論文95-32B-9-7

DLMS 알고리듬의 수렴에 관한 연구

(Almost-Sure Convergence of the DLMS Algorithm)

安 相 植 *

(Sang-Sik Ahn)

요 약

LMS 알고리듬을 구현할 때 궤환 딜레이가 발생하는 경우에 계수 갱신 알고리듬은 Delayed Least Mean Square (DLMS)가 된다. 이의 수렴특성에 대한 해석들이 발표되고 있지만 대부분은 연속적인 입력벡터끼리 통계적으로 독립한다는 비현실적인 가정 아래 Mean Convergence를 증명하고 있다. 본 논문에서는 Mixing condition과 Law of Large Numbers 가정아래, Decreasing Step Size를 갖는 DLMS 알고리듬의 Almost-Sure Convergence를 증명하고 응용에로 결정궤환 적응등화기에 적용하여 컴퓨터 시뮬레이션을 행하였다.

Abstract

In some practical applications of the LMS algorithm the coefficient adaptation can be performed only after some fixed delay. The resulting algorithm is known as the Delayed Least Mean Square (DLMS) algorithm in the literature. There exist analyses for this algorithm, but most of them are based on the unrealistic independence assumption between successive input vectors. In this paper we consider the DLMS algorithm with decreasing step size $\mu(n) = \frac{a}{n}, a > 0$ and prove the almost-sure convergence of the weight vector W(n) to the Wiener solution W_{opt} as $n \to \infty$ under the mixing input condition and the satisfaction of the law of large numbers. Computer simulations for decision-directed adaptive equalizer with decoding delay are performed to demonstrate the functioning of the proposed algorithm.

I. Introduction

The LMS algorithm can be implemented only under the assumption that we can measure the error signal and input vector at every iteration. In some practical situations the error signal can be obtained only after some fixed delay. We will encounter this kind of situation in at least two typical situations. First, when employing a decision-directed adaptive equalizer, if we use a decoding procedure such as the Viterbi algorithm, the desired signal, hence the error, is not available until several symbol intervals later because of decoding delay. We encounter the same problem in adaptive reference echo cancellation[1]. Second, in the high speed signal processing application, if the LMS algorithm is implemented using parallel

* 正會員, 高麗大學校 應用電子工學科

(Dept. of Applied Electronics, Korea Univ.)

接受日字:1995年4月29日, 수정완료일:1995年8月28日

architecture, such as a pipeline structure or systolic array, the error signal is generated only after some inherent processing delay[2][3]. In such applications the algorithm realized is the modified version of the LMS algorithm known as the Delayed LMS (DLMS) algorithm in the literature.

The convergence properties of the DLMS algorithm have been investigated by Kabal[4] Long, Ling, and Proakis[5][6] and Herzberg, Haimi-Cohen, and Be'ery[7] Kabal derived a stability bound to ensure convergence of the mean of the weight vector under the standard independence assumption, which assumes that the LMS weight vector is statistically independent of the input vector. Long et al. considered the special case when the input vector arises from a tapped delay line implementation, and derived a stability bound on the step size for convergence of the excess mean square error under the independence assumption. These analyses are interesting and give valuable insights into the convergence properties but, from a practical viewpoint, they are not enough to guarantee the correct performance of the particular realization with which the user must live. Indeed we need probability one convergence to guarantee convergence for almost all sample functions. For the normalized version of the DLMS (DNLMS) algorithm we[8] proved almost-sure convergence with a decreasing step size assuming the mixing input and the satisfaction of a certain law of large numbers instead of the independence of the input. The normalized algorithm has a nice geometrical property of projection which is benevolent to the convergence proof. From an implementation viewpoint, however, normalized algorithm has disadvantages of requiring additional multiplications to calculate the input vector norm square and a division, which is difficult to implement. In practice, therefore, the standard LMS algorithm is preferred for actual implementation to take full advantage of the simplicity of the LMS algorithm. The purpose of this paper is then to extend the previous analysis to the proof of the sample convergence of the DLMS algorithm with an additional assumption of uniformly bounded input.

