論 文

Statistical Convergence Properties of an Adaptive Normalized LMS Algorithm with Gaussian Signals

Sung Ho CHO*, Iickho SONG**, Kwang Ho PARK* *Regular Members*

가우시안 신호를 갖는 적응 정규화 LMS 앨고리듬의 통계학적 수렴 성질

正會員 趙 誠 鎬* 正會員 宋 翊 鎬** 正會員 朴 光 浩*

ABSTRACT This paper presents a statistical convergence analysis of the normalized least mean square (NLMS) algorithm that employs a single-pole lowpass filter. In this algorithm, the lowpass filter is used to adjust its output towards the estimated value of the input signal power recursively. The estimated input signal power so obtained at each time is then used to normalize the convergence parameter. Under the assumption that the primary and reference inputs to the adaptive filter are zero-mean, wide-sense stationary, and Gaussian random processes, and further making use of the independence assumption, we derive expressions that characterize the mean and mean-squared behavior of the filter coefficients as well as the mean-squared estimation error. Conditions for the mean and mean-squared convergence are explored. Comparisons are also made between the performance of the NLMS algorithm and that of the popular least mean square (LMS) algorithm . Finally, experimental results that show very good agreement between the analytical and empirical results are presented.

要 約 이 논문에서는 극이 하나 있는 저역 여파기를 쓰는 정규화된 LMS 앨고리듬의 통계학적 수렴을 분석하였다. 이 앨고리듬에서 저역 여파기는 출력이 입력 신호전력의 추정 값에 가깝도록 순환적으로 조정하는 데에 쓰인다. 또한 이때 얻은 입력신호 전력 추정값은 수렴 매개변수를 정규화하는데에 쓰인다. 적응여파기 입력값들이 독립이고 평균이 0이며 넓은 뜻에서 정상인 가우시안 확률과정이라는 가정아래에서 여파기 계수의 평균, 제곱 평균과 제곱 평균 추정오차의 성격을 나타내는 식을 얻었다. 평균과 제곱 평균수렴에 필요한 조건을 살펴보았으며 정규화된 LMS 앨고리듬의 성능과 LMS 앨고리듬의 성능을 견주어 보았다. 해석적 방법과 경험적 방법이 매우 잘 들어 맞는다는 것을 보여주는 실험결과도 보였다.

1. Introduction

The adaptive LMS algorithm[1][2] has received a great deal of attention during the last twc decades and now becomes very popular in variety of applications due to its simplicity. It is, however, well known that the statistical behavior and the convergence properties of the LMS algorithm depend strongly on the convergence parameter and the eigenvalue spread of the input autocorrelation matrix[3],[4]. In particular, we need to know the eigen-values of the matrix to choose an appropriate convergence parameter so that the mean convergence of the algorithm is ensured. Since these eigenvalues are gencerally unknown in practice and very difficult to estimate, we

*韓國電子通信研究所
ETRI
**韓國科學技術院 電氣및 電子工學科
Department of Electrical Engineering KAIST
論文番號：91－121(接受1991. 7. 2)

often compute the trace of the correlation matix instead(based on Gershgorin's theorem[5]), and use it to form a more practical, but more restrictive upper bound for the convergence parameter. Sometimes, this new bound is unnecessarily too tight, and thus makes us sacrifice the speed of adaptation. Furthermore, if the statistics of an environment changes with time, we may have to select the convergence parameter that is time-varying so as to permit the tracking capability.

One of the ways to overcome these problems is to use the NLMS algorithm. In general, there exist two kinds of NLMS algorithms : one of which is to normalize the convergence parameter by the sum of the squared tap inputs (i.e., L_2-norm of the tap input vector)[6][8], and the other by the estimated input signal power using a single-pole lowpass filter[9][10]. Even though convergnece properties of the former NLMS algorithm have been extensively studied, those for the latter have not yet been deeply investigated to the best of our knowledge.

The NLMS algorithm with a single-pole lowpass filter is particularly more useful than that by the L_2-norm of the tap input vector in applications where a fixed point arithmetic digital signal processor (DSP) chip, such as Motorola's DSP56001 or Texas Instruments' TMS320C25, is to be used whth large number of tap weights, since we do not have to worry about any arithmetic overflow for computing the normalization factor. In this paper, we investigate some important statistical properties of the NLMS algorithm when the single-pole lowpass filter is employed.

