Correspondence

Analysis of LMS and NLMS Algorithms with Delayed Coefficient Update Under the Presence of Spherically Invariant Processes

Markus Rupp and Rudi Frenzel

Abstract—**Certain conditions require a delay in the coefficient update of the least mean square (LMS) and normalized least mean square (NLMS) algorithms. This paper presents an in-depth analysis of these modificated versions for the important case of spherically invariant input processes (SIRP's), which are known as an excellent model for speech signals. Some derived bounds and the predicted dynamic behavior of the algorithms are found to correspond very well to simulation results and a real time implementation on a fixed-point signal processor. A modification of the algorithm is proposed to assure the well known properties of the LMS and NLMS algorithms.**

I. Introduction

In this paper, a modified version of the least mean square (LMS) algorithm is analyzed, the difference being a delay $D \in \mathbf{N}$ in the feedback path of the error signal $e(n)$. Such delays are inherent to many applications of adaptive filtering, e.g., Viterbi decoding [7] and active noise compensation [4]. Since even a simple one step delay can cause instability of the LMS algorithm, modifications of the algorithm are necessary. In the past analyses for these algorithms followed two paths: calculating the first moments (see [6], [9]) of the weight-error vector $\underline{\epsilon}(n) = \underline{w}(n) - \underline{w}_o$, i.e., the difference between the actual filter coefficients $\underline{w}(n)$ and the Wiener solution $\underline{w}_o$, and calculating the behavior of the average mean-squared error (see [7]). In spite of some corrections which have been published recently [8], the calculations in [7] involve some simplifications resulting in practical bounds. The analysis presented in this paper expands upon the results of [7] by using the classical LMS analysis [5], [10] and the large class of spherically invariant random processes (SIRP's) for excitation, resulting in a recursive vector equation of the weight-error vector variance matrix. Now, bounds and optimal step sizes are achievable with only numerical analyses. The results are validated by simulations and real time experiments.

II. Delayed LMS Algorithm

It is common to the algorithms under consideration that only a delayed version $e_d(n) = e(n-D)$ of the error signal $e(n) = d(n) - \underline{u}^T(n)\underline{w}(n)$ is available, where $d(n)$ is the disturbed output $y(n)$ of a plant G. All equations are in reference to the simplified adaptive transversal compensator structure in Fig. 1, where all sampling devices, anti-aliasing filters, and interpolator filters are neglected and the whole model is purely time-discrete. The design of a modified LMS algorithm that is stable for all excitations can be accomplished in several manners:

1) If an update of the coefficients is done only every $(D+1)$ steps, the equations for the LMS and NLMS algorithms are, respectively[1]

$$\underline{\mathbf{w}}(n+1) = \underline{\mathbf{w}}(n-D) + \mu \mathbf{e}_d(n)\underline{\mathbf{u}}(n-D) \quad \text{if } n \bmod (D+1) = 0. \tag{1}$$

$$\underline{\mathbf{w}}(n+1) = \underline{\mathbf{w}}(n-D) + \frac{\alpha}{\underline{\mathbf{u}}^T(n-D)\underline{\mathbf{u}}(n-D)}\mathbf{e}_d(n)\underline{\mathbf{u}}(n-D) \quad \text{if } n \bmod (D+1) = 0. \tag{2}$$

The vector $\underline{\mathbf{u}}(n)$ of dimension M is the input vector of the transversal filter, i.e., M samples of the input sequence $\mathbf{u}(n)$. The coefficients of the transversal filter compose the vector $\underline{\mathbf{w}}(n)$

$$\underline{\mathbf{u}}^T(n) = (\mathbf{u}(n), \mathbf{u}(n-1), \cdots, \mathbf{u}(n-M+1)) \tag{3}$$

$$\underline{\mathbf{w}}^T(n) = (\mathbf{w}_0(n), \mathbf{w}_1(n), \cdots, \mathbf{w}_{M-1}(n)). \tag{4}$$

Fig. 2 illustrates the time relationship between the input and error signals. An analysis of the algorithms shows that all properties are the same as in the classical case. Since the adaptation is done only every $(D+1)$ steps, the convergence rate is reduced by the same factor.

