\sigma_x^4 \operatorname{tr}\{K(n-1)\} \\ &\quad + \rho^2 \operatorname{tr}\{(2\sigma_x^4 K(n) + \sigma_e^2(n)\sigma_x^2 I) \\ &\quad \cdot (2\sigma_x^4 K(n-1) + \sigma_e^2(n-1)\sigma_x^2 I)\}. \end{aligned} \tag{29}$$

Details of the derivation are given in Appendix A.

C. Steady-State Properties of the Adaptive Filter

As discussed earlier, deriving conditions on ρ so as to guarantee convergence of the algorithm appears to be a difficult task. However, we can assure that the adaptive filter converges in the mean-square sense by restricting the range of $\mu(n)$ as discussed in the previous section. Assuming that the system of evolution equations derived above converges, we now proceed to study the steady-state behavior of the adaptive filter.

Let $\bar{\mu}_\infty$, $\overline{\mu^2_\infty}$, $\sigma_e^2(\infty)$, and K_∞ represent the steady-state values of $E\{\mu(n)\}$, $E\{\mu^2(n)\}$, $\sigma_e^2(n)$, and $K(n)$, respectively. Substituting these values for their counterparts in equations (25), (26), (28), and (29) will yield the following characterization of the steady-state behavior of the adaptive filter:

$$\begin{aligned} \bar{\mu}_\infty &= \bar{\mu}_\infty\{1 - \rho(N\sigma_e^2(\infty)\sigma_x^4 + 2\sigma_x^6 \operatorname{tr}(K_\infty))\} \\ &\quad + \rho\sigma_x^4 \operatorname{tr}(K_\infty) \end{aligned} \tag{30}$$

$$\begin{aligned} \overline{\mu^2_\infty} &= \overline{\mu^2_\infty}\{1 - 2\rho(N\sigma_e^2(\infty)\sigma_x^4 + 2\sigma_x^6 \operatorname{tr}(K_\infty))\} \\ &\quad + 2\rho\bar{\mu}_\infty\sigma_x^4 \operatorname{tr}(K_\infty) \end{aligned} \tag{31}$$

$$\begin{aligned} &+ \rho^2 \operatorname{tr}\{(2\sigma_x^4 K_\infty + \sigma_e^2(\infty)\sigma_x^2 I)^2\}, \end{aligned} \tag{32}$$

$$\sigma_e^2(\infty) = \xi_{\min} + \sigma_x^2 \operatorname{tr}(K_\infty)$$

and

$$\begin{aligned} K_\infty &= K_\infty - 2\bar{\mu}_\infty\sigma_x^2 K_\infty \\ &\quad + \overline{\mu^2_\infty}(2\sigma_x^4 K_\infty + \sigma_x^2\sigma_e^2(\infty)I) + \sigma_c^2 I. \end{aligned} \tag{33}$$

It is relatively straightforward to see from (25) that K_∞ is a diagonal matrix and that all of its diagonal elements are equal. Let

$$K_\infty = k_\infty I. \tag{34}$$

Then, substitution of (34) and (32) in (33) will give

$$\begin{aligned} k_\infty &= k_\infty - 2\bar{\mu}_\infty\sigma_x^2 k_\infty \\ &\quad + \overline{\mu^2_\infty}(2\sigma_x^4 k_\infty + \sigma_x^2(\xi_{\min} + N\sigma_x^2 k_\infty)) + \sigma_c^2. \end{aligned} \tag{35}$$

Solving for k_∞ in terms of the other parameters yields

$$k_\infty = \frac{\overline{\mu^2_\infty}\xi_{\min} + (\sigma_c^2/\sigma_x^2)}{2\bar{\mu}_\infty - \overline{\mu^2_\infty}(N+2)\sigma_x^2}. \tag{36}$$

Similarly,

$$\bar{\mu}_\infty = \frac{k_\infty}{\xi_{\min} + \sigma_x^2 k_\infty(N+2)} \tag{37}$$

and

$$\overline{\mu^2_\infty} = (\bar{\mu}_\infty)^2 + \frac{\rho}{2}(\xi_{\min} + (N+2)\sigma_x^2 k_\infty). \tag{38}$$

One could substitute (37) and (38) in (36) to obtain a cubic equation for k_∞ that is not coupled with $\bar{\mu}_\infty$ and $\overline{\mu^2_\infty}$. Unfortunately, solution of the cubic equation does not seem to give many insights into the dependence of the system behavior on the choice of ρ and the parameters of the operating environment. We now proceed to make a few simplifications in (36)–(38) which will enable us to get a much better insight into the behavior of the algorithm.

In most practical applications involving stationary environments, the excess mean-squared error is much

smaller than the minimum mean-squared estimation error so that

$$\xi_{\min} \gg (N+2)\sigma_x^2 k_\infty. \tag{39}$$

Then we can approximate $\overline{\mu}_\infty$ and $\overline{\mu_\infty^2}$ as

$$\overline{\mu_\infty} \approx \frac{k_\infty}{\xi_{\min}} \tag{40}$$

$$\overline{\mu_\infty^2} \approx (\overline{\mu_\infty})^2 + \frac{\rho}{2}\xi_{\min}. \tag{41}$$

Also, under the same circumstances, experience has shown that

$$\overline{\mu_\infty} \gg \overline{\mu_\infty^2}(N+2)\sigma_x^2. \tag{42}$$

The above inequality should be intuitively easy to agree with, since $\overline{\mu_\infty}$ itself is very small for stationary environments and $\overline{\mu_\infty^2}$ is of the order of $(\overline{\mu_\infty})^2$. Using (40)–(42) in (36) and setting $\sigma_c = 0$ for stationary environments, we get

$$k_\infty \approx \frac{\dfrac{k_\infty^2}{\xi_{\min}} + \dfrac{\rho}{2}\xi_{\min}^2}{\dfrac{2k_\infty}{\xi_{\min}}}. \tag{43}$$

Solving for k_∞ from (43) gives an approximate expression for the steady-state, mean-squared value of the coefficient fluctuations as

$$k_\infty \approx \sqrt{\frac{\xi_{\min}^3 \rho}{2}}. \tag{44}$$

The corresponding expression for the steady-state excess-mean-squared estimation error is given by

$$e_{ex} = N\sigma_x^2 k_\infty \approx N\sigma_x^2\sqrt{\frac{\xi_{\min}^3 \rho}{2}}. \tag{45}$$

ρ is usually chosen to be very small and this implies that k_∞ is also very small.