Ⅱ. Formulation of the Problem

The most common application of the LMS algorithm is to attack the following problem. Given a desired signal u(n) and a data vector X(n), the linear estimation of u(n) in terms of X(n) is characterized as a weight vector W of the same dimension N as the data vector such that the estimate $\hat{u}(n)$ of u(n) is obtained by

$$\hat{u}(n) = X(n)^T W \tag{1}$$

The problem is to find W which minimizes the mean square error ξ namely,

$$\xi = E\{[u(n)-X(n)^T W]^2\} \tag{2}$$

where E{·} denotes the statistical expectation of the braced quantity and the superscript T denotes the matrix transposition. Since ξ is a quadratic functional of W, it is well known that it has a unique minimizing vector obtained by choosing

$$W_{opt} = R^{-1} r \tag{3}$$

where

$$R = E\{X(n)X(n)^T\} \tag{4}$$

$$r = E\{u(n)X(n)\} \tag{5}$$

assuming the stationarity of X(n), u(n), and invertibility of R. In the absence of knowledge about the statistics of of u(n) and X(n), we can solve (3) iteratively by the well-known stochastic gradient algorithm:

$$W(n+1) = W(n)+\mu \{u(n)-X(n)^T W(n)\}X(n) \quad (6)$$

where W(n) is the weight vector for the n-th iteration cycle μ and is a constant gain which controls the rate of convergence and stability of the algorithm. This is the Widrow's well known LMS algorithm. The delayed update LMS algorithm arises when, in some practical situations the error signal can not be observed until after some fixed delay. In such applications the implemented algorithm, by analogy with (6), becomes

$$W(n+1)=W(n)+\mu \{n(n-d)-X(n-d)^T W(n-d)\}X(n-d)$$

This is the modified version of the LMS known as the Delayed LMS (DLMS) algorithm in the literature. Defining the weight error vector $Y(n)=W(n)-W_{opt}$ and time-varying step size μ(n) and describing the algorithm in terms of Y(n), we obtain

$$\begin{aligned} Y(n+1)= & Y(n)-\mu(n-d)X(n-d)X(n-d)^T Y(n-d) \\ & +\mu(n-d)\{u(n-d)-X(n-d)^T W_{opt}\}X(n-d) \end{aligned} \quad (7)$$

Ⅲ. Assumptions and Idea of Proof

In this section we summarize the required assumptions, basic idea of proof, and important technical lemmas, which are proved in the appendices.

(A1) The input vectors are mixing in the sense that there exist a finite integer T and a> 0 such that for any constant non-zero N-vector h, the following holds for all n

$$\frac{1}{T}\sum_{i=0}^{T-1}\left\{\frac{h^T X(n+i)}{\| X(n+i)\|}\right\} \geq a\| h\|^2$$

The random matrix sequences $\{X(n)X(n)^T\}$ and $\{u(n)X(n)\}$ satisfy the law of large numbers in the sense that

(A2-1) $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X(i)X(i)^T = E\{X(i)X(i)^T\}=R$ a.s

(A2-2) $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}u(i)X(i)=E\{u(i)X(i)\}=r$ a.s.

(A3) $0 < k\leq \| X(n)\|^2\leq K \ \forall n$

The mixing assumption (A1) is a deterministic constraint, and is very common assumption in the literature [11] for discussions of algorithm convergence. Basically it means that, over any time interval of length T, the components of X(n) have an average length of at least α in any direction. To examine the property more clearly, let us rewrite (A1) as

$$h^T\left\{\frac{1}{T}\sum_{i=0}^{T-1}\frac{X(n+i)X(n+i)^T}{\| X(n+i)\|^2}\right\}h\geq a\| h\|^2$$

and note that $\frac{X(n+i)X(n+i)^T}{\| X(n+i)\|^2}$ is a projection matrix. Therefore if a sequence of N-vector is restricted to any proper subspace of $\Re^N$, then it is non mixing since there exists an N vector h which is orthogonal to the subspace. By the same reasoning we can expect that T must be greater than or equal to dimension N. An equivalent rephrasing of the mixing condition is

$$\lambda_{\min}\left\{\frac{1}{T}\sum_{i=0}^{T-1}\frac{X(n+i)X(n+i)^T}{\| X(n+i)\|^2}\right\} \geq a$$

For the assumption (A2), it is well known that there exists a large class of stochastic processes following the law of large numbers such as the infinite terms moving average process [9] and the stationary and asymptotically independent process [10]. Finally, (A3) is nothing but a bounded input assumption and is not a constraint at all in practice. Notice that under (A2)

