ON THE CONVERGENCE AND MSE OF CHEN'S LMS ADAPTIVE ALGORITHM

Sau-Gee Chen, Yung-An Kao and Ching-Yeu Chen

Department of Electronics Engineering and Institute of Electronics
National Chiao Tung University
E-mail: sgchen@cc.nctu.edu.tw

ABSTRACT

The recently proposed Chen's LMS algorithm [1] costs only half multiplications that of the conventional direct-form LMS algorithm (DLMS). Despite of the merit, the algorithm lacked rigorous theoretical analysis. This work intends to characterize its properties and conditions for mean and mean-square convergences. Closed-form MSE are derived, which is slightly larger than that of DLMS algorithm. It is shown, under the condition that the LMS step size μ is very small and an extra compensation step size α is properly chosen, Chen's algorithm has comparable performance to that of the DLMS algorithm. For the algorithm to converge, a tighter bound for α than before is also derived. The derived properties and conditions are verified by simulations.

1. INTRODUCTION

The direct-form LMS algorithm (DLMS) [2], [3] is the most popular temporal-domain direct-form adaptive filtering algorithm due to its simplicity and robustness. Regarding the temporal-domain direct-form approaches, there exists many LMS variants in reducing the coefficient updating complexities such as the sign-error, sign-input and zero-forcing algorithms.

However, little improvements were done in reducing its filtering complexities. Recently, a so-called FELMS algorithm [3] was proposed to retain the same convergence properties as those of DLMS while reducing both filtering and updating complexities of LMS algorithm.

More recently, Chen et al. proposed a new LMS adaptive filtering algorithm [1] which has close to 50% reduction in filtering multiplications. Moreover, the algorithm can be combined with the FELMS algorithm in reducing its coefficient updating complexities. Despite the merits, the algorithm's properties have not been fully addressed.

Here, the properties of the convergence both in the mean and in the mean square are investigated in detail, verified by both MATLAB and C simulations. It is shown, under the condition that the LMS step size μ is very smaller and an extra compensation step size α is properly chosen, Chen's algorithm has comparable performance to that of the DLMS algorithm.

The paper is organized as follows. In the second section, the Chen's algorithm will be reviewed, followed by its stability analysis in the third section. The third section covers the issues of mean and mean-square convergence of the weights and an extra compensation factor, convergence bound for α and excess MSE of the algorithm. The final section draws a conclusion on issues to be further investigated.

2. REVIEW OF THE CHEN'S LMS ALGORITHM

Given an adaptive filter with its zero-mean input sequence $x(n)$, zero-mean desired signal $d(n)$ and coefficients $w_k(n)$'s to be adapted. The Chen's LMS algorithm is as shown below:

$$\begin{aligned} y'(n) &= \sum_{k=0}^{N/2-1}\{[x(n-2k)-w_{2k+1}(n)][x(n-2k-1) \\ &\quad + w_{2k}(n)]\} - h_N(n) - P(n) \\ &= \sum_{k=0}^{N-1} x(n-k)w_k(n) - [h_N(n) - C(n)] \\ &= y(n) - [h_N(n) - C(n)] \end{aligned} \quad , (1)$$

$$w_k(n+1) = w_k(n) + 2\mu e'(n)x(n-k), \quad k = 0, 1, \cdots, N-1, \quad (2)$$

$$h_N(n+1) = h_N(n) - \alpha e'(n), \quad (3)$$

where $C(n) = \sum_{k=0}^{N/2-1} w_{2k}(n)w_{2k+1}(n)$,

Copyright 1997 IEEE

$$\sum_{k=0}^{N/2-1} x(n-2k)x(n-2k-1) \equiv P(n)$$
$$= P(n-2) + x(n)x(n-1) - x(n-N)x(n-N-1) \ ,$$

$$\begin{aligned} e'(n) &= d(n) - y'(n) \\ &= d(n) - y(n) + [h_N(n) - C(n)] \\ &= e(n) + [h_N(n) - C(n)] \quad , \end{aligned} \tag{4}$$

and $x(n)=0$, $P(n)=0$ for $n<0$. It can be shown that the complexity for the convolution is $N/2+1$ multiplications and $3N/2+3$ adds.

3. STABILITY ANALYSIS OF THE ALGORITHM

For convenience, some definitions and notations are defined as follows.

