Tracking Analysis of the Sign Algorithm in Nonstationary Environments

SUNG HO CHO, MEMBER, IEEE, AND V. JOHN MATHEWS, SENIOR MEMBER, IEEE

Abstract—This paper presents a tracking analysis of the adaptive filters equipped with the sign algorithm and operating in nonstationary environments. Under the assumption that the nonstationarity can be modeled using a random disturbance, it is shown that the long-term time average of the mean-absolute error is bounded and that there exists an optimal choice of the convergence constant μ which minimizes this quantity. Using the commonly employed independence assumption and under the assumption that the nonstationarity is solely due to the time-varying behavior of the optimal coefficients, we then show that the distributions of the successive coefficient misalignment vectors converge to a limiting distribution when the adaptive filter is used in the "system identification" mode. Finally, under the additional assumption that the signals involved are zero mean and Gaussian, we derive a set of nonlinear difference equations that characterizes the mean and mean-squared behavior of the filter coefficients and the mean-squared estimation error during adaptation and tracking. Results of several experiments that show very good correlation with the theoretical analyses are also presented in this paper.

I. Introduction

THE least mean-square (LMS) adaptive filtering algorithm [23] is very popular because of its simplicity. However, there are many applications for which even simpler approaches are needed in order to implement the adaptive algorithm in real time. Consequently, the sign algorithm and its variants have been actively studied in recent years [4], [6], [8]–[10], [13], [14], [16], [18], [22], [26]. This paper presents a performance analysis of the sign algorithm operating in nonstationary environments. The contributions of this paper are mainly the following three:

1) By modeling the nonstationarity as a random disturbance, we show that the long-term average of the mean-absolute error is bounded for all positive values of the convergence parameter and that there exists an optimal choice of the convergence parameter that minimizes this bound.

2) By making use of the commonly employed independence assumption and further assuming that the random disturbance itself is stationary and that the adaptive filter is used in the "system identification" mode, we show that the coefficient misalignment vector converges in distribution for all positive values of the convergence parameter.

3) Finally, by further assuming that the signals involved are Gaussian and have zero-mean values, we develop a set of nonlinear difference equations that characterizes the mean and mean-squared behavior of adaptive filters equipped with the sign algorithm during adaptation and tracking.

The rest of the paper is organized as follows. The next section contains a statement of the problem and a brief discussion of past work done in this area. Section III contains the analysis of the sign algorithm that demonstrates the existence of an optimum value for the convergence parameter. The analysis that deals with the convergence in distribution of the coefficient misalignment vector is presented in Section IV. In Section V we develop the nonlinear difference equations that characterize the behavior of the sign algorithm during adaptation and tracking. Several simulation examples demonstrating the validity of the analytical results derived in this paper are included in Section VI. Finally, the concluding remarks are made in Section VII.

II. Problem Statement and Past Work

Consider the problem of adaptively estimating the desired response input signal $d(n)$ using the current and past $N-1$ samples of the reference input signal $x(n)$. The sign algorithm (SA) [4], [10], [16] updates the adaptive filter coefficient vector $H(n)$ of size N as

$$H(n+1) = H(n) + \mu \operatorname{sign}(e(n))\, X(n) \qquad (1)$$

where

$$e(n) = d(n) - H^T(n)\, X(n) \qquad (2)$$

denotes the estimation error at time n, $X(n)$ denotes the reference input vector to the adaptive filter defined as

$$X(n) = [x(n), x(n-1), \cdots, X(n-N+1)]^T \qquad (3)$$

$(\bullet)^T$ denotes the matrix transpose of $(\bullet)$, and μ denotes the convergence parameter.

Even though many people consider the SA as an approximation to the LMS algorithm, it really is a stochastic gradient adaptive algorithm that tries to minimize the absolute value of the estimation error at each time. Several variations of the SA are available in literature. The dual sign algorithm [13], [18] has essentially the same com-

Manuscript received April 15, 1989; revised December 19, 1989.

S. H. Cho was with the University of Utah, Salt Lake City, UT 84112. He is now with the Electronics and Telecommunications Research Institute, Section 0730, Daejeon 305-606, Korea.

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

IEEE Log Number 9038957.

putational complexity as the SA, but has faster convergence speeds. Multiplication-free implementation of adaptive filters that combines differential pulse-code modulation (DPCM) with the SA was introduced in [17], [19]. Adaptive lattice-filter structures employing the SA were studied in [11], [26].

Even though techniques similar to the SA were studied as early as 1964 (see [10] for some early references), it is only in the late 1970's, after Verhoeckx *et al.* [21] used the SA for echo cancellation, that there has been vigorous activity in the study of the SA. Claasen and Mecklenbrauker [4] compared the performance of the SA with that of the LMS algorithm when the input signal is zero mean and white. The results from this work were used for developing some design rules for the SA in [22]. Duttweiler [6] has considered the more general problem of updating the coefficient vector using arbitrary nonlinearities in the correlation multiplier. Bershad [2] has analyzed adaptive systems with an error-function type nonlinearity applied to the estimation error in the coefficient update equation. By varying a certain parameter of the error-function type nonlinearity, the performance of the adaptive filter could be studied when the estimation error sequence was represented using a finite number of bits in the update equation. All these studies, using the assumption that the input signal is white, have shown that nonlinearities in the correlation multiplier (especially when applied to the error term) in general slows down the convergence of the adaptive filter when compared with the basic LMS algorithm for fixed steady-state excess mean-squared estimation error.

Mathews and Cho [16] studied the SA for colored, but Gaussian, input signals. Their studies also showed that the convergence speed of the SA is slower than that of the LMS algorithm when both systems produce the same steady-state mean-squared estimation error. Kwong [13] showed that a simple modification of the SA (he termed the resultant structure the dual-sign algorithm (DSA)) can improve the convergence speed without increasing the computational complexity or degrading the steady-state performance significantly. Mathews [18] has analytically verified the good characteristics of the DSA.

Gersho [10] proved under very general conditions that the long-term average of the mean-absolute estimation error produced by adaptive filters equipped with the SA is bounded for any value of the convergence parameter μ. Under the commonly used independence assumption, he also showed that the distributions of the successive coefficient misalignment vector converges to a limiting distribution. Both of these results were proved for the case when the SA was operating in stationary environments. Eweda [8] recently proved almost sure convergence of the SA for correlated input signals when the convergence parameter was a monotone decreasing function of time. There has been limited work done in analyzing the performance of the SA in nonstationary environments. With the help of several simplifying assumptions, Eweda [9] has shown that there exists an optimal value of the convergence parameter that minimizes the long-term average of the absolute value of the estimation error. We will prove a similar result for the mean-absolute error in this paper. The only other work that the authors are aware of was recently done by Kwong [14], where he employed a control theoretic formulation for designing the SA in nonstationary environments.

This paper is a more comprehensive study than either of the above works. In fact, this work extends the results obtained in [10] and [16] for the SA operating in stationary environments to the case of the SA operating in nonstationary environments.

Before we start the analysis, let us introduce certain notations that we will use throughout the paper. Let $H_{\text{opt}}(n)$ denote the optimal coefficient vector at time n. Note that we have explicitly used the time index n for the optimal coefficient to stress the fact that it can be time varying. Also, let

$$C(n) = H_{\text{opt}}(n) - H_{\text{opt}}(n+1) \tag{4}$$

be the difference vector between the optimal coefficient values at times n and $n + 1$. It is also convenient to define the coefficient misalignment vector as

$$V(n) = H(n) - H_{\text{opt}}(n). \tag{5}$$

Finally, note that the optimal estimation error at time n is given by

$$e_{\min}(n) = d(n) - H_{\text{opt}}^T(n)\, X(n). \tag{6}$$

Substituting (5) and (6) in (2), we also have the following result:

$$e_{\min}(n) - e(n) = V^T(n)\, X(n). \tag{7}$$

III. Boundedness of the Long-Term Average of the Mean-Absolute Estimation Error and the Existence of an Optimum μ

In this section, under a set of very mild assumptions, we will show that the long-term time average of the mean-absolute estimation error is bounded for all positive values of μ and that there exists an optimal value of μ that minimizes it.