2) A delay D in the input vector $\underline{u}(n)$ is introduced. If a delay D is taken into account, more samples from the input sequence $u(n)$ and the error sequence $e_d(n)$ have to be stored. The respective equations of the LMS algorithm with delayed update (DLMS) and its normalized version (NDLMS) are as follows

$$\underline{\mathbf{w}}(n+1) = \underline{\mathbf{w}}(n) + \mu \mathbf{e}_d(n)\underline{\mathbf{u}}(n-D) \quad ; n \in \mathbf{Z} \tag{5}$$

$$\underline{\mathbf{w}}(n+1) = \underline{\mathbf{w}}(n) + \frac{\alpha}{\underline{\mathbf{u}}^T(n-D)\underline{\mathbf{u}}(n-D)} \mathbf{e}_d(n)\underline{\mathbf{u}}(n-D) \quad ; n \in \mathbf{Z}. \tag{6}$$

An analysis based on the classical approach is presented next. It provides a deeper insight to the behavior of the DLMS and NDLMS algorithm. The complexity of the algorithms is approximately $2M$.

3) An additional correction term, as described later, permits a correction in such a way that the resulting algorithm maintains essentially the same properties as the classical LMS algorithm. However, if M equals the order of the transversal filter W, an additional complexity of M operations is incurred.

In the following paragraphs the behavior of the algorithm according to point 2 is analyzed. In [10] the behavior of the LMS and NLMS algorithms has been analyzed for *spherically invariant processes* as excitation. Since these processes are excellent in modeling the statistical properties of speech signals [2], [3], the analysis of [10] has been modified in the following in order to describe the behavior of the DLMS algorithm.

[1] To emphasize random processes they are typed in bold faced letters.

Manuscript received July 23, 1991; revised March 31, 1993. The associate editor coordinating the review of this paper and approving it for publication was Prof. Monson Hayes.

M. Rupp was with the Institut für Netzwerk- und Signaltheorie, Darmstadt, Germany. He is now with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560.

R. Frenzel is with the Institut für Netzwerk- und Signaltheorie, Darmstadt, Germany.

IEEE Log Number 9214633.

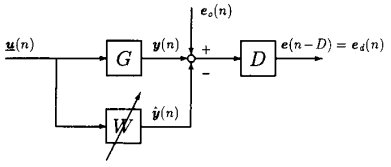

Fig. 1. Structure of an adaptive transversal filter with a time delay in the error feedback path.

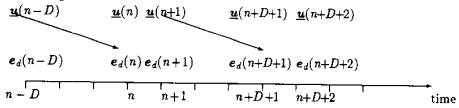

Fig. 2. Time relationship of the signals while updating the coefficients.

The first moments of the weight-error vector $\underline{\epsilon}(n) = \underline{\mathbf{w}}(n) - \underline{w}_o$ can be evaluated using some abbreviations. The vector $\underline{w}_o$ is the Wiener solution to which the algorithm should converge. For a linear, time-invariant plant G the Wiener solution $\underline{w}_o$ is equal to the first M impulse response steps of G if the autocorrelation function (ACF) matrix R_{uu} of the input process is regular. This leads to

$$\mathbf{e}(n) = \underline{\mathbf{u}}^T(n)\underline{\epsilon}(n) + \mathbf{e}_o(n) \tag{7}$$

$$\underline{\epsilon}(n+1) = \underline{\epsilon}(n) - \mu\underline{\mathbf{u}}(n-D)\underline{\mathbf{u}}^T(n-D)\underline{\epsilon}(n-D) + \mu\mathbf{e}_o(n-D)\underline{\mathbf{u}}(n-D) \tag{8}$$

$$E[\underline{\epsilon}(n+1)] = E[\underline{\epsilon}(n)] - \mu R_{uu}E[\underline{\epsilon}(n-D)]. \tag{9}$$

The error sequence $\mathbf{e}_o(n)$ is assumed to be statistical independent of the input sequence $\mathbf{u}(n)$ and $E[\mathbf{e}_o(n)] = 0$, $E[\mathbf{e}_o^2(n)] = J_{\min}$. The Wiener solution causes the minimum mean-squared error $J_{\min}$. In general, the excess mean-squared error $E[\mathbf{e}_\infty^2] = \lim_{n\to\infty} E[\mathbf{e}^2(n)]$ achieved by the algorithm exceeds $J_{\min}$. Since in (9) the 'Independence theorem' has been used, the results hold only for linear combinations of statistically independent input vectors but not for transversal filters. The results that are derived from this simplification must be considered to be approximations. A stability bound for the first moments $E[\underline{\epsilon}(n+1)]$ in dependence of the eigenvalues of R_{uu} and the delay D can be derived [6].

