CONVERGENCE ANALYSIS OF LMS AND NLMS ADAPTIVE ALGORITHMS

T. C. Hsia

Signal and Image Processing Laboratory
Department of Electrical and Computer Engineering
University of California, Davis
Davis, California 95616

ABSTRACT

The main contribution of this paper is the unified treatment of convergence analysis for both LMS and NLMS adaptive algorithms. The following new results are obtained: (i) necessary and sufficient conditions of convergence, (ii) optimal adjustment gains and optimal convergence rates, (iii) interrelationship between LMS and NLMS gains, and (iv) nonstationary algorithm design.

INTRODUCTION

Among various types of adaptive algorithms for real-time parameter estimation, the lease mean square (LMS) algorithm is very well known and widely used. Available in the literature are extensive studies regarding its convergence characteristics and performance properties [1-4]. The primary design parameter in applying LMS algorithm to practical problems is the judicial selection of the adaptive gain μ so that the algorithm is not only stable, but a good compromise is achieved between fast convergence rate and low parameter (or weight) noise in the adaptive filter [1-5].

Under the assumption that the input vectors of the adaptive filter are independent and stationary, and the reference system is time invariant, it was possible to establish the stability bounds for μ [1]. With regard to the optimum μ which gives fastest convergence rate, many different analyses have been reported [1-3]. These results are shown to agree with one another when the filter input vector elements are independent and identically distributed (iid) [5].

The ability to adapt in a nonstationary environment is an important function of an adaptive algorithm, and some investigation regarding design of adjustment gain μ for nonstationary LMS algorithm have been reported [1]. However, in-depth investigations of this important feature have been neglected, so that no further results are available [5]. One reason for what is happening is that LMS algorithm is ill-suited analytically for such analysis.

A companion algorithm to LMS is the normalized LMS (NLMS) algorithm which has been subjected to extensive investigation as well [4,6,7]. Due to its particular structure, this algorithm is more suited to more in-depth analytical study. Consequently, optimum convergence rate as well as nonstationary adaptation properties have been rigorously analyzed.

The objective of this paper is to reexamine the convergence properties of both LMS and NLMS algorithms, this time in a unified framework. This then allows us to explore the interrelationships between the two algorithms so that the optimum adjustment gains and optimum convergence rates can be compared. Through these relationships an LMS adjustment gain design suitable for nonstationary adaptation is derived from known results of nonstationary NLMS algorithm.

LMS AND NLMS ALGORITHMS

Let y_k be a scalar-valued stationary random process satisfying the relationship

$$y_k = x_k^T w^* \tag{1}$$

where x_k is an N dimension vector valued stationary random process, and w^* is a constant N dimensional parameter (weight) vector. Assume w^* is unknown, an estimate is to be recursively determined.

Suppose w_k is an estimate of w^*, then $\hat{y}_k = x_k^T w_k$ is the corresponding estimate of y_k. The recursive algorithms for updating w_k based on LMS and NLMS designs are

$$\text{LMS:}\quad w_{k+1} = w_k + \mu x_k e_k \tag{2}$$

$$\text{NLMS:}\quad w_{k+1} = w_k + \frac{\alpha}{||x_k||^2} x_k e_k \tag{3}$$

where μ and α are positive constants, and the error e_k is defined as

$$e_k = y_k - \hat{y}_k = y_k - x_k^T w_k . \tag{4}$$

Now define ε_k as the parameter error vector

$$\varepsilon_k = w_k - w^* . \tag{5}$$

The parameter error equation for both LMS and NLMS algorithm becomes

$$\varepsilon_{k+1} = (I - F_k^{(r)})\varepsilon_k \tag{6}$$

where

$$r = 1,\ F_k^{(1)} = \mu x_k x_k^T \qquad \text{for LMS} \tag{7}$$

15A.2

 CH1841-6/83/0000-0667 $1.00 © 1983 IEEE

$$r = 2,\ F_k^{(2)} = \alpha \frac{x_k x_k^T}{||x_k||^2} \quad \text{for NLMS} \tag{8}$$