The analysis in this section uses the following assumptions.

Assumption 1: $d(n)$ and $X(n)$ are zero-mean random processes and have finite-second moments. Furthermore,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} E\{X(k)\, X^T(k)\} \qquad \text{exists and is finite.}$$

Assumption 2: $C(n)$ is a zero-mean random process with bounded autocorrelation matrix, and is independent of $X(n)$. $C(n)$ is also independent of $C(k)$ if $n \neq k$. Moreover,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} E\{C(k)\, C^T(k)\} \qquad \text{exists and is finite.}$$

Assumption 3: The long-term time average of the mean-absolute value of the minimum estimation error $e_{\min}(n)$ exists and is finite.

Note that in Assumption 1, we are not restricting the nature of the autocorrelation matrix of $X(n)$ or the nature of the correlation of $X(n)$ and $X(k)$ for arbitrary n and k. The existence of the limit guarantees that the covariances do not grow in an unbounded manner. In Assumption 2, we are basically assuming that the nonstationarity can be modeled as a random disturbance. While this is not true in many situations, there are several practical applications in which this assumption is valid. Many of the analyses of the LMS adaptive algorithm in nonstationary environments make this assumption [1], [15]. Furthermore, the situations where the optimal coefficient process is assumed to belong to a first-order Markov process with the Markov parameter very close to one (for example, see [7], [24], [25]) are only slightly different from our situation and most of such results are exact only when the Markov parameter is one, which is the same as our model. Note that Assumption 2 does not in any way restrict the characteristics of the input signal to the adaptive filter.

Using (4) and (5) in (1), the coefficient update equation can be rewritten in terms of the coefficient misalignment vector as

$$V(n+1) = V(n) + \mu \operatorname{sign}(e(n)) X(n) + C(n). \tag{8}$$

Premultiplying both sides of (8) with their own transpose, we get

$$\begin{aligned}\|V(n+1)\|^2 = {} & \|V(n)\|^2 + \mu^2\|X(n)\|^2 + \|C(n)\|^2 \\ & + 2V^T(n)\, C(n) \\ & + 2\mu \operatorname{sign}(e(n))\, V^T(n)\, X(n) \\ & + 2\mu \operatorname{sign}(e(n))\, X^T(n)\, C(n) \end{aligned} \tag{9}$$

where $\|(\bullet)\|^2$ denotes the squared Euclidean norm of $(\bullet)$. Using (7) in (9), it follows that

$$\begin{aligned}\|V(n+1)\|^2 \le {} & \|V(n)\|^2 + \mu^2\|X(n)\|^2 + \|C(n)\|^2 \\ & + 2V^T(n)\, C(n) \\ & + 2\mu|e_{\min}(n)| - 2\mu|e(n)| \\ & + 2\mu \operatorname{sign}(e(n))\, X^T(n)\, C(n). \end{aligned} \tag{10}$$

Taking the statistical expectation of both sides of (10) and employing Assumptions 1 and 2, we obtain

$$\begin{aligned} & E\|V(n+1)\|^2 \\ & \quad \le E\|V(n)\|^2 + \mu^2 E\|X(n)\|^2 + E\|C(n)\|^2 \\ & \qquad + 2\mu E|e_{\min}(n)| - 2\mu E|e(n)|. \end{aligned} \tag{11}$$

The following inequality results from iterating (11) n times:

$$\begin{aligned} & E\|V(n+1)\|^2 \\ & \quad \le E\|V(1)\|^2 + \mu^2 \sum_{k=1}^{n} E\|X(k)\|^2 + \sum_{k=1}^{n} E\|C(k)\|^2 \\ & \qquad + 2\mu \sum_{k=1}^{n} E|e_{\min}(k)| - 2\mu \sum_{k=1}^{n} E|e(k)|. \end{aligned} \tag{12}$$

Since the left side of (12) is always nonnegative, it follows that

$$\begin{aligned} & 2\mu \sum_{k=1}^{n} E|e(k)| \\ & \quad \le E\|V(1)\|^2 + \mu^2 \sum_{k=1}^{n} E\|X(k)\|^2 + \sum_{k=1}^{n} E\|C(k)\|^2 \\ & \qquad + 2\mu \sum_{k=1}^{n} E|e_{\min}(k)|. \end{aligned} \tag{13}$$

Since the initial misalignment vector is bounded in all practical situations, we can obtain the following upper bound for the long-term time average of the mean-absolute estimation error, i.e.,

$$\begin{aligned} & \limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} E|e(k)| \\ & \quad \le \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left\{ \frac{\mu}{2} E\|X(k)\|^2 + \frac{1}{2\mu} E\|C(k)\|^2 \right. \\ & \qquad \left. + E|e_{\min}(k)| \right\}. \end{aligned} \tag{14}$$

Under Assumptions 1–3, the right-hand side of the inequality exists and is finite for any nonzero and finite value of μ.

The above result is intuitively appealing. It shows that the long-term time average of the excess mean-absolute estimation error is bounded by two terms: one of which is proportional to μ and depends on the statistical variability of the input signal and the other term that is inversely proportional to μ and depends on the statistical variability of the random disturbance that contributes to the nonstationarity. Note that a smaller value of μ will decrease the ability of the adaptive system to track nonstationary environments and thus the term that is inversely proportional to μ gives a quantitative measure for the tracking ability of the adaptive filter.

It is important to note that the result given in (14) is very general. It applies to all adaptive filters equipped with the sign algorithm regardless of the applications in which they are used. Furthermore, the long-term time average of the mean absolute estimation error is bounded regardless of how large μ is or how fast the unknown system changes. In this sense, the sign algorithm is much more robust than the LMS algorithm, which works well only if the convergence parameter is within a range specified by the statistics of the input signal.

It is straightforward to show that there exists an optimal value of μ that minimizes the right-hand side of (14) and that the value μ_{opt} is given by

$$\mu_{\text{opt}} = \sqrt{\frac{\operatorname{ave}\langle E\|C(n)\|^2 \rangle}{\operatorname{ave}\langle E\|X(n)\|^2 \rangle}} \tag{15}$$

where $\operatorname{ave}\langle(\bullet)\rangle$ denotes a long-term time average of $(\bullet)$. Note that if $X(n)$ and $C(n)$ are known to be stationary processes, μ_{opt} can be written as

$$\mu_{opt} = \sqrt{\frac{E\|C(n)\|^2}{E\|X(n)\|^2}}. \qquad (16)$$

Even though minimization of the bound does not necessarily minimize the left-hand side of the inequality in (14), all the experiments that we have done indicates that the long-term time average of the mean-absolute estimation error is minimized when μ is at or very close to the value suggested by (15). (See the results in Section VI.) Furthermore, we show in Section V under somewhat more restrictive assumptions that the mean-squared estimation error is indeed minimized when the convergence parameter is chosen to be as given by (16). Also, a comparison of the performances of the sign and LMS algorithms operating in the same nonstationary environment is given in Section V.

IV. Convergence in Distribution of the Coefficient Misalignment Vector

In this section we derive much stronger results for the performance of the sign algorithm in nonstationary environments. In order to do so, we need to restrict our attention to the "system identification mode" as depicted in Fig. 1. We will assume that the "plant" that is being identified has time-varying characteristics, but that the input signal $x(n)$ as well as the measurement noise $e_{min}(n)$ belong to stationary processes. While this setup is not representative of all situations, there are several practical applications in which the adaptive filter is used in an environment as described above and therefore the analysis using this setup will give useful information about the performance of the SA.