Consider the problem of adaptively estimating the primary input signal $d(n)$ using the reference input $x(n)$, Let $H(n)$ denote the adaptive filter coefficient vector of size N, and $e(n)$ denote the estimation error signal. Define the tap input vector $X(n)$ as

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

where $[\cdot]^T$ stands for the transpose of $[\cdot]$. The NLMS algorithm of interest updates the coefficient vector $H(n)$ using

$$H(n+1) = H(n) + \frac{\mu}{\hat{\sigma}_x^2(n)} X(n)\, e(n), \tag{2}$$

and

$$\hat{\sigma}_x^2(n) = \beta\, \hat{\sigma}_x^2(n-1) + (1-\beta)\, x^2(n), \tag{3}$$

where μ denotes the convergence parameter that controls the speed and stability of adaptation, β denotes the smoothing factor of the lowpass filter that is usually a positive number less than but close to one, and

$$e(n) = d(n) - H^T(n) X(n). \tag{4}$$

Note that $\hat{\sigma}_x^2(n)$, the estimated input signal power at time n, is obtained by processing $x^2(n)$ through the single-pole lowpass filter, where the location of the pole is at β on the real-axis inside the unit circle in the Z-plane. The lowpass filter adjusts its output towards the estimated value of the input signal power recursively. We often, but not necessarily always, choose the value for β as

$$\beta = 1 - \mu \tag{5}$$

in the sense that having the same time constant for the evolutions of both the adaptive process and the lowpass filter may yield better results[10].

In the next section, under the assumption that the signals involved are zero-mean, wide-sense stationary, and Gaussian, and further employing the independence assumption[11], we derive a set of nonlinear recursive difference equations that characterizes the mean and mean-squared behavior of the filter coefficients and the mean-squared estimation error. Conditions for the mean and mean-squared convergence are explored. Comparisons are also made between the performance of the NLMS algorithm and that of the popular LMS algorithm in this section. In Section 3, included are the experimental results that demonstrate the validity of the theoretical derivations. Finally, the concluding remarks are made in Section 4.

2. Convergence Analysis

In this section, 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 of the NLMS algorithm.

The following notations will be used in the analysis. Let H_{opt} denote the optimum coefficient vector given by

$$H_{opt} = R_{XX}^{-1} R_{dX}, \tag{6}$$

where

$$R_{XX} = E\left\{X(n)X^T(n)\right\} \tag{7}$$

and

$$R_{dX} = E\{d(n)X(n)\} \tag{8}$$

denote the autocorrelation matrix of $X(n)$ and the crosscorrelation vector of $d(n)$ and $X(n)$, respectively, and $E\{\cdot\}$ indicates the statistical expectation of $\{\cdot\}$.

Also, define the coefficient misalignment vector $V(n)$ as

$$V(n) = H(n) - H_{opt}, \tag{9}$$

and its autocorrelation matrix $K(n)$ as

$$K(n) = E\left\{V(n)V^T(n)\right\}. \tag{10}$$

Using (9) in (2), we get the update equation for the coefficient misalignment vector as

$$V(n+1) = V(n) + \frac{\mu}{\hat{\sigma}_x^2(n)} e(n)X(n). \tag{11}$$

The optimal estimation error $e_{min}(n)$ is given by

$$e_{min}(n) = d(n) - H_{opt}^T X(n). \tag{12}$$

Combining (4), (9), and (12) leads to

$$e(n) = e_{min}(n) - V^T(n)X(n). \tag{13}$$

Note that by the orthogonality principle,

$$E\{e_{min}(n)X(n)\} = 0. \tag{14}$$

Finally, let

$$\xi_{min} = E\left\{e_{min}^2(n)\right\} \tag{15}$$

denote the minimum mean-squared estimation error, and let

$$\rho(n) = E\left\{\hat{\sigma}_x^2(n)\right\} \tag{16}$$

and

$$\phi(n) = E\left\{\hat{\sigma}_x^4(n)\right\} \tag{17}$$

denote the first and the second moment of $\hat{\sigma}_x^2(n)$, respectively.

We now employ the following assumptions to make the analysis mathematically more tractable :

Assumption 1 : $d(n)$ and $X(n)$ are zero-mean, wide-sense stationary, and jointly Gaussian random processes.

Assumption 2 : The input pair $\{d(n), X(n)\}$ at time n is independent of $\{d(k), X(k)\}$ at time k if $n \neq k$.

A consequence of Assumption 1 is that the estimation error $e(n)$ given in (4) is also a zero mean and Gaussian when conditioned on the coefficient vector $H(n)$. Assumption 2 is the commonly employed "independence assumption" and hardly true in practice. It is, however, shown [11] that Assumption 2 is valid if μ is chosen to be sufficiently small. Also, the analysis using this assumption has produced results that accurately predict the behavior of the adaptive filters even in circumtances where the assumption is grossly violated[12]. One of the consequence of Assumption 2 is that $H(n)$ is independent of the input pair $\{d(n), X(n)\}$, since $H(n)$ depends only on inputs at time n-1 and before. It is very important to note in Assumption 2 that we do not restrict the nature of the matrix R_{XX}.

