CONVERGENCE ANALYSIS OF A MIXED CONTROLLED $L_2 - L_P$ ADAPTIVE ALGORITHM

Abdelmalek Zidouri

Department of Electrical Engineering,
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
Phone: + (966) 38603677,
Fax: + (966) 38603535,
Email: malek@kfupm.edu.sa
Web: faculty.kfupm.edu.sa/ee/malek

ABSTRACT

A new adaptive scheme for system identification is proposed. The derivation of the algorithm and its convexity property are detailed. Also, the first moment behaviour as well as the second moment behaviour of the weights are studied. Bounds for the step size on the convergence of the proposed algorithm are derived, as well as the steady-state analysis is carried out. Finally, simulation results are performed and are found to corroborate with the theory developed.

Keywords: Mixed norm, LMS Algorithm, l_p-norm, Adaptive Algorithm, Convergence.

1. INTRODUCTION

The least mean square (LMS) algorithm [1] is one of the most widely used adaptive schemes. Several works have been presented using the LMS or its variants, such as signed LMS [2]-[3], the least mean fourth (LMF) algorithm and its variants [4], or the mixed LMS-LMF [5]-[7] all of which are intuitively motivated.

The LMS algorithm is optimum only if the noise statistics are Gaussian. However, if these statistics are different from Gaussian other criteria, such as l_p-norm ($p \neq 2$), perform better than the LMS algorithm, [4], [8], [9]. The idea here is to use a mixed controlled $l_2 - l_p$ adaptive algorithm. This is similar to that given in [7]:

$$J_n = \alpha E[e_n^2] + (1-\alpha)E[e_n^4], \tag{1}$$

where the error is defined as follows:

$$e_n = d_n + w_n - \mathbf{c}_n^T \mathbf{x}_n, \tag{2}$$

d_n is the desired value, $\mathbf{c}_n$ is the filter coefficient of the adaptive filter, $\mathbf{x}_n$ is the input vector, w_n is the additive noise, and α is the mixing parameter between zero and one and set in this range to preserve the unimodal character of the cost function. It is clear from (1) that if $\alpha = 1$ the algorithm reduces to the LMS algorithm, if however, $\alpha = 0$ the algorithm is the least-meam fourth (LMF) [4]. Any choice for α in the interval (0,1) enhances the performance of the algorithm.

The algorithm for adjusting the tap coefficients, $\mathbf{c}_n$, is given by the following recursion:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mu\{\alpha + 2(1-\alpha)e_n^2\}e_n\mathbf{x}_n. \tag{3}$$

Adaptive filter algorithms designed through the minimization of equation (1) have a disadvantage when the absolute value of the error is greater than one. This will make the algorithm go unstable unless either a small value of the step size or a large value of the controlling parameter are chosen such that this unwanted instability disappears. l_p-norm based minimization algorithms for signal parameter estimation or minimization algorithm of mixed l_1 and l_2 norms can be found in literature, [12], [13].

Unfortunately, a small value of the step size will make the algorithm converge very slowly, and a large value of the controlling parameter will make the LMS algorithm essentially dominant.

The rest of the paper is organized as follows: Section 2 describes the proposed algorithm as well as its convexity property. In Section 3 the convergence analysis is detailed, while Section 4 the simulations results reports the performance behavior of the proposed algorithm. Finally, Section 5 concludes the work reported in this paper.

2. PROPOSED ALGORITHM

To overcome the above mentioned problem a modified approach is proposed where both constraints of the step size and the control parameter are eliminated. The proposed criterion consists of the cost function (1) where the l_p-norm is substituted for the l_4-norm. Ultimately, this should eliminate the instability in the l_4-norm, especially if $p < 4$, and retains the good features of (1), i.e., the mixed nature of the criterion. The proposed scheme is defined as:

$$J_n = \alpha E[e_n^2] + (1-\alpha)E[|e_n|^p], \quad p \geq 1, \tag{4}$$

If $p = 2$, the cost function defined by (4) reduces to the LMS algorithm whatever the value of α in the range [0,1] for which the unimodality of the cost function is preserved.

For $\alpha = 0$, the algorithm reduces to the l_p-norm adaptive algorithm, and moreover if $p = 1$ results in the familiar signed LMS algorithm.

