The Optimum Error Nonlinearity in LMS Adaptation with an Independent and Identically Distributed Input

Tareq Y. Al-Naffouri
Electrical Engineering Department
Stanford University
Stanford, CA 94305
USA

Azzedine Zerguine & Maamar Bettayeb
Electrical Engineering Department
King Fahd University of Petroleum & Minerals
Dhahran 31261
Saudi Arabia

Abstract

The class of LMS algorithms employing a general error nonlinearity is considered. The calculus of variations is employed to obtain the optimum error nonlinearity for an independent and identically distributed input. The nonlinearity represents a unifying view of error nonlinearities in LMS adaptation. In particular, it subsumes two recently developed optimum nonlinearities for arbitrary and Gaussian inputs. Moreover, several more familiar algorithms such as the LMS algorithm, the least-mean fourth (LMF) algorithm and its family, and the mixed norm algorithm employ (non)linearities that are actually approximations of the optimum nonlinearity.

1 Introduction

The least-mean square (LMS) algorithm [1] is one of the most widely used adaptive schemes. It has several desirable features and some limitations. As such, several LMS-variants have been proposed that trade some of the LMS features for an enhanced performance in some of its limitations. Of particular importance is the class of least-mean square algorithms that employ an error nonlinearity $f(e(k))$ instead of the (linear) error term in LMS adaptation. Examples include the sign-error algorithm [2], the least-mean fourth (LMF) algorithm and its family [3], and the least-mean mixed norm algorithm [4], all of which are *intuitively motivated.* Table 1 defines $f(e(k))$ for many famous algorithms. Remember that the nonlinearity $f(e(k)) = sign[e(k)]$ will not be considered among the nonlinearities in the analysis since it does not meet one of the assumptions, namely, the smoothness assumption that will be defined later. Also, mentioned in Table 1 is $f(e(k)) = \alpha e(k) + (1-\alpha)e^3(k)$ which is the error nonlinearity used in the mixed LMS-LMF algorithm [4]-[5] with α as the mixing parameter. This algorithm is found to result in better performance than either the LMS or the LMF algorithms in Gaussian and non-Gaussian environments [4]-[5].

Algorithm	$f(e(k))$
LMS	$e(k)$
NLMS	$\dfrac{e(k)}{\|\mathbf{x}_k\|^2}$
Sign-LMS	$sign[e(k)]$
LMF	$e^3(k)$
Mixed LMS-LMF	$\alpha e(k) + (1-\alpha)e^3(k)$

Table 1: Examples for $f(e(k))$.

In contrast, *rigorous* variational methods were used in [6] to optimize the choice of the error-nonlinearity f. In particular, it was shown that for a general input

$$f_{opt}(y) = -\sigma_x^2 \frac{p'(y)}{p(y)}, \tag{1}$$

and for a white Gaussian input

$$f_{opt}(y) = -\frac{p'_{e(k)}(y)}{p_{e(k)}(y) + \mu\sigma_x^2 p''_{e(k)}(y)}, \tag{2}$$

where $p(y)$ is the pdf of the additive noise and $p_{e(k)}(y)$ is the pdf of the output error. In this paper, we employ this variational approach to obtain the optimum error-nonlinearity for an *independent and identically distributed (iid)* input. The nonlinearity arrived at correlates strongly with the optimum nonlinearities (1) and (2). Moreover, other more familiar nonlinearities turn out to be approximations of the derived optimum nonlinearity.