Now, taking the statistical expectation on both sides of (2) gives.

$$E\{H(n+1)\} = E\{H(n)\} + \mu E\left\{\frac{1}{\hat{\sigma}_x^2(n)} e(n) X(n)\right\}. \tag{18}$$

In (18), $\hat{\sigma}_x^2(n)$ may be considered to be uncorrelated with $X(n)$ and $e(n)$ if β is chosen to be very close to one. We thus can rewrite (18) as

$$E\{H(n+1)\} \approx= E\{H(n)\} + \mu E\left\{\frac{1}{\hat{\sigma}_x^2(n)}\right\} E\{e(n) X(n)\}. \tag{19}$$

To evaluate the second expectation on the right-hand side (RHS) of (19), we first expand $1/\hat{\sigma}_x^2(n)$ in a Tayler series about $E\{\hat{\sigma}_x^2(n)\}$ to the first two terms and a remainder term. We then take the expectation, after discarding the remainder term, to get

$$E\left\{\frac{1}{\hat{\sigma}_x^2(n)}\right\} \approx \frac{1}{E\{\hat{\sigma}_x^2(n)\}} = \frac{1}{\rho(n)}, \tag{20}$$

where by substituting (3) in (16) and denoting $\sigma_x^2 = E\{x^2(n)\}$,

$$\begin{aligned}\rho(n) &= \beta\rho(n-1) + (1-\beta)\sigma_x^2 \\ &= \beta^n \rho(0) + (1-\beta^n)\sigma_x^2.\end{aligned} \tag{21}$$

Note that $\rho(n)$ converges to σ_x^2 in the limit as we wish.

The last expectation of (19) can be evaluated using(4) as well as Assumption 2, i.e.,

$$E\{e(n) X(n)\} = R_{dX} - R_{XX} E\{H(n)\}. \tag{22}$$

Therefore, substituting (20) and (22) in (19), we obtain the mean behavior of the NLMS algorithm as

$$E\{H(n+1)\} \approx \left[I_N - \frac{\mu}{\rho(n)} R_{XX}\right]$$
$$E\{H(n)\} + \frac{\mu}{\rho(n)} R_{dX}, \qquad (23)$$

where I_N denotes the $N \times N$ identity matrix. The above expression can be rewritten using the coefficient misalignment vector defined in (9) as

$$E\{V(n+1)\} \approx \left[I_N - \frac{\mu}{\rho(n)} R_{XX}\right] E\{V(n)\}. \qquad (24)$$

From (24), we can see that the mean behavior of the coefficient misalignment vector approaches the zero vector if the convergence parameter μ is selected to be

$$0 < \mu < \frac{2\,\rho(n)}{\lambda_i}, \quad \forall i \text{ and } \forall n, \qquad (25)$$

where λ_i denotes the i-th eigenvalue of the matrix R_{XX}. It is not difficult to show that a more restrictive and sufficient, but simpler and more practical condition for the convergence can be obtained as

$$0 < \mu < \frac{2}{N}, \qquad (26)$$

under the condition that the initial value for the single-pole lowpass filter is chosen to be greater than the input power to be estimated, i.e., $\rho(0) > \sigma_x^2$. Note here that knowledge of the input signal statistics is not now necessary in determining values for the convergence parameter. It must be pointed out that the two approximated expressions gien in (23) and (24) do not mean that the NLMS algorithm is biased in the mean sense. In fact, the NLMS algorithm does converge to the optimum coefficient vector in the limit in the mean sense as long as the condition in (26) is satisfied.

Even though the weight vector converges to the optimum value in the mean sense, we still have to guarantee its convergence in the mean squared sense as well. An expression for the mean squared estimation error $\sigma^2_{e(n)}$ is derived next. Squaring both sides of (13) and taking the expectation yield

$$\begin{aligned}\sigma_e^2(n) &= E\left\{e^2(n)\right\} \\ &= \xi_{min} + E\left\{V^T(n)\,X(n)\,X^T(n)\,V(n)\right\} \\ &\quad - 2\,E\left\{V^T(n)\,X(n)\,e_{min}(n)\right\},\end{aligned} \qquad (27)$$

where ξ_{min} is obtained by using (12) in (15) as

$$\xi_{min} = E\left\{d^2(n)\right\} - H_{opt}^T\, R_{dX}. \qquad (28)$$

Note that last expectation of (27) becomes zero by Assumption 2 and orthogonality principle. It thus follows that

$$\sigma_e^2(n) = \xi_{min} + tr\{K(n)\, R_{XX}\}, \qquad (29)$$

where $K(n)$ is defined in (10), and $tr\{\cdot\}$ denotes the trace of $\{\cdot\}$.