Under a certain set of assumptions, we will now show that the distribution functions of successive coefficient misalignment vectors converge to a limiting distribution. The analysis closely follows the method used in [10] for showing that the coefficient vector sequence of the SA converges in distribution when the filter operates in stationary environments. Obviously, in the nonstationary case, the coefficient vector sequence will not converge to a limiting value since the optimal coefficient vector sequence is time varying. Therefore, we modify the analysis to the case when the properties of the coefficient misalignment vector sequence are considered.

In addition to the assumptions made in Section III, we need the following assumptions to make the analysis tractable.

Assumption 4: The input pair $\{X(n), d(n)\}$ is independent of $\{X(k), d(k)\}$ if $n \neq k$.

Assumption 5: The joint probability density function of $\{X(n), e_{min}(n)\}$ is continuous and strictly positive. The probability density function of $C(n)$ is also continuous and strictly positive. Moreover, $X(n)$, $C(n)$, and $e_{min}(n)$ all belong to stationary random processes.

Assumption 4 is the commonly used "independence assumption" and is seldom true in practice. However, analysis using this assumption has produced results in the past that accurately predict the behavior of adaptive filters in circumstances where the assumption is grossly violated. For a single-coefficient LMS adaptive filter with Gaussian input signals and operating in stationary environments, Bershad and Qu [3] have shown that the analytical results obtained with and without the independence assumption differ significantly only when the inputs are extremely correlated (correlation coefficient > 0.99). Futhermore, if the coefficient updates are made at time intervals longer than one sampling interval, it is possible that the correlation between successive input vectors used for coefficient updates is negligible.

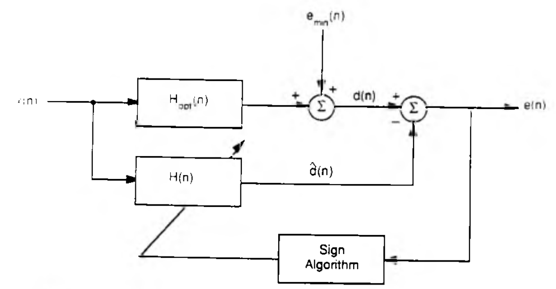

Fig. 1. Adaptive system identification using the sign algorithm.

One of the consequences of Assumption 4 is that this, in conjunction with Assumption 2, implies that the coefficient misalignment vector belongs to a Markov process. Furthermore, by defining a cost function ψ as the mean-absolute value of the estimation error process conditioned on all past misalignment values, we have

$$\begin{aligned}\psi(V(n), V(n-1), \cdots, V(1)) &= E\{|e(n)|\,|V(n), V(n-1), \cdots, V(1)\} \\ &= E\{|e_{min}(n) - V^T(n)X(n)|\,|V(n), \\ &\quad V(n-1), \cdots, V(1)\}. \qquad (17)\end{aligned}$$

Since $V(n)$ belongs to a Markov process, (17) implies that

$$\begin{aligned}\psi(V(n), V(n-1), \cdots, V(1)) \\ = E\{|e(n)|\,|V(n)\} = \psi(V(n)). \qquad (18)\end{aligned}$$

Also, note that

$$E\{\psi(V(n))\} = E\{|e(n)|\}. \qquad (19)$$

A consequence of the first part of Assumption 5 is that for any n, $\psi(V(n))$ becomes unbounded as $V(n)$ becomes unbounded. The stationarity of the processes $X(n)$ and $e_{min}(n)$ implies that the nonstationarity of the environment is solely due to the time-varying nature of the optimal coefficient values. The increment process $C(n)$ is also assumed to be stationary, thus further restricting the behavior of the optimal coefficient process. The properties of the LMS algorithm were analyzed in [1], [15] using a similar framework to describe the nonstationarity.

A. Convergence Analysis

Let S denote a Borel measurable subset of an N-dimensional Euclidean space and let $\mathfrak{F}_n\{S\}$ denote the probability distribution function of the coefficient misalignment vector at time n being in S, i.e.,

$$\mathfrak{F}_n\{S\} = P\{V(n) \in S\} \tag{20}$$

where $P\{(\bullet)\}$ denotes the probability of the event $(\bullet)$. Also let

$$\mathcal{G}_n\{S\} = \frac{1}{n}\sum_{k=1}^{n} \mathfrak{F}_k\{S\}. \tag{21}$$

It is straightforward to show that $\mathcal{G}_n\{(\bullet)\}$ is also a probability distribution function. It is shown in Appendix A-1 that $\mathcal{G}_n\{S\}$ is a sequence of stochastically bounded distribution functions. Stochastic boundedness of $\mathcal{G}_n\{S\}$ implies that there exists a convergent subsequence of ditributions $\mathcal{G}_{n_i}\{S\}$ such that

$$\lim_{i\to\infty} \mathcal{G}_{n_i}\{S\} = \mathcal{G}\{S\} \tag{22}$$

where $\mathcal{G}\{S\}$ denotes a limiting distribution.

Now, since $V(n)$ belongs to a Markov process, there exists a linear transformation operator $\mathfrak{I}$ such that

$$\mathfrak{F}_{n+1}\{S\} = \mathfrak{I}\{\mathfrak{F}_n\{S\}\}. \tag{23}$$

Since by Assumption 5, $X(n)$, $e_{\min}(n)$, and $C(n)$ all belong to stationary processes, the transition operator is time invariant and it depends on μ, and the statistics of $X(n)$, $e_{\min}(n)$, and $C(n)$. (Notice from (7) and (8) that $V(n+1)$ depends on $V(n)$, $X(n)$, $e_{\min}(n)$, and $C(n)$, and the last three quantities are stationary random processes.) Due to linearity and time invariance of $\mathfrak{I}$, we obtain

$$\mathfrak{I}\{\mathcal{G}_{n_i}\{S\}\} = \frac{1}{n_i}\sum_{k=1}^{n_i} \mathfrak{I}\{\mathfrak{F}_k\{S\}\} \tag{24a}$$

$$= \frac{1}{n_i}\sum_{k=1}^{n_i} \mathfrak{F}_{k+1}\{S\} \tag{24b}$$

$$= \mathcal{G}_{n_i}\{S\} + \frac{1}{n_i}[\mathfrak{F}_{n_i+1}\{S\} - \mathfrak{F}_1\{S\}]. \tag{24c}$$

Taking the limit as i goes to infinity, we get

$$\lim_{i\to\infty} \mathfrak{I}\{\mathcal{G}_{n_i}\{S\}\} = \mathcal{G}\{S\}. \tag{25}$$

It is shown in Appendix A-2 that

$$\mathfrak{I}\{\mathcal{G}\{S\}\} = \mathcal{G}\{S\}. \tag{26}$$

The above implies that the coefficient misalignment vector sequences (which is Markovian) has a stationary distribution $\mathcal{G}\{S\}$. Then by Doob's theorem [5], it follows that the distributions $\mathfrak{F}_n\{S\}$ of the coefficient misalignment vector converge to $\mathcal{G}\{S\}$. Once again, the details are given in Appendix A-3.

Finally, it is shown in Appendix A-4 that the expected value of $\psi(V)$ under the limiting distribution $\mathcal{G}\{S\}$ is bounded by the right-hand side of (14).

What we have shown here under somewhat more rigid constraints than in Section III, is that the SA is very well behaved even under nonstationary conditions. In order to complete the analysis, we need a description of the filter characteristics as the adaptive process evolves. This is done in the next section where we develop a set of nonlinear difference equations that characterizes the mean and mean-squared behavior of the adaptive filter coefficient.

V. Mean and Mean-Squared Behavior of the Sign Algorithm

In this section we will derive a set of nonlinear difference equations that characterizes the mean and mean-squared behavior of the filter coefficients and the mean-squared error during tracking as well as in the limit at which the distribution of the coefficient misalignment vector possesses the limiting distribution. For this analysis, the system identification model is still under consideration along with those assumptions made earlier. The analysis, however, concerned with only the first- and second-order statistics of the filter parameters and, therefore, the results obtained are not as strong as those in the previous section. We will also need certain additional assumptions for our analysis here.