$\mathbf{w}(n) = [w_0(n), w_1(n), \cdots, w_{N-2}(n), w_{N-1}(n)]^T_{1\times N}$
= the tap weight vector,

$\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^T_{1\times N}$
= the filter input vector,

$\mathbf{w}^* = [w_0^*, w_1^*, \cdots, w_N^*]^T_{1\times N}$
= the Wiener solution,

$\xi_w(n) = w(n) - w^*$.

It is also emphasized that the *independence assumption* and the *independence theory* [2] are adopted here. That is: $h_N(n+1)$, $\mathbf{w}(n+1)$ are independent of both $\mathbf{x}(n+1)$ and $d(n+1)$, and they depend only on four inputs: (1) The previous sample vectors of the input process, $\mathbf{x}(n)$, $\mathbf{x}(n-1)$, $\cdots$, $\mathbf{x}(1)$; (2) The previous samples of the desired response, $d(n)$, $d(n-1)$, $\cdots$, $d(1)$; (3) The initial value of the tap vector, $w(0)$; and (4) The initial value of extra coefficient, $h_N(0)$. Therefore

$$E[\mathbf{x}(n)\mathbf{x}^T(k)] = \mathbf{0}, \ k = 0,1,\cdots,n-1,$$
$$E[x(n)d(k)] = 0, \ k = 0,1,\cdots,n-1.$$

(a) The convergence of w(*n*):

It can be shown that

$$\begin{aligned} \xi_\mathbf{w}(n+1) &= (\mathbf{I} - \mu\mathbf{x}(n)\mathbf{x}^T(n))\xi_\mathbf{w}(n) \\ &+ \mu\mathbf{x}(n)e_{opt}(n) + \mu\mathbf{x}(n)[h_N(n) - C(n)], \end{aligned} \tag{5}$$

where $e_{opt}(n)=d(n)-\mathbf{w}^{*T}\mathbf{x}(n)$.

Here, by invoking the *direct-averaging method* described by Kushner [2], [5] and assuming a small step size μ, the weight error equation can be replaced by the following stochastic difference equation:

$$\begin{aligned} \xi_\mathbf{w}(n+1) &= (\mathbf{I} - \mu\mathbf{R})\xi_\mathbf{w}(n) + \mu\mathbf{x}(n)e_{opt}(n) \\ &+ \mu\mathbf{x}(n)[h_N(n) - C(n)] \quad , \end{aligned} \tag{6}$$

where $\mathbf{R}= E[\mathbf{x}(n)\mathbf{x}^T(n)]$ is the correlation matrix of $\mathbf{x}(n)$. Then, based on the mentioned independence assumption and orthogonal property $E[x(n)e_{opt}(n)]=0$, we have

$$\begin{aligned} &E[\xi_\mathbf{w}(n+1)] \\ &= (\mathbf{I} - \mu\mathbf{R})E[\xi_\mathbf{w}(n)] + \mu E[\mathbf{x}(n)]E[h_N(n) - C(n)] \\ &= (\mathbf{I} - \mu\mathbf{R})E[\xi_\mathbf{w}(n)] \quad . \end{aligned} \tag{7}$$

The convergence equation is the same as that of DLMS, and the properties of weight in mean will be same as that of DLMS accordingly. So is the step size μ has the constraint of $0<\mu<2/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of the correlation matrix **R**.

(b) The Convergence of $h_N(n)$

Let
$$\begin{aligned} \xi_{hc}(n+1) &= h_N(n+1) - C(n+1) \\ &= h_N(n) - C(n+1) - \alpha[d(n) - \\ &\quad \mathbf{w}(n)^T\mathbf{x}(n) + h_N(n) - C(n)] \ . \end{aligned} \tag{8}$$

For very small μ or large n, $C(n)$ is varying slowly so that we can assume that $C(n+1)\approx C(n)$, then

$$\begin{aligned} \xi_{hc}(n+1) &= \xi_{hc}(n) - \alpha[d(n) - \mathbf{w}^T(n)\mathbf{x}(n) + \xi_{hc}(n)] \\ &= (1-\alpha)\xi_{hc}(n) - \alpha[d(n) - \mathbf{w}^T(n)\mathbf{x}(n)] \ . \end{aligned}$$

Therefore

$$\begin{aligned} E[\xi_{hc}(n+1)] &= (1-\alpha)E[\xi_{hc}(n)] \\ &= (1-\alpha)^{n+1}E[\xi_{hc}(0)] \ . \end{aligned} \tag{9}$$

For the convergence of $h_N(n)$, it is required that $|1-\alpha| < 1$ such that $E[\xi_{hc}(n)] = 0$ as $n \to \infty$. This rough bound can be further tightened later.