For $p < 2$, l_p gives less weight for larger error and this tends to reduce the influence of aberrant noise, while it gives relatively larger weight to smaller errors and this will improve the tracking capability of the algorithm [14].

 ISSN 2076-1465

2.1 Convex property of performance function

We can prove that the performance function

$$J(c) = \alpha E[e_n^2] + (1-\alpha)E[|e_n|^p] \tag{5}$$

is a convex function defined on $\mathbf{R}^{(N_1+N_2)}$ for $p \geq 1$.

2.2 Analysis of the error surface

• Case $p = 2$:

Let the input autocorrelation matrix be $\mathbf{R} = E[\mathbf{x}_n \mathbf{x}_n^T]$, and the crosscorrelation vector that describes the cross-correlation between the received signal ($\mathbf{x}_n$) and the desired data ($\hat{d}_n$), $\mathbf{p} = E[\mathbf{x}_n d_n]$. The error function can be more conveniently expressed as follows:

$$J_n = \sigma_x^2 - 2\mathbf{c}_n^T \mathbf{p} + \mathbf{c}_n^T \mathbf{R} \mathbf{c}_n. \tag{6}$$

It is clear from (6) that the mean-square error (MSE) is precisely a quadratic function of the components of the tap coefficients, and the shape associated with it is hyperparaboloid. The adaptive process continuously adjusts the tap coefficients, seeking the bottom of this hyperparaboloid.

$$\mathbf{c}_{opt} = \mathbf{R}^{-1}\mathbf{p} \tag{7}$$

• Case $p \neq 2$:

It can be shown as well that the error-function for the feedback section will have a global minimum since the latter one is a convex function. As in the feedforward section, the adaptive process will continuously seek the bottom of the error-function of the feedback section.

2.3 The updating scheme

The updating scheme is given by:

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mu[\alpha e_n + p(1-\alpha)|e_n|^{(p-1)} sign(e_n)]\mathbf{x}_n, \tag{8}$$

and sufficient condition for convergence in the mean of the proposed algorithm can be shown to be given by:

$$0 < \mu < \frac{2}{\{\alpha + p(p-1)(1-\alpha)E[|w_n|^{p-2}]\}tr\{\mathbf{R}\}} \tag{9}$$

where $tr\{\mathbf{R}\}$ is the trace operation of the autocorrelation matrix $\mathbf{R}$.

In general, the step size is chosen small enough to ensure convergence of the iterative procedure and produce less misadjustment error. In the ensuing, the convergence analysis of the proposed $l_2 - l_p$ is carried out.

3. CONVERGENCE ANALYSIS

Usual assumptions [8], [10], [11], [9] that can be found in literature and which can also be justified in several practical instances are used during the convergence analysis of the proposed mixed controlled $l_2 - l_p$ algorithm.

3.1 First Moment Behavior of the Weights

We start by evaluating the statistical expectation of both sides of (8) which looks after subtracting $\mathbf{c}_{opt}$ of both sides to give:

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mu[\alpha e_n + (1-\alpha) sign(e_n)]\mathbf{x}_n. \tag{10}$$

After substituting the error e_n defined by Equation (2) in the above equation and taking the expectation of both sides results in:

$$E[\mathbf{v}_{n+1}] = [\mathbf{I} - \alpha\mu\mathbf{R}]\, E[\mathbf{v}_n] + \mu(1-\alpha)E\left[\mathbf{x}_n sign(e_n)\right] \tag{11}$$

It is to show that the mis-alignment vector will converge to the zero vector if the step-size, μ, is given by

$$0 < \mu < \frac{2}{\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi J_{min}}}\right]\lambda_{max}}, \tag{12}$$

where λ_{max} is the largest eigenvalue of the autocorrelation matrix $\mathbf{R}$, since in general $tr\{\mathbf{R}\} >> \lambda_{max}$, and J_{min} is the minimum MSE.