To evaluate $\sigma_e^2(n)$, an expression for $K(n)$ is also needed. Substituting (11) in (10) and employing the same approximations as those in (19) and (20) once again lead to

$$
\begin{aligned}
K(n+1) \approx\ & K(n) \\
&+ \frac{\mu}{\rho(n)} E\left\{e(n)\, V(n)\, X^T(n)\right\} \\
&+ \frac{\mu}{\rho(n)} E\left\{e(n)\, X(n)\, V^T(n)\right\} \\
&+ \frac{\mu^2}{\phi(n)} E\left\{e^2(n)\, X(n)\, X^T(n)\right\}, \qquad (30)
\end{aligned}
$$

where $\rho(n)$ and $\phi(n)$ are defined in (16) and (17), respectively. Here, using (13) and employing Assumption 2 as well as orthogonality principle once again, we have

$$
\begin{aligned}
& E\left\{e(n)\, V(n)\, X^T(n)\right\} \\
&= E\left\{ \left[e_{min}(n) - V^T(n)\, X(n)\right] V(n)\, X^T(n)\right\} \\
&= -E\left\{V(n)\, V^T(n)\, X(n)\, X^T(n)\right\} \\
&= -K(n)\, R_{XX}. \qquad (31)
\end{aligned}
$$

Similarly,

$$E\left\{e(n)\, X(n)\, V^T(n)\right\} = -R_{XX}\, K(n). \qquad (32)$$

We now invoke the Gaussian signal assumption (Assumption 1) to evaluate the last expectation on the RHS of (30) since the fourth-order expectation of Gaussian random variables can be decomposed into their second-order expectations. That is, for zero-mean, Gaussian random variables X_1, X_2, X_3, and X_4,

$$
\begin{aligned}
E\{X_1 X_2 X_3 X_4\} &= E\{X_1 X_2\}\, E\{X_3 X_4\} \\
&+ E\{X_1 X_3\}\, E\{X_2 X_4\} \\
&+ E\{X_1 X_4\}\, E\{X_2 X_3\} \qquad (33)
\end{aligned}
$$

Using (13) once again, it is straightforward to show that

$$
\begin{aligned}
& E\left\{e^2(n)\, X(n)\, X^T(n)\right\} \\
&= \xi_{min} R_{XX} + tr\{K(n)\, R_{XX}\}\, R_{XX} \\
&\quad + 2\, R_{XX}\, K(n)\, R_{XX} \\
&= [\xi_{min} + tr\{K(n)\, R_{XX}\}]\, R_{XX} \\
&\quad + 2\, R_{XX}\, K(n)\, R_{XX} \\
&= \left[\sigma_e^2(n)\, I_N + 2\, R_{XX}\, K(n)\right] R_{XX}. \qquad (34)
\end{aligned}
$$

We also need an expression for $\phi(n)$ for complete evaluation of (30). Squaring (3) and using (17), it follows that

$$
\begin{aligned}
\phi(n) &= E\left\{ [\beta\, \hat{\sigma}_x^2(n-1) + (1-\beta)\, x^2(n)]^2 \right\} \\
&= \beta^2\, \phi(n-1) + 2\beta\,(1-\beta)\, \sigma_x^2\, \rho(n-1) \\
&\quad + 3\,(1-\beta)^2\, \sigma_x^4. \qquad (35)
\end{aligned}
$$

In (35), we have made use of the fact that $E\{x^4(n)\} = 3\sigma_x^4$.

Substituting (31), (32), and (34) in (30), we therefore achieve the mean-squared behavior of the coefficients as

$$
\begin{aligned}
K(n+1) &= K(n) - \frac{\mu}{\rho(n)} [K(n)\, R_{XX} \\
&\quad + R_{XX}\, K(n)] + \frac{\mu^2}{\phi(n)} [\sigma_e^2(n)\, I_N \\
&\quad + 2\, R_{XX}\, K(n)]\, R_{XX}, \qquad (36)
\end{aligned}
$$

for which $\rho(n)$, $\phi(n)$, and $\sigma_e^2(n)$ are given in (21), (35), and (29), respectively.