In the nonstationary environments, the inequalities in (39) and (42) hold only if σ_c is very small. Under such circumstances

$$k_\infty \approx \frac{\overline{\mu_\infty^2}\,\xi_{\min} + \sigma_c^2/\sigma_x^2}{2\overline{\mu_\infty}} \tag{46}$$

Substituting (40) and (41) into (46) and solving, we get the following approximate expression for k_∞:

$$k_\infty \approx \sqrt{\frac{\rho}{2}\xi_{\min}^3 + \frac{\sigma_c^2}{\sigma_x^2}\xi_{\min}}. \tag{47}$$

The excess mean-square error in this case is given by

$$e_{ex} = N\sigma_x^2 k_\infty \approx N\sqrt{\frac{\rho}{2}\xi_{\min}^3\sigma_x^4 + \sigma_c^2\sigma_x^2\xi_{\min}}. \tag{48}$$

It is very instructive to compare the above expression with the optimal value of the excess mean-squared error (say, $e^*_{\text{ex, LMS}}$) for the LMS algorithm obtained when the convergence parameter is chosen to minimize the excess mean-squared estimation error. $e^*_{\text{ex, LMS}}$ is approximately given by [14]

$$e^*_{\text{ex, LMS}} \approx N\sqrt{\xi_{\min}\sigma_x^2\sigma_c^2}. \tag{49}$$

(The above result assumes that the optimum value of μ is within the range that guarantees convergence of the algorithm. Otherwise, $e^*_{\text{ex, LMS}}$ will be larger.) Comparison of (48) with (49) shows that the excess mean-squared error for our algorithm is always larger than the best performance of the LMS adaptive filter. (This is to be expected since there is an extra adaptation involved in our algorithm.) However, the extra error term within the square root is proportional to $\rho\xi_{\min}^3$ and this, in general, will be very small. What this means is that it is possible to get arbitrarily close to the optimal performance of the LMS adaptive filter by choosing ρ to be appropriately small. The key result is that we do not have to have prior knowledge about the statistics of the environment to get close to the best performance.

If σ_c^2 is large (implying that the level of nonstationarity is high), we will have to resort to a different set of approximations. In order to develop these approximations, let us define

$$\Delta = \xi_{\min} + \sigma_x^2 k_\infty(N+2). \tag{50}$$

Substituting (50), (37), and (38) in (36) and manipulating the resultant equation, we get

$$\frac{2k_\infty^2}{\Delta} - \left(\frac{k_\infty^2}{\Delta^2} + \frac{\rho}{2}\Delta\right)(N+2)\sigma_x^2 k_\infty = \left(\frac{k_\infty^2}{\Delta^2} + \frac{\rho}{2}\Delta\right)\xi_{\min} + \frac{\sigma_c^2}{\sigma_x^2}. \tag{51}$$

Further simplification will give

$$\frac{k_\infty^2}{\Delta} - \frac{\rho}{2}\Delta^2 = \frac{\sigma_c^2}{\sigma_x^2}. \tag{52}$$

Since

$$\overline{\mu}_\infty = \frac{k_\infty}{\Delta} \tag{53}$$

we have that

$$(\overline{\mu}_\infty)^2 = \frac{\rho}{2}\Delta + \frac{\sigma_c^2}{\sigma_x^2\Delta}. \tag{54}$$

Comparing (54) with (38), we see that

$$(\overline{\mu}_\infty)^2 > \frac{\rho}{2}\Delta = \frac{\rho}{2}(\xi_{\min} + \sigma_x^2 k_\infty(N+2)) \tag{55}$$

and that if

$$\sigma_c^2 \gg \frac{\rho}{2}\Delta^2\sigma_x^2 = \frac{\rho}{2}(\xi_{\min} + \sigma_x^2 k_\infty(N+2))^2\sigma_x^2 \tag{56}$$

it follows that

$$\overline{\mu_\infty^2} \approx \overline{\mu_\infty}^2. \tag{57}$$

Note that the right-hand side of (56) contains k_∞. If we assume that the excess mean-squared error does not dominate the minimum mean-squared error, we can use the approximation in (57) whenever $\sigma_c^2 >> \rho/2\,\xi_{\min}^2\,\sigma_x^2$. What we are saying is that the fluctuations in $\mu(n)$ are small compared with its mean value itself. This approximation leads to the following two coupled equations that describe the steady-state behavior of the algorithm:

$$k_\infty \approx \frac{\overline{\mu}_\infty^2 \xi_{\min} + \sigma_c^2/\sigma_x^2}{2\overline{\mu}_\infty - \overline{\mu}_\infty^2 (N+2)\sigma_x^2} \tag{58}$$

and

$$\overline{\mu}_\infty \approx \frac{k_\infty}{\xi_{\min} + \sigma_x^2 (N+2) k_\infty}. \tag{59}$$

Substituting (59) in (58) will result in a cubic equation in k_∞ with three real roots. The only positive solution of the equation gives an approximate expression for k_∞ as

$$k_\infty \approx \frac{\sigma_c^2 (N+2)}{2} + \sqrt{\frac{(N+2)^2 \sigma_c^4}{4} + \frac{\sigma_c^2 \xi_{\min}}{\sigma_x^2}}. \tag{60}$$

