A KURTOSIS-DRIVEN VARIABLE STEP-SIZE LMS ALGORITHM

Dimitrios I. Pazaitis and A.G. Constantinides

Signal Processing Section, Department of Electrical and Electronic Engineering,
Imperial College, Exhibition Road, London SW7 2BT, ENGLAND
e-mail : {d.pazaitis,a.constantinides}@ic.ac.uk

ABSTRACT

In this contribution a new technique for adjusting the step-size of the LMS algorithm is introduced. The proposed method adjusts the step-size sequence utilising the kurtosis of the estimation error, reducing therefore performance degradation due to the existence of significant gaussian-distributed noise. The algorithm's behaviour is analysed and equations regarding the evolution of the weight-error correlation matrix and stability of the algorithm are established. The obtained theoretical results are shown to agree well with the experimental ones.

Furthermore, the performance of the proposed algorithm compared to that of LMS and other existing time-varying step-size algorithms is found superior in terms of tracking speed and steady-state error.

1. INTRODUCTION.

Stochastic gradient adaptive filters using the Least Mean Squares (LMS) algorithm [12], enjoy great popularity due to their inherent simplicity. A significant aspect of LMS is that its performance (convergence rate) is often hampered by wide disparity in the eigenvalues (eigenvalue spread) of the autocorrelation matrix of the input vector.

An adaptive (feedback) constant step-size μ governs the stability of the algorithm, as well as the rate of convergence and the steady state excess mean-squared error in relation to the optimal Wiener solution. The convergence time is inversely proportional to μ whereas the misadjustment is proportional to μN , where N equals the number of the weights of the adaptive filter [12]. When a nonstationary environment is considered, the "lag misadjustment" is proportional to μ^{-1} . Thus, for a fixed step-size there is a trade-off between convergence rate and steady-state excess error.

To overcome this problem time varying step-size sequences have been proposed (e.g. [1, 3, 5, 6, 8]). The main rationale behind these approaches is to sense in some way the distance from the optimum and correspondingly to adapt the value of the step-size μ . It seems to us however, that the performance of such methods is adversely affected by the presence of significant noise in the system. The *LMS+F* algorithm proposed in [11] can be also considered as a time varying step-size algorithm.

This work was supported by a scholarship from the State Scholarship Foundation (I.K.Y) of Hellas

In this paper a new time varying step-size selection method, which exhibits robustness under Gaussian distributed noise conditions, is considered. To achieve this we resort to higher order statistics and more precisely to the fourth-order cumulant, known as the kurtosis. The reason for using cumulants is twofold. Cumulants are additive in their arguments, i.e., the sum of two signals has a cumulant equal to the sum of the individual cumulants, and secondly, all cumulants are blind to any kind of zero mean Gaussian process [10]. Thus by introducing cumulants effectively we immunize the step-size selecting algorithm against gaussian or gaussian-like noise. In a real environment, even if the noise is not gaussian, the existence of a large number of individual noise sources "moves" the p.d.f of the noise - according to the central limit theorem - towards gaussianity.

2. THE VARIABLE STEP-SIZE LMS ALGORITHM

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

$$\mathbf{H}_{n+1} = \mathbf{H}_n + \mu_n \mathbf{X}_n e(n), \quad (1)$$

where μ_n is the time-varying step-size, $\mathbf{H}_n$ is the filter coefficient vector and $\mathbf{X}_n$ is the input vector, all at time instant n . In its original form [12] the step-size assumes a constant value. In this contribution we propose the step-size to be a function of the kurtosis (fourth-order cumulant) of the error and to be updated as follows

$$\mu_n = \mu_{max} \left(1 - e^{-\alpha |C_4^e(n)|} \right), \quad (2)$$

where

$$C_4^e(n) = E\{e^4(n)\} - 3E^2\{e^2(n)\} \quad (3)$$

is the kurtosis of the (assumed zero mean) real error signal, α is a positive scaling factor and $|\cdot|$ denotes the absolute value. As seen from the update formula, the step-size sequence μ_n is self-constrained by an upper bound value μ_{max} . The value of μ_{max} can be chosen equal to the maximum allowable value of μ_n that ensures convergence. The use of the exponential function for determining μ_n was firstly proposed by Karni and Zeng in [5], where the norm

of the instant gradient was used in the index rather than the kurtosis as above.