3.2 Second Moment Behavior of the Weights

From Equation (10) we get the following expression:

$$\begin{aligned}\mathbf{v}_{n+1}\mathbf{v}_{n+1}^T &= \mathbf{v}_n\mathbf{v}_n^T + \mu\left[\alpha e_n + (1-\alpha)sign(e_n)\right] \\ &\times \left[\mathbf{v}_n\mathbf{x}_n^T + \mathbf{x}_n\mathbf{v}_n^T\right] + \mu^2 \\ &\times \left[\alpha^2 e_n^2 + 2\alpha(1-\alpha)|e_n| + (1-\alpha)^2\right]\mathbf{x}_n\mathbf{x}_n^T. \end{aligned} \tag{13}$$

Let $\mathbf{K}_n = E[\mathbf{v}_n\mathbf{v}_n^T]$ define the second moment of the misalignment vector, therefore the above equation, after taking the expectation of both of its sides, becomes as follows:

$$\begin{aligned}\mathbf{K}_{n+1} &= \mathbf{K}_n + \mu\alpha\left\{E\left[\mathbf{v}_n\mathbf{x}_n^T e_n\right] + E\left[\mathbf{x}_n\mathbf{v}_n^T e_n\right]\right\} \\ &+\mu(1-\alpha)\left\{E\left[\mathbf{v}_n\mathbf{x}_n^T sign(e_n)\right] + E\left[\mathbf{x}_n\mathbf{v}_n^T sign(e_n)\right]\right\} \\ &+\mu^2\left\{\alpha^2 E\left[\mathbf{x}_n\mathbf{x}_n^T e_n^2\right] + 2\alpha(1-\alpha)E\left[\mathbf{x}_n\mathbf{x}_n^T|e_n|\right]\right. \\ &\left.+(1-\alpha)^2\mathbf{R}\right\}. \end{aligned} \tag{14}$$

Two cases can be considered for the step size μ so that the weight vector converges in the mean square sense.

• First case $i \neq j$:

In this case, a sufficient condition for mean square convergence can be shown to be given by the following:

$$0 < \mu < \frac{1}{\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi J_{min}}}\right] tr\{\mathbf{R}\}}. \tag{15}$$

• Second case $i = j$:

Whereas here, the convergence in the mean square sense will be given by:

$$0 < \mu < \frac{2\left[\alpha + (1-\alpha)\sqrt{\frac{2}{\pi}\frac{1}{\sigma_{e_n}}}\right]}{\left[\alpha^2 - 2\alpha(1-\alpha)\sqrt{\frac{2}{\pi}\frac{1}{\sigma_{e_n}}}\right]\lambda_i}. \tag{16}$$

Discussion:

Note that if $\alpha = 0$ will result in zero in the denominator of expression (16) and therefore will make μ take any value in the range of positive numbers, a contradiction with the ranges of values for the step sizes of LMS and LMF algorithms. Moreover, any value for α in $]0, 1[$ will make of the step size μ set by (16) less than zero, also this condition is discarded. This concludes that it is safer to use a more realistic range for the step size μ for convergence in the mean square dictated by the range of (15) which will guarantee stability regardless of the value of α, and therefore will be considered here.

4. SIMULATION RESULTS

In this section, the performance analysis of the proposed mixed controlled $l_2 - l_p$ adaptive algorithm is investigated in an unknown system identification problem for different values of p and different values of the mixing parameter α. The simulations reported here are based on an FIR channel system identification. Furthermore, we have considered the following channel:

$$\mathbf{c}_{opt} = [0.227, 0.460, 0.688, 0.460, 0.227]^T.$$

Three different noise environments have been considered namely gaussian, uniform, and laplacian. The length of the adaptive filter is the same as that of the unknown system. The learning curves are obtained by averaging 600 independent runs. Two scenarios are considered for the case of the value of p, i.e., $p = 1$ and $p = 4$. The signal-to-noise ratio (SNR) is set to 10 dB throughout the simulations. Finally, the performance measure most appropriate to system identification problem considered here is the normalised weight error norm defined

$$10log_{10}\frac{||\mathbf{c}(n) - \mathbf{c}_{opt}||^2}{||\mathbf{c}_{opt}||^2},$$

where $\mathbf{c}_{opt}$ is the impulse response of the unknown system. The learning curves obtained are the average of 600 runs.

Figure 1 depicts the convergence behavior of the proposed algorithm for different values of α and $p = 1$ in a white Gaussian noise, Laplacian noise, and uniform noise, respectively. As can be depicted from this figure the best performance is obtained when $\alpha = 0.8$. This makes sense as the resultant algorithm is steering towards the LMS algorithm.