(It is relatively straightforward to show that the negative roots are located at

$$-\xi_{\min}/(N+2)\sigma_x^2 \quad \text{and} \quad \frac{\sigma_c^2 (N+2)}{2} - \sqrt{\frac{(N+2)^2 \sigma_c^4}{4} + \frac{\sigma_c^2 \xi_{\min}}{\sigma_x^2}}.)$$

$$e_{\text{ex}} \approx \frac{N(N+2)}{2} \sigma_x^2 \sigma_c^2 + N \sqrt{\frac{(N+2)\sigma_x^4 \sigma_c^4}{4} + \sigma_x^2 \sigma_c^2 \xi_{\min}}. \tag{61}$$

Comparing (61) with (49), we find that (61) will always be larger than $e_{\text{ex, LMS}}^*$. However, when the term involving $\sigma_x^2 \sigma_c^2 \xi_{\min}$ dominates the other two terms, the performance of our filter will be very close to that of the best performance of the LMS adaptive filter. Note that this term will dominate the other two when σ_c^2 is very small compared with $\xi_{\min}$ and σ_x^2, a situation that would be true if the environment is only very slowly varying and the observation noise is relatively large. Note also that the equations (60) and (61) do not depend on ρ. This implies that the steady-state behavior of our algorithm is relatively insensitive to the choice of ρ when the environment is highly nonstationary.

Similar comparisons can also be made about the exponentially weighted recursive least squares filters. From [2], the excess mean-squared estimation error for the nonstationarity model we have is given by (for white input signals)

$$e_{\text{ex}} = \frac{1-\lambda}{1+\lambda} \xi_{\min} N + \frac{1}{2(1-\lambda)} N \sigma_x^2 \sigma_c^2 \tag{62}$$

where λ is the parameter of the exponential weighting factor and $0 < \lambda \leq 1$. e_{ex} is minimized when λ is chosen as

$$\lambda_{\text{opt}} = \frac{1-\beta}{1+\beta} \tag{63}$$

where

$$\beta = \left(\frac{\sigma_x^2 \sigma_c^2}{4\xi_{\min}}\right)^{1/2} \tag{64}$$

and the optimal value (say, $e_{\text{ex, RLS}}^*$) is given by

$$e_{\text{ex, RLS}}^* = N \sqrt{\sigma_x^2 \sigma_c^2 \xi_{\min}} + \frac{N}{4} \sigma_x^2 \sigma_c^2. \tag{65}$$

In a large class of practical situations, the second term on the right-hand side of (65) is negligible when compared with the first term. (The results for the LMS algorithm make use of approximations of the same order.) We can see that the optimal tracking performances of the LMS and RLS adaptive filters are very comparable if the optimal choice of the step size for the LMS algorithm is within the stability bounds. Otherwise (this happens for long filter lengths) the RLS algorithm will have a performance advantage [2]. Comparisons between the adaptive step-size algorithm and the RLS adaptive filter can now be made, similar to the comparisons with the LMS adaptive filter and similar conclusions can be reached, i.e., our algorithm tracks almost as well as the RLS adaptive filter with optimal choice of λ, even when the parameters of the filter are not very carefully chosen.

IV. Experimental Results

In this section, we will present the results of several experiments that demonstrate the good properties of the algorithm described in the paper and also compare the performance against that of the normalized LMS (NLMS) adaptive filters. Mikhael *et al.* [9] have shown that the normalized LMS algorithm with step-size $\mu = 0.5$ will provide the fastest speed of convergence among all gradient adaptive algorithm (with one step-size sequence) attempting to minimize the squared estimation error. The results of the experiments presented in this section will demonstrate that our algorithm's speed of convergence is comparable to the "fastest possible speed." All the simulation results presented are averages over 50 independent runs.

Example 1: In this example, we consider identifying a five-point FIR filter with coefficients

$$\{h_i;\, i = 0, 1, 2, 3, 4\} = \{0.1, 0.3, 0.5, 0.3, 0.1\}. \tag{66}$$

The input signal $x(n)$ is a pseudorandom, zero-mean, and Gaussian process obtained as the output of the all-pole

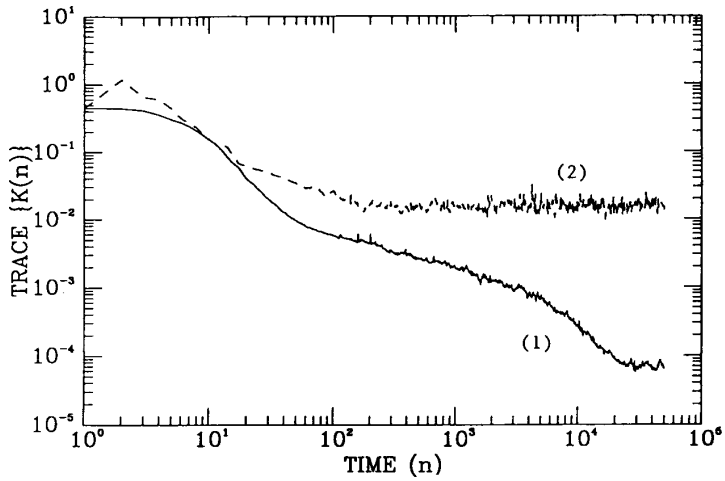


Fig. 1. Comparison of the performance of the gradient adaptive step-size algorithm and the normalized LMS adaptive filter: (*1*) Adaptive step-size algorithm and (*2*) NLMS filter. The algorithm described in the paper has initial convergence speeds similar to those of the NLMS adaptive filter, but its steady-state behavior is far superior.

filter with transfer function

$$A(z) = \frac{0.44}{1 - 1.5z^{-1} + z^{-2} - 0.25z^{-3}} \tag{67}$$

when the input to the filter was zero-mean, white, and pseudo-Gaussian noise with unit variance. Note that the variance of the resultant signal is aprpoximately one. The desired response signal $d(n)$ was obtained by corrupting the output of the system in (66) (when the input was $x(n)$) with zero-mean, white, and pseudo-Gaussian additive measurement noise. The measurement noise was uncorrelated with the input signal and its variance was 0.01. The adaptive filter was run using five coefficients (note that the eigenvalue spread of the input autocorrelation matrix is more than 140) and all the coefficients were initialized to have zero values.

