DELAYED ADAPTIVE LMS FILTERING: CURRENT RESULTS

R. Haimi-Cohen[1], H. Herzberg[2] and Y. Be'ery[2]

1. The DSP Group (ISRAEL) Ltd., 18 Ben-Gurion st., Giv'at Shmuel, ISRAEL.
2. Department of Electronic Systems, Tel-Aviv University, Tel-Aviv, ISRAEL.

ABSTRACT

In some practical situations the coefficients adaptation of LMS adaptive filters can be done only after a fixed delay. The resulting algorithm, known as delayed LMS (DLMS), may behave quite differently than the LMS unless the step size is kept below a certain threshold. This work presents conditions for convergence and estimates of convergence rate, both for the mean of the DLMS filter coefficients and for its excess mean square error. The results provide an insight into the manner of operation of the DLMS adaptive filter and may be used in the design of such filters.

1. INTRODUCTION

In some practical situations, the adaptation step in the LMS algorithm can be performed only after a fixed delay. For example, in systolic array implementation the filter output is generated after a delay which equals to the number of taps [1][2]; in decision directed adaptive equalization, the desired signal is available only after the delay introduced by the receiver in making its decision [3]. In such applications the implemented algorithm is often a modified version of the LMS algorithm, known as the Delayed LMS (DLMS) [3-4], in which coefficient adaptation is performed after a delay. This work presents the main results of our analysis of the convergence and the steady state behavior of the DLMS algorithm, with the attempt of providing useful insight which may be helpful in the design of such filters. Section 2 defines the problem and presents some basic definitions. Conditions for convergence, convergence rate and limits are discussed: In section 3 for the mean of the filter coefficients and in section 4 for the excess mean square error (EMSE). Section 5 concludes this report with some remarks about the implications of the results on the design of DLMS adaptive filters. A full account of this work, including proofs, will be published elsewhere [5]. Some previously reported results may be found in [1][3-4][6].

2. PROBLEM FORMULATION

The DLMS algorithm for an N-taps adaptive filter is defined by the following equations:

$$r(t) = g(t) - \underline{x}(t)^T\underline{w}(t) \qquad (1)$$

$$\underline{w}(t+1) = \underline{w}(t) + \mu\underline{x}(t-d)r(t-d) \qquad (2)$$

Where $\underline{w}(t)$ is the N-dimensional vector containing the coefficients of the adaptive filter at time t, $\underline{x}(t)$ is the N-dimensional reference signal at time t, g(t) is the desired signal. r(t) is the residual (prediction error), μ is the step size (a small positive number), and d is the fixed delay (a positive integer). Note that if d=0, Eq. (1)-(2) define the LMS algorithm.

In the analysis of the DLMS algorithm we assume that the sequence of N+1 dimensional vectors $[\underline{x}(t)^T;g(t)]^T$, t=0,1,... is a sequence of independent, zero mean, identically distributed random vectors. The independence assumption is generally false, but it is a conventional useful approximation in the analysis of adaptive filters [7]. An immediate consequence of this assumption and eq. (2) is that $\underline{x}(t),\underline{x}(t-d)$ are independent of $\underline{w}(t),\underline{w}(t-d)$. We also assume that $R=E\{\underline{x}(t)\underline{x}(t)^T\}$ is positive definite. The analysis of the EMSE involves higher moments of $\underline{x}(t)$, hence further assumptions should be made regarding the distribution of $\underline{x}(t)$. We consider here two important special cases: (1) $\underline{x}(t)$ is Gaussian (not necessarily white); (2) $\underline{x}(t)$ is white binary (WB), i.e. the elements of $\underline{x}(t)$ are independent and get each of the values {1, -1} with probability 0.5. Let

$$J(\underline{w}) = E\{[g(t) - \underline{x}(t)^T\underline{w}]^2\} \qquad (3)$$

$J(\underline{w})$ has a single (global) minimum at the Wiener filter coefficients $\underline{w}^*=R^{-1}\underline{p}$, where $\underline{p}=E\{\underline{x}(t)g(t)\}$. Let $\phi(t)=g(t)-\underline{x}(t)^T\underline{w}^*$ be the output of this filter. Then $E\{\phi(t)^2\}=J(\underline{w}^*)$. The excess mean square error (EMSE) of the adaptive filter at time t is the difference between the actual power of the residual to the minimal obtainable one:

CH2847-2/90/0000-1273 $1.00 © 1990 IEEE

$$\begin{aligned}EMSE(t) &= E\{[g(t)-\underline{x}(t)^T\underline{w}]^2\}-E\{\phi(t)^2\}\\ &= E\{(\underline{w}(t)-\underline{w}^*)^T\underline{xx}^T(\underline{w}(t)-\underline{w}^*)\\ &= tr\{RE\{(\underline{w}(t)-\underline{w}^*)(\underline{w}(t)-\underline{w}^*)^T\}\}\end{aligned} \quad (4)$$

We are interested in the convergence properties of $E\{\underline{w}(t)\}$ and EMSE(t). The analysis is performed in a translated and rotated space [7]. Let $R=QMQ^T$, where $M=diag(m_1,\ldots,m_N)$, $m_1 \geq m_2 \geq \ldots \geq m_N > 0$ and the columns of Q are a set of orthonormal eigenvectors of R, corresponding to the eignevalues $m_1,\ldots,m_N$. We define $\underline{u}(t)=Q^T(\underline{w}(t)-\underline{w}^*)$, $\underline{y}(t)=Q^T\underline{x}(t)$. Using these definitions and substituting eq. (1) into eq. (2) we get:

$$\underline{u}(t+1) = \underline{u}(t) - \mu\underline{y}(t-d)\underline{y}(t-d)^T\underline{u}(t-d) + \mu\underline{y}(t-d)\phi(t-d) \quad (5)$$

Using the same substitution, eq. (4) becomes:

$$EMSE(t) = tr\{ME\{\underline{u}(t)\underline{u}^T(t)\}\} \quad (6)$$

Eqs. (5-6) will be the basis for the starting point for the analysis in the following sections.

3. CONVERGENCE OF THE COEFFICIENTS MEAN

Taking the expectation of both sides of eq. (5) we get:

$$E\{\underline{u}(t+1)\} - E\{\underline{u}(t)\} + \mu ME\{\underline{u}(t-d)\} = \mu E\{y(t-d)\phi(t-d)\} = Q^T[\underline{p} - R\underline{w}^*] = 0 \quad (7)$$

Since M is a diagonal matrix the recurrence for each entry in $\{\underline{u}(t)\}$ is independent of the other entries. Hence, the solution to eq. (7) is given by:

$$E\{u_i(t)\} = \sum_{j=1}^{d+1} c_{ij}z_{ij}^t \qquad i=1,\ldots,N \quad (8)$$

where z_{ij}, $j=1,\ldots,d+1$ are the roots of the characteristic equations (the roots are simple except for some singular cases discussed below):

$$z^d(1-z) = \mu m_i \qquad i=1,\ldots,N \quad (9)$$

Therefore convergence and convergence behavior of $E\{u_i(t)\}$ depend on the location of the roots of eq. (9) and particularly on their maximal magnitude. The properties of these roots are summarized in the following lemma:

<u>Lemma 1</u>: Consider the equation:

$$z^d(1-z) = \sigma \qquad (d>0) \quad (10)$$

if z_i, z_j are roots of eq. (10) and $|arg(z_i)|<|arg(z_j)|$ $(|arg(z)|\leq\pi)$ then $|z_i|>|z_j|$. The roots of maximal magnitude of eq. (10) are inside the unit circle if and only if:

$$0 < \sigma < T_0 \equiv 2\sin[\pi/(4d+2)] \quad (11)$$

in which case the absolute value of the argument of these roots is less than $\pi/(2d+1)$. Eq. (10) has real positive roots if and only if

$$\sigma \leq T_1 \equiv d^d/(d+1)^{d+1} \quad (12)$$

If eq. (12) holds and $\sigma>0$ there are two positive roots x_1, x_2 which satisfy:

$$\sigma^{1/d}<x_2<[\sigma(d+1)]^{1/d}<d/(d+1)<x_1<1-\sigma \quad (13)$$

and

$$\lim_{\sigma\to 0} \frac{x_1}{1-\sigma} = 1 \quad (14)$$

The roots of eq. (10) are always simple except for the trivial case $\sigma=0$ and when $\sigma=T_1$, in which case $x_1=x_2$ is a double root. The maximal magnitude of the roots is a decreasing function of σ when $\sigma<T_1$, and an increasing function of σ when $\sigma>T_1$.