(c) The Convergence in the Mean Square

With the independence assumption [2], [5] the MSE $J'(n)=E[e'(n)e'(n)]$ of Chen's algorithm can be reduced to

$$J'(n) = J(n) + E[\xi_{hc}^2(n)], \ J(n) \equiv E[e^2(n)] \ .$$

Next, by applying the orthogonal property, $J(n)$ can be reduced to:

$$\begin{aligned} J(n) &= E[(e_{opt}(n) - \xi_\mathbf{w}^T(n)\mathbf{x}(n))^2] \\ &= J_{\min} + E[\xi_\mathbf{w}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\xi_\mathbf{w}(n)], \end{aligned} \tag{10}$$

where $J_{\min} = E[e_{opt}^2(n)]$.

Since $\xi_\mathbf{w}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\xi_\mathbf{w}(n)$ is a scalar, by applying the independence assumption [2], [5] we may rewrite it as

$$\begin{aligned} &E[\xi_\mathbf{w}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\xi_\mathbf{w}(n)] \\ &= \mathrm{tr}\{E[\xi_\mathbf{w}(n)\xi_\mathbf{w}^T(n)\}E\{\mathbf{x}(n)\mathbf{x}^T(n)]\} \\ &= \mathrm{tr}\{E[\xi_\mathbf{w}(n)\xi_\mathbf{w}^T(n)]\mathbf{R}\} \quad . \end{aligned} \tag{11}$$

And many cross terms arising from the multiplication

Copyright 1997 IEEE

$\xi_{\mathbf{w}}(n)\xi_{\mathbf{w}}^T(n)$ reduce to zero matrices. Besides, as it was shown before that for sufficiently large n, $E[\xi_{hc}(n)] \approx 0$. As a result, for $n \to \infty$,

$$\mathbf{K}(n+1) \equiv \mathrm{E}[\xi_{\mathbf{w}}(n+1)\xi_{\mathbf{w}}^T(n+1)]$$
$$= (\mathbf{I}-\mu\mathbf{R})\mathbf{K}(n)(\mathbf{I}-\mu\mathbf{R}) + \mu^2 J_{\min}\mathbf{R}$$
$$+ \mu^2 E[\xi_{hc}^2(n)\mathbf{R}] \quad . \quad (12)$$

By applying unitary transformation to $\mathbf{K}(n)$:

$$\mathbf{K}(n)=\mathbf{Q}\mathbf{U}(n)\mathbf{Q}^T,$$

we have

$$E[\xi_{\mathbf{w}}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\xi_{\mathbf{w}}(n)] = \mathrm{tr}[\mathbf{K}(n)\mathbf{R}]$$
$$= \mathrm{tr}[\mathbf{Q}\Lambda\mathbf{Q}^T\mathbf{Q}\mathbf{U}(n)\mathbf{Q}^T] = \mathrm{tr}[\mathbf{Q}\Lambda\mathbf{U}(n)\mathbf{Q}^T]$$
$$= \mathrm{tr}[\mathbf{Q}^T\mathbf{Q}\Lambda\mathbf{U}(n)] = \mathrm{tr}[\Lambda\mathbf{U}(n)] = \sum_{i=0}^{N-1}\lambda_i u_i(n), \quad (13)$$

where $u_i(n)$, i=0,1,$\cdots$,N-1, are the elements of the diagonal matrix $\mathbf{U}(n)$, and λ_i are the eigenvalues of the correlation matrix $\mathbf{R}$. Further

$$\mathbf{U}(n+1) = (\mathbf{I}-\mu\Lambda)\mathbf{U}(n)(\mathbf{I}-\mu\Lambda)\} + \mu^2 J_{\min}\Lambda$$
$$+ \mu^2 \Lambda E[\xi_{hc}(n)^2] \quad , (14)$$

therefore,

$$u_i(n+1) = (1-\mu\lambda_i)^2 u_i(n) + \mu^2\lambda_i(J_{\min} + E[\xi_{hc}^2(n)]) \ ,(15)$$

$$\therefore\ u_i(\infty) = \mu(J_{\min} + E[\xi_{hc}^2(\infty)])/2-\mu\lambda_i. \quad (16)$$