In Fig. 1, we have plotted the sum of the mean-squared deviations of each coefficient from its mean value (mean-squared norm of the coefficient error vector) as a function of time for the first 50 000 iterations for both the algorithms. The parameters used were $\rho = 0.0008$ and $\mu(0) = 0.06$. The maximum possible value of $\mu(n)$ was set to $2/15$ in accordance with the discussion in Remark 2 of Section II. The step-size was limited to the same maximum value in all the experiments described in this paper.

Also plotted in Fig. 1 is the performance measure of the normalized LMS algorithm with $\mu = 0.5$. We can see that the algorithm presented in the paper has an initial convergence speed that is similar to that of the NLMS adaptive filter. However, the squared norm of the misalignment vector is more than 20 dB smaller for our algorithm after 50 000 iterations.

In Fig. 2, we have plotted the mean behavior of the convergence sequence $\mu(n)$ for the same problem. We can see that $\mu(n)$ goes up very fast initially and then comes down slowly and smoothly. This behavior explains the fast convergence and low misadjustment associated with the algorithm.

In Fig. 3, the mean-squared norm of the coefficient error vector is plotted for several values of $\mu(0)$ when $\rho = 0.0008$. As we would expect, the speed of convergence is slightly better for larger values of $\mu(0)$. However, we note that even for zero value for $\mu(0)$, the convergence rate is very good. Also, note that the steady-state behavior does not depend on the initial values of the step-size sequence. The behavior of $E\{\mu(n)\}$ is documented in Fig. 4 for the same experiment. Note that for each value of $\mu(0)$, the step size increases very quickly to some peak value and then decreases. Because of this type of behavior, speed of convergence of the algorithm is not very sensitive to the initialization of $\mu(n)$. Fig. 5 displays curves similar to those in Fig. 3 for several values of ρ and fixed $\mu(0) = 0.08$. Since the initial value of μ was relatively large in each case, the initial speed of convergence is more or less insensitive to the choice of ρ. As explained in Section III, the steady-state behavior does depend on ρ, but it will take a very large number of iterations (a few hundred thousands) before the differences show up in a significant fashion in this example. On the basis of these two figures, it is reasonable to infer that it is advisable to select a $\mu(0)$ that is close to the upper bound in (9). This would ensure very fast initial convergence speeds. ρ can be chosen to provide the desired level of steady-state performance.

Example 2: In this example, we study the performance of the adaptive filter when the operating environment changes abruptly. The nonstationarity model considered in the analysis will be investigated in Example 4. The input signals and optimal coefficient set were the same as in the previous example from time $n = 0$ to $n = 10000$.

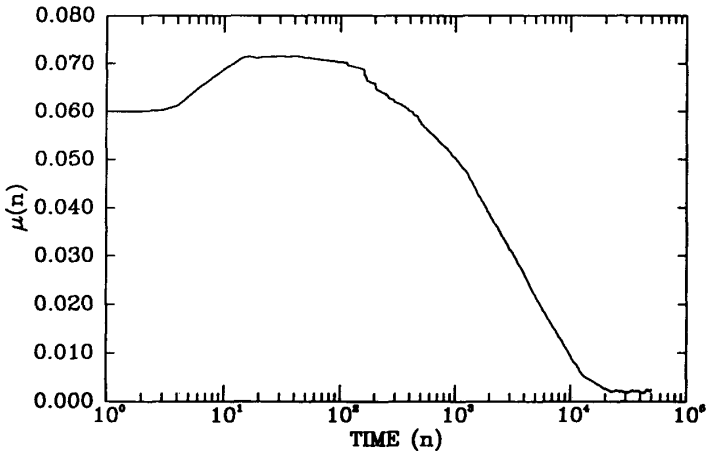


Fig. 2. Mean behavior of $\mu(n)$ goes up very quickly and then smoothly descends to very small values in stationary environments. This behavior accounts for the very fast convergence speed and small misadjustment exhibited by the algorithm.

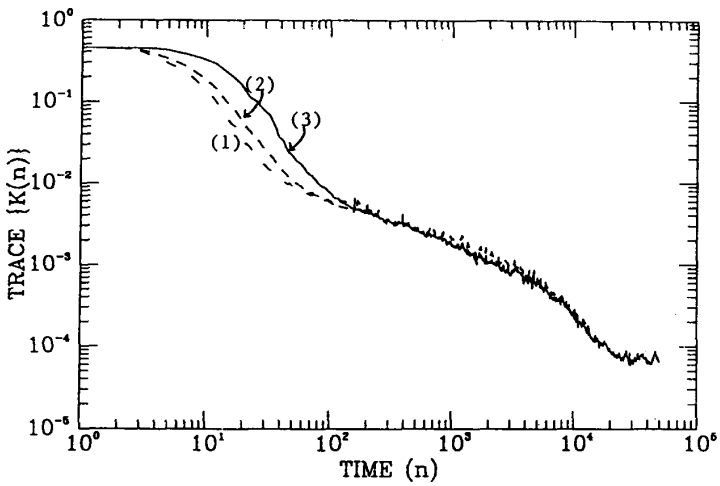


Fig. 3. Performance of the adaptive filter for different values of $\mu(0)$ and fixed $\rho = 0.0008$: (*1*) $\mu(0) = 0.08$, (*2*) $\mu(0) = 0.04$, and (*3*) $\mu(0) = 0.00$. Even for $\mu(0) = 0.0$, the algorithm exhibits very good convergence speed.

At $n = 10001$, the coefficients of the unknown system were all changed to their corresponding negative values. The mean-squared coefficient behavior and the mean step-size behavior are tabulated in Figs. 6 and 7, respectively. We observe that the step size increases very quickly immediately after the environment changes and therefore the algorithm is able to track abrupt changes in the operating environments very well. (Because of this change of scales in the time axis, the time indices 10000 and 10001 are represented by the same point in both Figs. 6 and 7.)