Lemma 1 holds also for the LMS case (d=0) with $T_0=2$, $T_1=d^d|_{d=0}=1$. Denote the maximum magnitude root of Eq. (9) by z_i, $i=1,\ldots,N$ (if there is a conjugate pair of such roots z_i is the one with a non-negative imaginary part). The values of $(1-|z_i|)(d+1)$ as a function of σ/T_0 are shown in fig. 1 for several values of d. It is clear that after the normalization of σ by T_0 and $1-|z_i|$ by (d+1), the result is relatively insensitive to changes in d, except for the sharp change between d=0 and $d\neq0$. Using lemma 1 we can summarize the convergence properties of the mean of the coefficients:

<u>Theorem 1</u>: $\underline{u}(t)=Q^T[\underline{w}(t)-\underline{w}^*]$ converges if and only if $0<\mu m_1<T_0$, in which case the limit of $\underline{u}(t)$ is 0. Convergence of each of the components of $\underline{u}(t)$ is linear with a convergence ratio of $|z_i|$, $i=1,\ldots,N$. The maximal possible convergence ratio is d/(d+1) which is obtained when $\mu m_i=T_1$. If $\mu m_i>T_1$ the convergence of u_i is oscillatory with oscillation period of at least 2(d+1). If $\mu m_i<T_1$ the convergence of u_i is non-oscillatory with a convergence ratio smaller than $1-\mu m_i$, but as $\mu m_i\to0$, the ratio between this expression and the actual convergence ratio tends to unity.

The necessary and sufficient condition for convergence was first reported by Kabal [3]. The term "non-oscillatory" applies only as t becomes large; in the beginning of the convergence some oscillations will be caused by the none positive roots of eq. (9), but these oscillations disappear as convergence proceeds, since their convergence ratio is smaller than the convergence ratio of the sequence: If $\mu m_i = \epsilon T_1$, $0<\epsilon<1$ then for any $1<j\le d+1$, $|z_{ij}|^d/|z_i|^d<\epsilon$. We note that for moderately large d, $T_0 \approx \pi(2d+1)^{-1}$ and $T_1 \approx e^{-1}(d+1)^{-1}$. Hence the non-oscillatory convergence range is less than a quarter of the total convergence range, because

$$\lim_{d\to\infty} T_1/T_0 = 2(\pi e)^{-1} \approx 0.234 \tag{15}$$

If d=0 then $z_i = \mu m_i$ and the theorem states the well known results about the LMS algorithm. Note that the LMS case is special in the sense that when $\mu m_i = T_1$ convergence is completed in one step (convergence ratio = 0) for the corresponding component, while as in the DLMS the convergence ratio is at least $d^d/(d+1)^{d+1}$. The convergence rate of $\underline{w}(t)$ is determined by the element of $\underline{u}(t)$ with the largest convergence ratio. Therefore the fastest convergence is obtained when μ is selected so that T_1 is located between the μm_N and μm_1. However, setting $\mu m_1 > T_1$ would cause a very high EMSE and undesirable oscillations in the convergence. Therefore, for practical purposes, μ should be selected so that $\mu m_1 < T_1$, in which case $d/(d+1)<z_1<\ldots<z_N<1$.

4. CONVERGENCE OF THE EMSE

In the analysis of the EMSE we assume that $\mu m_1 \ll T_1$, because otherwise the EMSE is too large for most applications. Let $O(t)=E\{\underline{u}(t)\underline{u}(t)^T\}$. From eq. (5) we get:

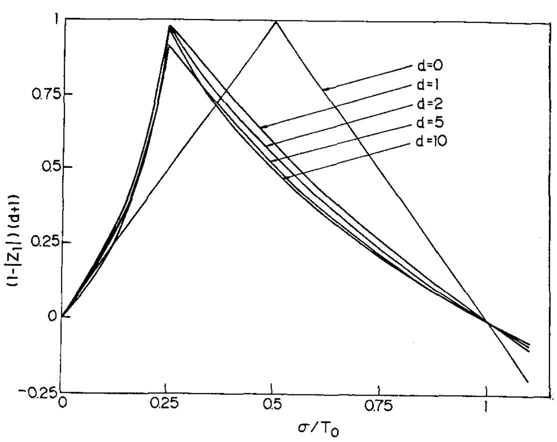

fig. 1: $(1-|z_1|)(d+1)$ versus σ/T_0.