In adaptive filtering the kurtosis of the error equals the sum of the noise kurtosis and the kurtosis of the difference between the filter output and the desired (without noise) signal. Under the gaussian distributed noise assumption, the first term equals zero and the second tends to zero as the filter converges to its optimum values. Thus the algorithm starts with a high value for μ_n and as it approaches the optimum this value decreases permitting theoretically exact convergence to the optimum.

3. CONVERGENCE OF THE ALGORITHM

To facilitate our analysis we introduce the commonly used assumption that the various input vectors come from mutually independent zero-mean gaussian distributed sequences [2, 4, 8, 12]. Although this is never true in our system identification setup, since consecutive input vectors share $N - 1$ entries, it is widely accepted to capture the first-order behaviour and is extensively used in the literature to simplify the analysis producing at the same time reliable results [9]. We also assume that the step-size sequence μ_n is statistically independent of X_n , H_n and $e(n)$ [8, 6]. In the ideal version of the algorithm, where the actual kurtosis is utilised, this is true. In practical applications however, where the error sequence $\{e(n)\}$ is considered ergodic and a finite-length window is used to estimate the kurtosis, this assumption does not really hold. Experimental results have shown nevertheless, that by increasing the time-span of the window the validity of the assumption is strengthened.

The unknown system (optimal weight vector) H^{opt} is considered time-varying in our application and its variations are modeled as

$$H_{n+1}^{opt} = H_n^{opt} + A(n), \quad (4)$$

where $A(n)$ is zero mean and white vector disturbance process with a covariance matrix equal to $\sigma_A^2 I$.

By letting G_n be a vector constituted by the diagonal terms of the matrix $\Lambda E\{V_n V_n^T\}$, and 1 be a column vector of 1's of the same length as G_n , we obtain the following set of equations, which governs the convergence behavior of the algorithm [4]

$$G_{n+1} = [I - 2\Lambda E\{\mu_n\} + \Lambda^2 E\{\mu_n^2\}(2I + 11^T)]G_n + E\{\mu_n^2\}\Lambda^2 1\xi_{min} + \Lambda 1\sigma_A^2, \quad (5)$$

$$E\{\mu_n\} = E\{\mu_{max}(1 - e^{-\alpha|C_4^*(n)|})\}, \quad (6)$$

where V_n' stands for the shifted weight-error vector $V_n = H_n - H_n^{opt}$, $\Lambda = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_n]$ is the diagonal matrix of eigenvalues of the input correlation matrix R and ξ_{min} is the minimum mean squared error. Adopting an approach similar to [2] and proceeding with some straightforward but tedious manipulations it is shown that the necessary and sufficient conditions for convergence are satisfied when

$$\frac{E\{\mu_n^2\}}{2E\{\mu_n\}} \leq \frac{1}{3 \text{tr}[R]}. \quad (7)$$

Under this condition G_n converges to [2]

$$G_\infty = (2E\{\mu_n\}I - \Lambda E\{\mu_n^2\}(2I + 11^T))^{-1} \cdot (\sigma_A^2 1 + E\{\mu_n^2\}\Lambda^2 1\xi_{min}). \quad (8)$$

By using (2) we see that $E\{\mu_n^2\} = E^2\{\mu_n\} = \mu_n^2$. Thus the convergence condition posed on μ_n (7) becomes

$$E\{\mu_n\} = \mu_n \leq \frac{2}{3 \text{tr}[R]}. \quad (9)$$

In the worst case, selecting μ_{max} according to

$$\mu_{max} \leq \frac{2}{3 \text{tr}[R]} \quad (10)$$

ensures convergence of the algorithm. It is observed that the upper bound on μ_{max} coincides with that obtained in [2]. This is due to the fact that when the actual value of the kurtosis is used, $\mu_n = E\{\mu_n\}$.

In a stationary environment, the algorithm reduces its step-size as it approaches the optimum set of filter coefficients, permitting an exact identification of the unknown system. It can be shown in fact, that by using the above equations, the step-size μ_n tends to zero, thus eliminating completely the excess mean squared error.