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\{u(i)X(i)-X(i)X(i)^T W_{opt}\}=r-RW_{opt}=0 \quad (8)$$

since $W_{opt}=R^{-1}r$. Now, with (2-7), if we define

$$P(n) = X(n)X(n)^T \quad (9)$$

and

$$Z(n) = u(n)X(n) - X(n)X(n)^T W_{opt}, \tag{10}$$

then it can be written in the simple form

$$Y(n+1) - Y(n) = -\mu(n-d)\{P(n-d)Y(n-d) - Z(n-d)\} \tag{11}$$

Now, adding up d terms of the above, and solving for Y(n-d) yields

$$\begin{aligned} Y(n-d) = \ & Y(n) + \sum_{i=0}^{d-1}\mu(n-2d+i)P(n-2d+i)Y(n-2d+i) \\ & - \sum_{i=0}^{d-1}\mu(n-2d+i)Z(n-2d+i) \end{aligned} \tag{12}$$

and substituting (12) into (11), we have

$$\begin{aligned} Y(n+1) = \ & [I - \mu(n-d)P(n-d)]\, Y(n) \\ & - \mu(n-d)P(n-d)\sum_{i=0}^{d-1}\mu(n-2d+i)P(n-2d+i)Y(n-2d+i) \\ & + \mu(n-d)P(n-D)\sum_{i=0}^{d-1}\mu(n-2d+i)Z(n-2d+i) + \mu(n-d)Z(n-d) \end{aligned}$$

where I denotes the identity matrix of dimension N. Again, for simplicity, let us define Q(n), Γ(n) (n), and L(n) as follows:

$$\begin{aligned} Q(n) &= I - \mu(n-d)P(n-d) \\ \Gamma(n) &= -\mu(n-d)P(n-d)\sum_{i=0}^{d-1}\mu(n-2d+i)P(n-2d+i)Y(n-2d+i) \\ L(n) &= \Gamma(n) + \mu(n-d)P(n-d)\sum_{i=0}^{d-1}\mu(n-2d+i)Z(n-2d+i) \\ & \quad + \mu(n-d)Z(n-d) \end{aligned}$$

We then have the following simple form of the DLMS algorithm

$$Y(n+1) = Q(n)Y(n) + L(n)$$

The solution of the above recursive equation is given by

$$\begin{aligned} Y(n+1) = \ & \\ & \Psi(n+1,1)Y(1) + \sum_{j=1}^{n}\Psi(n+1,j+1)L(j) \\ & \Psi(n+1,1)Y(1) + \sum_{j=1}^{n}\Psi(n+1,j+1)\Gamma(j) \\ & + \Sigma + j = 1^n\mu(j-d)\Psi(n+1,j+1)P(j-d)\sum_{i=0}^{d-1}\mu(j-2d+i)Z(j-2d+i) \\ & + \sum_{j=1}^{n}\mu(j-d)\Psi(n+1,j+1)Z(j-d) \end{aligned} \tag{13}$$

where the transition matrix is defined as

$$\begin{aligned} \Psi(n,k) &= Q(n-1)Q(n-2)\ \ldots\ Q(k+1)Q(k)\ ; \ n > k \\ &= I \qquad ; \ n = k \end{aligned} \tag{14}$$

Now taking the norm of (13) with $\mu(n) = \frac{a}{n}$ a>0, we have

$$\begin{aligned} \| Y(n+1) \| \le & \left\| \Psi(n+1,1)Y(1) = \sum_{j=1}^{n}\Psi(n+1,j+1)\Gamma(j) \right\| \\ & + a^2 \left\| \sum_{j=1}^{n}\frac{\Psi(n+1,j+1)}{j-d}P(j-d)\sum_{i=0}^{d-1}\frac{Z(j-2d+i)}{j-2d+i} \right\| \\ & + a \left\| \sum_{j=1}^{n}\frac{\Psi(n+1,j+1)}{j-d}Z(j-d) \right\| \end{aligned} \tag{15}$$

Thus, if we can show that each term of (15) converges to zero as n tends to infinity, then $W(n) \to W_{opt}$ as $n \to \infty$.

Ⅳ. Proof of the Convergence

Lemma 1: Let $\Psi(\cdot,\cdot)$ be the transition matrix of the LMS algorithm with a decreasing step size. Under (A1) and (A3), there exist $0 < \beta < 1$ and N_0 that

$$\| \Psi(n, n_0) \| \ e^{-KA\beta T\sum_{i=1}^{m}\mu(n_0+iT)} \quad \text{for} \quad n_0 > N_0$$

where $m = \left[\frac{n-n_0}{T}\right]$ and [x] denotes the greatest integer function of x. The proof of this lemma is an extension of the convergence proof of the NLMS algorithm given by Weiss and Mitra [11] to the LMS algorithm.