Let Q be an orthogonal matrix that diagonalizes the matrix R_{XX} with the property that

$$Q Q^T = I_N \quad \text{or} \quad Q^{-1} = Q^T. \tag{37}$$

By pre and postmultiplying both sides of (36) by Q^T and Q, respectively, a transformed coordinate version of $K(n)$ becomes

$$\begin{aligned} K'(n+1) &= K'(n) - \frac{\mu}{\rho(n)}[K'(n)\Lambda \\ &+ \Lambda K'(n)] + \frac{\mu^2}{\phi(n)}[\sigma_e^2(n) I_N \\ &+ 2\Lambda K'(n)]\Lambda, \end{aligned} \tag{38}$$

where

$$K'(n) = Q^T K(n) Q, \tag{39}$$

and

$$\Lambda = Q^T R_{XX} Q = \text{diag}[\lambda_1, \lambda_2, \cdots \lambda_N]. \tag{40}$$

Note that λ's are all nonnegative real values since R_{XX} is always a nonnegative definate matrix.

The (i,j)-th element $k'_{ij}(n)$ of the matrix $K'(n)$ can be identified using (38) as

$$\begin{aligned} k'_{ij}(n+1) &= \left[1 - \frac{\mu(\lambda_i + \lambda_j)}{\rho(n)} \right. \\ &\left. + \frac{2\mu^2 \lambda_i \lambda_j}{\phi(n)}\right] k'_{ij}(n) + \frac{\mu^2 \lambda_i \sigma_e^2(n)}{\phi(n)} \delta(i-j), \end{aligned} \tag{41}$$

where

$$\delta(i-j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{42}$$

In particular, the diagonal terms of $K'(n)$ become

$$\begin{aligned} k'_{ii}(n+1) &= \left[1 - \frac{2\mu\lambda_i}{\rho(n)} + \frac{2\mu^2\lambda_i^2}{\phi(n)}\right] k'_{ii}(n) \\ &+ \frac{\mu^2 \lambda_i \sigma_e^2(n)}{\phi(n)}. \end{aligned} \tag{43}$$

Note here that from (29)

$$\begin{aligned} \sigma_e^2(n) &= \xi_{min} + tr\{K'(n)\Lambda\}, \\ &= \xi_{min} + \sum_{i=1}^{N} \lambda_i k'_{ii}(n). \end{aligned} \tag{44}$$

Since, by Schwartz inequality,

$$k'^2_{ij}(n) \le k'_{ii}(n) k'_{jj}(n), \quad \forall i, j \text{ and } \forall n, \tag{45}$$

the convergence of the main diagonal elements of the matrix $K'(n)$ ensures that of the off diagonal elements. Note that all the off-diagonal elements of $K'(n)$ becomes zero once the convergence of $K'(n)$ takes place since, according to (41), they contain only homogeneous exponential parts in their difference equations. Note also that the convergence of $K'(n)$ guarantees the convergence of $\sigma_e^2(n)$.

Recall that conditions for the mean-squared convergence of the coefficient misalignment vector of the LMS algorithm were derived in an elegant way in [4]. Here, we modify the results in [4] to obtain the following conditions for the mean squared convergence of the

NLMS algorithm :

$$0 < \mu < \frac{1}{N}, \tag{46}$$

and for all n

$$0 < \frac{\mu\,\rho(n)}{2\,\phi(n)} \sum_{i=1}^{N} \frac{\lambda_i}{1 - \mu\,\lambda_i\,\rho(n)/\phi(n)} < 1. \tag{47}$$

We are now ready to compute the steady-state behavior of the mean-squared estimation error. Let $\rho(\infty)$, $\phi(\infty)$, $K'(\infty)$, and $\sigma_e^2(\infty)$ denote the limiting values of $\rho(n)$, $\phi(n)$, $K'(n)$, and $\sigma_e^2(n)$, respectively. To evaluate $\sigma_e^2(\infty)$, $K'(\infty)$ needs to be computed first. Taking the limit as n goes to infinity on both sides of (38) yields

$$\begin{aligned} K'(\infty) &= K'(\infty) - \frac{\mu}{\rho(\infty)}\,[K'(\infty)\,\Lambda \\ &\quad + \Lambda\,K'(\infty)] + \frac{\mu^2}{\phi(\infty)}\,[\sigma_e^2(\infty)\,I_N \\ &\quad + 2\,\Lambda\,K'(\infty)]\,\Lambda, \end{aligned} \tag{48}$$

where, using(21) and (35) respectively, we obtain

$$\rho(\infty) = \sigma_x^2, \tag{49}$$

$$\phi(\infty) = \frac{3-\beta}{1+\beta}\,\sigma_x^4. \tag{50}$$

Notice that if β is chosen to be less than one but very close to one, it follows that

$$\phi(\infty) \approx (2-\beta)\,\sigma_x^4 \approx \sigma_x^4. \tag{51}$$

Now, define a constant α as

$$\alpha = \frac{\rho(\infty)}{\phi(\infty)}. \tag{52}$$

Then

$$\alpha = \frac{1+\beta}{(3-\beta)\,\sigma_x^2} \approx \frac{1}{(2-\beta)\,\sigma_x^2} \approx \frac{1}{\sigma_x^2}. \tag{53}$$