Example 3: This example and the next one are intended to demonstrate the validity of the analysis in Section III. We also demonstrate some attractive properties of the adaptive filter using numerical evaluation of some of the analytical results. We again consider identifying the same system as in Example 1, but with a zero-mean and white Gaussian input signal with unit variance. The measurement noise variance was 0.01 and the parameters of the adaptive filter were $\rho = 0.006$ and $\mu(0) = 0.06$. In Fig. 8, we have plotted the trace of the second moment matrix of the coefficient misalignment vector obtained from the theoretical analysis presented in this section and also from simulation experiments. Note that the experimental and analytical results match very well, in spite of the several simplifying approximations made in the analysis. As before, note that the algorithm shows very fast initial convergence behavior and then slowly "improves" the estimates so as to get very low misadjustment values in the steady state. Fig. 9 displays the empirical and analytical behavior of the mean step-size sequence for this

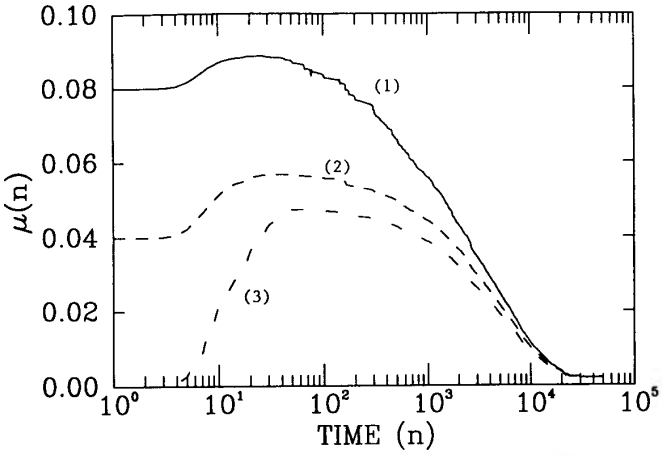


Fig. 4. Mean behavior of $\mu(n)$ for different initializations and fixed $\rho = 0.0008$. (*1*) $\mu(0) = 0.08$, (*2*) $\mu(0) = 0.04$, and (*3*) $\mu(0) = 0.00$.

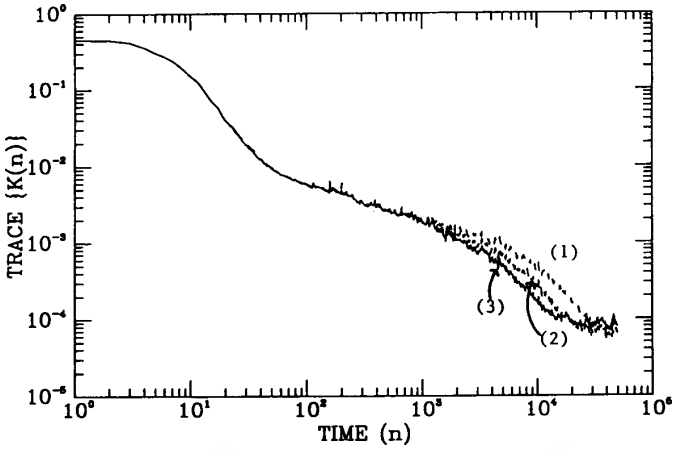


Fig. 5. Performance of the adaptive filter for $\mu(0) = 0.08$ and (*1*) $\rho = 0.0012$, (*2*) $\rho = 0.0008$, and (*3*) $\rho = 0.0005$. By choosing $\mu(0)$ to be relatively large, we can get very good initial convergence speeds for a large range of values of ρ.

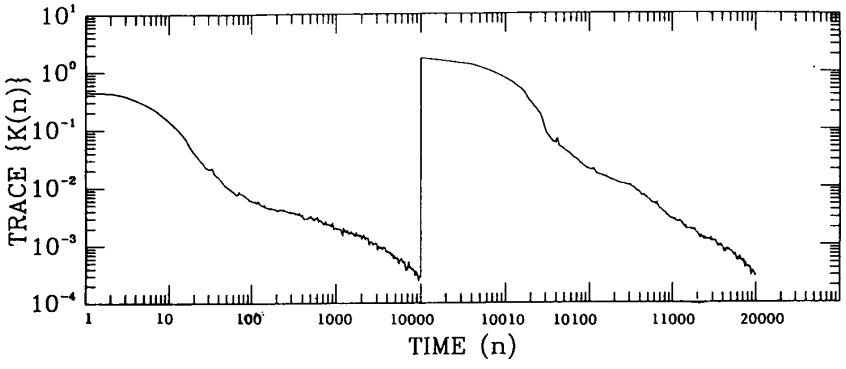


Fig. 6. Response of the adaptive filter to an abrupt change in the environment. Note that the time axis uses different scales before and after $n = 1000$.

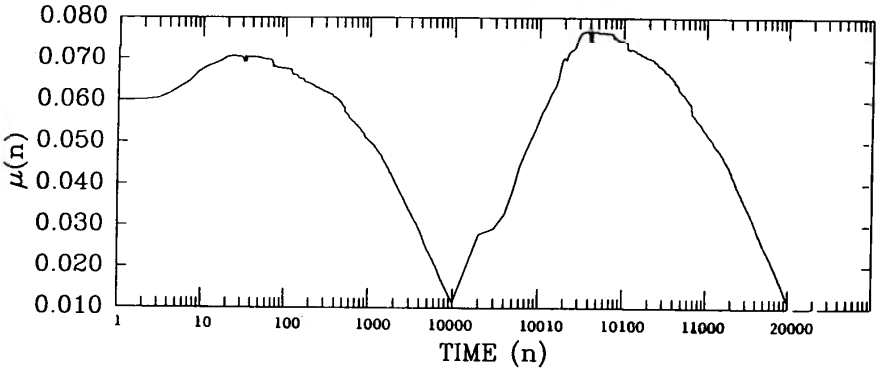


Fig. 7. Mean behavior of $\mu(n)$ when there is an abrupt change in the environment. Again note that the time scales are different before and after $n = 10000$.