Proof: See Appendix I

Lemma 2: Under (A1), (A2), (A3) and $\mu(n) = \frac{a}{n}, a > 0$ the weight vector W(n) in the LMS algorithm converges to W_{opt} almost surely as n tends to infinity.

Proof: See Appendix Ⅱ

Theorem: Under (A1), (A2), (A3) and $\mu(n) = \frac{a}{n}, a > 0$, W(n) in the DLMS algorithm converges to W_{opt} almost surely as n tends to infinity.

Proof of the Theorem:

We prove the convergence of the DLMS

algorithm by demonstrating that each term of (15) converges to zero. The first term in (15) represents the homogeneous DLMS algorithm. Therefore

$$\lim_{n\to\infty}\left\| \Psi(n+1,1)Y(1)+\sum_{j=1}^{n}\Psi(n+1,j+1)\Gamma(j)\right\| = 0$$

due to Lemma 1 and [12]. For the second term of (3-8), observe that

$$\begin{aligned}
&\left\| \sum_{j=1}^{n}\frac{\Psi(n+1,j+1)}{j-d}P(j-d)\sum_{i=0}^{d-1}\frac{Z(j-2d+i)}{j-2d+i}\right\| \\
&\le \sum_{i=0}^{d-1}\left\| \sum_{j=1}^{n}\frac{\Psi(n+1,j+1)}{j-d}P(j-d)\frac{Z(j-2d+i)}{j-2d+i}\right\| \\
&\le \sum_{i=0}^{d-1}\left\| \sum_{j=1}^{n_0-1}\frac{\Psi(n+1,j+1)}{j-d}P(j-d)\sum_{i=0}^{d-1}\frac{Z(j-2d+i)}{j-2d+i}\right\| \\
&+\sum_{i=0}^{d-1}\left\| \sum_{j=n_0}^{n_0-1}\frac{\Psi(n+1,j+1)}{j-d}P(j-d)\sum_{i=0}^{d-1}\frac{Z(j-2d+i)}{j-2d+i}\right\|
\end{aligned} \tag{16}$$

The first term above converges to zero since it contains a finite number of terms, each of which converges to zero by Lemma 1. Therefore, we only need to show the convergence of the second term of (16) and the third term of (15) to zero as n tends to infinity. To finish the proof, we may follow the same steps of Appendix II from (II-5) to the end.

Ⅴ. Computer Simulation

In this simulation we study the use of DLMS algorithm for decision-directed adaptive equalizer with decoding delay. Figure 1 shows the block diagram of the system used to carry out the simulation, which is a modified version of Haykin's model[13] which is widely accepted for the simulation of the adaptive equalizer.

The random data generator provides the test signal, a(n), used for probing the channel, whereas the random noise generator v(n) serves as the source of additive white noise that corrupts the channel output. These two random number generators are independent of each other. The adaptive equalizer has the task of correcting the distortion produced by the channel in the presence of the additive white noise. The random signal generator, after suitable delay plus decoding delay, also supplies the desired signal applied to the adaptive equalizer to simulate a decoder with decoding delay. The random sequence {a(n)} applied to the channel is in bi-polar form: $a(n)=\pm 1$ with equal probability, so the sequence {a(n)} has zero mean and σ_a^2 . The impulse response of the channel is described by the raised cosine:

$$h(n)=\begin{cases}\frac{1}{2}\left[\ 1+\cos\left(\frac{2\pi}{\Theta}\ (n-2)\right)\right] \\ 0, \quad otherwise\end{cases} \tag{17}$$

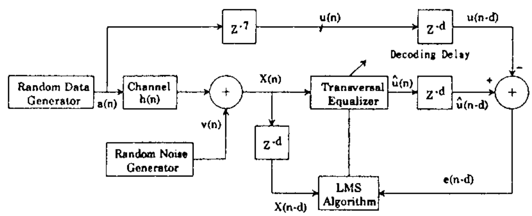

Fig. 1. Block Diagram for Adaptive Equalizer Simulation.

where the parameter Θ controls the amount of amplitude distortion produced by the channel and eigenvalue spread $\chi(R)$ of the auto- correlation matrix of the input to the equalizer. The sequence {v(n)}, produced by the second random generator, has a normal distribution with zero mean and variance $\sigma_v^2=0.001$. The equalizer has N=11 taps. Since the channel has an impulse response {h(n)} that is symmetrical about time n=2, it follows that the optimum tap weights W_{opt} of the equalizer are likewise symmetric, about time n=5. Therefore, when there is no decoding delay the channel input {a(n)} must be delayed by 7 symbols to provide the correct desired signal to the equalizer. In this simulation the adaptive algorithm will be DLMS due to the decoding delay d. We have chosen d=20 and a=1 for our simulation.