2 Analysis Model and Assumptions

Consider the class of least-mean adaptive algorithms with a general error nonlinearity f. 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

$$\begin{aligned} \boldsymbol{v}_{k+1} &= \boldsymbol{v}_k + \mu f(e(k))\boldsymbol{x}_k && (3) \\ e(k) &= n_k - \boldsymbol{v}_k^T \boldsymbol{x}_k, && (4) \end{aligned}$$

where $\boldsymbol{x}_k$ is the input data in the filter memory at time k of length L, $e(k)$ is the output error, and n_k is the additive noise. The performance of an algorithm of this class is mainly determined by the mean-square behavior of $\boldsymbol{v}_k$, i.e., by $E\left[\boldsymbol{v}_k^T \boldsymbol{v}_k\right]$. Thus, a successful formulation of the optimization problem hinges on obtaining a recursion for $\eta_k = E\left[\boldsymbol{v}_k^T \boldsymbol{v}_k\right]$ that applies for a *general* error nonlinearity. This in turn motivates the following assumptions which are quite similar to what is usually assumed in literature and which can also be justified in several practical instances:

A1 The input x_k is a zero-mean iid process.

A2 The noise n_k is a zero-mean iid process with a symmetric pdf $p(y)$, and is independent of the input process.

A3 The error nonlinearity f is odd-symmetric and sufficiently smooth.

A4 The step size is small enough for the independence assumption to be valid.

A5 The excess error $E\left[\left(\boldsymbol{v}_k^T \boldsymbol{x}_k\right)^2\right]$ is small enough for $f(e(k))$ to be approximated by a $2nd$-order Taylor series about the noise sample n_k, i.e.,

$$\begin{aligned} f(e(k)) &= f\left(n_k - \boldsymbol{v}_k^T \boldsymbol{x}_k\right) \\ &\simeq f(n_k) - f'(n_k)\left(\boldsymbol{v}_k^T \boldsymbol{x}_k\right) \\ &+ \frac{1}{2} f''(n_k)\left(\boldsymbol{v}_k^T \boldsymbol{x}_k\right)^2. \end{aligned}$$

While assumptions **A1-A3** can be justified in several practical instances, assumptions **A4** and **A5** can only be attained asymptotically. Many authors have made assumptions similar to **A2** and **A3** whether studying a general or a specific error nonlinearity. The popularity of the two assumptions can be understood by noting that they imply

$$E\left[f(n_k)\right] = E\left[f''(n_k)\right] = 0, \tag{5}$$

which makes it easier to carry out convergence analysis.

Assumptions **A2-A5** were used in [6] to characterize the performance of (3)-(4) and subsequently obtain the optimum error nonlinearity for an *arbitrary* input x_k. Assumption **A1** was also employed in [7] to more accurately study the convergence and performance of (3)-(4).

To formulate the optimization problem, we need only focus on the the second order behavior of $\boldsymbol{v}_k$ as reflected by $\eta_k = E[\boldsymbol{v}_k^T \boldsymbol{v}_k]$. As shown in [7]:

$$\begin{aligned} \eta_{k+1} = &\Big(1 + \lambda\mu^2 E[f'^2 + f f''] \\ &-2\sigma_x^2 \mu E[f']\Big)\,\eta(k) + \sigma_x^2 L E[f^2], \end{aligned} \tag{6}$$

where

$$\lambda = m_{x,4} + (L-1)\sigma_x^4 \tag{7}$$

and where the expectations $E[f^2], E[f']$, and $E[f'^2 + f f']$ are taken with respect to the noise n_k, σ_x^2 is the power of the input signal, and $m_{x,4}$ denotes the fourth order moment of the input signal.

3 The Optimum Solution

Recursion (6) is of the form

$$\eta_{k+1} = a\eta_k + b, \tag{8}$$

in which $(1-a)$ controls the convergence rate of $\eta(k)$ and, together with b, determines its steady-state value given by

$$\overline{\eta} = \lim_{k\to\infty} \eta_k = \frac{b}{1-a}. \tag{9}$$

Interestingly, Bershad in [8] arrived at a recursion that has the same form as (9) when f is the *data* rather than the *error* nonlinearity. He defined the optimum error nonlinearity as that which guarantees fastest transient behavior for a fixed steady state value. Instead, we choose the optimum nonlinearity as that which minimizes $\overline{\eta}$ for a fixed steady-state behavior. This can be done by minimizing b for a fixed value of $(1-a)$, i.e.,

$$\min_f \int_{-\infty}^{\infty} f^2(y)p(y)dy \tag{10}$$

subject to

$$\int_{-\infty}^{\infty} \Big[\mu\lambda(f'^2(y) + f(y)f''(y)) - 2\sigma_x^2 f'(y)\Big] p(y)dy = C. \tag{11}$$