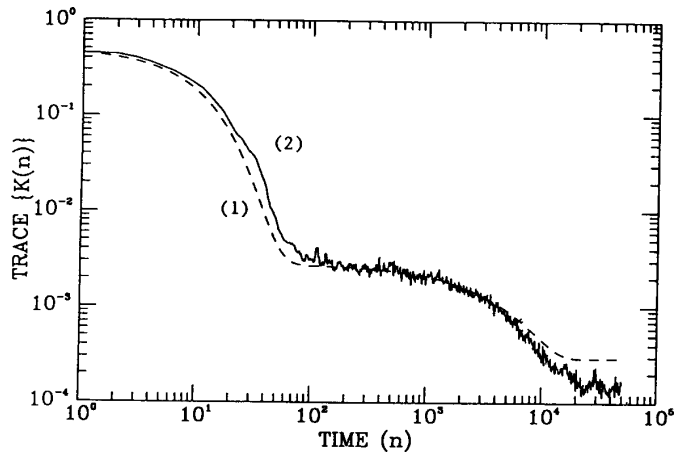


Fig. 8. Comparison of empirical and analytical results for the mean-squared behavior of the adaptive filter coefficients: (*1*) Analytical (from (25)), and (*2*) empirical curve.

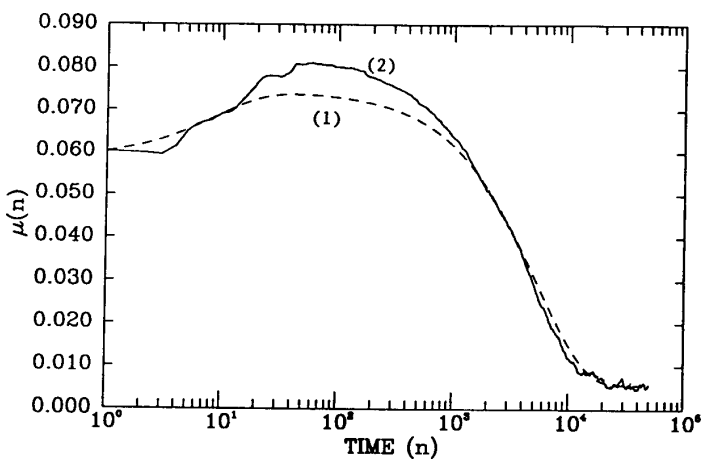


Fig. 9. Comparison of empirical and analytical results for the mean behavior of the step-size sequence: (*1*) Analytical (from (28)), and (*2*) empirical curve.

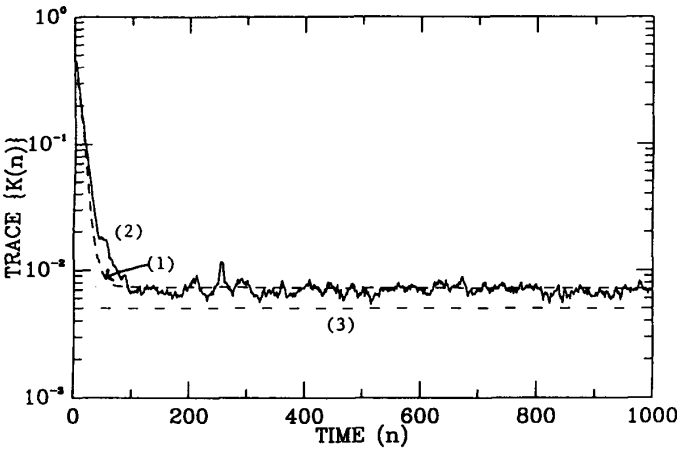


Fig. 10. Comparison of the mean-squared coefficient behavior of the adaptive filter in a nonstationary environment. (*1*) Analytical result, (*2*) empirical result, and (*3*) the best performance that can be achieved by the LMS adaptive filter.

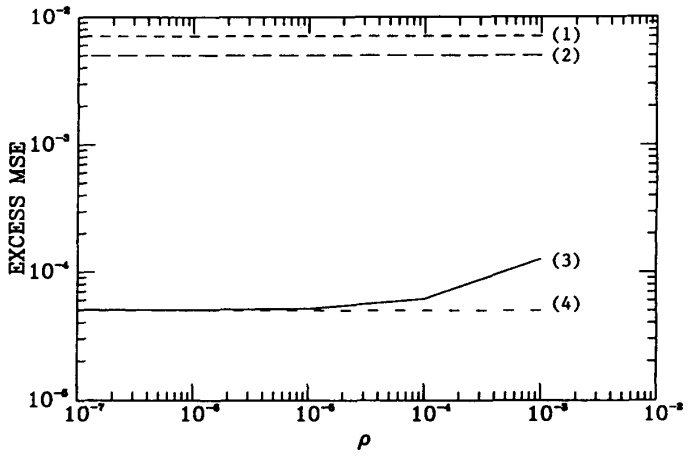


Fig. 11. Steady-state performance on the adaptive step-size algorithm and the best performance of the LMS algorithm in two nonstationary environments. (*1*) Adaptive step-size algorithm, $\sigma^2 = 10^{-4}$. (*2*) Best performance of the LMS adaptive filter, $\sigma^2 = 10^{-4}$. (*3*) Adaptive step-size algorithm, $\sigma_c^2 = 10^{-8}$. (*4*) Best performance of the LMS filter $\sigma_c^2 = 10^{-8}$.

problem. Once again, observe that there is very good match between the two curves.

Example 4: In this example, we consider the identification of a time-varying system. The time-varying coefficients $H_{\text{opt}}(n)$ of the system are modeled using a random disturbance process

$$H_{\text{opt}}(n) = H_{\text{opt}}(n-1) + C(n), \tag{68}$$

where $C(n)$ is a zero-mean white vector process with covariance matrix

$$\sigma_c^2 I = 10^{-4}\, I. \tag{69}$$

The initial values of the optimal coefficients were as in Example 1. The sum of the mean-squared coefficient misalignment values ($\text{tr}\{K(n)\}$) obtained using analysis and experimentation are plotted in Fig. 10. All the parameters of the experiments with the exception of the actual coefficients of the time-varying system were the same as in the previous example. Note the close match between the two curves. Also note that the steady-state behavior is only slightly worse than the best possible steady-state performance of the LMS algorithm.

