A NEW ROBUST ADAPTIVE STEP SIZE LMS ALGORITHM

Dimitrios I. Pazaitis* and Anthony G. Constantinides
Signal Processing Section, Department of Electrical and Electronic Engineering,
Imperial College, Exhibition Road, London SW7 2BT, England
Tel.: +44 171 5946231 Fax: +44 171 5946234
e-mail : {d.pazaitis, a.constantinides}@ic.ac.uk

ABSTRACT

In this contribution a new robust technique for adjusting the step size of the Least Mean Squares (LMS) adaptive algorithm is introduced. The proposed method exhibits faster convergence, enhanced tracking ability and lower steady state excess error compared to the fixed step size LMS and other previously developed variable step size algorithms, while retaining much of the LMS computational simplicity.

A theoretical behaviour analysis is conducted and equations regarding the evolution of the weight error vector correlation matrix together with convergence bounds are established. Extensive simulation results support the theoretical analysis and confirm the desirable characteristics of the proposed algorithm.

1 INTRODUCTION

Stochastic gradient adaptive filters using the Least Mean Squares (LMS) algorithm [10], enjoy great popularity due to their inherent simplicity and robustness. A constant step size, also known as convergence factor, μ governs the convergence properties, the stability and the steady state excess error of the algorithm in relation to the optimal Wiener solution. The time constant of the algorithm is inversely proportional to the step size μ, whereas the misadjustment is proportional to μN, where N equals the number of the weights of the adaptive filter [10]. When a nonstationary environment is considered, the *lag misadjustment* is proportional to μ^{-1} [10]. Consequently, the optimal selection of the step size is dictated by a trade-off between convergence rate and steady state excess error. To meet the conflicting requirements of good tracking ability and low steady state excess error, time varying step size sequences have been proposed (e.g.,[1, 2, 4, 6]). The main rationale behind these approaches is to sense the distance from the optimum and correspondigly adapt the step size value. It seems to us, however, that the performance of such methods is adversely affected by the presence of strong noise in the system and is thus susceptible to degradation. Recently, a kurtosis driven step size sequence was proposed [7], reducing significantly performance degradation due to gaussian distributed noise.

In this paper a new robust time varying step size selection method is proposed that combines, even under adverse noise conditions, fast tracking ability and low steady state error regardless of the noise type and retains the LMS desirable characteristic of low computational complexity.

*This work was supported by a scholarship from the State Scholarships Foundation (I.K.Y) of the Hellenic Republic.

2 A ROBUST ADAPTIVE STEP SIZE LMS ALGORITHM

The popular Least Mean Squares (LMS) adaptive algorithm is a steepest descent gradient search type algorithm, which uses at each iteration the instantaneous value of the gradient and seeks to minimize the mean square error $E\{e^2(n)\}$. The weight vector is updated according to the following equation

$$\boldsymbol{H_{n+1}} = \boldsymbol{H_n} + \mu_n \boldsymbol{X_n} e(n), \tag{1}$$

$$e_n \stackrel{\triangle}{=} w_n - \boldsymbol{V_n^T X_n} \tag{2}$$

where μ_n is the time varying step size, $\boldsymbol{H_n}$ and $\boldsymbol{H_n^{opt}}$ are the filter coefficient (weight) and the optimal weight vectors respectively, $\boldsymbol{X_n}$ is the input vector, w_n is the additive noise and $\boldsymbol{V_n} = \boldsymbol{H_n} - \boldsymbol{H_n^{opt}}$ is the weight error vector, all at time instant n.

In its original form [10] the step size assumes a constant value. In this contribution we propose to adaptively adjust the step size according to the follwing rule

$$\mu_{n+1} = \alpha |e_n| E\{e_n^2\}, \tag{3}$$

where α is a scaling constant and $|\cdot|$ stands for the absolute value of the bracketed term. The averaged square error term reduces the vulnerability of the algorithm to spikey or burstly interferences, whereas the form of the update, i.e., a product, increases the flexibility of the algorithm.

3 CONVERGENCE OF THE ALGORITHM

To make our analysis tractable we introduce the assumption that the various input vectors come from mutually independent zero mean gaussian distributed sequences. Though far from reality in most practical applications, this assumption is commonly found in the literature (e.g., [3, 4, 5, 10]) and is widely accepted to capture the first order behaviour of the algorithm, validating thus its use under slow convergence conditions. To strengthen the above argument, practice has shown that the results obtained under this assumption are in excellent agreement with the ones obtained from experiments and simulations.