We may write Eq. (13) as

$$E[\xi_w^T(\infty)\mathbf{x}(\infty)\mathbf{x}^T(\infty)\xi_w(\infty)] = \sum_{i=0}^{N-1}\lambda_i u_i(\infty)$$
$$= (J_{\min} + E[\xi_{hc}^2(n)])\sum_{i=0}^{N-1}\mu\lambda_i/(2-\mu\lambda_i) \quad , \quad (17)$$

$$J(\infty) = (1+\sum_{i=0}^{N-1}\frac{\mu\lambda_i}{2-\mu\lambda_i})J_{\min} + \sum_{i=0}^{N-1}\frac{\mu\lambda_i}{2-\mu\lambda_i}E[\xi_{hc}^2(\infty)] .(18)$$

The final task is to evaluate $E[\xi_{hc}^2(n)]$. Obviously,

$$E[\xi_{hc}^2(n)] = (1-\alpha)^2 E[\xi_{hc}(n-1)] + \alpha^2 J(n-1)$$
$$-2\alpha(1-\alpha)E[\xi_{hc}(n-1)(d(n-1)-\mathbf{w}(n-1)^T\mathbf{x}(n-1))] .(19)$$

Based on the independence assumption again, a simplified form is obtained as follows,

$$E[\xi_{hc}^2(\infty)] = \frac{\alpha J(\infty)}{2-\alpha}. \quad (20)$$

As a result, $J(\infty)$ is equal to

$$J(\infty) = J_{\min}\frac{(2-\alpha)(1+\sum_{i=0}^{N-1}\frac{\mu\lambda_i}{2-\mu\lambda_i})}{2-(1+\sum_{i=0}^{N-1}\frac{\mu\lambda_i}{2-\mu\lambda_i})\alpha}. \quad (21)$$

Finally,

$$J'(\infty) = J(\infty) + \frac{\alpha J(\infty)}{2-\alpha}$$
$$= J_{\min}\frac{2(1+\sum_{i=0}^{N-1}\frac{\mu\lambda_i}{2-\mu\lambda_i})}{2-(1+\sum_{i=0}^{N-1}\frac{\mu\lambda_i}{2-\mu\lambda_i})\alpha}. \quad (22)$$

Note that when $\alpha \approx 0$, $J'(\infty)$ becomes

$$J'(\infty) \approx (1+\sum_{i=0}^{N-1}\frac{\mu\lambda_i}{2-\mu\lambda_i})J_{\min} = J_{\mathrm{LMS}}, \quad (23)$$

which is the same as the result derived by Haykin [2] for DLMS algorithm under the condition of small μ. When μ is sufficiently small, we have

$$J'(\infty) \approx \frac{2}{2-\alpha}J_{\min}. \quad (24)$$

Since $\mathbf{R}$ is non-negative definite, λ_i>0 and $2-\mu\lambda_i$>0 for all i, for the convergence of $\mathbf{w}(n)$. Therefore

$$\sum_{i=0}^{N-1}\frac{\mu\lambda_i}{2-\mu\lambda_i} > 0. \quad (25)$$

In addition, since $J'(n)$ is positive for all n, the condition $2-\alpha[1+\sum_{i=0}^{N-1}\mu\lambda_i/(2-\mu\lambda_i)] > 0$ must hold. Consequently we can locate a more conservative upper bound than before for α:

$$0 < \alpha < 2/(1+\sum_{i=0}^{N-1}\frac{\mu\lambda_i}{2-\mu\lambda_i}). \quad (26)$$

4. SIMULATION RESULTS

In this section, we summarize the equalizer simulation results for the property verifications of the Chen's LMS algorithm. The impulse response of the channel to be equalized is assumed $C(n)$=[-1/2.1, 1, 1/2.1]. The channel input signal is assumed a white Gaussian, zero-mean noise with variance=1. The tap number of the adaptive equalizer equals to 4. Fig. 1(a), 1(b) show the MSE (average of 500 runs) with μ=0.0001 and μ=0.03 respectively as a function of α in steady state. Fig. 2(a), (b) show the MSE (average of 500 runs) of DLMS and new algorithm with μ=0.0001 and some α values.

The upper bound of α is very close to 2 when μ is very small from Eq. (26). Large α should be avoided as suggested by both theoretical and simulation results. As depicted in [2] and [3], the rate of convergence of $\mathbf{w}$ is dominated by $1-\mu\lambda_{\min}$ (where $\lambda_{\min}$ is the smallest eigenvalue of $\mathbf{R}$). The rate of convergence of h_N is dominated by 1-α. However, as α is extreme small, $h_N(n)$ will track $C(n)$ very slowly. In such a case, the rate of convergence of MSE becomes much slower than that of

Copyright 1997 IEEE

DLMS. Consequently, a comparably larger α than $\mu\lambda_{\min}$ is preferred.

