Convergence Analysis of the LMS Algorithm with a General Error Nonlinearity and an IID Input

Tareq Y. Al-Naffouri
Dept. of Electrical Eng.
Stanford University
Stanford, CA 9305
U.S.A

Azzedine Zerguine
Dept. of Physics
KFUPM
Dhahran 31261
Saudi Arabia

Maamar Bettayeb
Dept. of Electrical Eng.
KFUPM
Dhahran 31261
Saudi Arabia

Abstract

The class of least mean square (LMS) algorithms employing a general error nonlinearity is considered. A linearization approach is used to characterize the convergence and performance of this class of algorithms for an independent and identically distributed (iid) input. The analysis results are entirely consistent with those of the LMS algorithm and several of its variants. The results also encompass those of a recent work that considered the same class of algorithms for arbitrary and Gaussian inputs.

1 Introduction

The LMS algorithm is one of the most widely used adaptive schemes. It has many desirable features, e.g., lower order of complexity and easy implementation. However, to overcome some of its limitations, several LMS-variants have been proposed and analyzed. These limitations are generally due to noise distributions and spectral coloration of the input signal. In the case of the noise distributions, the LMS algorithm is optimal only if this distribution happens to be Gaussian. However, the least mean fourth (LMF) [1] outperforms the LMS algorithm only when the distribution is not Gaussian.

In many situations in which adaptive filters are used, the dominant source of interference is not Gaussian and instead has impulsive character. For example, lightning and switching transients can cause impulsive interference on telephone lines. In these kinds of operating environments, echo cancellers and adaptive equalizers which use the LMS algorithm may suffer from poor performance due to high variance gradient estimates resulting from non-Gaussian noise.

Of particular importance is the class of least mean adaptive algorithms with a general error nonlinearity. The optimum error nonlinearity will result in the best performance of the adaptive algorithm. In this work, the convergence analysis for stochastic gradient algorithms with a general error nonlinearity and an iid input is considered. The optimum nonlinearity for such a scenario is reported in [2].

After presenting the proposed algorithm in Section 2, the convergence analysis of the algorithm is presented in Section 3, and finally a conclusion to this work is given.

2 Proposed algorithm

For analysis purposes, it is more convenient to describe this class in terms of the weight error vector $\boldsymbol{V}(k)$ which is updated according to

$$\boldsymbol{V}(k+1) = \boldsymbol{V}(k) + \mu f(e(k))\boldsymbol{X}(k), \qquad (1)$$

0-7803-5148-7/98/$10.00©1998 IEEE

where $\boldsymbol{X}(k)$ is the input data in the filter memory at time k, f is an error nonlinearity, and μ is the step size. In a system identification model, as it is depicted in Figure 1, the error $e(k)$ is related to $\boldsymbol{V}(k)$ and to the additive noise $n(k)$ through the filtering equation

$$e(k) = n(k) - \boldsymbol{V}^t(k)\boldsymbol{X}(k). \tag{2}$$

Several familiar algorithms are obtained by a proper choice of the nonlinearity f. Examples include the LMS algorithm, the sign-error algorithm, the LMF algorithm and its family [1], and the mixed-norm algorithm [3], [4].

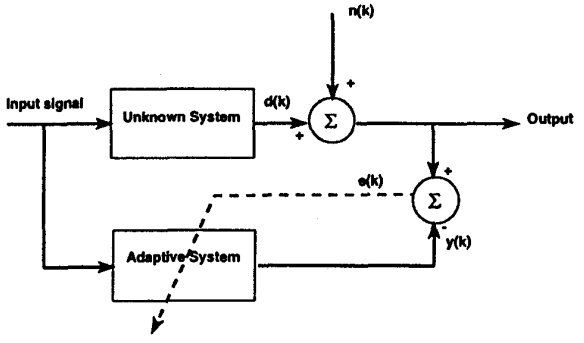

Figure 1: Adaptive system identification.