Since $K'(\infty)$ is a diagonal matrix, (45) simplifies to

$$K'(\infty) = \frac{\mu}{2}\,\alpha\,\sigma_e^2(\infty)\,[\,I_N - \mu\,\alpha\,\Lambda\,]^{-1}, \tag{54}$$

and thus the diagonal terms of $K'(\infty)$ becomes

$$k'_{ii}(\infty) = \frac{\mu}{2}\,\frac{\alpha\,\sigma_e^2(\infty)}{1 - \mu\,\alpha\,\lambda_i}. \tag{55}$$

Taking the limit on both sides of (44) and using (55), we get

$$\sigma_e^2(\infty) = \xi_{min} + \frac{\mu}{2}\,\alpha\,\sigma_e^2(\infty) \sum_{i=1}^{N} \frac{\lambda_i}{1 - \mu\,\alpha\,\lambda_i} \tag{56}$$

Solving (56) for $\sigma_e^2(n)$ yields the steady-state mean squared estimation error as

$$\begin{aligned} \sigma_e^2(\infty) &= \frac{\xi_{min}}{1 - \frac{\mu\,\alpha}{2} \sum_{i=1}^{N} \frac{\lambda_i}{1 - \mu\,\alpha\,\lambda_i}} \\ &= \xi_{min} \sum_{m=0}^{\infty} \left\{\frac{\mu\,\alpha}{2}\right\}^m \\ &\quad \left\{\sum_{i=1}^{N} \frac{\lambda_i}{1 - \mu\,\alpha\,\lambda_i}\right\}^m \end{aligned} \tag{57}$$

Note that it is possible to expand (56) as in (57) only when the convergence parameter μ satisfies the condition given in (47).

Now, for small values of μ, $\sigma_e^2(\infty)$ in (57) can be approximated by considering only the first two terms of the power series such that

$$\sigma_e^2(\infty) \approx \xi_{min} + \frac{\mu\alpha}{2}\xi_{min}\sum_{i=1}^{N}\frac{\lambda_i}{1-\mu\alpha\lambda_i}. \qquad (58)$$

Employing the "small values of μ" approximation once again and using (53), the above expression can be further simplified as

$$\sigma_e^2(\infty) \approx \xi_{min} + \frac{\mu}{2\sigma_x^2}\xi_{min}\sum_{i=1}^{N}\lambda_i. \qquad (59)$$

Under the same set of conditions and assumptions, it is not difficult to show that the limiting mean squared error of the LMS algorithm can be written as

$$\sigma_e^2(\infty) \approx \xi_{min} + \frac{\mu}{2}\xi_{min}\sum_{i=1}^{N}\lambda_i. \qquad (60)$$

Comparing(59) and (60), we can make the following two remarks about performances of the LMS and NLMS algorithms : First, attaining the value of μ for each algorithm in such a way that the steady state mean squared error for both algorithms becomes the same, and using it in the coefficient update equation of each algorithm, we can see that the speed of mean convergence for the NLMS algorithm is more or less the same as that for the LMS algorithm. Second, if the same value of μ is employed for both algorithms, the NLMS algorithm is preferred in environments where the input signal power is greater than one, since in this case the steady-state mean squared error of the NLMS algorithm. is less than that of the LMS algorithm. Remember though that these comparisons are meaningful only when μ is fairly small and β is less than but very close to one.

3. Experimental Results

We now present some of the experimental results to demonstrate the validity of our analysis. For this, a third-order adaptive predictor is used. The primary input signal $d(n)$ is modeled as the output of the third-order autoregressive filter in such a way that

$$d(n) = \eta(n) + 0.9\,d(n-1) - 0.1\,d(n-2) - 0.2\,d(n-3). \qquad (61)$$

The input $\eta(n)$ to the autoregressive filter is a zero mean, white Gaussia pseudorandom sequence with variance such that the variance of $d(n)$ is one. The eigenvalue spread ratio of the signal $d(n)$ is approximately 16.3. The results presented are comparisons of the theoretical curves with those obtained from the simulation. The ensemble averages are obtained by averaging over 400 independent runs with 10,000 samples each. The constants μ and β are chosen to be 0.005 and 0.995, respectively.