When summing the second moments of the weight-error vector elements, the system mismatch is obtained

$$S(n) = E[\underline{\epsilon}^T(n)\underline{\epsilon}(n)]. \tag{10}$$

For the investigation of $S(n)$ the variance matrix $K(n) = E[\underline{\epsilon}(n)\underline{\epsilon}^T(n)]$ of the weight-error vector will be used. The mean-squared error $E[\mathbf{e}^2(n)]$ is easily derived from $K(n)$ by using

$$E[\mathbf{e}^2(n)] = E[\underline{\epsilon}^T(n)R_{uu}\underline{\epsilon}(n)] + J_{\min} \tag{11}$$

$$= \text{trace}\,(K(n)R_{uu}) + J_{\min}. \tag{12}$$

The following expression is obtained directly when using (8) and the definition of $K(n)$

$$\begin{aligned} K(n+1) &= E\Big[(\underline{\epsilon}(n) - \mu\underline{\mathbf{u}}(n-D)\underline{\mathbf{u}}^T(n-D)\underline{\epsilon}(n-D)) \\ &\quad (\underline{\epsilon}(n) - \mu\underline{\mathbf{u}}(n-D)\underline{\mathbf{u}}^T(n-D)\underline{\epsilon}(n-D))^T\Big] \\ &\quad + \mu^2 J_{\min}R_{uu} \\ &= K(n) - \mu R_{uu}E[\underline{\epsilon}(n-D)\underline{\epsilon}^T(n)] \\ &\quad - \mu E[\underline{\epsilon}(n)\underline{\epsilon}^T(n-D)]R_{uu} \\ &\quad + \mu^2 E[\underline{\mathbf{u}}(n-D)\underline{\mathbf{u}}^T(n-D)K(n-D)\underline{\mathbf{u}}(n-D) \\ &\quad \times \underline{\mathbf{u}}^T(n-D)] + \mu^2 J_{\min}R_{uu}. \end{aligned} \tag{13}$$

To simplify the expression, the following abbreviation is used

$$K_{ij}(n) = E[\underline{\epsilon}(n-i)\underline{\epsilon}^T(n-j)]. \tag{14}$$

Equation (13) with $K_{00}(n-D) = K_{DD}(n)$ leads to

$$\begin{aligned} K_{00}(n+1) &= K_{00}(n) - \mu R_{uu}K_{D0}(n) - \mu K_{0D}(n)R_{uu} \\ &\quad + \mu^2 E[(\underline{\mathbf{u}}(n-D)\underline{\mathbf{u}}^T(n-D)K_{DD}(n) \\ &\quad \times \underline{\mathbf{u}}(n-D)\underline{\mathbf{u}}^T(n-D)] + \mu^2 J_{\min}R_{uu}. \end{aligned} \tag{15}$$

Without any loss of generality, the correlated random process $\underline{\mathbf{u}}(n)$ can be considered as a linear transformation A of an uncorrelated process $\underline{\mathbf{x}}(n)$ of the same order M: $\underline{\mathbf{u}}(n) = A\underline{\mathbf{x}}(n)$. The white process $\underline{\mathbf{x}}(n)$ can be chosen in such a way that its ACF matrix is the identity matrix I (without any loss of generality). The ACF matrix of the colored process is then: $R_{uu} = AA^T$. The matrix Q can be chosen not only to diagonalize the ACF matrix but also to: $Q^TA = \Lambda^{1/2}$. The diagonalization of the ACF matrix ($Q^TR_{\mathrm{uu}}Q = \Lambda$) in (15) leads to $Q^TK_{\mathrm{ij}}(n)Q = C_{\mathrm{ij}}(n)$ and