In our case, the assumption can be relaxed, in that we only require that the input sequence is uncorrelated with the filter weights. Other than that, no restriction applies to the nature of the input autocorrelation matrix $\boldsymbol{R}$.

The estimation error e_n is assumed to follow a gaussian distribution and conditional expectation terms of the form $E\left\{e_n^2 \middle| \boldsymbol{V_n}\right\}$ are approximated with the unconditional mean square estimation error [5].

Letting $\boldsymbol{K_n}$ denote the correlation matrix of the weight error vector $\boldsymbol{V_n}$ at time instant n, the following time evolution formula is obtained

$$\begin{aligned} \boldsymbol{K_{n+1}} &= \boldsymbol{K_n} + E\{\mu_n e_n \boldsymbol{X_n} \boldsymbol{V_n^T}\} + E\{e_n \mu_n \boldsymbol{V_n} \boldsymbol{X_n^T}\} \\ &+ E\{\mu_n^2 e_n^2 \boldsymbol{X_n} \boldsymbol{X_n^T}\}. \end{aligned} \quad (4)$$

Introducing (2) in the above equation, we get

$$\begin{aligned} \boldsymbol{K_{n+1}} = \boldsymbol{K_n} &+ E\left\{\alpha \sigma_{e_n}^2 sgn\{e_n\} e_n^2 \left(\boldsymbol{X_n V_n^T} + \boldsymbol{V_n X_n^T}\right)\right\} \\ &+ E\left\{\alpha^2 \sigma_{e_n}^4 e_n^4 \boldsymbol{X_n X_n^T}\right\}, \end{aligned} \quad (5)$$

where

$$\sigma_{e_n}^2 = \sigma_w^2 + tr[\boldsymbol{R K_n}], \quad (6)$$

denotes the variance of the assumed zero mean gaussian error, σ_w^2 stands for the noise variance, $sgn\{\cdot\}$ is the signum function and $tr[\cdot]$ denotes the trace of the bracketed term.

Taking conditional on $\boldsymbol{V_n}$ expectation of the second term on the right hand part of the above equation and applying Price's theorem, we obtain

$$\begin{aligned} &E\left\{sgn\{e_n\} e_n^2 \left(\boldsymbol{X_n V_n^T} + \boldsymbol{V_n X_n^T}\right) \middle| \boldsymbol{V_n}\right\} = \\ &= -2\sigma_{e_n} \sqrt{\frac{2}{\pi}} \left(\boldsymbol{R V_n V_n^T} + \boldsymbol{V_n V_n^T R}\right). \end{aligned} \quad (7)$$

Taking the expectation of both sides - this time over $\boldsymbol{V_n}$- we get

$$\begin{aligned} &E\left\{sgn\{e_n\} e_n^2 \left(\boldsymbol{X_n V_n^T} + \boldsymbol{V_n X_n^T}\right)\right\} = \\ &= -2\sigma_{e_n} \sqrt{\frac{2}{\pi}} \left(\boldsymbol{R K_n} + \boldsymbol{K_n R}\right). \end{aligned} \quad (8)$$

The term $E\left\{e_n^4 \boldsymbol{X_n X_n^T}\right\}$ is similarly evaluated and is given by [8]

$$E\{e_n^4 \boldsymbol{X_n X_n^T}\} = 3\sigma_e^4 \boldsymbol{R} + 12\sigma_e^2 \boldsymbol{R K_n R} \quad (9)$$

We may now combine all the previous results, substitute (8) and (9) in (5) to obtain

$$\begin{aligned} \boldsymbol{K_{n+1}} &= \boldsymbol{K_n} - 2\alpha \sqrt{\frac{2}{\pi}} \sigma_{e_n}^3 \left(\boldsymbol{R K_n} + \boldsymbol{K_n R}\right) \\ &+ \alpha^2 \sigma_e^4 \left(3\sigma_e^4 \boldsymbol{R} + 12\sigma_e^2 \boldsymbol{R K_n R}\right). \end{aligned} \quad (10)$$

The matrix difference equation (10) provides us a useful tool for determining the transient behaviour of the mean square error of the variable step size LMS algorithm. We next proceed to establish the stability conditions.