Figure 1 shows plots of the theoretical and empirical mean behavior of the filter coefficients, where $E\{h_i(n)\}$, $i=1,2,3$, represents the i-th element of the vector $E\{H(n)\}$. In each plot of the figure, curve 1 is for the simulation result and curve 2 is for the theoretical result predicted by (23). (21) is required also.) It can be seen that

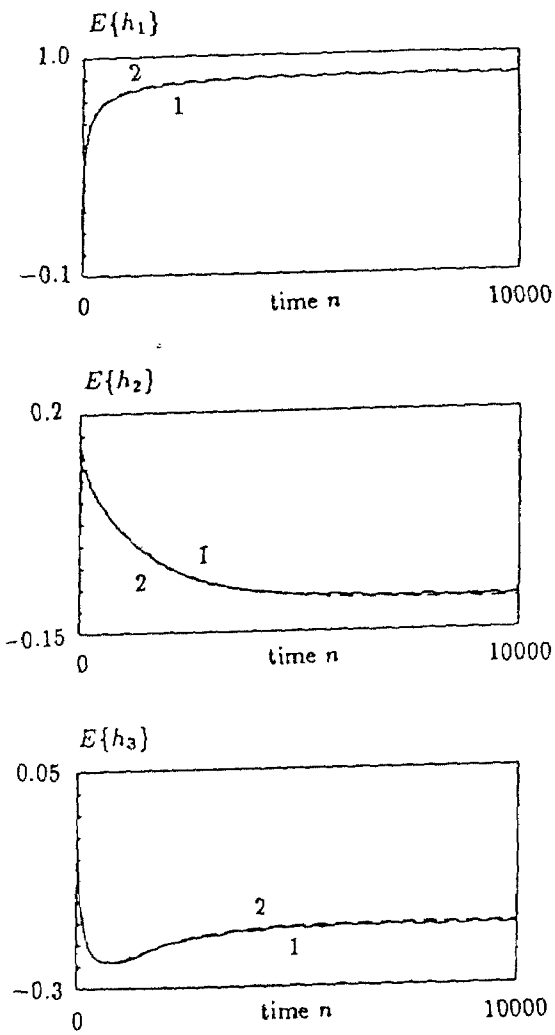

Fig.1 Mean behavior of the coefficients : 1=simulation result, 2=theoretical result.

the theoretical results for the mean behavior agree with the simulation results fairly well.

Next, the diagonal elements of the matrix $K(n)$ predicted by (36) are compared with their simulation counterparts and depicted in Figure 2, in which $K_{ii}(n)$ denotes the i-th diagonal element of $K(n)$. Recall that (21), (29), and (33) are necessary to evaluate $K(n)$. Once again, the theoretical curves show very good

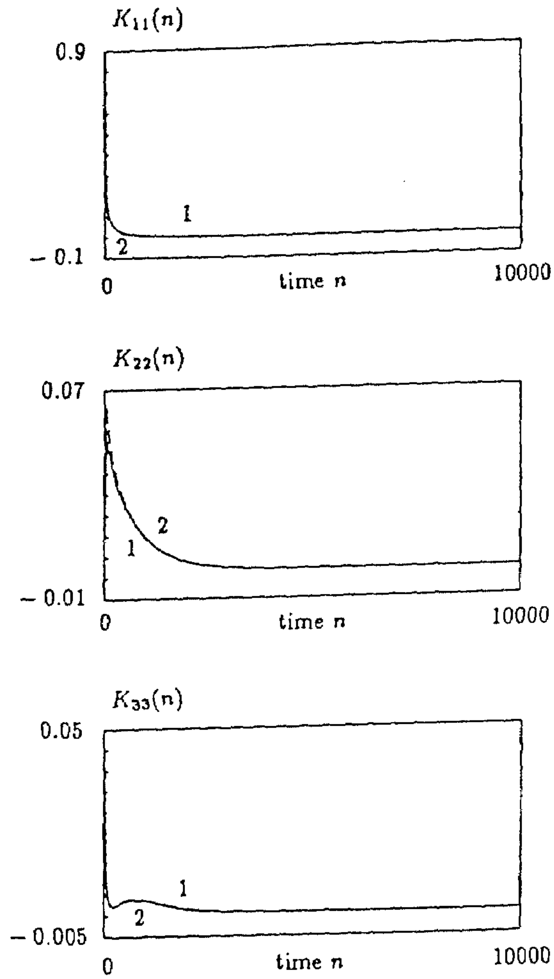

Fig.2. Mean-squared behavior of the coefficients : 1=simulation result, 2=theoretical result.

agreement with empirical results. Note that all the curves in Figure 2 are nonnegative. Negative values for the verticval axis are only for ease of dispalying the results.

4. Conclusion

In this paper, we have presented a convergence analysis of the NLMS algorithm using a single-pole lowpass filter. Under the assumption that the primary and reference inputs to the adaptive filter are zero-mean, wide-sense stationary, and Gaussian random processes, and further making use of the independence assumption, we derive a set of difference equations that characterizes the mean and mean-squared behavior of the filter coefficients as well as the mean-squared estimation error. Conditions for the mean and mean-squared convergence are also explored. We then derive an expression for the steady stute mean squared estimation error, and compare the performance of the NLMS algorithm with that of LMS algorithm. Finally, experimental results that show very good agreement between the analytical and empirical results are presented.