$$\begin{aligned} C_{00}(n+1) &= C_{00}(n) - \mu\Lambda C_{D0}(n) - \mu C_{0D}(n)\Lambda \\ &\quad + \mu^2 E[\Lambda^{1/2}\underline{\mathbf{x}}(n-D)\underline{\mathbf{x}}^T(n-D)\Lambda^{1/2}C_{\mathrm{DD}}(n)\Lambda^{1/2} \\ &\quad \times \underline{\mathbf{x}}(n-D)\underline{\mathbf{x}}^T(n-D)\Lambda^{1/2}] + \mu^2 J_{\min}\Lambda \end{aligned} \tag{16}$$

$$C_{0D}(n+1) = C_{0D-1}(n) - \mu\Lambda C_{DD-1}(n) \tag{17}$$

result. The trace of the weight-error vector variance matrix $C_{00}(n)$ equals the system mismatch $S(n)$. Due to the invariance of the trace to diagonalization, only the elements on the diagonal of $C_{00}(n)$ are of interest. From step n to $n+1$ only the diagonal elements of $C_{ij}(n)$ influence the elements of $C_{kl}(n+1)$. Setting up a new variance vector $\underline{c}_{ij}(n)$ for $i,j = 0\cdots D$ with only the diagonal elements of $C_{\mathrm{ij}}(n)$, the following simpler presentation for $\underline{c}_{00}(n)$ and $\underline{c}_{0D}(n)$ is obtained

$$\begin{aligned} \underline{c}_{00}(n+1) &= \underline{c}_{00}(n) - 2\mu\Lambda\underline{c}_{0D}(n) + \mu^2 m_x^{(2,2)}(\underline{\lambda}\underline{\lambda}^T + 2\Lambda^2) \\ &\quad \times \underline{c}_{DD}(n) + \mu^2 J_{\min}\underline{\lambda} \end{aligned} \tag{18}$$

$$\underline{c}_{0D}(n+1) = \underline{c}_{0D-1}(n) - \mu\Lambda\underline{c}_{DD-1}(n). \tag{19}$$

The term $S = \mu^2 m_x^{(2,2)}(\underline{\lambda}\underline{\lambda}^T + 2\Lambda^2)$ has been calculated in [10]. There, the joint moment of order four has been used

$$m_x^{(2,2)} = E[\mathbf{x}^2(n-i)\mathbf{x}^2(n-j)] \qquad \text{for } i \neq j. \tag{20}$$

The vector $\underline{\lambda}$ consists of all eigenvalues of R_{uu}, which are the diagonal elements of Λ.

For the delays $D = 1$ to $D = 3$, (18) and (19) are described in a matrix form in which the disturbance term $\mu^2 J_{\min}\underline{\lambda}$ has been suppressed.

For $D = 1$:

$$\begin{pmatrix} \underline{c}_{00}(n+1) \\ \underline{c}_{01}(n+1) \\ \underline{c}_{11}(n+1) \end{pmatrix} = \begin{pmatrix} I & -2\mu\Lambda & S \\ I & -\mu\Lambda & \mathbf{0} \\ I & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \underline{c}_{00}(n) \\ \underline{c}_{01}(n) \\ \underline{c}_{11}(n) \end{pmatrix}. \tag{21}$$

For $D = 2$:

$$\begin{pmatrix} \underline{c}_{00}(n+1) \\ \underline{c}_{01}(n+1) \\ \underline{c}_{02}(n+1) \\ \underline{c}_{11}(n+1) \\ \underline{c}_{12}(n+1) \\ \underline{c}_{22}(n+1) \end{pmatrix} = \begin{pmatrix} I & \mathbf{0} & -2\mu\Lambda & \mathbf{0} & \mathbf{0} & S \\ I & \mathbf{0} & -\mu\Lambda & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} & \mathbf{0} & -\mu\Lambda & \mathbf{0} \\ I & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \underline{c}_{00}(n) \\ \underline{c}_{01}(n) \\ \underline{c}_{02}(n) \\ \underline{c}_{11}(n) \\ \underline{c}_{12}(n) \\ \underline{c}_{22}(n) \end{pmatrix}. \tag{22}$$