$$O(t+1) = O(t) - \mu[E\{\underline{u}(t)\underline{u}(t-d)^T\}M + ME\{\underline{u}(t-d)\underline{u}(t)^T\}] + \mu^2 C(t-d) + \mu^2 ME\{\phi(t-d)^2\} + \mu E\{\phi(t-d)[\underline{y}(t-d)(E\{\underline{u}(t)\}-\mu\underline{y}(t-d)\underline{y}(t-d)^T E\{\underline{u}(t-d)\})^T + (E\{\underline{u}(t)\}-\mu\underline{y}(t-d)\underline{y}(t-d)^T E\{\underline{u}(t-d)\})\underline{y}(t-d)^T]\} \tag{16}$$

where (the index t is omitted for clarity):

$$C_{ij} = \sum_{k=1}^{N}\sum_{l=1}^{N} O_{kl}E\{y_i y_j y_k y_l\} \tag{17}$$

C(t-d) is a complicated expression involving the fourth moments of $\underline{x}(t-d)$. However, in the Gaussian and the WB cases it gets the following simple form:

$$C_{ij}(t) = \begin{cases} tr\{MO(t)\} + \alpha O_{ii}(t)m_i m_i & i=j \\ 2m_i m_j O_{ij}(t) & i\neq j \end{cases} \tag{18}$$

where $\alpha=2$ for the Gaussian case and $\alpha=0$ for the WB case. In order to compute $E\{\underline{u}(t)\underline{u}(t-d)^T\}$, let,

$$E\{\underline{u}(t)\underline{u}(t-d)^T\} = E\{E\{\underline{u}(t)\underline{u}(t-d)^T|\underline{u}(t-d)\}\}$$

$$= E\{E\{\underline{u}(t)|\underline{u}(t-d)\}\underline{u}(t-d)^T\} \tag{19}$$

Since we assumed that $\mu m_1 \ll T_1$ the convergence of $E\{\underline{u}(t)\}$ depends mainly on the maximum magnitude roots of eq. (9). Hence we may use the results of the previous section and approximate $E\{\underline{u}(t)|\underline{u}(t-d)\}$ by $Z^d\underline{u}(t-d)$ where $Z=diag(z_1,\ldots z_N)$. Therefore, by (19),

$$E\{\underline{u}(t)\underline{u}(t-d)^T\} = Z^d O(t-d) \tag{20}$$

The last term in eq. (16) contains $\underline{u}(t-d)$, $\underline{u}(t)$, which converge to zero, as multiplicative factors. Therefore this term may be ignored as t gets large. Using these results and some algebra eq. (16) is transformed into

$$O(t+1) = O(t) - \mu[Z^d O(t-d)M + MO(t-d)Z^d] + \mu^2[C(t-d) + \Phi M] \tag{21}$$

where $\Phi=E\{\phi(t)^2\}$.

<u>Theorem 2</u>: If the EMSE(t) converges, its limit is given by:

$$EMSE(\infty) = \frac{\mu\, tr\{MZ^{-d}\}}{2-\mu[tr\{MZ^{-d}\}+\alpha\beta]}\Phi = \frac{tr\{I-Z\}}{2-\alpha\beta - tr\{I-Z\}}\Phi \tag{22}$$

where $1-z_N \le \beta \le 1-z_1$.

In the Gaussian case (where $\alpha \neq 0$), the expression for the steady state EMSE provides only upper and lower bounds, since it depends on the unknown value of β. However, in most practical situations the difference between these bounds is very small (since $0<1-z_N \leq \beta \leq 1-z_1<(d+1)^{-1}$) and the conservative approximation $\beta=(d+1)^{-1}$ is usually sufficient.

Let $\underline{o}(t)=(o_{11}(t),\ldots,o_{NN}(t))^T$ and $\underline{m}=(m_1,\ldots,m_N)^T$ be the vectors consisting of the diagonal elements of O(t) and M respectively. Define

$$B = 2\mu Z^d M - \mu^2 \alpha M^2 - \mu^2 \underline{m}\underline{m}^T \quad (23)$$