It is hoped that the results obtained here furnish additional design criteria for implementation of adaptive filters in practice.

Referencess

1. B. Widrow, et al., "Adaptive noise cancelling : Principles and applications," *Proc. of IEEE*, Vol.63, pp. 1692-1716, Dec. 1975.
2. B. Widrow, et al., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. of IEEE*, Vol.64, pp.1151-1162, Aug. 1976.
3. L.L. Horowitz and K.D. Senne, "Performance advantage of complex LMS for controlling narrow band adaptive arrays", *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol.ASSP-29, pp.722-736, June 1981.
4. A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data", *IEEE Trans. Acoust., Speech, and Signal Proc.* Vol. ASSP-33, pp.222-229, Feb. 1985.
5. A. Ralston, *A First Course in Numerical Analysis*, McGraw Hill, New York, 1965.
6. T.C. Hsia, "Convergence analysis of LMS and NLMS adaptive algorithms", *Proc. of the IEEE Int. Conf. Acoust. Speech. and Signal Proc.* pp.667-670, 1983.
7. N.J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian input," *IEEE Trans. Acoust., Speech, and Signal Proc.*, Vol.ASSP-34, pp. 793-806, Aug. 1986.
8. M. Tarrab and A. Feuer, "Convergence and Performance Analysis of the Normalized LMS Algorithm with Uncorrelated Gaussian Data," *IEEE Trans. on Inform. Theory*, Vol.IT-34, pp.680-691, July 1988.
9. L.J. Griffiths, "An adaptive lattice structure for noise cancelling applications," *Proc. of the IEEE Int. Conf. Acoust., Speech, and Signal Proc.* pp.87-90, 1978.
10. D.H. Youn, N.Ahmed, and G.C. Carter, "Magnitude squared coherence function estimation : An adaptive approach," *IEEE Trans. Acoust., Speech, and Signal Proc.*, Vol.ASSP-31, pp.137-142, Feb. 1983.
11. J.E. Mazo, "On the independence theory of the equaliz er convergence," *Bell Sys. Tech. J.*, Vol.58, pp.96e-993, May-June 1979.
12. S.H. Cho and V.J. Mathews, "Tracking analysis of the sign algorithm in nonstationary environments," *IEEE Trans. Acoust., Speech, and Signal Proc.* Vol .ASSP- 38, pp.2046-2057, Dec. 1990.

趙 誠 鎬(Sung Ho CHO) 종신회원
1960년 2월 21일생
1982년 2월 : 한양대학교 전자공학과 졸업
1984년 12월 : Univ. of Iowa 전기 및 컴퓨터공학과 졸업(M.S.)
1989년 8월 : Univ. of Utah 전기공학과 졸업(Ph.D.)
1989년 8월~현재 : 한국전자 통신연구소 선임연구원
※주관심분야 : Adaptive Filtering Algorithms and Structures, Optimal Estination Theory, Stochastic Process, Digital Communication Mobile Telecommunications

朴 光 浩(Kwang Ho PARK) 正會員
1980년 2월 : 한국항공대학 전자공학과 졸업
1986년 2월 : 한국항공대학 대학원 전자공 학과 졸업(M.S.)
1980년 2월~현재 : 한국전자통신연구소 선임연구원, 실장
※주관심분야 : Data Communication Speech Coding

宋 翊 鎬(Iickho SONG) 正會員
1960년 2월 20일생
1982년 2월 : 서울대학교 전자공학과(공학사, magna cum laude)
1984년 2월 : 서울대학교 대학원 전자공학과(공학석사)
1985년 8월 : Univ. of Pennsylvania, Dept. of EE(M.S.E.)
1987년 5월 : Univ. of Pennsylvania, Dept. of EE(Ph.D.)
1982년 9월~1983년 8월 : IEEE 학생활동위원회 서울대학교 지부 학생지부장
1989년 1월~1989년12월 : IEEE 한국지회 재무
1990년 1월~1991년 4월 : 한국음향학회 편집위원
1984년 1월~1987년 2월 : Univ. of Pennsylvania 연구조교
1987년 3월~1988년 2월 : Bell Communications Research (Morristown)연구원
1988년 3월~1991년 8월 : 한국과학기술원 전기 및 전자공학과 조교수
1991년 9월~현재 : 한국과학기술원 전기 및 전자공학과 부교수
연구관심분야 : 검파와 추정, 통계학적 신호(화상)처리와 통신이론