For $D = 3$: (23)

$$\begin{pmatrix} \underline{c}_{00}(n+1) \\ \underline{c}_{01}(n+1) \\ \underline{c}_{02}(n+1) \\ \underline{c}_{03}(n+1) \\ \underline{c}_{11}(n+1) \\ \underline{c}_{12}(n+1) \\ \underline{c}_{13}(n+1) \\ \underline{c}_{22}(n+1) \\ \underline{c}_{23}(n+1) \\ \underline{c}_{33}(n+1) \end{pmatrix} =$$

$$\begin{pmatrix} I & 0 & 0 & -2\mu\Lambda & 0 & 0 & 0 & 0 & 0 & S \\ I & 0 & 0 & -\mu\Lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & -\mu\Lambda & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & -\mu\Lambda & 0 \\ I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \end{pmatrix} \begin{pmatrix} \underline{c}_{00}(n) \\ \underline{c}_{01}(n) \\ \underline{c}_{02}(n) \\ \underline{c}_{03}(n) \\ \underline{c}_{11}(n) \\ \underline{c}_{12}(n) \\ \underline{c}_{13}(n) \\ \underline{c}_{22}(n) \\ \underline{c}_{23}(n) \\ \underline{c}_{33}(n) \end{pmatrix}.$$

Block matrices with zero elements are indicated by $\mathbf{0}$. The eigenvalues of these matrices determine the dynamic behavior of the algorithm. By varying the step-size μ, it is possible to find an optimum value for which the absolute value of the largest eigenvalue is minimal. Furthermore, it is also possible to find the critical step size for which the algorithm becomes unstable. Since the matrices are of no special structure (i.e., symmetrical), some eigenvalues may be complex. This means that an oscillation can occur in the learning curve, and simple exponential behavior cannot be expected.

The analysis of the NDLMS algorithm can be treated similarly. In [10] it is shown that instead of using the eigenvalues λ_i of the ACF matrix, normalized eigenvalues λ_{ij}^{NN} must be considered. Accordingly, joint moments are replaced by their normalized counterparts. Because of the protracted calculations the details are not repeated here. The main results will be summarized at the end of the paper.

A deterministic description [11], however, leads to a bound for the normalized step-size α, similiar to that of the first moments [6]

$$0 < \alpha < 2\sin\left(\frac{\pi}{2(2D+1)}\right). \tag{24}$$

Surprisingly, this bound depends only on the delay D and not on the system order M or on any other parameters.

In the following an improved algorithm will be derived. As already mentioned, in most applications the delay D is caused by an 'analog summation point' (e.g., in active noise compensation). Therefore, only the following signals are available for the compensator processing units at time n:

- the delayed error $\mathbf{e}(n-D) = \mathbf{e}_d(n)$,
- the estimated system output $\hat{\mathbf{y}}(n)$, and
- all previous input vectors $\underline{\mathbf{u}}(n-k)$, $k = 0 \cdots D$.

Both adaptation equations (1) and (2) indicate that the error $\mathbf{e}(n)$ caused by the input $\underline{\mathbf{u}}(n)$ to the plant G and the compensator W does not influence the following updates until D time steps later. In other words, each update is done without any information learned from the former $(D-1)$ updates. This causes the drawbacks of the convergence behavior of the DLMS and NDLMS algorithms.

To show how this 'lack of information' can be overcome, the error $\mathbf{e}(n-D)$ is first split into its components

$$\begin{aligned} \mathbf{e}(n-D) &= \mathbf{y}(n-D) - \hat{\mathbf{y}}(n-D) \\ &= \mathbf{y}(n-D) - \underline{\mathbf{w}}^T(n-D)\underline{\mathbf{u}}(n-D). \end{aligned} \tag{25}$$

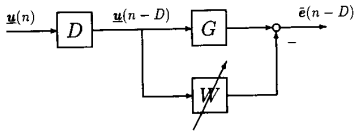

Fig. 3. Analysis structure of an adaptive transversal filter with the modified LMS or NLMS algorithm.