Then by eq. (21)

$$\underline{o}(t+1) = \underline{o}(t) - B\underline{o}(t-d) \quad (24)$$

Let $B=FDF^T$ where F is unitary ($FF^T=I$) and $D=\mathrm{diag}(\delta_1,\ldots,\delta_1)$ where $\delta_1 \geq \ldots \geq d_N \geq 0$ are the eigenvalues of B. Using lemma 1 and the same type of analysis that was done for $E\{\underline{w}(t)-\underline{w}^*\}$, we conclude that:

$$(F^T\underline{o}(t))_i = \sum_{j=1}^{d+1} h_{ij}\Omega_{ij}^t \quad i=1,\ldots,N \quad (25)$$

where h_{ij} are constant coefficients and $\Omega_{i1} \geq \ldots \geq \Omega_{i(d+1)}>0$ are the solutions to eq. (10) with $\sigma=\delta_i$. $\underline{o}(t)$ converges if and only if all the eigenvalues of B are positive and less than T_0. Its convergence is non-oscillatory if and only if the maximal magnitude eigenvalue of B is less than T_1. The following theorem states when these conditions are satisfied:

<u>theorem 3</u>: A necessary condition for EMSE(t) to converge is that

$$tr\{I-Z\} < 2 - \alpha(1-z_N) \quad (26)$$

A sufficient condition for EMSE(t) to converge is that

$$tr\{I-Z\} < 2 - \alpha(1-z_1) \quad (27)$$

If convergence exists, a sufficient condition for it to be non-oscillatory (in the sense of theorem 1) is that

$$2\mu m_1 \leq T_1 \quad (28)$$

In the WB case $\alpha=0$ hence the necessary and the sufficient conditions (26) and (27) coincide. Even in the Gaussian case the difference between them is small since $0 \leq 1-z_N \leq 1-z_1<(d+1)^{-1}$ and both $\alpha(1-z_1)$ and $\alpha(1-z_N)$ vanish as $\mu m_1 \to 0$. Note that these conditions are automatically satisfied whenever $N<2(d+1)-\alpha$ (i.e. they are implied by the requirement $\mu m_1<T_1$). The convergence of EMSE(t) is linear with a convergence ratio which depends on the eigenvalues of B. B is dominated and very close to the matrix $2\mu M$, hence the eigenvalues of B are near but smaller than $2\mu m_i$, i=1,...,N. Therefore the convergence rate of the EMSE is linear with a convergence rate which is approximately twice the convergence rate of the mean of the coefficients, as is also the case in the LMS algorithm. The theorem is proved under the assumption that $\mu<<T_1$ [5], however, using it for a qualitative analysis we see that as μ is increased the convergence rate reaches a maximum when $\delta_i=T_1$ (with $\Omega_{i1}=d/(d+1)$), and for larger values of μ the convergence rate decreases and becomes oscillatory.

5. CONCLUSION

The behavior of the DLMS algorithm was summarized in theorems 1-3. These theorems show the major role of the delay dependent parameter T_1 in the analysis of the behavior of the algorithm. For obtaining a practically acceptable EMSE it is necessary that $\mu m_1<<T_1$. Under this condition theorems 1-3 are immediate generalizations of the theory of LMS filtering, with μm_i replaced by $1-z_i$, whereas in the LMS case these two values are equal. Note also that when d is sufficiently large so that $N<2(d+1)-\alpha$, the conditions of theorem 3 for convergence of the EMSE are automatically satisfied whenever the basic condition for convergence of the mean of the coefficients, $\mu m_i<T_1$, is satisfied.

REFERENCES

1. H. Herzberg, R. Haimi-Cohen and Y. Be'ery, "A systolic-array for adaptive filters: Design and performance analysis". Proc. 15-th conference of IEEE in Israel, paper no. 4.1.3, pp. 1-4, Apr. 1987.
2. H. Herzberg, R. Haimi-Cohen and Y. Be'ery, "A systolic-array realization of adaptive filters and the effects of delayed adaptation", submitted to IEEE Trans. Acoustics, Speech and Signal Processing.
3. P. Kabal, "The stability of adaptive minimum mean square error equalizers using delayed adjustment", IEEE Trans. Comm., vol. COM-31. pp. 430-432, 1983.
4. G. Long, F. Ling and J.G. Proakis, "Adaptive transversal filters with delayed coefficient adaptation", Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing, 1987.
5. R. Haimi-Cohen, H. Herzberg and Y. Be'ery, "LMS adaptive filtering with delayed coefficients adaptation", submitted to IEEE Trans. Acoustics, Speech and Signal Processing.
6. G. Long, F. Ling and J.G. Proakis, "The LMS algorithm with delayed coefficient adaptation", IEEE Trans. Acoustics, Speech, Signal Processing, vol. 37(9), pp.1397-1405, 1989.
7. S. Haykin, Adaptive Filter Theory, ch. 5, Prentice Hall, 1986.