In practice, however, the real kurtosis of the error signal is unknown and, as mentioned before, has to be estimated. This is usually done by assuming ergodicity of the error sequence and thus replacing ensemble-averaging by time-averaging over a finite window. In reality therefore, the estimated kurtosis is a random variable and its steady state variance is a function of the window length. As a consequence, the step-size sequence μ_n , as a function of the absolute value of the kurtosis, reaches a steady state value which is different from zero. Its distance from zero decreases asymptotically to this value, as the window length increases to infinity.

Nevertheless, the condition for convergence given by (7) still holds. Substituting the values for μ_n given by (2) and expanding the square, we obtain

$$\frac{(1 + E\{(e^{-\alpha|C_4^*(n)|})^2\} - 2E\{e^{-\alpha|C_4^*(n)|}\})}{2(1 - E\{e^{-\alpha|C_4^*(n)|}\})} \leq \frac{\mu_{max}^{-1}}{3 \text{tr}[R]}. \quad (11)$$

It can be readily seen that the left hand term is always lower than or equal to one. Thus by constraining μ_{max} to be

$$\mu_{max} \leq \frac{1}{3 \text{tr}[R]} \quad (12)$$

convergence can be guaranteed.

The above result was produced under the independence assumption of successive input vectors. Ljung in [7] has shown that the the LMS algorithm is almost always convergent, under the realistic assumption of correlated input vectors, by replacing the constant step-size μ with a decreasing step-size sequence μ_n tending to zero. These conditions hold with respect to our algorithm as well, and similar arguments as in [7] could be used in relation to its convergence under the correlated observations assumption.

Figure 1: Performance curves of the LMS (dotted) and the proposed Variable step-size (solid) algorithm. $M = 50$.

4. SIMULATION RESULTS

In this section we present and analyse the results obtained from simulations. The algorithm was applied in a system identification setup, where the system to be identified was considered non stationary. Its coefficients assumed the following initial values $\mathbf{H}^{opt} = [1.0, 0.268, 0.083, -0.501, 0.941, 0.011, -0.819, -0.376, 0.701, 0.159]$ and, after that, experienced random disturbances. The null vector ($\mathbf{0}$) was chosen as the initial vector $\mathbf{H}_0$, and all the results were obtained by averaging over an ensemble of 200 runs. All signals were real and zero mean. The input sequence is assumed to be zero mean i.i.d gaussian sequence with variance equal to unity. The kurtosis of the error ($C_4^e(n)$) was estimated by using a moving finite length rectangular window of size $M = 50$. Alternatively, the kurtosis can be estimated from its consisting moments as follows

$$E\{e_n^k\} = \beta E\{e_{n-1}^k\} + (1 - \beta)e_n^k, \quad (13)$$

where β is the forgetting factor in the interval $[0, 1]$ and $k = 2, 4$. The values of the forgetting factor β and the window length control the memory of the estimator and should be chosen according to the application characteristics and requirements. Their selection is dictated by a trade-off between tracking speed and steady-state variance of the estimated value.

Figure 1 depicts the behaviour of the proposed variable step-size algorithm in comparison with that of the fixed step-size LMS algorithm. The parameters of the algorithms were chosen so that the two algorithms exhibit similar convergence rate and are as follows : $\mu_{max} = 0.07, \mu_{LMS} = 0.044$ and $\alpha = 5.0$. The sudden changes in the system were due to zero mean uniformly-distributed disturbances with variance $\sigma_A^2 = 0.2$ at random the first time but fixed afterwards (i.e., for the following runs) time instants. The Signal to Noise ratio (SNR) equals 10 dB ($\sigma_w^2 = 0.1$ and as a performance measure we select the *System mismatch* ($E\{\mathbf{V}_n^T \mathbf{V}_n\}$).

The performance of the adaptive step-size sequence algorithms is clearly superior to that of LMS which is outper-

formed 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 (and hence its tracking speed) whenever far from the optimum.

Figure 2: Comparison between theoretical (dashed) and simulation (solid) results. SNR = 10 dB

Figure 3: Comparison between theoretical (dashed) and simulation (solid) results. SNR = 10 dB

It should be also noted that the selection of the parameter α (similarly to the window-length selection) is a compromise between convergence speed and steady state misadjustment. This influence of the parameter α is shown in Figure 4, where the performance curve is depicted for $\alpha = 2$ and $\alpha = 5$. Thus, in nonstationary environments a higher value for α would be more beneficial since it would improve the response speed of the algorithm.