To solve for the optimum nonlinearity, we resort to the calculus of variations. Thus, as demonstrated in the Appendix, the optimum nonlinearity is given by

$$f_{opt}(y) = -\frac{2\sigma_x^2\gamma p'(y)}{2p(y) + \mu\gamma\lambda p''(y)}, \tag{12}$$

where γ is a Lagrange multiplier associated with the constraint (11). To determine γ, we can substitute (12) into (11) and subsequently solve for γ. A simpler approach though follows by substituting f_{opt} into the adaptation equation (3) (see [6]). Notice then that γ and the step size μ always appear multiplied by each other. Thus, γ can be absorbed into the design parameter μ and the optimum nonlinearity becomes effectively

$$f_{opt}(y) = -\frac{\sigma_x^2 p'(y)}{p(y) + \frac{\mu\lambda}{2}p''(y)}. \tag{13}$$

Remarks

1. Note that for small values of μ, f_{opt} can be approximated as

   $$f_{opt}(y) = -\sigma_x^2\frac{p'(y)}{p(y)}, \tag{14}$$

   which is the same as the optimum error-nonlinearity (1) obtained in [6] for an arbitrary input.

2. If the input x_k is further restricted to be Gaussian, then $m_{x,4} = 3\sigma_x^4$ and the nonlinearity reads

   $$f_{opt}(y) = -\frac{\sigma_x^2 p'(y)}{p(y) + \mu\sigma_x^4\frac{L+2}{2}p''(y)}. \tag{15}$$

   Upon replacing $\mu\frac{\sigma_x^2(L+2)}{2}$ by μ, and by noting that for large k, $e(k) \simeq n_k$ [3], [6] so that $p(y) \simeq p_{e(k)}(y)$, f_{opt} becomes

   $$f_{opt}(y) \simeq -\sigma_x^2\frac{p'_{e(k)}(y)}{p_{e(k)}(y) + \mu\sigma_x^2 p''_{e(k)}(y)}. \tag{16}$$

   This is the optimum nonlinearity (2) (up to a scalar multiple) developed in [6] by a conditional analysis approach.

3. The optimum nonlinearity (13) can also be related to more familiar nonlinearities. This can be shown, as demonstrated in [9], by approximating the pdf $p(y)$ in (13) by a Gram-Charlier series. As a result, f_{opt} reads approximately

   $$f_{opt}(y) \simeq c_1 y + c_3 y^3 + c_5 y^5 + c_7 y^7, \tag{17}$$

   where the c_i's are a function of the noise moments. Thus, the error (non)linearities in such algorithms as the LMS, the LMF algorithm and its family [3], and the mixed norm algorithm [4] are simply approximations of the optimum nonlinearity f_{opt}. This justifies the use of these algorithms.

4. The optimum nonlinearity (13) is difficult to implement because the pdf $p(y)$ is usually unknown or time varying. The approximation (17) does away with this problem by trading pdf estimation for the estimation of the c_i's. This in turn is easier to carry out as the c_i's can be explicitly expressed in terms of the noise moments (see [9]).

4 Conclusion

In this paper, the class of LMS algorithms with a general error nonlinearity was considered. The optimum error nonlinearity for an iid input was derived. The simplifying iid assumption made it possible to arrive at a more accurate description of the nonlinearity. As a result, the optimum nonlinearity subsumes two recently developed optimum nonlinearities. Moreover, the LMS algorithm and several of its error-modified variants employ nonlinearities that are actually approximations of the optimum nonlinearity. Thus, the optimum nonlinearity represents a unifying view of error nonlinearities (optimum or otherwise) in LMS adaptation.