Due to the nonlinear form of (10) an exact convergence condition is difficult to find. We thus introduce the concept of the distance T_n [9],

$$T_n = tr[\boldsymbol{R K_n}] = E\left\{\left(\boldsymbol{V_n^T X_n}\right)^2\right\}, \quad (11)$$

which equals the excess mean squared error and can be thought of as a measure of the distance from the optimum. Following the approach in [3], the following distance dependent sufficient condition for convergence of the algorithm in the mean square sense is derived

$$0 < \alpha_{max} < \frac{4\sqrt{\frac{2}{\pi}}\left[\sigma_w^2 + T_n\right]^{\frac{3}{2}}}{I \cdot tr[\boldsymbol{R}]}, \quad (12)$$

where

$$I = 3(\sigma_w^2 + T_n)^3 + 3\sigma_w^2(3\sigma_w^4 + 3\sigma_w^2 T_n + T_n^2). \quad (13)$$

4 SIMULATION RESULTS

In this final section we present and analyse the results obtained from simulations. The algorithm is applied to a system identification problem, where the system to be identified is considered non stationary. The optimum filter coefficients assume the following initial values $\boldsymbol{H_0^{opt}} = [0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0,\ 1.0,\ 0.8,\ 0.6,\ 0.4,\ 0.2]$, and, after that, experience random disturbances. The null vector ($\boldsymbol{0}$) is chosen as the initial vector - starting point - $\boldsymbol{H_0}$, and all the results are obtained by averaging over an ensemble of 200 runs. Both noise and input sequences are assumed to be zero mean i.i.d. gaussian sequences with input and noise variance equal to *unity* and 0.1 respectively.

For the esimation of the mean square error the following formula is used

$$E\{e_n^2\} = \beta E\{e_{n-1}^2\} + (1-\beta) e_n^2, \quad (14)$$

where the constant β, $\beta \in [0, 1]$, is a memory controlling factor. The larger the value of β the "stronger" the memory of the system. Alternatively, a finite length moving window could be used.

As is the case with all the estimators, the estimator (14) produces a random value. In a stationary environment, the larger the value of β, the smaller the difference between the estimated and the actual parameters. Under non stationary conditions, however, the choise of β is dictated by a trade-off betwen the adaptation speed and the variance of the estimator. In our simulations, β was chosen equal to 0.985.

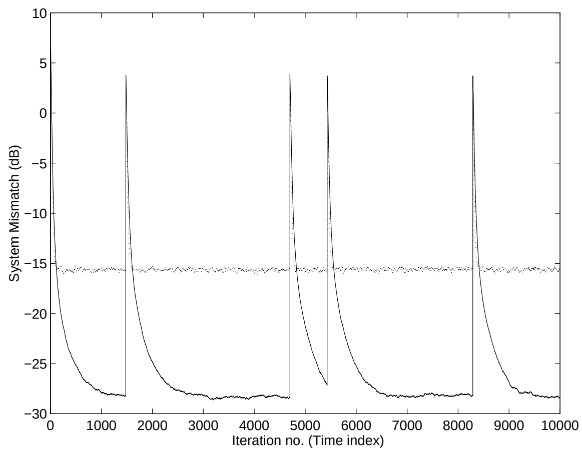

Figure 1: Performance comparison between the LMS (dotted line) and the proposed algorithm (solid line); $\beta = 0.985$, $\alpha = 0.05$, $\mu_{LMS} = 0.045$.

Figure 1 depicts the behaviour of the variable step size algorithm in comparison with the LMS algorithm. The algorithms' parameters were chosen so that the two algorithms exhibit similar convergence rate and are as follows : $\alpha = 0.05$, $\mu_{LMS} = 0.045$. The selection of the parameter α is dictated by a compromise between convergence speed and steady state excess error. The selection of the parameter α is thus application dependent. In non stationary environments, where abrupt changes are frequently experienced, it is advisable to choose a high value for α, so as to improve the algorithm's response speed and tracking ability. On the other hand, in the presence of long stationary periods, rarely interrupted by changes, a small value for α is doubtless more beneficial. The sudden changes that experiences the system are due to zero mean uniformly distributed disturbances with variance $\sigma_A^2 = 0.25$, at random the first time but fixed afterwards (i.e., for the following runs) time instants. The *System Mismatch*, $E\{\boldsymbol{V_n^T V_n}\}$, is chosen as the performance measure and is expressed in dB.

The performance of the adaptive step size sequence algorithm is clearly superior to that of LMS, which is outperformed by more than 10 dB. This is due to the time varying nature of μ_n, which allows the system to increase its step size whenever far from the optimum.