Finally, we evaluate the steady-state excess mean-squared error predicted by our analysis in Section III numerically for several values of ρ for the same system identification problem. These quantities, which were obtained from equations (36)–(38), are plotted against the optimal performance of the LMS adaptive filter in Fig. 11. As predicted in Section III, we note that the performance of the adaptive step-size algorithm is very close to that of the best performance of the LMS algorithm for a large range of values of ρ. Note also that the steady-state be-

havior seems to be only very weakly dependent on ρ for a large range of values. This is a very attractive property since we can make the design decisions based more or less on the speed of convergence alone in the nonstationary tracking problems. Even though curves (*1*) and (*3*) in the figure were obtained using (36)–(38), it must be pointed out that the difference between curve (*1*) and the approximate value given by (61) was very small (much smaller than 1% for all the values of ρ in the figure. Similarly, the differences between curve (*3*) and the approximate results given by (48) were also very small—less than 1% for $\rho \leq 10^{-4}$ and 2.3% for $\rho = 10^{-3}$. These comparisons demonstrate the usefulness of the approximations that were made to obtain the simplified results.

V. Concluding Remarks

This paper presented a stochastic gradient adaptive filtering algorithm with time-varying step sizes. The algorithm is different from traditional methods involving time-varying step sizes in that the changes in the step-sizes were also controlled by a gradient algorithm designed to minimize the squared estimation error. We presented a theoretical performance analysis of the algorithm. Experimental results showed that 1) the initial convergence rate of the adaptive filters is very fast. After an initial period when the step size increases very rapidly, the step size decreases slowly and smoothly, giving rise to small misadjustment errors; and, 2) in the case of nonstationary environments, the algorithms seek to adjust the step-sizes in such a way as to obtain close-to-the-best-possible performance. The steady-state performance of the gradient, adaptive step-size adaptive filter is often close to the best possible performance of the LMS and RLS algorithms, and is relatively independent of the choice of ρ in many nonstationary environments. The good properties and the computational simplicity associated with the algorithm makes us believe that it will be used consistently and successfully in several practical applications in the future.

Appendix

Derivation of (25), (28) and (29)

Taking the statistical expectation of both sides of (24), we get

$$K(n+1) = E\{(I - \mu(n)X(n)X^T(n))K(n) \cdot (I - \mu(n)X(n)X^T(n))\} + E\{\mu^2(n)\}\,\xi_{\min}\sigma_x^2 I + \sigma_c^2 I. \quad \text{(A1)}$$

In order to obtain (A1), we have made the use of the independent assumption and also the uncorrelatedness of $\mu(n)$ with the other quantities involved. Expanding the first term on the right-hand side, (A1) transforms to

$$K(n+1) = K(n) - 2E\{\mu(n)\}\,\sigma_x^2 K(n) + E\{\mu^2(n)\}E\{X(n)X^T(n)K(n)X(n)X^T(n)\} + E\{\mu^2(n)\}\,\xi_{\min}\sigma_x^2 I + \sigma_c^2 I. \quad \text{(A2)}$$

Realizing that the entries of $X(n)$ are zero-mean and white Gaussian random variables and that fourth-order expectations of Gaussian variables can be expressed as a sum of products of second-order expectations [12], we can simplify the fourth-order expectation in (A2) as

$$E\{X(n)X^T(n)K(n)X(n)X^T(n)\} = 2\sigma_x^4 K(n) + \sigma_x^4 \operatorname{tr}\{K(n)\}I. \quad \text{(A3)}$$

Substituting (A3) in (A2) and simplifying using (26) results in (25).

In order to develop the evolution equations for the mean behavior of $\mu(n)$, we start with taking the statistical expectation of (5). This yields

$$E\{\mu(n)\} = E\{\mu(n-1)\} + \rho E\{e(n)e(n-1)X^T(n)X(n-1)\}. \quad \text{(A4)}$$

Expanding $e(n)$ as in (20) and substituting for $V(n)$ from (21) in the second expectation on the right-hand side of (A4) will lead to the following:

$$\begin{aligned} E\{e(n)e(n-1)X^T(n)X(n-1)\} &= -\xi_{\min}E\{\mu(n-1)\}E\{X^T(n)X(n-1) \cdot X^T(n-1)X(n)\} + E\{X^T(n-1)V(n-1)X^T(n) \\ &\quad \cdot V(n-1)X^T(n-1)X^T(n-1)X(n)\} \\ &\quad - E\{\mu(n-1)\}E\{X^T(n)X(n-1)X^T(n-1) \\ &\quad \cdot V(n-1)X^T(n-1)V(n-1)X^T(n-1)X(n)\}. \end{aligned} \quad \text{(A5)}$$

Some straightforward calculations will show that

$$E\{X^T(n)X(n-1)X^T(n-1)X(n)\} = N\sigma_x^4 \quad \text{(A6)}$$

$$E\{X^T(n-1)V(n-1)X^T(n)V(n-1)X^T(n-1)X(n)\} = \sigma_x^4 \operatorname{tr}\{K(n-1)\} \quad \text{(A7)}$$

and

$$E\{X^T(n)X(n-1)X^T(n-1)V(n-1)X^T(n-1) \cdot V(n-1)X^T(n-1)X(n)\} = (2+N)\sigma_x^6 \operatorname{tr}\{K(n-1)\}. \quad \text{(A8)}$$

Substituting (A6), (A7), and (A8) in (A4) and simplifying using (26) will give (28).

