A UNIFYING VIEW OF ERROR NONLINEARITIES IN LMS ADAPTATION

Tareq Y. Al-Naffouri
School of Electrical & Computer Eng.
Georgia Institute of Technology
Atlanta, Ga 30332
E-mail: gte696@prism.gatech.edu

Azzedine Zerguine
Dept. of Physics
KFUPM
Dhahran 31261
Saudi Arabia

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

Abstract

This paper presents a unifying view of various error nonlinearities that are used in least mean square (LMS) adaptation such as the least mean fourth (LMF) algorithm and its family and the least-mean mixed-norm algorithm. Specifically, it is shown that the LMS algorithm and its error-modified variants are approximations of two recently developed optimum nonlinearities which are expressed in terms of the additive noise probability density function (pdf). This is demonstrated through an approximation of the optimum nonlinearities by expanding the noise pdf in a Gram-Charlier series. Thus a link is established between intuitively proposed and theoretically justified variants of the LMS algorithm. The approximation has also a practical advantage in that it provides a trade-off between simplicity and more accurate realization of the optimum nonlinearities.

1 Introduction and Motivation

The least mean squares (LMS) algorithm is one of the most widely used adaptive schemes because of its simplicity, efficiency, robustness, and numerical stability [1, p. 53]. Moreover, it has been very thoroughly investigated over a period of time that it is now very well understood. Therefore, many modifications of this algorithm have been suggested and analyzed. Of particular importance is the class of least-mean algorithms employing an error nonlinearity. This class is described by

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{W}(k) + \mu q(e(k))\mathbf{X}(k) \\ e(k) &= d(k) - \mathbf{W}^t(k)\mathbf{X}(k), \end{aligned}$$

where $\mathbf{W}(k)$ is the adaptive filter vector of coefficients, $\mathbf{X}(k)$ is the data input vector, $d(k)$ is the desired response, and $q(e(k))$ is the error nonlinearity. The choice

$$q(e(k)) = e^3(k) \tag{1}$$

yields the least mean fourth (LMF) algorithm [2], which is one of the most popular variants of the LMS algorithm. Other members of the LMF family are obtained by choosing [2]

$$q(e(k)) = e^{(2l-1)}(k), \ (l \geq 3). \tag{2}$$

The recently proposed least-mean mixed-norms algorithm [3] combines the advantages of the LMS and LMF algorithms by choosing

$$q(e(k)) = \lambda e(k) + (1-\lambda)e^3(k), \ 0 \leq \lambda \leq 1. \tag{3}$$

Here a natural question arises: are these nonlinearities theoretically justified, or is intuition used to support the choice of a certain nonlinearity? Unfortunately, more often than not, the latter approach is used.

This was not the case in [4] were the calculus of variations was used to arrive at the optimum error nonlinearity. Specifically, it was shown there that under the system identification model

$$d(k) = \mathbf{W_0}^t \mathbf{X}(k) + n(k), \qquad (4)$$

the optimum error nonlinearity for an arbitrary input is given by

$$q_{opt}(x) = -\frac{p'(x)}{p(x)}, \qquad (5)$$

where $p(x)$ is the probability density function of the additive noise $n(k)$. In the independent input case, a more accurate description of the nonlinearity is possible [5]

$$q_{opt}(x) = -\frac{p'(x)}{p(x) + \alpha\mu p''(x)}. \qquad (6)$$

Both of these optimum nonlinearities are valid for a symmetrically distributed white noise and for sufficiently small step size μ. The optimum nonlinearities in turn give rise to two questions, a theoretical and a practical one.

1. How do the theoretically justified nonlinearities (5) and (6) relate to the LMS algorithm and its error-modified variants (1)-(3)?

2. How is it possible to have a practical and inexpensive implementation of the optimum nonlinearities, especially that they are expressed in terms of the noise pdf which is usually unknown?

Both of these questions are answered by expanding the pdf $p(x)$ in a Gram-Charlier series. This will be demonstrated for the nonlinearity $-p'(x)/p(x)$ only although the discussion applies equally to the other optimum nonlinearity (6).