In figure 2 the variable step size algorithm is compared with two other variable step size algorithms developed by Harris *et. al* [2] and Mathews and Xie [6]. The parameters of the algorithms were chosen as recommended in the corresponding papers, i.e., $\alpha = 2.0, m_0 = m_1 = 2, \mu_{max} = 0.07$ and $\mu_{min} = 1E-8$ for Harris' algorithm and $\rho = 0.0008$, $\mu_{max} = 0.1$ and $\mu_0 = 0.06$ for Mathews' algorithm. The system disturbances are uniformly dis-

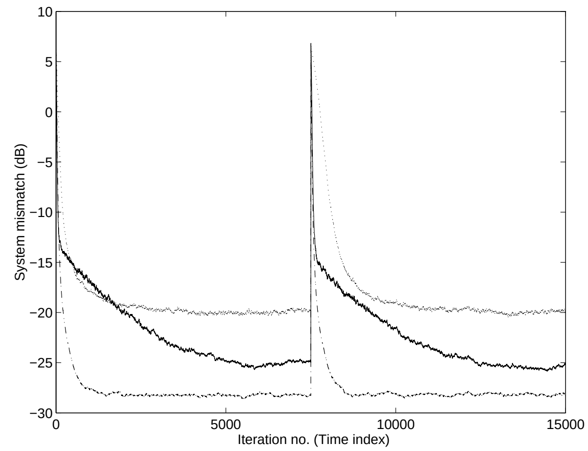

Figure 2: Comparison between variable step size algorithms; the proposed algorithm: dash dot line, Mathews: solid line, Harris: dotted line.

tributed with variance equal to $\sigma_A^2 = 0.5$. It is evident that, in the presence of strong interference (noise), the performance of the other variable step size algorithms is severely degraded, whereas the proposed algorithm exhibits a desirable robustness.

Now we concentrate on the theoretical behaviour of the algorithm and compare the theoretically obtained results with the experimental ones. Figure 3

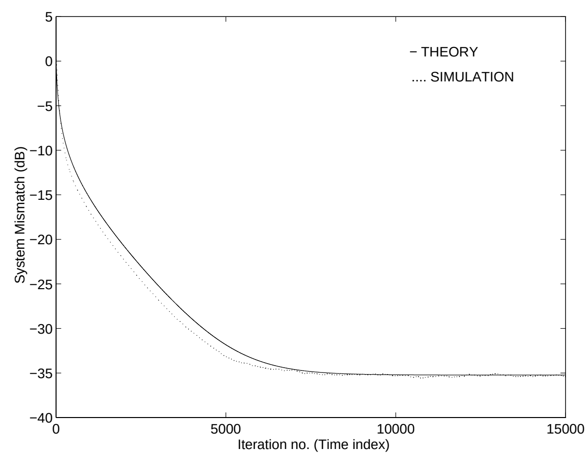

Figure 3: System mismatch evolution comparison between theoretical (solid line) and simulation (dotted line) results; $\alpha = 0.05$, $\beta = 0.985$

compares the expected system mismatch evolution of the adaptive step size algorithm, as given by equation (10), with that obtained from simulations. The parameter α was chosen equal to 0.05 and the system to be identified was chosen to be $\boldsymbol{H_0^{opt}} = [0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.5,\ 0.4,\ 0.3,\ 0.2,\ 0.1]$.

The theoretically obtained step size sequence is also compared with the experimentally observed one in figure 4. To obtain the expected step size, we have used the fact that the average absolute value of a zero mean

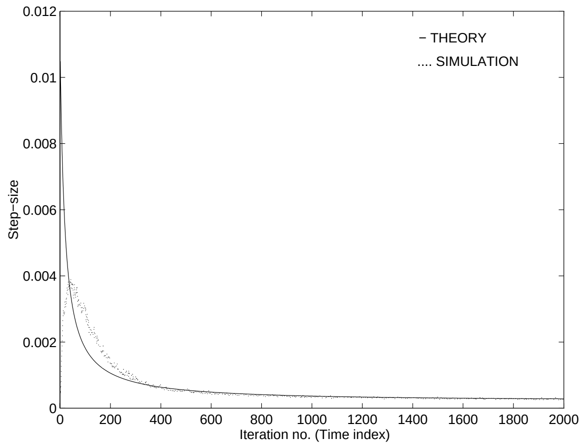

Figure 4: Step size sequence evolution comparison between theoretical (solid line) and simulation (dotted line) results; $\alpha = 0.05$, $\beta = 0.985$

gaussian variable with variance σ^2 is equal to $\sqrt{2/\pi}\sigma$. The figures show a very good match between the theoretical and experimental curves, supporting our analysis and strengthening the validity of the underlying assumptions.