Also, as can be seen from these figures that the best performance, as far as the noise statistics are concerned, is obtained when the noise environment is Laplacian, then gaussian and finally uniform. This makes sense as the update algorithm is biased to the sign error LMS algorithm as α approaches zero. If one compares the performance of the proposed algorithm when the noise statistics are Laplacian, one sees clearly that an enhancement in performance is obtained and about a 2dB improvement is achieved for all values of α.

The situation changes when $p = 4$ as reported in Fig. 2 which depicts the convergence behavior of the proposed algorithm for different values of α in a white Gaussian noise, Laplacian noise, and uniform noise, respectively. As can be depicted from this figure, the best performance is obtained when $\alpha = 0.2$ as the proposed algorithm is mostly LMF in this case. More importantly, the best noise statistics for this scenario is when the noise is uniformly distributed. Similarly as above if one compares the performance of the proposed algorithm when the noise statistics are uniform, one sees clearly that an enhancement in performance is obtained and about a 2dB improvement is achieved for all values of α. Also, one can notice that the worst performance is obtained when the noise is Laplacian distributed.

Next, to assess further the performance of the proposed algorithm for the same steady-state value, two different cases are considered, for $p = 1$ and $p = 4$ when $\alpha = 0.8$. Figure 3 illustrates the learning behavior of the proposed algorithm for $p = 1$. As can be seen from this figure that the best performance is obtained with Laplacian noise while the worst performance is obtained with Uniform noise environment.

In the case of $p = 4$, as reported in Fig. 4 which depicts the learning behavior of the proposed algorithm in the different noise environments, it can be seen that the best performance is obtained with Uniform noise environment. The Laplacian noise results in the worst performance when compared to Gaussian and Uniform noise environments.

5. CONCLUSION

A new adaptive scheme for system identification has been introduced. The derivation of the algorithm is worked out and the first moment behaviour as well as the second moment behaviour of the weights are studied. Bounds for the step size on the convergence of the proposed algorithm are derived. Finally, simulations are found to be in good agreement with the theory developed here.

Acknowledgments

The author acknowledges the support of King Fahd University of Petroleum & Minerals.

REFERENCES

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

[2] E. Eweda, "Convergence of the sign algorithm for adaptive filtering with correlated data," *IEEE Transactions on Information Theory*, pp. 1450 - 1457, vol 37, Sept. 1991.

[3] E. Eweda, "Convergence analysis of the sign algorithm without the independence and gaussian assumptions," *IEEE Transactions on Signal Processing* Volume 48, Issue 9, Sept. 2000 Page(s):2535 - 2544

[4] E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Trans-*

actions on Information Theory, vol. 30, no. 2, pp. 275-283, February 1984.

[5] 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.

[6] A. Zerguine, C. F. N. Cowan and M. Bettayeb , "LMS-LMF Adaptive Scheme for Echo Cancellation," *Electron. Lett.*, 32 (19), pp. 1776-1778, September 1996.

[7] A. Zerguine, C. F. N. Cowan and M. Bettayeb , "Adaptive echo cancellation using least mean mixed-norm algorithm," *IEEE transactions on Signal Processing*, Vol. 45, no. 5, pp. 1340–1342, May 1997.

[8] S. Haykin, *Adaptive Filter Theory*, 4th ed., Prentice-Hall, Upper-Saddle River, NJ, 2002.

[9] A. H. Sayed, *Fundamentals of Adaptive Filtering*, New York: Wiley-Interscience, 2003.

[10] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.,* No. 58, pp. 963-993, May-June 1979.

[11] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission*, West Sussex, England: John Wiley & Sons Ltd, 1995.

[12] S., Alliney and S.A., Ruzinsky, "an algorithm for the minimization of mixed l1 and l2 norms with application to Bayesian estimation" *IEEE Transactions on Signal Processing* Volume 42, Issue 3, Pages:618 - 627, March 1994.

[13] H. Zhang and Y. Peng; "Lp-norm based minimisation algorithm for signal parameter estimation" *Electronics Letters* Volume 35, Issue 20, Pages:1704 - 1705, 30 September 1999.