Assumption 6: $\{H_{\text{opt}}(n)\}$ is a sequence of nonstationary random vectors and is independent of $\{X(n), e_{\min}(n)\}$. (Note that independence of $C(n)$ and $X(n)$ does not automatically imply independence of $H_{\text{opt}}(n)$ and $X(n)$.)

Assumption 7: $d(n)$ and $X(n)$ are jointly Gaussian random processes.

Let

$$R_{XX} = E\{X(n)\, X^T(n)\} \tag{27}$$

denote the autocorrelation matrix of the stationary input $X(n)$, and let

$$P(n) = E\{d(n)\, X(n)\} \tag{28}$$

denote the cross-correlation vector of $d(n)$ and $X(n)$ at time n. Also, define the following autocorrelation matrices:

$$J(n) = E\{H_{\text{opt}}(n)\, H_{\text{opt}}^T(n)\} \tag{29}$$

$$K(n) = E\{V(n)\, V^T(n)\} \tag{30}$$

and

$$L = E\{C(n)\, C^T(n)\}. \tag{31}$$

Note that under Assumption 5, R_{XX} and L do not change with time, while $P(n)$ does. Note also that the expected value of the optimum filter coefficient vector is given by

$$E\{H_{\text{opt}}(n)\} = R_{XX}^{-1} P(n). \tag{32}$$

Derivation of the evolution equations for $K(n)$ that are included at the later part of this section can be easily done for a more general situation with time-varying autocorrelation matrices $R_{XX}(n)$ and $L(n)$, but we have not done so since we feel that the more restricted framework will give much more insight into the behavior of the sign algorithm operating in nonstationary environments.

Now, by evaluating the statistical expectation of (1), we have

$$E\{H(n+1)\} = E\{H(n)\} + \mu E\{X(n) \operatorname{sign}(e(n))\}. \tag{33}$$

Note that $e(n)$ is a zero-mean and Gaussian sequence when conditioned on $H(n)$. We can, therefore, simplify the expectation involving $e(n)$ using Price's theorem [20] or a lemma found in [12]. We reproduce only sketches of the derivation here (we refer the readers to [16] for details):

$$\begin{aligned} E\{X(n) \operatorname{sign}(e(n))\} &= E\{E[X(n) \operatorname{sign}(e(n)) \mid H(n)]\} \\ &= E\left\{\sqrt{\frac{2}{\pi}} \frac{1}{\sigma_{e|H}(n)} E[X(n)\, e(n) \mid H(n)]\right\} \\ &= \sqrt{\frac{2}{\pi}} E\left\{\frac{1}{\sigma_{e|H}(n)} [P(n) - R_{XX} H(n)]\right\} \\ &\approx \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_e(n)} [P(n) - R_{XX} E\{H(n)\}] \end{aligned} \tag{34}$$

where

$$\sigma^2_{e|H}(n) = E\{e^2(n) \mid H(n)\} \tag{35a}$$

and

$$\sigma^2_e(n) = E\{e^2(n)\} \tag{35b}$$

denote the conditional (on $H(n)$) and unconditional mean-squared estimation error at time n, respectively. In order to derive (34), we have made use of Assumptions 4 and 7 as well as the approximation that

$$\sigma_{e|H}(n) \approx \sigma_e(n). \tag{36}$$

This approximation is valid for small values of μ and has been employed to obtain useful analytical results in the past [16]–[18].

Substituting (34) in (33) gives

$$E\{H(n+1)\} = \left[I - \sqrt{\frac{2}{\pi}} \frac{\mu}{\sigma_e(n)} R_{XX}\right] E\{H(n)\} + \sqrt{\frac{2}{\pi}} \frac{\mu}{\sigma_e(n)} P(n) \tag{37}$$

where I denotes an identity matrix. Using (4), (5), and (32) in (37), and applying the zero-mean assumption for $C(n)$ (Assumption 2), we obtain an expression for the mean behavior of the coefficient misalignment vector as

$$E\{V(n+1)\} = \left[I - \sqrt{\frac{2}{\pi}} \frac{\mu}{\sigma_e(n)} R_{XX}\right] E\{V(n)\}. \tag{38}$$

Now, let

$$\xi_{\min} = E\{e^2_{\min}(n)\} \tag{39}$$

denote the minimum mean-squared estimation error. Note that $\xi_{\min}$ is constant over time under Assumption 5. We derive an expression for the mean-squared estimation error next. From (7), we have

$$e(n) = e_{\min}(n) - V^T(n)\, X(n). \tag{40}$$

Squaring both sides of (40) and taking the statistical expectation, we get

$$\sigma^2_e(n) = \xi_{\min} + \operatorname{tr}\{K(n) R_{XX}\} - 2E\{V^T(n)\, X(n)\, e_{\min}(n)\} \tag{41}$$

where tr $\{\bullet\}$ denotes the trace of $\{\bullet\}$. Since $X(n)$ is independent of $e_{\min}(n)$ and $V(n)$, and is assumed to be zero mean (Assumption 1), the last term of the right-hand side of (41) becomes zero. Therefore, we get

$$\sigma^2_e(n) = \xi_{\min} + \operatorname{tr}\{K(n) R_{XX}\}. \tag{42}$$

To complete our analysis, we need an expression for the second-moment behavior of the misalignment vector. Using (8), (30), and (31), we have

$$\begin{aligned} K(n+1) &= K(n) + \mu^2 R_{XX} + L \\ &\quad + \mu E\{V(n)\, X^T(n) \operatorname{sign}(e(n))\} \\ &\quad + \mu E\{X(n)\, V^T(n) \operatorname{sign}(e(n))\} \\ &\quad + E\{[V(n) + \mu X(n) \operatorname{sign}(e(n))] C^T(n)\} \\ &\quad + E\{C(n)[V^T(n) + \mu X^T(n) \operatorname{sign}(e(n))]\}. \end{aligned} \tag{43}$$

The fourth term on the right-hand side can be evaluated using some results in [16] as

$$\begin{aligned} E\{V(n)\, X^T(n) \operatorname{sign}(e(n))\} &= E\{V(n)\, E[X^T(n) \operatorname{sign}(e(n)) \mid V(n), H_{\text{opt}}(n)]\} \\ &= E\left\{V(n) \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_{e|H}(n)} \cdot E[e(n)\, X^T(n) \mid V(n), H_{\text{opt}}(n)]\right\}. \end{aligned} \tag{44}$$

Now,

$$\begin{aligned} E\{e(n)\, X^T(n) \mid V(n), H_{\text{opt}}(n)\} &= E\{d(n)\, X^T(n) - [V^T(n) + H^T_{\text{opt}}(n)] X(n)\, X^T(n) \mid V(n), H_{\text{opt}}(n)\} \\ &= E\{d(n)\, X^T(n) - [V^T(n) + E\{H^T_{\text{opt}}(n)\}] X(n)\, X^T(n) \mid V(n)\} \\ &= P^T(n) - V^T(n) R_{XX} - E\{H^T_{\text{opt}}(n)\} R_{XX} \\ &= -V^T(n) R_{XX}. \end{aligned} \tag{45}$$

Substituting (45) in (44) and using the approximation in (36) once again gives

$$E\left\{V(n)\, X^T(n) \text{ sign }(e(n))\right\} \approx -\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_e(n)} K(n)R_{XX}. \quad (46)$$

Similarly, the fifth term on the right-hand side of (43) can be written as

$$E\left\{X(n)\, V^T(n) \text{ sign }(e(n))\right\} \approx -\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_e(n)} R_{XX}K(n). \quad (47)$$

Also, under Assumption 2, we can see that the last two terms of the right-hand side of (43) become zero. Therefore, using (46) and (47) in (43), we get the following nonlinear difference equation for the autocorrelation matrix of the coefficient misalignment vector:

$$K(n+1) = K(n) + \mu^2 R_{XX} + L - \sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_e(n)}\left\{K(n)R_{XX} + R_{XX}K(n)\right\}. \quad (48)$$

Equations (38), (42), and (48) together describe the mean and mean-squared behavior of the parameters of the adaptive filter during the time it is tracking or adapting to a nonstationary environment. The validity of these results, as well as some others in the earlier sections, are demonstrated by computer simulations in the next section.