All the results summarized are based on the independence theory. But the *shifting property* of input data introduces statistical dependence results [2]. The results make $E[x(n)(h_N(n)-C(n))] \neq 0$ even when $x(n)$ is zero-mean. That is, $E[x(n)(h_N(n)-C(n))]$ will yield a constant upon convergence. Accordingly, every converged weight is equal to the sum of Wiener solution and a DC bias. The magnitude of DC bias is directly proportional to α from the simulation results. The DC bias approaches to zero when α approaches to zero. The rate of convergence of Chen's algorithm is the same as the DLMS, and the DC bias can be ignored when μ=0.0001 and α=0.01 in the example.

5. CONCLUSION

The properties of Chen's algorithm have been characterized in the paper. The simulation results match the derived properties closely. Proper bounds for α and μ are also given to facilitate the new algorithm's pratical usage. In this theoretical analysis, small μ is assumed. The future work to be done is to assume a large condition.

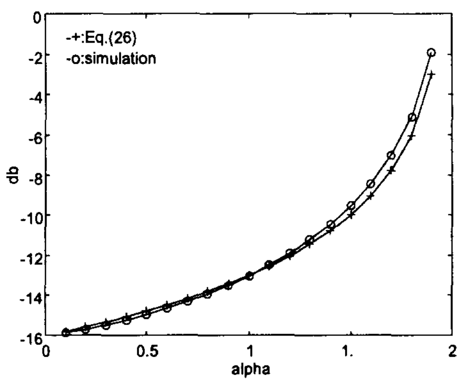

Fig. 1(a). The mean square errors in steady state as a function of α; μ=0.0001, 4 taps.

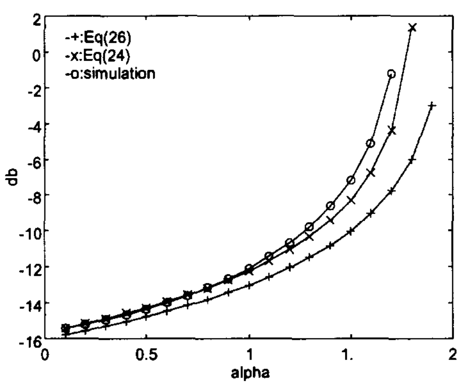

Fig. 1(b). The mean square errors in steady state as a function of α; μ=0.03, 4 taps.

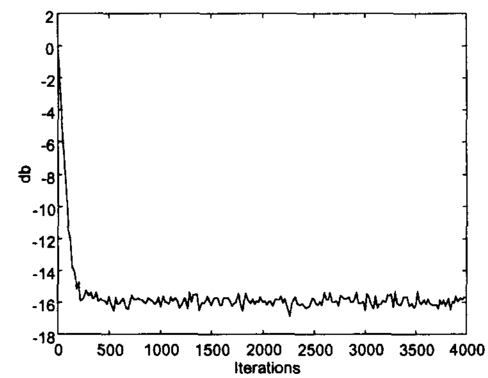

Fig. 2(a). The mean square error of LMS, μ=0.0001.

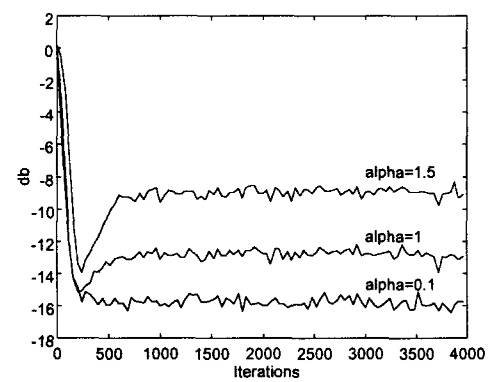

Fig. 2(b). The mean square error of Chen's algorithm, for μ=0.0001 and various α values.

REFERENCES

[1] S. G. Chen, et al., "A New Efficient LMS Adaptive Filtering Algorithm," *IEEE Trans. Circuits Syst.-II: Analog and Digital Signal Processing*, vol. 43, pp. 372-378, May 1996.

[2] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall,1991, 3rd ed..

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

[4] J. Benesty and P. Duhamel, "A Fast Exact Least Mean Square Adaptive Algorithm," *IEEE Trans. Signal Process.*, vol. 40, pp. 2904-2920, Dec. 1992.

[5] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Process with Applications to Stochastic System Theory*, MIT Press, Cambridge, Mass.

Copyright 1997 IEEE