Derivation of the evolution equation for $E\{\mu^2(n)\}$ requires some more simplifications. With the help of the set of assumptions stated in Section III, it is relatively easy to show that

$$E\{e(n)e(n-1)\} = 0. \quad \text{(A9)}$$

(This simplification was not used in deriving (28).) Squaring both sides of (5) and taking the expectations gives

$$E\{\mu^2(n)\} = E\{\mu^2(n-1)\} + 2\rho E\{\mu(n-1)\} \cdot E\{e(n-1)e(n)X^T(n-1)X(n)\} + \rho^2 E\{e^2(n-1)e^2(n)X^T(n-1) \cdot X(n)X^T(n)X(n-1)\}. \quad \text{(A10)}$$

The second term on the right-hand side of the above equation has been simplified using Assumption iv. Only the third term above remains to be evaluated. Using the independence assumption and the uncorrelatedness of $e(n)$ and $e(n-1)$,

$$E\{e^2(n-1)e^2(n)X^T(n-1)X(n)X^T(n)X(n-1)\} = E\{e^2(n-1)X^T(n-1) \cdot E\{e^2(n)X(n)X^T(n)\}X(n-1)\}. \quad \text{(A11)}$$

The inner expectation can be evaluated by recognizing that $e(n)$ is a zero-mean and Gaussian signal when conditioned on the coefficient vector $H(n)$. With the help of the approximation in (18), we can derive the following result:

$$\begin{aligned} E\{e^2(n)X(n)X^T(n)\} &\approx E\{e^2(n)X(n)X^T(n)|H(n)\} \\ &= \xi_{\min}\sigma_x^2 I + 2\sigma_x^4 n K(n) + \sigma_x^4 \operatorname{tr} K(n) I \\ &= \sigma_x^2\sigma_e^2(n) I + 2\sigma_x^4 K(n). \end{aligned} \quad \text{(A12)}$$

Let

$$\beta(n) = \sigma_x^2\sigma_e^2(n) I + 2\sigma_x^4 K(n). \quad \text{(A13)}$$

It is straightforward to show using the same approach as above that

$$E\{e^2(n-1)e^2(n)X^T(n-1)X(n)X^T(n)X(n-1)\} = \operatorname{tr}\{\beta(n-1)\beta(n)\}. \quad \text{(A14)}$$

Equation (29) follows immediately.

References

[1] P. Bragard and G. Jourdain, "A fast self-optimized LMS algorithm for nonstationary identification—application to underwater equalization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, NM, Apr. 1990, pp. 1425–1428.

[2] E. Eleftheriou and D. D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 5, pp. 1097–1110, Oct. 1986.

[3] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 1, pp. 222–230, Feb. 1985.

[4] R. W. Harris, D. M. Chabries, and F. A. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 2, pp. 309–316, Apr. 1986.

[5] S. Karni and G. Zeng, "A new convergence factor for adaptive filters," *IEEE Trans. Circuits Syst.*, vol. 36, no. 7, pp. 1011–1012, July 1989.

[6] C. P. Kwong, "Dual sign algorithm for adaptive filtering," *IEEE Trans. Commun.*, vol. COM-34, no. 12, pp. 1272–1275, Dec. 1986.

[7] V. J. Mathews and H. S. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 4, pp. 450–454, Apr. 1987.

[8] V. J. Mathews and Z. Xie, "Stochastic gradient adaptive filters with gradient adaptive step sizes," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, NM, Apr. 1990, pp. 1385–1388.

[9] W. B. Mikhael *et al.*, "Adaptive filters with individual adaptation of parameters," *IEEE Trans. Circuits Syst.*, vol. CAS-33, no. 7, pp. 677–685, July 1986.

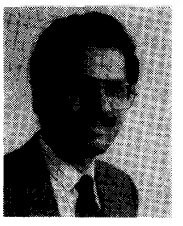
[10] T. J. Shan and T. Kailath, "Adaptive algorithms with an automatic gain control feature," *IEEE Trans. Circuits Syst.*, vol. CAS-35, no. 1, pp. 122–127, Jan. 1988.

[11] Y. K. Shin and J. G. Lee, "A study on the fast convergence algorithm for the LMS adaptive filter design," *Proc. IEEE*, vol. 19, no. 5, pp. 12-19, Oct. 1985.

[12] H. Urkowitz, *Signal Theory and Random Processes*. Dedham, MA: Artech House, 1983.

[13] V. Weerackody, S. A. Kassam, and K. R. Laker, "Convergence analysis for an algorithm for blind equalization," *IEEE Trans. Communi.*, vol. 39, no. 6, pp. 856–865, June 1991.

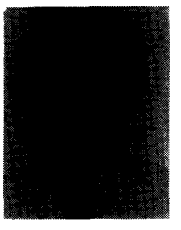
[14] B. Widrow *et al.*, "Stationary and nonstationary learning characteristics of the LMS adaptive filters," *Proc. IEEE*, vol. 64, no. 8, pp. 1151–1162, Aug. 1976.



V. John Mathews (S'82–M'84–SM'90) was born in Nedungadappally, Kerala, India, in 1958. He received the B.E. (hons.) degree in electronics and communication engineering from the University of Madras, India, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Iowa, Iowa City, in 1980, 1981, and 1984, respectively.

From 1980 to 1984 he held a Teaching Research Fellowship at the University of Iowa, where he also worked as a Visiting Assistant Professor with the Department of Electrical and Computer Engineering from 1984 to 1985. He is currently an Associate Professor with the Department of Electrical Engineering, University of Utah, Salt Lake City. His research interests include adaptive filtering, spectrum estimation, and data compression.

Dr. Mathews was an Associate Editor of the IEEE Transactions on Signal Processing during 1989–1991. He is a member of the Digital Signal Processing Technical Committee of the IEEE Signal Processing Society.



Zhenhua Xie (S'84–M'90) was born in Zhejiang, People's Republic of China, in July 1958. He received the B.S. degree from Zhejiang University, China, in 1982 and the M.S. and Ph.D. degrees from the University of Utah, in 1985, 1989, respectively, all in electrical engineering.

From 1989 to 1990 he worked as a postdoctoral fellow with the Department of Electrical Engineering, University of Utah. He is currently a Member of the Technical Staff with the Technical Center, R. R. Donnelley & Sons Company, Lisle, IL. His research interests include electronic publication, color science and color reproduction, human vision modeling, multiuser communications, and adaptive filtering.