with the symmetry property that

$$F_k^{(r)} = F_k^{(r)T} \qquad r = 1,2\ . \tag{9}$$

Let us assume that the vector sequence $\{x_k\}$ are statistically independent (most frequently used assumption). Then $F_k^{(r)}$ and ε_k are independent. Define an error covariance matrix P_k as

$$P_k = E\{\varepsilon_k \varepsilon_k^T\} \tag{10}$$

Then we have from (6)

$$P_{k+1} = P_k - 2E_F\{F_k^{(r)}\}P_k + E_F\{F_k^{(r)} P_k F_k^{(r)T}\}\ . \tag{11}$$

CONDITIONS FOR CONVERGENCE

We shall now introduce some new results on the convergence of w_k to w^*. Our convergence analysis is based on that of the error covariance P_k. Define an N^2 vector q_k

$$q_k = [p_k(1,1),\ldots,p_k(1,N),p_k(2,1),\ldots,p_k(2,N),\ \ldots,p_k(N-1),\ldots,p_k(N,N)]^T \tag{12}$$

where $p_k(i,j)$ are elements of P_k. To show that $\lim_{k\to\infty} P_k = 0$ is equivalent to

$$\lim_{k\to\infty} q_k = 0. \tag{13}$$

Theorem 1

If $\{x_k\}$ is an independent stationary random sequence, and $E[x_k] = 0$, then

(i) The necessary and sufficient condition of convergence of the LMS algorithm in the sense of mean square is that the magnitudes of all eigenvalues of the $N^2 x N^2$ matrix

$$I_{N^2} - 2\mu E\{I_N \otimes xx^T\} + \mu^2\{E[(x \otimes x)(x^T \otimes x^T)]\} \tag{14}$$

are less than 1.

(ii) The necessary and sufficient condition of convergence for NLMS algorithm in the sense of mean square is that the magnitudes of all eigenvalues of the $N^2 x N^2$ matrix

$$I_{N^2} - 2\alpha E\left\{I_N \otimes \frac{xx^T}{||x||^2}\right\} + \alpha^2 E\left[\frac{1}{||x||^4}(x \otimes x)(x^T \otimes x^T)\right] \tag{15}$$

are less than 1.

We note that the notation $\otimes$ denotes tensor product. Proof: The proofs of parts (i) and (ii) are algebraically cumbersome, we shall only give a brief sketch of the proofs. Apply (7) to (11) yielding the following equation

$$P_{k+1} = P_k - 2\mu E[x_k x_k^T]P_k + \mu^2 E\{\underbrace{\sum_{m\ell}\sum x_i x_j x_\ell x_m p_k(\ell,m)}_{(i,j)^{th}\ \text{element}}\} \tag{16}$$

Then apply (12) to (16) and simplify, we get

$$q_{k+1} = q_k - 2\mu E\{I \otimes (xx^T)\}q_k + \mu^2 E\{(x \otimes x)(x^T \otimes x^T)\}q_k \tag{17}$$

Thus (17) is asymptotically stable if and only if (14) has all its eigenvalues to be less than one in magnitude. We note that the matrix (14) is symmetric so all its eigenvalues are real. In a similar manner, we can proof part (ii) of Theorem 1.

SIMPLIFIED CONVERGENCE CONDITIONS

The general conditions stated in Theorem 1 are cumbersome to apply, so now we wish to present a simplified version of Theorem 1 on the assumption that, in addition to the statistical independence of $\{x_k\}$, the elements in x_k are independent and identically distributed (iid). The simplification due to iid assumption further enable us to establish optimum rate of convergence and optimum adjustment gain.