Now, we can obtain expressions for the mean-squared estimation error behavior in the limit at which the coefficient misalignment vector is governed by the stationary limiting distribution. Let $\sigma_e^2(\infty)$ and $K(\infty)$ denote the limiting responses of $\sigma_e^2(n)$ and $K(n)$, respectively. Convergence in distribution of the coefficient misalignment vector under Assumptions 1–5, as seen in Section IV, implies that $\sigma_e^2(\infty)$ and $K(\infty)$ exist and are unique.

To evaluate $\sigma_e^2(\infty)$, we need to compute $K(\infty)$ first. Taking the limit as n goes to infinity on both sides of (48), it follows that

$$\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_e(\infty)}\left[K(\infty)R_{XX} + R_{XX}K(\infty)\right] = \mu^2 R_{XX} + L. \quad (49)$$

Let Q denote an orthonormal matrix which diagonalizes R_{XX}. Premultiplying and postmultiplying both sides of (49) by Q^T and Q, respectively, we get

$$\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_e(\infty)}\left[K'(\infty)\Lambda + \Lambda K'(\infty)\right] = \mu^2\Lambda + L' \quad (50)$$

or equivalently

$$\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_e(\infty)}\left[K'(\infty) + \Lambda K'(\infty)\Lambda^{-1}\right] = \mu^2 I + L'\Lambda^{-1} \quad (51)$$

where

$$K'(\infty) = Q^T K(\infty) Q \quad (52)$$

$$L' = Q^T L Q \quad (53)$$

$$\Lambda = Q^T R_{XX} Q = \text{diag}\,[\lambda_1, \lambda_2, \cdots, \lambda_N] \quad (54)$$

and λ_i, $1 \le i \le N$, are eigenvalues of R_{XX}. The (i, j)th element of (51) becomes

$$\sqrt{\frac{2}{\pi}}\frac{\mu}{\sigma_e(\infty)}\left[1 + \frac{\lambda_i}{\lambda_j}\right]k'_{ij} = \mu^2\delta(i-j) + \frac{1}{\lambda_j} r'_c(i, j) \quad (55)$$

where k'_{ij} and $r'_c(i, j)$ denote the (i, j)th elements of $K'(\infty)$ and L', respectively, and

$$\delta(i-j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases} \quad (56)$$

Therefore, it follows that

$$k'_{ij} = \sqrt{\frac{\pi}{2}}\frac{\sigma_e(\infty)}{\lambda_i + \lambda_j}\left[\mu\lambda_j\delta(i-j) + \frac{r'_c(i, j)}{\mu}\right]. \quad (57)$$

The diagonal elements of the primed second-moment matrix are then given by

$$k'_{ii} = \sqrt{\frac{\pi}{2}}\frac{\sigma_e(\infty)}{2}\left[\mu + \frac{1}{\mu\lambda_i} r'_c(i, i)\right]. \quad (58)$$

Now, taking the limit as n goes to infinity in (42), we get

$$\sigma_e^2(\infty) = \xi_{\min} + \text{tr}\left\{K(\infty)R_{XX}\right\}. \quad (59)$$

The second term on the right-hand side of (59) can be rewritten using (58) as

$$\begin{aligned}\text{tr}\left\{K(\infty)R_{XX}\right\} &= \text{tr}\left\{K'(\infty)\Lambda\right\} \\ &= \sum_{i=1}^{N}\lambda_i k'_{ii} \\ &= \sqrt{\frac{\pi}{2}}\frac{\sigma_e(\infty)}{2}\sum_{i=1}^{N}\left[\mu\lambda_i + \frac{r'_c(i, i)}{\mu}\right] \\ &= \sqrt{\frac{\pi}{2}}\frac{\sigma_e(\infty)}{2}\left[\mu\,\text{tr}\left\{R_{XX}\right\} + \frac{\text{tr}\left\{L\right\}}{\mu}\right].\end{aligned} \quad (60)$$

Substituting this in (59), we get

$$\sigma_e^2(\infty) = \xi_{\min} + \alpha\sigma_e(\infty) \quad (61)$$

where

$$\alpha = \frac{1}{2}\sqrt{\frac{\pi}{2}}\left[\mu \operatorname{tr}\{R_{XX}\} + \frac{\operatorname{tr}\{L\}}{\mu}\right]. \tag{62}$$

Solving (61) for $\sigma_e(\infty)$ and retaining the positive root, the limiting value of the root mean-squared estimation error is given by

$$\sigma_e(\infty) = \tfrac{1}{2}[\alpha + \sqrt{\alpha^2 + 4\xi_{\min}}]. \tag{63}$$

Substitution of (63) in (61) gives an expression for the mean-squared estimation error of the sign algorithm in the limit at which the coefficient misalignment vector possesses the limiting distribution

$$\sigma_e^2(\infty) = \xi_{\min} + \frac{\alpha}{2}[\alpha + \sqrt{\alpha^2 + 4\xi_{\min}}]. \tag{64}$$

It is straightforward to show that μ_{opt} as given by (16) also minimizes (64).

Under the same set of assumptions, it is not difficult to show that the limiting mean-squared estimation error of the LMS algorithm is given by

$$\sigma_e^2(\infty) = \frac{\xi_{\min} + \dfrac{1}{2\mu}\left(\displaystyle\sum_{i=1}^{N} \frac{r_c'(i,i)}{1-\mu\lambda_i}\right)}{1 - \dfrac{\mu}{2}\left(\displaystyle\sum_{i=1}^{N}\frac{\lambda_i}{1-\mu\lambda_i}\right)}. \tag{65}$$

For very small values of μ, (65) can be approximated as

$$\begin{aligned}\sigma_e^2(\infty) &\approx \left[\xi_{\min} + \frac{1}{2\mu}\left(\sum_{i=1}^{N}\frac{r_c'(i,i)}{1-\mu\lambda_i}\right)\right] \cdot \left[1 + \frac{\mu}{2}\sum_{i=1}^{N}\frac{\lambda_i}{1-\mu\lambda_i}\right] \\ &\approx \left[\xi_{\min} + \frac{\operatorname{tr}\{L\}}{2\mu}\right]\left[1 + \frac{\mu}{2}\operatorname{tr}\{R_{XX}\}\right]. \end{aligned} \tag{66}$$

Similar, but more simplified result than (66) can be found in [25]. Now, let $\mu_{\text{opt, SA}}$ and $\mu_{\text{opt, LMS}}$ denote the optimal values of μ that minimize the limiting mean-squared estimation errors given in (64) for the sign algorithm and in (66) for the LMS algorithm, respectively. Since μ_{opt} given in (16) also minimizes $\sigma_e^2(\infty)$ for the sign algorithm, we have

$$\mu_{\text{opt, SA}} = \sqrt{\frac{E\|C(n)\|^2}{E\|X(n)\|^2}} = \sqrt{\frac{\operatorname{tr}\{L\}}{\operatorname{tr}\{R_{XX}\}}}. \tag{67}$$

Also, it can be easily shown from (66) that $\sigma_e^2(\infty)$ for the LMS algorithm has its minimum value when the convergence parameter is chosen to be

$$\mu_{\text{opt, LMS}} = \sqrt{\frac{\operatorname{tr}\{L\}}{\xi_{\min}\operatorname{tr}\{R_{XX}\}}}. \tag{68}$$

Note that $\mu_{\text{opt, LMS}}$ depends on $\xi_{\min}$. Note also that the same result can be found in [25].