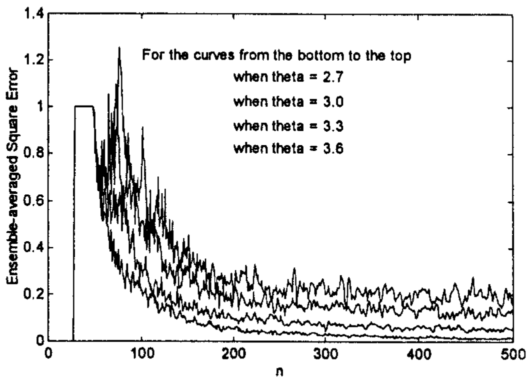

Fig. 2. Learning Curves of the DLMS algorithm.

Simulation result are shown in Figure 2 We took ensemble average over 100 samples to obtain this result. From figure, we can observe the effect of eigenvalue spread of the input auto-correlation matrix on the learning curve of the algorithm; the performance of the DLMS algorithm is getting worse as the eigenvalue spread gets large, which is a characteristic of the LMS algorithm.

Ⅵ. Conclusion

This paper has been concerned with the convergence of the LMS algorithm when it is constrained to operate with coefficient update delay. The main result has been to prove the almost-sure convergence of the DLMS algorithm with decreasing step size assuming the mixing input and the satisfaction of a certain law of large numbers. The analysis does not require unrealistic independence assumption. Computer simulations are performed for decision-directed adaptive equalizer with decoding delay for completeness.

Appendix I

To prove the Lemma 1, we start with the following recursive equation:

$$Y(n+1) = Q(n)Y(n) = [I - \mu(n)X(n)X(n)^T]\,Y(n) \qquad \text{(I-1)}$$

Under (A3), to extend the result of [11], defining

$$\theta(n) = \mu(n)\,\|X(n)\|^2 \qquad \text{(I-2)}$$

we get

$$Y(n+1) = Q(n)Y(n) = \left[\, I - \theta(n)\frac{X(n)X(n)^T}{\|X(n)\|^2}\right] Y(n) \qquad \text{(I-3)}$$

The solution of this recursive equation is given by

$$Y(n+T) = \Psi(n+T, n)Y(n) \qquad \text{(I-4)}$$

where the transition matrix is defined in (3-7). Taking the norm of (I-4), we get

$$\|Y(n+T)\| \le \|\Phi(n+T, n)\|\,\|Y(n)\| \qquad \text{(I-5)}$$

Now, to use the mixing assumption, substituting Y(n) for h in (A1)

$$\begin{aligned} aT\|Y(n)\|^2 &\le \sum_{i=0}^{T-1}\left\{\frac{Y(n)^T X(n+i)}{\|X(n+i)\|}\right\}^2 \\ &= \sum_{i=0}^{T-1}\left[\{Y(n)-Y(n+i)+Y(n+i)\}^T\frac{X(n+i)}{\|X(n+i)\|}\right]^2 \\ &= \sum_{i=0}^{T-1}\left\{\frac{Y(n+i)^T X(n+i)}{\|X(n+i)\|}\right\}^2 + \sum_{i=0}^{T-1}\left\{\{Y(n)-Y(n+i)\}^T\frac{X(n+i)}{\|X(n+i)\|}\right\}^2 \\ &\quad +2\sum_{i=0}^{T-1}\{Y(n)-Y(n+i)\}^T\frac{X(n+i)X(n+i)^T}{\|X(n+i)\|^2}\,Y(n+i) \end{aligned} \qquad \text{(I-6)}$$

(I-6) is the fundamental inequality from which the basic result is developed. Bounds on each term of (I-6) can be obtained by utilizing the decreasing property of μ(n) sequence as follows:

$$\sum_{i=0}^{T-1}\left\{\frac{Y(n+i)^T X(n+i)}{\|X(n+i)\|}\right\}^2 \le \frac{1}{k\mu(n+T-1)(2-K\mu(n))}$$

$$(\|Y(n)\|^2 - \|Y(n+T)\|^2)$$

$$\sum_{i=0}^{T-1}\left\{\{Y(n)-Y(n+i)\}^T\frac{X(n+i)}{\|X(n+i)\|}\right\}^2$$

$$\le \frac{T(T-1)}{2}\frac{\mu(n)^2K^2}{k\mu(n+T-1)(2-K\mu(n))}(\| Y(n)\|^2-\| Y(n+T)\|^2)$$