2 Gram-Charlier Approximation of the Optimum Nonlinearity

Let γ_i denote the ith cumulant of the noise $n(k)$ and let σ^2 denote its variance. The Gram-Charlier expansion of the pdf $p(x)$ is given by [6]

$$p(x) = \frac{1}{\sigma}\phi(\frac{x}{\sigma})\sum_{i=0}^{\infty} a_{2i}\mathrm{He}_{2i}(\frac{x}{\sigma}), \qquad (7)$$

where ϕ is the standard Gaussian pdf, He_{2i} is the Hermite polynomial of degree $2i$, and a_{2i} is a function of the cumulants of $n(k)$. In particular,

$$a_0 = 1,\; a_2 = 0,\; a_4 = \frac{\gamma_4}{4!\sigma^4}, \qquad (8)$$

$$a_6 = \frac{\gamma_6}{6!\sigma^6},\; \text{and}\; a_8 = \frac{1}{8!\sigma^8}(\gamma_8 + 35\gamma_4^2). \qquad (9)$$

With this expansion the optimum error nonlinearity, $-p'(x)/p(x)$ can be expressed as

$$-\frac{p'(x)}{p(x)} = \frac{x}{\sigma^2} - \frac{1}{\sigma}\frac{\frac{d}{dx}\left(\sum_{i=0}^{\infty} a_{2i}\mathrm{He}_{2i}(\frac{x}{\sigma})\right)}{\sum_{i=0}^{\infty} a_{2i}\mathrm{He}_{2i}(\frac{x}{\sigma})}. \qquad (10)$$

To simplify the discussion, only the first 4 terms of the summation in (10) are retained, which amounts to approximating $p(x)$ by the first four terms of Gram-Charlier expansion (7). It is also more convenient to express this sum as a polynomial in x

$$\sum_{i=0}^{4} a_{2i}\mathrm{He}_{2i}(\frac{x}{\sigma}) = A_0\left(1 + \sum_{i=1}^{4} A_{2i}x^{2i}\right), \qquad (11)$$

where

$$A_0 = (3a_4 - 15a_6 + 105a_8), \qquad (12)$$

$$A_2 = \frac{1}{\sigma^2 A_0}(-64a_4 + 45a_6 - 420a_8), \quad (13)$$

$$A_4 = \frac{1}{\sigma^4 A_0}(a_4 - 15a_6 + 210a_8), \qquad (14)$$

$$A_6 = \frac{1}{\sigma^6 A_0}(a_6 - 28a_8), \; \text{and} \qquad (15)$$

$$A_8 = \frac{1}{\sigma^8 A_0}a_8. \qquad (16)$$

Thus, the nonlinearity $-p'(x)/p(x)$ can be approximated as

$$-\frac{p'(x)}{p(x)} \simeq \frac{x}{\sigma^2} - \frac{1}{\sigma}\frac{8A_8 x^7 + 6A_6 x^5 + 4A_4 x^3 + 2A_2 x}{A_8 x^8 + A_6 x^6 + A_4 x^4 + A_2 x^2 + 1}.$$

We can also approximate the rational part of the last equation by its 8th-order Taylor series so that the nonlinearity $-p'(x)/p(x)$ finally reads

$$-\frac{p'(x)}{p(x)} \simeq \left(c_1 + \frac{1}{\sigma^2}\right)x + c_3 x^3 + c_5 x^5 + c_7 x^7, \quad (17)$$

where

$$c_1 = \frac{-2A_2}{\sigma}, \quad (18)$$

$$c_3 = \frac{2}{\sigma}\left(A_2^2 - 2A_4\right), \quad (19)$$

$$c_5 = \frac{2}{\sigma}\left(A_2^3 + 3A_2A_4 - 3A_6\right), \text{ and} \quad (20)$$

$$c_7 = \frac{2}{\sigma}\left(-A_2^4 + 2A_4^2 - 4A_2^2A_4 + 3A_2A_6 - 4A_8\right)(21)$$

The relation sets (8)-(9),(12)-(16), and (18)-(21) serve to show that the coefficients c_i of the approximation (17) can be explicitly written as a function of the noise cumulants.