We will now compare the performances of the two algorithms by evaluating the minimum excess mean-squared estimation errors, denoted as $\xi_{\text{ex, SA}}$ for the sign algorithm and $\xi_{\text{ex, LMS}}$ for the LMS algorithm. Since both the analyses are valid for small μ's, it is important to note that this comparison is valid only when the optimal values of the convergence constants have small magnitudes, or equivalently when the increment process $C(n)$ has very small mean-squared values.

Substitution of (67) in (62) and (64) gives

$$\xi_{\text{ex, SA}} = \frac{\pi}{4}\operatorname{tr}\{R_{XX}\}\operatorname{tr}\{L\} + \sqrt{\frac{\pi}{2}\xi_{\min}\operatorname{tr}\{R_{XX}\}\operatorname{tr}\{L\} + \frac{\pi^2}{16}\operatorname{tr}^2\{R_{XX}\}\operatorname{tr}^2\{L\}}. \tag{69}$$

Similarly, using (68) in (66) will give

$$\xi_{\text{ex, LMS}} = \tfrac{1}{4}\operatorname{tr}\{R_{XX}\}\operatorname{tr}\{L\} + \sqrt{\xi_{\min}\operatorname{tr}\{R_{XX}\}\operatorname{tr}\{L\}}. \tag{70}$$

It is clear that the minimum excess mean-squared estimation error generated by the sign algorithm is always larger than that generated by the LMS algorithm. However, in many applications, this additional error produced by the sign algorithm may be relatively small and the reduction in computational complexity over the LMS algorithm may justify the extra error. For the LMS algorithm, it is also possible that $\mu_{\text{opt, LMS}}$ is outside the permissible range of convergence parameters for which mean-squared convergence is assured, for the case when $\operatorname{tr}\{L\}/\xi_{\min}$ is very large. Since the sign algorithm does not diverge for any choice of the convergence parameter, it is guaranteed that the optimal performance can be achieved.

VI. Experimental Results

In this section, we present the results of several experiments in order to demonstrate the validity of our analyses. The example problem considered here is that of identifying a nonstationary linear system as illustrated in Fig. 1, where $X(n)$, $C(n)$, and $e_{\min}(n)$ are assumed to be stationary processes. The time-varying impulse response $H_{\text{opt}}(n)$ of the system is constructed using (4) such that

$$H_{\text{opt}}(n+1) = H_{\text{opt}}(n) - C(n) \tag{71}$$

where $C(n)$ is a zero-mean, stationary, and white Gaussian pseudorandom sequence, and the initial value $H_{\text{opt}}(1)$ is selected to be

$$H_{\text{opt}}^T(1) = [0.1, 0.3, 0.5, 0.7, 0.9, 0.7, 0.5, 0.3, 0.1]. \tag{72}$$

Two sets of the reference input signals, $x_1(n)$ and $x_2(n)$, were used in our experiments and they were generated as the outputs of the autoregressive filters so that

$$x_1(n) = \zeta_1(n) + 1.5x_1(n-1) - x_1(n-2) + 0.25x_1(n-3) \tag{73}$$

and

$$x_2(n) = \zeta_2(n) + 1.79x_2(n-1) - 1.9425x_2(n-2) + 1.27x_2(n-3) - 0.5x_2(n-4) \tag{74}$$

where the input processes to the autoregressive filter $\zeta_1(n)$ and $\zeta_2(n)$ were zero-mean, stationary, and white Gaussian random sequences with variances such that the variances of $x_1(n)$ and $x_2(n)$ were 1. Note that the ratios of the maximum and minimum eigenvalues of the autocorrelation matrices of $x_1(n)$ and $x_2(n)$ are more than 200 and 670, respectively. The corresponding primary input signals $d_1(n)$ and $d_2(n)$ were produced by processing $x_1(n)$ and $x_2(n)$ individually through the time-varying system with the coefficient vector $H_{\text{opt}}(n)$, and then corrupting the system output using a zero-mean, stationary, and white Gaussian random sequence $e_{\min}(n)$ with variance 0.01.

The first set of experiments deals with the bound on the long-term time average of the mean-absolute estimation error sequence. For this experiment, the variance of the disturbance process $C(n)$ was chosen to be 10^{-4}. Note that according to our analysis, the optimum value of the convergence parameter μ is 0.01 in this case (see (16)). The time-averages of the absolute estimation error over 100 000 samples were computed and averaged over 50 independent runs for several values of the convergence parameter around $\mu_{\text{opt}} = 0.01$. The results are tabulated in Table I. Also tabulated are the upper bound for the long-term time average of the mean-absolute estimation error predicted by (14). We can see that the long-term average of the mean-absolute value of the estimation error does achieve its minimum value at $\mu = 0.01$, which is the same value as μ_{opt} that minimizes the upper bound.

We next compare the evolution equations derived in Section V with simulation results. In order to make the presentation compact, we compare the trace of the second-order moment matrix of the misalignment vector (tr $\{K(n)\}$) as predicted by our analysis and obtained from simulations. The results presented were averaged over 50 independent runs with 15 000 samples for the two sets of the reference input signals described earlier. For each set of signals, we used two different values (10^{-4} and 10^{-6}) for the variance of the disturbance process $C(n)$ and the convergence parameter μ was selected to be the optimum value given by (15). The performance measures that are presented in Figs. 2 and 3 are in decibels and have been normalized so that the initial value is 0 dB. We can see that the empirical and theoretical curves in each case show very good agreement, despite the fairly large eigenvalue spreads for the input autocorrelation matrices and the obvious violation of the independence assumption.

TABLE I
THE TIME AVERAGES OF THE MEAN-ABSOLUTE ESTIMATION ERROR: (a) WHEN THE EIGENVALUE SPREAD IS MORE THAN 200, (b) WHEN THE EIGENVALUE SPREAD IS MORE THAN 670, AND (c) UPPER BOUNDS OF THOSE PREDICTED IN (14) FOR BOTH CASES. NOTE THAT $\mu_{\text{opt}} = 0.01$

μ	(a)	(b)	(c)
0.005	0.11450	0.11640	0.12250
0.006	0.10210	0.10470	0.11200
0.007	0.09538	0.09770	0.10580
0.008	0.09187	0.09343	0.10225
0.001	0.09064	0.09069	0.10050
0.01	0.09021	0.08944	0.10000
0.011	0.09057	0.08964	0.10041
0.012	0.09213	0.09073	0.10150
0.013	0.09405	0.09265	0.10312
0.014	0.09584	0.09462	0.10514
0.015	0.09833	0.09691	0.10750

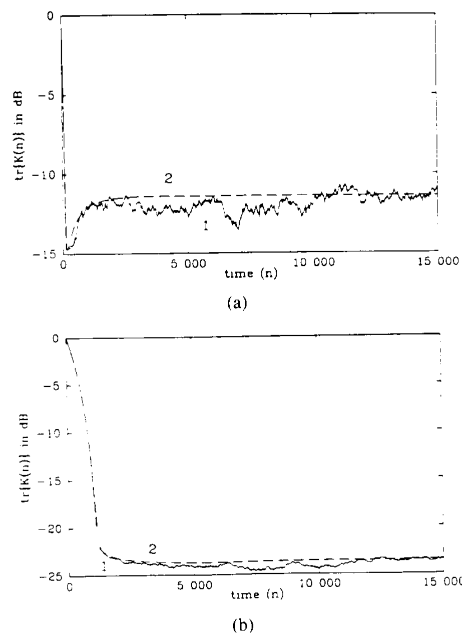

Fig. 2. Normalized tr $\{K(n)\}$ when the reference input signal is $x_1(n)$ given by (73). (Eigenvalue spread is more than 200.) (a) variance of the disturbance process = 10^{-4} and (b) variance of the disturbance process = 10^{-6}: (1) empirical curve and (2) theoretical curve.