[since $\| Y(n)\|$ is non-increasing if $0<\mu(n)\frac{2}{K}$]

$$2\sum_{i=0}^{T-1}\{Y(n)-Y(n+i)\}^T\frac{X(n+i)X(n+i)^T}{\| X(n+i)\|^2}Y(n+i)$$

$$\le \frac{(2T^2-T-1)\mu(n)K}{k\mu(n+T-1)(2-K\mu(n))}(\| Y(n)\|^2-\| Y(n+T)\|^2)$$

Now substituting above bounds into (I-6) yields

$$aT\|(Y(n)\|^2\le\frac{1}{k\mu(n+T-1)(2-K\mu(n))}(\| Y(n)\|^2-\| Y(n+T)\|^2)$$
$$+\frac{T(T-1)}{2}\frac{\mu(n)^2K^2}{k\mu(n+T-1)(2-K\mu(n))}(\| Y(n)\|^2-\| Y(n+T)\|^2)$$
$$+\frac{2T^2-T-1)\mu(n)K}{k\mu(n+T-1)(2-K\mu(n))}(\| Y(n)\|^2-\| Y(n+T)\|^2) \quad \text{(I-7)}$$

Multiplying $2k\mu(n+T-1)(2-K\mu(n))$ on both sides of the above, we will have

$$\frac{2k\mu(n+T-1)(2-K\mu(n)aT\| Y(n)\|^2}{\|(n+T)\|^2}$$
$$\le\{2+2(2T^2-T-1)K\mu(n)+T(T-1)K^2\mu(n)^2\}$$

Finally, arranging for $\| Y(n+T)\|^2$, we obtain the following bound:

$$\| Y(n+T)\|^2<\left[1-\frac{2k(2-K\mu(n))\mu(n+T-1)aT}{2+2(2T^2-T-1)K\mu(n)+T(T-1)K^2\mu(n)^2}\right]\| Y(n)\|^2$$

At this point, comparing the above with (I-5), we get

$$\|\Phi(n+T,n)\|\le\sqrt{1-\frac{2K(2-2-K\mu(n))\mu(n+T-1)T}{2+2(2T^2-T-1)K\mu(n)+T(T-1)K^2\mu(n)^2}}$$

Now using the decreasing property of $\mu(n)$, we obtain the exponential bound on $\|\Phi(n+T,n)\|$ Equation as follows:

$$\|\Phi(n+T,n)\|\le\sqrt{1-\frac{2kaT\mu(n+T-1)\{2-K\mu(n)\}}{2-K\mu(n)+(4T^2-2T-1)K\mu(n)+T(T-1)K^2\mu(n)^2}}$$
$$\le\sqrt{1-2KaT\mu(n+T-1)\beta(n)}$$

where

$$\beta(n)=\frac{2-K\mu(n)}{2-K\mu(n)+(4T^2-222T-1)K\mu(n)+T(T-1)K^2\mu(n)^2}$$

Notice that $0<\beta(n)<1$ if $0<\mu(n)<\frac{2}{K}$. Thus, there exists N_0 such that for $n>N_0$

$$0<\mu(n)<\frac{2}{K},\ 2KaT\mu(n+T-1)<1$$

Finally, we obtain

$$\|\Phi(n+T,n)\|\le 1-Ka\beta T\mu(n+T-1)$$
$$\le 1-Ka\beta T\mu(n+T)\ [\text{since } \mu(n) \text{ is decreasing}]$$
$$\le\exp[-ka\beta T\mu(n+T)]$$

where $\beta=\inf_n\beta(n)<0$. Q.E.D.

Appendix II

From (2-7), with $\mu(n)=\frac{a}{n}$ and d=0, we have the following LMS algorithm in terms of weight error vector Y(n):

$$Y(n+1)=\left[\frac{I-a}{n}X(n)X(n)^T\right]Y(n)+\frac{a}{n}\{\mu(n)-X(n)^TW_{opt}\}X(n).$$

In this appendix we show that $Y(n)\to 0\ as\ n\to\infty$ assuming (A1), (A2) and (A3). Using the transition matrix, (3-7), the solution of the above is given by

$$Y(n+1)=\Psi(n+1,1)Y(1)+a\sum_{i=1}^{n}\frac{\Psi(n+1,i+1)}{i}\{\mu(i)X(i)-X(i)X(i)^TW_{opt}\}$$