Theorem 2

If $\{x_k\}$ is an independent random sequence and the elements of x_k, $x(i)$ are iid with

$$\left.\begin{aligned} &E[x(i)x(j)] = \sigma^2 \delta_{i-j} \\ &E[x^3(i)x(j)] = \xi \delta_{i-j} \\ &E[x(j)] = 0 \\ &\gamma(\text{kurtosis}) = \frac{E[x^4(i)]}{\{E[x^2(i)]\}^2} = \frac{\xi}{\sigma 4} \end{aligned}\right\} \tag{18}$$

Then (i) LMS algorithm is convergent in mean square if and only if

$$0 < \mu < \frac{2}{(N+\gamma-1)\sigma^2} \tag{19}$$

(ii) the NLMS algorithm is convergent in mean square if and only if

$$0 < \alpha < 2 \tag{20}$$

Proof: The iid assumption of x_k reduces convergence proof of q_k to that of trace P_k. From (18), we can show that trP_k results in $trP_k = E[||\varepsilon_k||^2]$

$$E[||\varepsilon_{k+1}||^2] = \{1-2\mu\sigma^2+\mu^2(\xi+\sigma^4(N-1))\}E[||\varepsilon_k||^2] \tag{21}$$

The necessary and sufficient condition for (21) to converge is $|1-2\mu\sigma^2+\mu^2(\xi+\sigma^4(N-1))| < 1$.

This yields the result (19). For part (ii), we use (8),(11) and (18) to obtain

$$trP_{k+1} = trP_k - (2\alpha-\alpha^2)E\left|\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{x(i)x(j)}{||x||^2}p_k(i,j)\right| \quad (22)$$

Apply the following approximation to (22)

$$E\left|\frac{x(i)x(j)}{||x||^2}\right| \cong \frac{E[x(i)x(j)]}{E||x||^2} = \frac{1}{N}\delta_{ij}$$

we get

$$E[||\varepsilon_{k+1}||^2] = \{1-(2\alpha-\alpha^2)\frac{1}{N}\}\ E[||\varepsilon_k||^2] \quad (23)$$

Asymptotic stability of (23) implies satisfaction of (20) by α.

OPTIMUM CONVERGENCE RATE

Equations (21) and (23) are amenable for convergence rate optimization if we define the following convergence ratio g_k:

$$g_k = \frac{E[||\varepsilon_{k+1}||^2]}{E[||\varepsilon_k||^2]} \quad (24)$$

Theorem 3

If x_k satisfies the conditions as stated in Theorem 2, the optimum adjustment gains and optimum convergence rates for LMS and NLMS algorithms are

(i) LMS: $\mu^* = \dfrac{1}{(N+\gamma-1)\sigma^2}$, $g^*(\mu) = 1-\dfrac{1}{\gamma+(N-1)}$ (25)

(ii) NLMS: $\alpha^* = 1$, $g^*(\alpha) = 1 - \dfrac{1}{N}$ (26)

Proof: Minimization of g_k (24) with respect to μ and α based on (21) and (23) results in (25) and (26) directly. Remark: Comparing $g^*(\mu)$ and $g^*(\alpha)$, NLMS converges faster than LMS does, a fact long noticed in computer simulation studies, but never theoretically proven. The optimum gain μ^* is also a new result (first reported in [8]). When x_k is Gaussian the kurtosis γ is $\gamma = 3$ and

$$\mu^* = \frac{1}{(N+2)\sigma^2}\ . \quad (27)$$

This result is close to the existing results $\mu^* = \dfrac{1}{N\sigma^2}$ under the same assumptions on x_k [5], but there is definitely a difference unless N is large. The result in (27) has been suggested in [5], but it's optimality was not justified as it was here. Computer simulation has shown (27) is indeed optimal.

INTERRELATIONSHIP BETWEEN μ AND $\alpha/||x||^2$

Let us define the adjustment gain for NLMS (3) as $\beta(\alpha)$

$$\beta(\alpha) = \frac{\alpha}{||x_k||^2} \quad (28)$$

where β is a random variable greater than zero. When vector x_k is Gaussian, independent with iid elements, the probability distribution can be evaluated [5]. Let $\beta^* = \beta(\alpha^*) = 1/||x_k||^2$ and β_M^* be the mode of β^*. The following theorem was proved in [5].