In Figures 2 and 3 both the theoretically obtained step-size sequence and system steady-state error are compared with the corresponding experimentally obtained ones. In this experiment $\alpha = 5.0$ and an upper bound on μ , $\mu_{max} = 0.07$, was applied. It can be observed from the above figures, that the simulation results support well the analysis of the proposed algorithm.

As mentioned before, the exact values of the kurtosis were

Figure 4: The effect of α in the system performance. ($\alpha = 5.0$ (dash-dot), $\alpha = 2.0$ (dotted))

Figure 5: Comparison between variable step-size algorithms (Harris : dotted, Mathews : dash-dot, Proposed : solid).

unknown and had to be estimated using a moving finite length window. Consequently, although in theory the step-size sequence tends to zero, in practice it tends to a small but finite positive value. The wider the window, the smaller this value tends to be.

Finally, in Figure 5 the performance of variable step-size algorithm is compared with that of two other variable step-size algorithms developed by Harris *et. al* [3] and Mathews and Xie [8]. The parameters for Harris's algorithm were chosen as : $\alpha = 2.0$, $m_0 = m_1 = 2$, $\mu_{max} = 0.07$, $\mu_{min} = 1e-8$, whereas for the algorithm proposed by Mathews and Xie we chose $\rho = 0.0008$, $\mu_{max} = 0.1$, $\mu_0 = 0.06$. These values correspond to those proposed by the respective authors. For our variable step-size algorithm we have used $\alpha = 5.0$ and $\mu_{max} = 0.07$. The disturbances of the system in the simulations occur at random times (fixed after the first run) with variance $\sigma_A^2 = 0.2$ and are uniformly distributed. It can be easily observed that the performance of the other variable step-size algorithms severely deteriorates in the presence of significant noise and is characterised by slower convergence and higher misadjustment.

5. CONCLUSIONS

A new variable step-size LMS adaptive filtering algorithm was proposed in this paper. The algorithm is different from previous techniques involving time-varying step-sizes in that it utilises the kurtosis of the estimation error signal to adapt its step-size. Consequently, it is characterised by a reduced sensitivity to any gaussian-type noise. The proposed algorithm exhibits improved tracking ability and lower misadjustment compared to the fixed step-size LMS algorithm and other previously developed variable step-size algorithms. The theoretical behaviour of the algorithm was analysed and the obtained results were verified by simulations.

6. REFERENCES

- [1] S. C. Douglas. Generalised Gradient Adaptive Step Sizes for Stochastic Gradient Adaptive Filters. In *Proc. Int. Conf. Acoust., Speech and Signal Processing*, volume 2, pages 1396-1399, Detroit, Michigan U.S.A, May 1995.
- [2] A. Feuer and E. Weinstein. Convergence analysis of LMS filters with uncorrelated gaussian data. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-33(1):222-229, February 1985.
- [3] 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.
- [4] L. L. Horowitz and K.D. Senne. Performance advantage of complex LMS for controlling narrow-band adaptive arrays. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-29(3):722-735, June 1981.
- [5] S. Karni and G. Zeng. A new convergence factor for adaptive filters. *IEEE Trans. Circuits and Systems*, 36(7):1011-1012, July 1989.
- [6] R. H. Kwong and E. W. Johnston. A variable step size LMS algorithm. *IEEE Trans. Signal Processing*, 40(7):1633-1642, July 1992.
- [7] L. Ljung. Analysis of recursive stochastic algorithms. *IEEE Trans. Automatic Control*, AC-22(4):551-575, August 1977.
- [8] 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.
- [9] J.E. Mazo. On the Independence Theory of Equalizer Convergence. *The Bell Systems Technical Journal*, 58(5):963-993, 1979.
- [10] C. L. Nikias and A. P. Petropoulou. *Higher-Order Spectral Analysis : A Nonlinear Signal Processing Framework*. Prentice-Hall, Inc, 1993.
- [11] D. I. Pazaitis and A. G. Constantinides. LMS+F Algorithm. *Electronics Letters*, 31(17):1423-1424, August 1995.
- [12] Bernard Widrow and Samuel D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, New Jersey, 1985.