[14] S. Siu, G. J. Gibson, and C. F. N. Cowan, "Decision Feedback Equalisation using Neural Network Structures and Performance Comparison with Standard Architecture," *IEE* Proc. I, vol. 137, No. 4, pp. 221-225, August 1990.

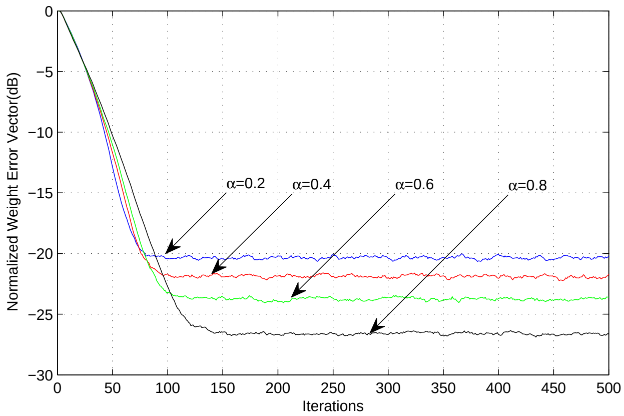

Figure 1: Effect of α on the learning curves of the proposed algorithm in an AWGN noise environment scenario for $p = 1$.

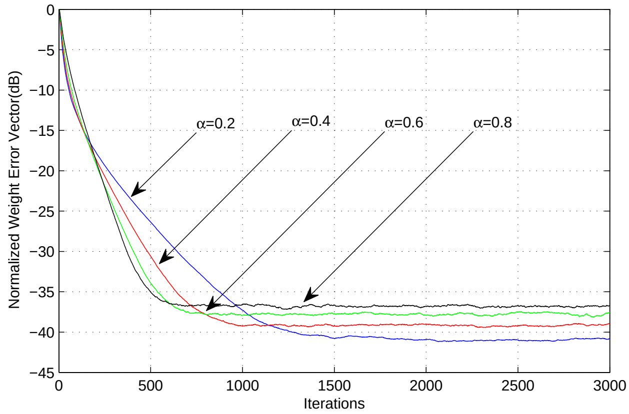

Figure 2: Effect of α on the learning curves of the proposed algorithm in an AWGN noise environment scenario for $p = 4$.

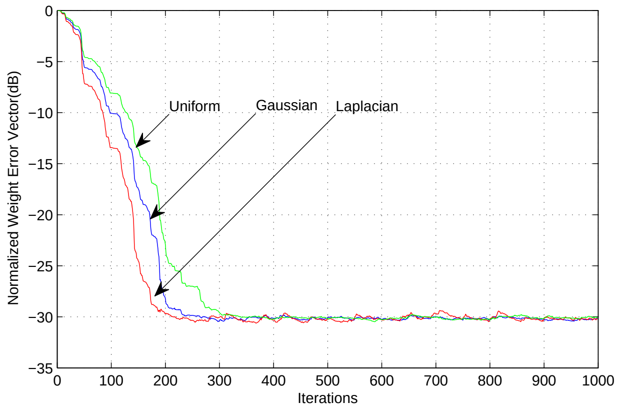

Figure 3: Learning behavior of the proposed algorithm in the different noise environments scenario for $p = 1$ and $\alpha = 0.8$.

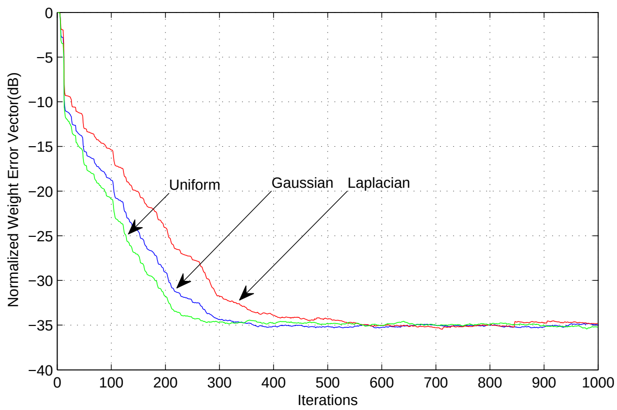

Figure 4: Learning behavior of the proposed algorithm in the different noise environments scenario for $p = 4$ and $\alpha = 0.8$.