3   Relationship between the Optimum and other Error Nonlinearities

The special polynomial approximation (17) shows that the LMS algorithm and its variants (1)-(3) are simply approximations of the optimum nonlinearity. To start with, the lowest order term in (17) is that which appears in LMS adaptation, so that the LMS is a first order approximation of the optimum nonlinearity (5). This explains the robustness that the LMS algorithm enjoys in different noisy environments. The second term of (17) corresponds to the error nonlinearity of the LMF algorithm while the higher-order terms correspond to error nonlinearities in the LMF family.

This approximation also suggests that a mixture of the LMS algorithm and the LMF family of algorithms will outperform the performance of each of the individual algorithms as this mixture provides a better approximation of the optimum nonlinearity. The LMS-LMF mixture was actually studied and simulated in [3] and [9] and was found to have better performance compared to both of the algorithms. The polynomial approximation (17) not only justifies such mixtures but also identifies the optimum mix in terms of the polynomial coefficients, which can in turn be expressed in terms of the noise cumulants.

4   Practical Implementation of the Optimum Nonlinearity

The nonlinearity $-p'(x)/p(x)$ is difficult to implement because a different nonlinearity must be implemented for each type of noise. Moreover, $p(x)$ is usually unknown and must be estimated. The approximation in (17) does away with both of these problems by preserving a general form of the nonlinearity which requires a small number of operations. The general form provides a method for a trade-off between simplicity of implementation and more accurate approximation of the optimum nonlinearity. Moreover, pdf estimation is traded for cumulant estimation which is straightforward to implement given a sample of the noise process. The cumulants can also be estimated on-line giving rise to a time varying nonlinearity.

Cumulant estimation might be complicated by the fact that the noise $n(k)$ is unobservable. Only the output error

$$e(k) = \mathbf{V}^t(k)\mathbf{x} + n(k)^1 \quad (22)$$

is accessible, and it can actually be used for cumulant estimation. This is justified by observing that under slow adaptation conditions, which is a requirement for the nonlinearity $-p'(x)/p(x)$ to be optimum, $e(k)$ consists mainly of $n(k)$ [2],[4]. The central limit theorem provides another justification in that $\mathbf{v}^t(k)\mathbf{x}$ is approximately Gaussian [7], [8], so that the cumulants of the output error are approximately equal to those of the additive noise.

---

[1]$\mathbf{V(k)}$ is the weight-error vector representing the diffrence between $\mathbf{W_0}$ and $\mathbf{W(k)}$.

5 Conclusion

The LMS algorithm and several of its error-modified variants are approximations of the optimum error nonlinearities for LMS adaptation. This link between intuitively proposed and theoretically derived least mean algorithms justifies the use of the former ones and actually provides a means for optimizing their performance. The general approximation arrived at also serves as a practical implementation of the optimum nonlinearities. Although the results of this paper were demonstrated for one optimum nonlinearity (5), they apply equally to the optimum nonlinearity (6).

References

[1] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission*, New York: Wiley, 1995.

[2] E. Walach and B. Widrow, "The Least Mean Fourth (LMF) Adaptive Algorithm and its Family", *IEEE Trans. Inf. Theory*, vol. IT-30, pp. 275–283, Feb. 1984.

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

[5] T. Y. Al-Naffouri, *Adaptive Filtering Using the Least-Mean Mixed-Norms Algorithm and its Application to Echo Cancellation*", Master thesis, King Fahd University of Petroleum and Minerals, June 1997.

[6] A. H. Nuttall, "Evaluation of densities and distributions via Hermite and generalized Laguerre series employing higher-order expansion coefficients determined recursively via moments or cumulants", Tech. Rep. Tr 7377, Naval Under-Water Systems Center, Feb. 1985.

[7] D. L. Duttweiler, "Adaptive Filter Performance with Nonlinearities in the Correlation Multiplier", *IEEE Trans. Acoust., Speech, Signal Processing*, vol ASSP-30, no. 4, pp. 578–586, Aug. 1982.

[8] N. Bershad and M. Bonnet, "Saturation Effects in LMS Adaptive Echo Cancellation for Binary Data", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-38, no. 10, pp. 1687–1696, Oct. 1990.

[9] O. Tanrikulu and J. A. Chambers, "Convergence and Steady-State Properties of the Least-Mean Mixed-Norm (LMMN) Adaptive Algorithm", *IEE Proc.-Vis Image Signal Process.*, vol. 143, no. 3, pp.137–142, June 1996.