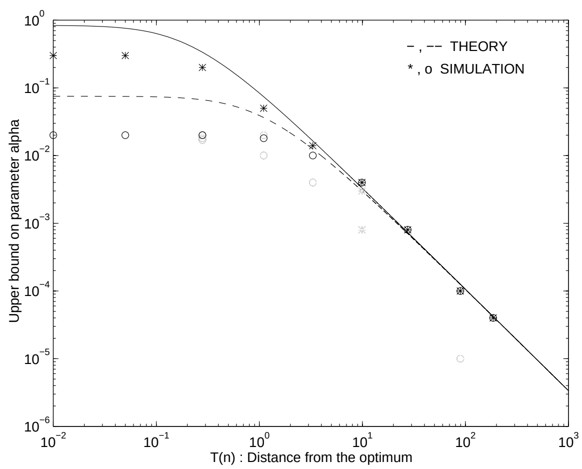

Figure 5: Comparison between the theoretical (solid line: $\sigma_w^2 = 0,1$, dashed line: $\sigma_w^2 = 0.5$) and experimental results (star line: $\sigma_w^2 = 0.1$, circle line: $\sigma_w^2 = 0.5$) obtained on the upper bound of α.

Finally, in figure 5 the maximum allowable value of α obtained from theory (Eq. (12)) is compared with that observed in simulations. The upper bound is drawn as a function of the distance (T_n) from the optimum and the comparison is made for two cases, for $\sigma_w^2 = 0.1$ and $\sigma_w^2 = 0.5$. It is easily observed that the simulation results support the adopted theoretic approach, especially when the starting point lies far from the optimal solution. The observed difference between the curves near the optimum is justified, if we consider that the theoretical results are obtained using averages, whereas in simulations, we deal with the instantaneous values of the stochastic variables. Thus, when around the optimum, where the noise is the main driving force, the absolute value of the error at time instant n is higher than the mean absolute one, constraining further the upper stability bound α_{max} on α.

5 CONCLUSIONS

In this contribution a new robust, yet simple technique for adjusting the step size of the LMS algorithm was presented. The convergence behaviour of the resulting algorithm was investigated and stability bounds together with error evolution equations were established.
Simulation results illustrated the desirable characteristics of the proposed method and verified its superior performance compared to the fixed step size LMS and other existing variable step size algorithms. Furthermore, the experimentally obtained curves and bounds were shown to agree very well with the theoretic ones, supporting therefore the presented analysis.

References

[1] S. C. Douglas. Generalised Gradient Adaptive Step Sizes for Stochastic Gradient Adaptive Filters. In *Proc. Int. Conf. Acoust.,Speech and Signal Proc.*, vol. 2, pp. 1396–1399, Detroit, Michigan, May 1995.

[2] R. W. Harris, D. M. Chabries, and F. A. Bishop. A variable step (VS) adaptive filter algorithm. *IEEE Trans. Acoustics,Speech and Signal Processing*, ASSP-34(2):309–316, April 1986.

[3] Simon Haykin. *Adaptive Filter Theory*. Englewood Cliffs : Prentice-Hall, 2nd edition, 1991.

[4] R. H. Kwong and E. W. Johnston. A variable step size LMS algorithm. *IEEE Trans. Signal Processing*, 40(7):1633–1642, July 1992.

[5] V. J. Mathews. Performance Analysis of Adaptive Filters Equipped with the Dual Sign Algorithm. *IEEE Trans. Signal Processing*, 39(1):85–91, January 1991.

[6] V. J. Mathews and Z. Xie. A stochastic gradient adaptive filter with gradient adaptive step size. *IEEE Trans. Signal Processing*, 41(6):2075–2087, June 1993.

[7] D. I. Pazaitis and A. G. Constantinides. A Kurtosis-Driven Variable Step-Size LMS Algorithm. In *Intern. Conf. Acoust., Speech and Signal Processing*, Atlanta, Georgia, 1996.

[8] D.I. Pazaitis. *Adaptive Signal Processing Techniques for Improved Convergence and Noise Robustness*. MPhil/PhD Transfer Report, Imperial College, March 1996.

[9] O. Tanrikulu and A.G. Constantinides. Mean square convergence analysis and variable step-size of the LMF adaptive algorithm. *Submitted in Trans. Circuits and Systems*, 1995.

[10] Bernard Widrow and Samuel D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, New Jersey, 1985.