VII. CONCLUDING REMARKS

This paper presented a fairly comprehensive study of the tracking properties of the adaptive sign algorithm when used in nonstationary environments. By modeling the nonstationarity as that due to a random disturbance, it was shown that the long-term time average of the mean-absolute estimation error is bounded for any positive value of μ and that there exists an optimum value of μ which minimizes this quantity. Under a set of commonly used assumptions, we then showed that the probability distribution functions of the successive coefficient misalign-

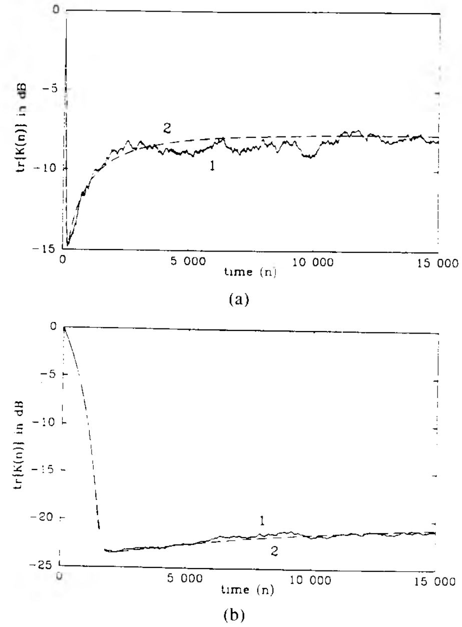

Fig. 3. Normalized tr $\{K(n)\}$ when the reference input signal is $x_2(n)$ given by (74). (Eigenvalue spread is more than 670.) (a) variance of the disturbance process $= 10^{-4}$ and (b) variance of the disturbance process $= 10^{-6}$: (1) empirical curve and (2) theoretical curve.

ment vectors converge to a limiting stationary distribution when the adaptive filter is used in the "system identification" mode. Under the assumption that the signals involved are zero mean and Gaussian, we also derived a set of nonlinear evolution equations that describes the mean and mean-squared behavior of the filter during the adapting and tracking operations. Experimental results were presented to demonstrate the validity of our analyses, and the empirical results matched those predicted by our analyses very well. It is hoped that the results of this work will provide additional design tools for adaptive filters equipped with the sign algorithm and operating in nonstationary environments.

While we do believe that our results are the most comprehensive to date in the area of tracking analysis of the sign algorithm, it must be pointed out that several aspects of the problem still need investigation. The results in Sections IV and V were obtained for the case of the "system identification" applications only. Performance analyses for several other applications (for example, inverse modeling) still need to be done. Relaxation of the independence assumption employed in Sections IV and V, as well as consideration of other models of nonstationarity are also important topics that need further attention.

Appendix A

This Appendix provides details of the derivations in Section IV. As mentioned earlier, the analysis applies the methods in [10] to the coefficient misalignment process rather than the coefficient process as done for the stationary case.

1. Stochastic Boundedness of $\mathcal{G}_n\{S\}$

A sequence of random vectors and the corresponding sequence of distributions are said to be stochastically bounded if for any probability ϵ there exists a distance R such that each random vector of the sequence has probability less than ϵ of having length greater than R.

Applying (19) to (14) and using the stationarity of $X(n)$, $C(n)$, and $e_{\min}(n)$ (Assumption 5), we have

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} E\left\{\psi(V(k))\right\} \le \frac{\mu}{2} E\left\|X(n)\right\|^2 + \frac{1}{2\mu} E\left\|C(n)\right\|^2 + E\left|e_{\min}(n)\right|. \tag{A.1}$$

The above inequality implies that there exists a constant t such that for any finite and nonzero value of μ

$$\frac{1}{n}\sum_{k=1}^{n} E\left\{\psi(V(k))\right\} \le t, \qquad \forall n. \tag{A.2}$$

Since $\psi(V(n))$ is nonnegative for any n, we can use the Chebyshev inequality to get

$$P\left\{\psi(V(n)) \ge M\right\} \le \frac{1}{M} E\left\{\psi(V(n))\right\} \tag{A.3}$$

for any positive value of M. Combining (A.2) and (A.3) gives

$$\frac{1}{n}\sum_{k=1}^{n} P\left\{\psi(V(k)) \ge M\right\} \le \frac{t}{M}. \tag{A.4}$$

Since $\psi(V(n))$ becomes unbounded as $V(n)$ becomes unbounded, it follows that there exists some positive value R (that depends on M) such that

$$P\left\{\left\|V(n)\right\| > R\right\} \le P\left\{\psi(V(n)) \ge M\right\}. \tag{A.5}$$

Combining (A.4) and (A.5), we have the following result:

$$\frac{1}{n}\sum_{k=1}^{n} P\left\{\left\|V(k)\right\| > R\right\} \le \frac{t}{M}. \tag{A.6}$$

The left-hand side of the inequality is nothing but $\mathcal{G}_n\{\|V(k)\| > R\}$, implying that

$$\mathcal{G}_n\left\{\left\|V(k)\right\| > R\right\} \le \frac{t}{M}, \qquad \forall k,\ 1 \le k \le n \tag{A.7}$$

which, in turn, means that the sequence of distributions $\mathcal{G}_n$ is stochastically bounded.

2. Existence of a Stationary Distribution for the Markov Process $V(n)$

In this part, we show that $\mathfrak{I}\{\mathcal{G}\{S\}\} = \mathcal{G}\{S\}$ and therefore that $\mathcal{G}$ is a stationary distribution for the Markov process $V(n)$.

Let $\pi(V, S)$ denote the probability transition function associated with the Markov process, i.e.,

$$\pi(V, S) = P\left\{V(n+1) \in S \mid V(n) = V\right\}. \tag{A.8}$$

One consequence of Assumption 5 (continuous distributions) is that $\pi(V, S)$ is a bounded continuous function in V for each S.

For any distribution $\mathcal{B}$, $\mathfrak{I}\{\mathcal{B}\{S\}\}$ can be rewritten as

$$\mathfrak{I}\{\mathcal{B}\{S\}\} = E_{\mathcal{B}}\{\pi(W, S)\} \tag{A.9}$$

where $E_{\mathcal{B}}\{\bullet\}$ denotes the statistical expectation under the distribution function $\mathcal{B}$ and W is a random vector distributed according to $\mathcal{B}$. Now,

$$\mathfrak{I}\{\mathcal{G}_{n_i}\{S\}\} = E_{\mathcal{G}_{n_i}}\{\pi(V_{n_i}, S)\} \tag{A.10}$$

where V_{n_i} has distribution $\mathcal{G}_{n_i}$. Since $\pi(V, S)$ is bounded and continuous in V, we can take the limiting value of both sides of (A.10) as i goes to infinity and find that

$$\begin{aligned}\lim_{i\to\infty} \mathfrak{I}\{\mathcal{G}_{n_i}\{S\}\} &= \lim_{i\to\infty} E_{\mathcal{G}_{n_i}}\{\pi(V_{n_i}, S)\}\\ &= E_{\mathcal{G}}\{\pi(V, S)\}\\ &= \mathfrak{I}\{\mathcal{G}\{S\}\}\end{aligned} \tag{A.11}$$

where V is a random process with distribution $\mathcal{G}$. Therefore, combining (25) and (A.11), we get

$$\mathfrak{I}\{\mathcal{G}\{S\}\} = \mathcal{G}\{S\} \tag{A.12}$$

implying that $\mathcal{G}$ is a stationary distribution of the Markov process $V(n)$.

3. *The Limiting Distribution for $\mathfrak{F}_n$ is $\mathcal{G}$*

To prove this part, we use a version of Doob's theorem [5] as given in Gersho [10]. Doob's theorem, as stated in [10], is reproduced here.