3 Convergence analysis of the algorithm

For an effective use of such algorithms, it is important to characterize their convergence and quantify their performance. For an arbitrary distribution of the input and noise, this is usually done by linearizing the nonlinearity f about some operating point. This linearization trick was explicitly realized in [5] and was used to study the convergence and performance of the class of algorithms (1)-(2) for an *arbitrary* input. Here, we employ this linearization approach to study the same class of algorithms for an *iid* input. The simplifying iid assumption makes it possible to arrive at tighter convergence conditions and a more accurate description or the algorithm performance. The nonlinearity f can be expanded in a third order Taylor series about the noise sample $n(k)$. Thus, as the error $e(k)$ is related to $n(k)$ by the filtering equation (2), it follows that

$$\begin{aligned} f(e(k)) &= f(n(k)) - f'(n(k))\left(\boldsymbol{V}^T(k)\boldsymbol{X}(k)\right) \\ &\quad + \frac{1}{2}f''(n(k))\left(\boldsymbol{V}^T(k)\boldsymbol{X}(k)\right)^2. \end{aligned} \tag{3}$$

With this linearization, the adaptation equation (1) becomes

$$\begin{aligned} \boldsymbol{V}(k+1) &= \boldsymbol{V}(k) + \mu f(n(k))\boldsymbol{X}(k) \\ &\quad - \mu f'(n(k))\left(\boldsymbol{V}^T(k)\boldsymbol{X}(k)\right)\boldsymbol{X}(k) \\ &\quad + \frac{1}{2}\mu f''(n(k)) \\ &\quad \times \left(\boldsymbol{V}^T(k)\boldsymbol{X}(k)\right)^2 \boldsymbol{X}(k). \end{aligned} \tag{4}$$

The above equation is called the linearized adaptation equation.

3.1 Convergence in the mean and the mean-square error

Starting with the adaptation equation (4) and taking expectation gives:

$$\begin{aligned} E[\boldsymbol{V}(k+1)] &= E[\boldsymbol{V}(k)] \\ &\quad - \mu E[f'(n(k))]E[\boldsymbol{V}^T(k)\boldsymbol{X}(k)\boldsymbol{X}(k)] \\ &\quad + \mu E[f(n(k))]E[\boldsymbol{X}(k)] \\ &\quad + \frac{1}{2}\mu E[f''(n(k))] \\ &\quad \times E[(\boldsymbol{V}^T(k)\boldsymbol{X}(k))^2\boldsymbol{X}(k)], \end{aligned} \tag{5}$$

the convergence in the mean is ensured by the following proposition [6]:

Proposition 1 *The inequality*

$$0 < \mu < \frac{2}{\sigma_x^2 E[f']} \tag{6}$$

represents a necessary and sufficient condition for convergence of (1)-(2) in the mean.

While, the convergence in the mean-square of the proposed algorithm is governed by the convergence of the following:

$$\begin{aligned} E\left[\boldsymbol{V}(k+1)\boldsymbol{V}^T(k+1)\right] &= E\left[\boldsymbol{V}\boldsymbol{V}^T\right] \\ &+\mu^2 E[f^2]E\left[\boldsymbol{X}\boldsymbol{X}^T\right] \\ &-\mu E[f']E\left[\boldsymbol{X}\boldsymbol{V}^T(\boldsymbol{V}^T\boldsymbol{X})\right] \\ &+\mu E[f']E\left[\boldsymbol{V}\boldsymbol{X}^T(\boldsymbol{V}^T\boldsymbol{X})\right] \\ &+\mu^2 E[(f'^2+ff'')] \\ &\times E\left[\boldsymbol{X}\boldsymbol{X}^T(\boldsymbol{V}^T\boldsymbol{X})^2\right], \quad (7) \end{aligned}$$

equivalently [6]:

Proposition 2 *The inequality*

$$0 < \mu < \frac{2\sigma_x^2 E[f']}{m_{x,4} + (L-1)\sigma_x^4]E[f'^2 + ff'']} \quad (8)$$

gives a sufficient condition for convergence of (1)-(2) in the mean square sense.

The expectations $E[f^2]$, $E[f']$, and $E[f'^2 + ff']$ are taken with respect to the noise $n(k)$, and $m_{x,4}$ is the fourth moment of x.