Taking the norm of the above yields, for any n_0,

$$\| Y(n+1)\|\le\|\Phi(n+1,1)\|\,\| Y(1)\|+$$
$$a\left\|\sum_{i=1}^{n}\Phi(n+1,\frac{i+1)}{i}\{u(i)X(i)-X(i)X(i)^TW_{opt}\}\right\|$$
$$\le\|\Phi(n+1,1)\|\,\| Y(1)\|+ \quad \text{(II-1)}$$
$$a\left\|\sum_{i=1}^{n_0-1}\frac{\Phi(n+1,i+1)}{i}\{u(i)X(i)-X)i)X(i)^TW_{opt}\}\right\|$$
$$+a\left\|\sum_{i=n_0}^{n}\frac{\Phi(n+1,i+1)}{i}\{u(i)X(i)-X)i)X(i)^TW_{opt}\}\right\|$$

Note that the first term in (II-1) converges to zero as $n\to\infty$ by Lemma 1, and the second term also converges to zero for any fixed n_0 because it contains a finite number of terms, each of which converges to zero by Lemma 1. To complete the proof, it only remains to be demonstrated that the third term in (II-1) can be upper bounded by an arbitrarily small number by choosing n_0 sufficiently large. To this end, we introduce

$$T(n)=\sum_{i=1}^{n}\{u(i)X(i)-X(i)X(i)^T W_{opt}\} \qquad \text{(II-2)}$$

Then from (A2) $\lim_{n\to\infty}\frac{\| T(n)\|}{n}=0$ a.s. (II-3)

Hence, for any $\delta > 0$ there exists N_1 such that for $n > N_1$

$$\frac{\| T(n)\|}{n} < \delta. \qquad \text{a.s.} \qquad \text{(II-4)}$$

Now choose $n_0 > max(N_0, N_1(\omega))$ where N_0 is the integer lower bound on n_0 required for the exponential bound on $\|\Phi(n,n_o)\|$ of Lemma 1. To proceed, for notational simplicity, let us define

$$LHS \equiv \left\| \sum_{i=n_o}^{n} \frac{\Phi(n+1,i+1)}{i}\{u(i)X(i)-X(i)X(i)^T W_{opt}\}\right\| \qquad \text{(II-5)}$$

which, using (II-2), can be expressed as

$$LHS = \left\| \sum_{i=n_0}^{n} \frac{\Phi(n+1,i+1)}{i}\{T(i)-T(i-1)\}\right\|$$

At this point applying Abel's partial summation formula for a sequence

$$\sum_{i=m}^{n} a(i)\{S(i)-S(i-1)\} = \sum_{i=m-1}^{n-1}\{a(i)-a(i+1)\}S(i) + a(n)S(n)-a(m-1)S(m-1),\ S(0)=0 \qquad \text{(II-6)}$$

we obtain

$$LHS = \left\| \sum_{i=n_0-1}^{n-1}\left\{\frac{\Phi(n+1,i+1)}{i}-\frac{\Phi(n+1,i+2)}{i+1}\right\}T(i) + \frac{\Phi(n+1,n+1)T(n)}{n}-\frac{\Phi(n+1,n_0)T(n_0-1)}{n_0-1}\right\| \qquad \text{(II-7)}$$

Now, using the definition of the transition matrix

$$\Psi(n+1,i+1)=\Psi(n+1,i+2)\Psi(i+2,i+1),$$

$$\Psi(i+2,i+1)=Q(i+1)=I-\frac{a}{i+1}X(i+1)X(i+1)^T,$$

and using (II-4), for $n_0 > N_1$ we obtain the bound on (II-7) as

$$LHS \le \left\| \sum_{i=n_0-1}^{n-1}\left[\frac{1}{i}\left\{I-\frac{a}{i+1}X(i+1)X(i+1)^T\right\} - \frac{I}{i+1}\right]\Psi(n+1,i+2)T(i)\right\| + \delta + \varepsilon\,\delta$$

[where $\varepsilon = \|\Psi(n+1,n_0)\|$]

$$\le \left\| \sum_{i=n_0-1}^{n-1}\left(\frac{1}{i}-\frac{1}{i+1}\right)\Psi(n+1,i+2)T(i)\right\|$$

$$+a\left\| \sum_{i=n_0-1}^{n-1}\frac{1}{i(i+1)}X(i+1)X(i+1)^T\Psi(n+1,i+2)T(i)\right\| + (1+\varepsilon)\delta$$

$$\le \sum_{i=n_0-1}^{n-1}\frac{\|\Psi(n+1,i+2)\|\ \|T(i)\|}{i+1} + aK\sum_{i=n_0-1}^{n-1}\frac{\Psi(n+1,i+2)\|}{i+1}\frac{\|T(i)\|}{i} + (1\varepsilon)\delta$$

$$\le \left\{(1+aK\sum_{i=n_0-1}^{n-1}\frac{\Phi(n+1,i+2)\|}{i+1} + 1+\varepsilon\right\}\delta$$