If

(a) the Markov process possesses a stationary distribution $\mathcal{G}$,

(b) $P\{V(n) \in S\}$ approaches zero as n goes to infinity for any initial misalignment $V(1)$ and for any region S to which $\mathcal{G}$ assigns zero probability,

(c) there is no region S to which $\mathcal{G}$ assigns probability less than one with the property that if $V(n) \in S$, then $V(n+1) \in S$ with probability 1, and

(d) there is no region S with the property that $P\{V(n+1) \in S \mid V(n) \in S\} = 0$, but $P\{V(n+m) \in S \mid V(n) \in S\} = 1$ for some integer $m > 1$;

then the successive distributions $\mathfrak{F}_n$ converge to $\mathcal{G}$ for any initial misalignment vector $V(1)$.

We have already shown that hypothesis (a) holds. Because of the continuity assumption of probability distributions, it follows that

$$\pi(V, S) > 0 \tag{A.13}$$

for any region S of nonzero volume, and

$$\pi(V, S) = 0 \tag{A.14}$$

for any region S of zero volume. This implies that $\mathcal{G}$ assigns zero probability to only regions with zero volume. This in turn implies that hypothesis (b) holds. To show that hypothesis (c) holds, suppose that $\mathcal{G}$ assigns probability less than 1 to the region S. This implies that the probability assigned to the complimentary region $\bar{S}$ is nonzero and therefore $\bar{S}$ has nonzero volume. But if $P\{V(n+1) \in S\} = 1$ and $P\{V(n+1) \in \bar{S}\} = 0$, then $\bar{S}$ has zero volume and this is a contraction. We can show that hypothesis (d) also holds in a similar manner. Therefore, by Doob's theorem

$$\lim_{n\to\infty} \mathfrak{F}_n\{S\} = \mathcal{G}\{S\}. \tag{A.15}$$

4. *$E_{\mathcal{G}}|e(n)|$ is Bounded by the RHS of (14)*

Even though we have shown that a limiting distribution does exist for the misalignment vectors, it is possible that the expected cost function under this distribution is unbounded. We now show that this is not possible in our situation and that the expected cost function under the limiting distribution is indeed bounded by the same bound for the long-term time average of the mean-absolute estimation error.

From (14) using Assumptions 1–3, we know

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} E_{\mathfrak{F}_k}\{\psi(V(k))\} \le t \tag{A.16}$$

for some positive t. This implies that there exists a subsequence $V(n_i)$ of misalignment vectors such that

$$E_{\mathfrak{F}_{n_i}}\{\psi(V(n_i))\} \le t. \tag{A.17}$$

Taking the limit as i goes to infinity

$$\lim_{i\to\infty} E_{\mathfrak{F}_{n_i}}\{\psi(V(n_i)) = E_{\mathcal{G}}\{\psi(V)\} \le t \tag{A.18}$$

where V is a random process with distribution function $\mathcal{G}$. Since

$$E_{\mathcal{G}}\{\psi(V)\} = E_{\mathcal{G}}|e(n)| \tag{A.19}$$

we have from (A.18) that the mean-absolute estimation error under the limiting distribution is bounded also. Moreover, by choosing the convergence parameter to be μ_{opt} as given in (15), this bound can be minimized.

Acknowledgment

The authors thank Prof. N. Bershad for his careful review of this paper. They also thank Dr. J. Tugnait and Dr. R. Short for several valuable comments.

References

[1] A. Benveniste and G. Ruget, "A measure of the tracking capability of recursive stochastic algorithms with constant gains," *IEEE Trans. Automat. Contr.*, vol. AC-27, no. 3, pp. 639–649, June 1982.

[2] N. J. Bershad, "On error-saturation nonlinearities in LMS adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, no. 4, pp. 440–452, Apr. 1988.

[3] N. J. Bershad and L. Z. Qu, "LMS adaptation with correlated data: A scalar example," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, no. 4, pp. 695–700, Aug. 1984.

[4] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "Comparison of the convergence of two algorithms for adaptive FIR digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 3, pp. 670–678, June 1981.

[5] J. L. Doob, "Asymptotic properties of Markoff transition probabilities," *Trans. Amer. Math. Soc.*, vol. 63, pp. 393–421, May 1948.

[6] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, no. 4, pp. 578-586, Aug. 1982.

[7] 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.

[8] E. Eweda, "Almost sure convergence of a decreasing gain sign algorithm for adaptive filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, no. 10, pp. 1669-1671, Oct. 1988.

[9] E. Eweda, "Optimum step size of sign algorithm for nonstationary adaptive filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 11, pp. 1897-1901, Nov. 1990.

[10] A. Gersho, "Adaptive filtering with binary reinforcement," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 2, pp. 191-199, Mar. 1984.

[11] Y. Iiguni, H. Sakai, and H. Tokumaru, "Convergence properties of simplified gradient adaptive lattice algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 6, pp. 1427-1434, Dec. 1985.

[12] T. Koh and E. J. Powers, "Efficient methods to estimate correlation functions of Gaussian processes and their performance analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 4, pp. 1032-1035, Aug. 1985.

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

[14] C. P. Kwong, "Control-theoretic design of the LMS and the sign algorithms in nonstationary environments," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 2, pp. 253-259, Feb. 1990.

[15] O. Macchi, "Optimization of adaptive identification for time-varying filters," *IEEE Trans. Automat. Contr.*, vol. AC-31, no. 3, pp. 283-287, Mar. 1986.

[16] V. J. Mathews and S. H. 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.

[17] V. J. Mathews, "An efficient FIR adaptive filter using DPCM and the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, no. 1, pp. 128-133, Jan. 1989.

[18] V. J. Mathews, "Performance analysis of adaptive filters equipped with the dual sign algorithm," *IEEE Trans. Signal Processing*, to be published, Jan. 1991.

[19] T. H. Park, D. H. Youn, and I. W. Cha, "Multiplication-free adaptive digital filters," *Proc. IEEE*, vol. 76, no. 5, pp. 632-634, May 1988.

[20] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 69-72, June 1958.

[21] N. A. M. Verhoeckx, H. C. Eizen, F. A. M. Snijders, and P. J. Gerwen, "Digital echo cancellation for baseband data transmission," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, no. 6, pp. 768-781, Dec. 1979.

[22] N. A. M. Verhoeckx and T. A. C. M. Claasen, "Some considerations on the design of adaptive digital filters equipped with the sign algorithm," *IEEE Trans. Commun.*, vol. COM-32, no. 3, pp. 258-266, Mar. 1985.

[23] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[24] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, no. 8, pp. 1151-1162, Aug. 1976.

[25] B. Widrow and E. Walach, "On the statistical efficiency of the LMS algorithm with nonstationary inputs," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 2, pp. 211-221, Mar. 1984.

[26] D. H. Youn, V. J. Mathews, and S. H. Cho, "An efficient lattice predictor algorithm for instantaneous frequency estimation," *Signal Processing*, vol. 10, no. 1, pp. 75-81, Jan. 1986.

Sung Ho Cho (S'86–M'88) was born in Seoul, Korea, on February 21, 1960. He received the B.E. degree in electronic engineering from Hanyang University, Seoul, Korea, in 1982, the M.S. degree in electrical and computer engineering from the University of Iowa, Iowa City, in 1984, and the Ph.D. degree in electrical engineering from the University of Utah, Salt Lake City, in 1989.

He is currently a Senior Member of the Technical Staff of the Electronics and Telecommunications Research Institute (ETRI), Dae Duk Science Town, Daejeon, Korea. His current research interests include adaptive filtering algorithms and structures, spectrum estimation, stochastic processes, and digital communications.

Dr. Cho is a member of Eta Kappa Nu and Tau Beta Pi.

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, in 1980, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Iowa, Iowa City, in 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 Assistant 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 is an Associate Editor of the IEEE Transactions on Acoustics, Speech, and Signal Processing.