Theorem 4

If x_k is a zero-mean independent Gaussian vector with iid elements, then

$$\mu^* = \frac{1}{(N+2)\sigma^2} = \beta_M^* \quad (29)$$

β_M^* is the mode of $\dfrac{1}{||x_k||^2}$.

This result is very interesting and useful in that we can now formally establish a direct link between the two adjustment gains (μ and β). The result (29) allows us to derive a nonstationary LMS optimal gain μ^* from known β^* of nonstationary NLMS algorithm.

NONSTATIONARY LMS OPTIMAL ADJUSTMENT GAIN

The nonstationary LMS gain μ^* given by [1] is a useful rule of thumb, but it lacks a rigorous foundation. A design approach was suggested in [5] which was derived based on the relationship (29). The nonstationary system model assumed was

$$y_k = x_k^T w_k^* + n_k \quad (30)$$

where x_k has the same assumptions as before, n_k is a stationary white noise with zero mean and variance σ_n^2, w_k^* is a random weight vector satisfying the

$$w_{k+1}^* = a\ w_k^* + \xi_k \quad (31)$$

ξ_k is a stationary white noise with zero mean and variance σ_ξ^2, and $0 \le a \le 1$.

It was shown in [5] that, under appropriate assumptions, the optimum α in NLMS (3) in the nonstationary case is

$$\alpha^* = \sqrt{2N(N-2)}\frac{\sigma_\xi\sigma}{\sigma_n} \quad (32)$$

In view of (28) and (29), we can obtain a nonstationary LMS optimal gain

$$\mu^* = \beta_M(\alpha^*) = \frac{\alpha^*}{(N+2)\sigma_x^2} = \frac{\sqrt{2N(N-2)}}{N+2}\frac{\sigma_\xi}{\sigma_x\sigma_n} \quad (33)$$

For large N, μ^* is further simplified to

$$\mu^* = \sqrt{2}\ \frac{\sigma_\xi}{\sigma_x\sigma_n} \quad (34)$$

This result compares favorably with Widrow's result [1] which is (after some simplifications)

15A.2

$$\mu^*_{Widrow} = \frac{\sigma_\xi}{2\sigma_x\sigma_\eta} \quad (35)$$

However, (34) can be rigorously justified through the interrelationship between LMS and NLMS algorithms. Thus, the result (34) is an important alternative to (35).

CONCLUSION

We have presented a systematic convergence analysis of LMS and NLMS algorithms and established the linkage between the two adjustment gains. The LMS optimal gains for both stationary and nonstationary adaptations are new and are very useful guides to adaptive system designers.

REFERENCES

[1] 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, pp. 1151-1162, 1976.

[2] J.E. Mazo, "On the independence theory of equilizer theory," Bell Sys. Tech. J., Vol. 58, pp. 963-993, 1979.

[3] R.D. Gitlin and S.B. Weinstein, "On the required tap-weight precision for digitally implemented, adaptive, mean-squared equalizers," Bell Sys. Tech. J., Vol. 58, pp. 963-993, 1979.

[4] R.R. Bitméad and B.D.O. Anderson, "Performance of adaptive estimation algorithms in dependent random environments," IEEE Trans. Automat. Contr., Vol.AC-25, pp. 788-794, 1980.

[5] T.C. Hsia and S. Sugimoto, "An investigation of adjustment gain design in stationary and nonstationary LMS algorithms," Proc. of IFAC, pp. 1196-1201, June 1982.

[6] J. Nagumo and A. Noda, "A learning method for system identification," IEEE Trans. Automat. Contr., Vol. AC-12, pp. 282-287, 1967.

[7] A. Noda, "Effects of noise and parameter-variation on the learning identification method," J. of SICE of Japan, Vol. 8, pp. 303-312, 1969.

[8] W.A. Gardner, "Bounds and approximations for learning characteristics of the stochastic-gradient-decent algorithm for independent non-Gaussian training vectors," Dept. of ECE, University of California, Davis, CA, 1981.