Finally, due to Lemma 1

$$\sum_{i=n_0-1}^{n-1}\frac{\|\Phi(n+1,i+2)\|}{i+1} \le C < \infty, \forall n$$

Therefore

$Y(n)\to 0$, i.e., $W(n)\to W_{opt}$ as $n\to\infty$

References

[1] D. D. Falconer, "Adaptive Reference Echo cancellation," IEEE Trans. Comm., Vol. COM-30, pp. 2083-2094, Sept. 1982.

[2] K. Konstantinides and N. Kanopoulos, "Design alternatives for adaptive digital lattice filters and a new bit-serial architecture, "IEEE Trans. Circuits and Systems, Vol. CAS-34, pp. 737-742, July 1987.

[3] H. Herzberg, R. Haimi-Cohen, and Y. Be'ery, "A systolic Array Realization of an LMS Adaptive Filter and the Effects of Delayed Adaptation," IEEE Trans. Signal Processing, Vol. 40, No. 11, pp. 2799-2803, Nov. 1992.

[4] P. Kabal, "The Stability of Adaptive Minimum Mean Square Error Equalizers Using Delayed Adjustment," IEEE Trans. Comm., Vol. 31, No. 3, pp. 430-432 Mar. 1983.

[5] G. Long, F. Ling, and J. Proakis, "The LMS Algorithm with Delayed Coefficient

Adaptation," IEEE Trans. ASSP. Vol. 37. No. 9. pp. 1397-1405. Sept. 1989.

[6] G. Long. F. Ling. and J. Proakis. "Corrections to 'The LMS Algorithm with Delayed Coefficient Adaptation.'" IEEE Trans. Signal Processing. Vol. 40. pp. 230-232. Jan. 1992.

[7] R. Haimi-Cohen. H. Herzberg. and Y. Be'ery. "Delayed Adaptive LMS Filtering: Current Results." Proc. IEEE Int. Conf. ASSP. pp. 1273-1276. Apr. 1990.

[8] Sang-Sik Ahn. Peter J. Voltz. "Almost-Sure Convergence of the Non-homogeneous DNLMS Algorithm with Decreasing Step Size." Proc. IEEE Int. Conf. ASSP. Vol. 3 pp. 401-404. April 1994.

[9] E. Eweda and O. Macchi. "Convergence of an adaptive linear estimation algorithm." IEEE Trans. Automatic Control Vol. 7. No. 5. pp. 435-442. 1986.

[10] A. Krieger and E. Masry. "Convergence analysis of adaptive linear estimation for dependent stationary process." IEEE Trans. Information Theory. Vol. 34 No. 4. pp. 642-654. 1988.

[11] A. Weiss and D. Mitra. "Digital Adaptive Filters: Conditions for Convergence. Effects of Noise and Errors Arising from the Implementation." IEEE Trans. on Information Theory. Vol. 25. No. 6. pp. 637-652 Nov. 1979.

[12] P.Voltz. "Sample Convergence of the Normed LMS Algorithm with Feedback Delay." Proc. IEEE Int. Conf. ASSP. pp 2129-2132. May 1991.

[13] S. Haykin. "Adaptive Filter Theory." Prentice Hall. Englewood Cliffs. NJ. 1991.

저 자 소 개

安 相 植(正會員)

1959년 11월 15일생. 1983년 2월 고려대학교 전자공학과(공학사). 1985년 2월 고려대학교 전기공학과(공학석사). 1984년 12월 ~ 1987년 8월 금성 중앙연구소(주임연구원). 1992년 1월 Polytechnic Univ. 전기공학과(M. S.). 1991년 1월 ~ 1993년 6월 Polytechnic Univ.(Teaching Fellow). 1994년 1월 Polytechnic Univ. 전기공학과 (Ph.D.). 1994년 2월 ~ 1995년 2월 금성중앙연구소 (책임연구원). 1995년 3월 ~ 현재 고려대학교 응용전자공학과 (조교수) 주관심분야는 DSP 알고리즘. 적응시스템. VLSI 구현 등임.