Appendix

Here we solve the variational problem (10)-(11). The associated composite functional is

$$\psi(y,f,f',f'') = \gamma\left[\mu\lambda\left(f(y)f''(y)+f'^2(y)\right) - 2\,\sigma_x^2 f'(y)\right]p(y) + f^2(y)p(y), \quad (18)$$

where γ is a Lagrange multiplier corresponding to the constraint (11). The desired nonlinearity is obtained by solving the the following Euler-Lagrange differential equation

$$-\frac{d^2\psi_{f''}}{dy^2} + \frac{d\psi_{f'}}{dy} - \psi_f = 0. \quad (19)$$

By differentiating (18), we get

$$\begin{aligned}
\psi_f &= \mu\gamma\lambda f''(y)p(y) + 2f(y)p(y),\\
\psi_{f'} &= 2\mu\gamma\lambda f'(y)p(y) - 2\gamma\sigma_x^2 p(y),\\
\frac{d\psi_{f'}}{dy} &= 2\mu\gamma\lambda\left(f''(y)p(y) + f'(y)p'(y)\right) - 2\gamma\sigma_x^2 p(y),\\
\psi_{f''} &= \mu\gamma\lambda f(y)p(y),\\
\frac{d\psi_{f''}}{dy} &= \mu\gamma\lambda\left(f'(y)p(y) + f(y)p'(y)\right),\\
\frac{d^2\psi_{f''}}{dy^2} &= \mu\gamma\lambda\left(f''(y) + 2f'(y)p'(y) + f(y)p''(y)\right),
\end{aligned}$$

and upon substituting the relevant terms into (19) we obtain

$$\left[2p(y) + \mu\gamma\lambda p''(y)\right] f(y) + 2\gamma\sigma_x^2 p'(y) = 0. \quad (20)$$

Solving for $f(y)$, we obtain the optimum nonlinearity

$$f_{opt}(y) = -\frac{2\gamma\sigma_x^2 p'(y)}{2p(y) + \mu\gamma\lambda p''(y)}. \quad (21)$$

Acknowledgement: The authors would like to acknowledge KFUPM for its support. The second author acknowledges KFUPM for the support received for the University project **PH/ECHO193**.

References

[1] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter," *Proc. IEEE, vol.* 64, No. 8, pp. 1151-1162, Aug. 1976.

[2] H. Sari, "Performance Evaluation of three Adaptive Equalization Algorithms ," *ICASSP*, pp. 1385-1389, May 1983.

[3] 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.

[4] O. Tanrikulu and J. A. Chambers, "Convergence and Steady-State Properties of the Least mean mixed-norm (LMMN) adaptive filtering," *IEE Proc.-Vis. Image Signal Process., Vol.* 143, No. 3, pp. 137-142, June 1996.

[5] T. Aboulnasr and A. Zerguine, "Variable Weight Mixed-Norm LMS-LMF Adaptive Algorithm," *Proc. 33rd Annual Asilomar Conf. Signals, Syst., Comput.*, (Pacific Grove, CA), Oct. 1999.

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

[7] T. Y. Al-Naffouri, A. Zerguine, and M. Bettayeb, "Convergence Analysis of the LMS Algorithm with a General Error Nonlinearity and an IID Input", *Proc. 32nd Annu Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, pp. 556-559, Nov. 1998.

[8] N. J. Bershad, "On the Optimum Data Nonlinearity in LMS Adaptation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 1, pp. 69-76, Feb. 1986.

[9] T. Y. Al-Naffouri, A. Zerguine, and M. Bettayeb, "A Unifying View of Error Nonlinearities in LMS Adaptation", Proc. *ICASSP*, Seattle, pp. 1697-1700, May 12-15, 1998.