3.2 Misadjustment

A measure of how far the excess mean-square error (MSE) to its minimum is measured through the use of the misadjustment factor, and for this algorithm is given by [6]:

Proposition 3 *The misadjustment for the algorithm (1)-(2) is given by*

$$M = \mu L \frac{\sigma_x^2}{\sigma_n^2} \frac{E[f^2]}{2E[f']} \left[1 + \mu \frac{m_{x,4} + (L-1)\sigma_x^4}{2\sigma_x^2 E[f']} E[f'^2 + ff'']\right]. \quad (9)$$

Remarks

1. The performance analysis results of the LMS algorithm, the LMF algorithm and its family, and the mixed-norm algorithm can be immediately recovered from (6), (8), and (9) by a proper choice of the nonlinearity f.
2. From Table 1, we can compare the performance results of [5] when specialized to an iid input with those reported in this paper. We notice that the arbitrary input results approximate, to the first order and for sufficiently small step size, the results of the independent input case.
3. If the input is restricted to be iid Gaussian, then the mean and mean-square recursions obtained through the linearization approach are in agreement with the corresponding recursions obtained in [5] through a conditional analysis approach (for details, see [6]).

4 Conclusion

The convergence analysis for stochastic gradient algorithms with a general error nonlinearity and an iid input is treated in this work. The performance analysis results of several algorithm can be immediately recovered by a proper choice of the error nonlinearity.

Acknowledgement: The authors acknowledge KFUPM for its support.

References

[1] E. Walach and B. Widrow, "The Least Mean Fourth (LMF) Adaptive Algorithm and its Family", *IEEE Trans. Inf. Theory*, vol. IT-30, no. 2, pp. 275–283, Feb. 1984.

[2] T. Y. Al-Naffouri, A. Zerguine, and M. Bettayeb, "A Unifying View of Error Nonlinearities In LMS Adaptation," in Proc. *ICASSP*, Seattle, U.S.A, 1998.

[3] J. A. Chambers, O. Tanrikulu, and A. G. Constantinides, "Least mean mixed-norm adaptive filtering," *Electr. lett., vol.* 30, No. 19, pp. 1574-1575, Sep. 1994.

[4] A. Zerguine, C. F. N. Cowan, and M. Bettayeb, "Adaptive Echo Cancellation Using Least Mean

	Arbitrary Input	Independent Input
Conditions for convergence of the mean	$0 < \mu < \frac{2}{\lambda_{max} E[f']}$	$0 < \mu < \frac{2}{\sigma_x^2 E[f']}$
Time constants	$\tau_i = \frac{1}{\mu \lambda_i E[f']}$	$\tau = \frac{1}{\mu \sigma_x^2 E[f']}$
Conditions for convergence in the mean-square	$0 < \mu < \frac{1}{\lambda_{max} E[f']}$	$0 < \mu < \frac{2\sigma_x^2 E[f']}{[m_{x,4} + (L-1)\sigma_x^4] E[f'^2 + f f'']}$
Misadjustment	$\mu \frac{\sigma_x^2}{\sigma_n^2} \frac{E[f^2]}{2E[f']} L$	$\mu \frac{\sigma_x^2}{\sigma_n^2} \frac{E[f^2]}{2E[f']} \left[1 + \mu \frac{m_{x,4} + (L-1)\sigma_x^4}{2\sigma_x^2 E[f']} E[f'^2 + f f'']\right] L$

Table 1: A summary of the convergence analysis results for the algorithms with a correlated input and an iid input, respectively.

Mixed-Norm Algorithm," *IEEE Trans. Signal Processing, vol.* SP-45, No. 5, pp. 1340-1343, May, 1997.

[5] S. C. Douglas and T. H. Y Meng, "Stochastic Gradient Adaptation under General Error Criteria", *IEEE Trans. Signal Processing*, vol. SP-42, no. 6, pp. 1352–1365, June 1994.

[6] T. Y. Al-Naffouri, *Adaptive Filtering Using the Least-Mean Mixed-Norms Algorithm and its Application to Echo Cancellation*, Master thesis, KFUPM, June 1997.