It can be seen that if the old filter vector $\underline{\mathbf{w}}(n-D)$ in (25) is replaced by the current vector $\underline{\mathbf{w}}(n)$, the resulting *modified* error

$$\bar{\mathbf{e}}(n-D) = \mathbf{y}(n-D) - \underline{\mathbf{w}}^T(n)\underline{\mathbf{u}}(n-D) \tag{26}$$

carries all the information about the previous updates up to time n. If the pure system output $\mathbf{y}(n-D)$ is not available (e.g., as in active noise compensation applications), the modified error must be computed according to

$$\bar{\mathbf{e}}(n-D) = \mathbf{e}(n-D) + \hat{\mathbf{y}}(n-D) - \underline{\mathbf{w}}^T(n)\underline{\mathbf{u}}(n-D). \tag{27}$$

With this, the modified adaptation equations for both algorithms are obtained

$$\underline{\mathbf{w}}(n+1) = \underline{\mathbf{w}}(n) + \mu\bar{\mathbf{e}}(n-D)\underline{\mathbf{u}}(n-D). \tag{28}$$

$$\underline{\mathbf{w}}(n+1) = \underline{\mathbf{w}}(n) + \alpha\frac{\bar{\mathbf{e}}(n-D)}{\underline{\mathbf{u}}^T(n-D)\underline{\mathbf{u}}(n-D)}\underline{\mathbf{u}}(n-D). \tag{29}$$

The output equations remain in both cases the same as in (25). The analysis of the new algorithm becomes easy if the plant G is assumed to be time-invariant. In this case, the plant G and the delay D may be exchanged, and the two delay blocks can be replaced by one (see Fig. 3). Obviously, the new algorithms (modified DLMS or NDLMS version) work like their well known original counterparts driven by a delayed input signal. From this it is concluded that all known properties of the initial convergence hold further on. The same approximately applies for the tracking behavior of the DLMS and the NDLMS algorithms after short, burst-like changes of the plant G.

The modified algorithms, in comparison to the originals, require a larger amount of computation. The same holds for the storage capacity, since the previous compensator outputs up to $\hat{\mathbf{y}}(n-D)$ must be stored. Furthermore, the additional inner product $\underline{\mathbf{w}}^T(n)\underline{\mathbf{u}}(n-D)$ has to be computed in each sampling interval. Simulations as well as implementation in active noise cancellation proved the superior behavior of the modified algorithm.

III. Simulations and an Implementation Example

In this section the theoretical results are compared with simulations. The algorithms were excited by samples of a spherically invariant K_0 process. According to [2], [3] this type of process has similar properties as speech signals. For simulations a white and a colored version were used. The colored process was generated by AR(1) filtering of the white process. The resulting eigenvalue ratio was chosen to be greater than 1000, since it is known to be typical for speech signals. In the following simulations a SIRP signal of variance one was used. Additionally, white noise of variance $J_{\min} = 10^{-6}$ was added to the system's output. Fig. 4 depicts the learning curve of the relative system mismatch $S_{\text{rel}}(n)$

$$S_{\text{rel}}(n) = \frac{E[\underline{\epsilon}^T(n)\underline{\epsilon}(n)]}{E[\underline{\epsilon}^T(0)\underline{\epsilon}(0)]}. \tag{30}$$

For Gaussian as well as K_0 processes and for white and colored versions the theoretical results and simulations agreed very well. Only minor differences of less than 1 dB were observed. Typically, a

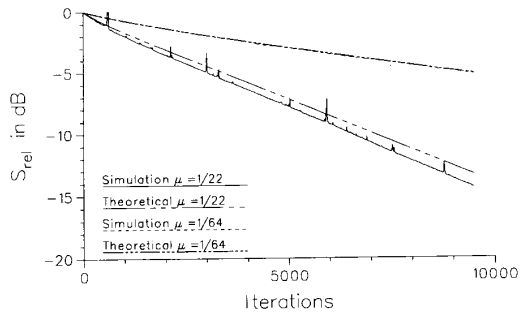

Fig. 4. Learning curve of the DLMS algorithm with a colored K_0 process.

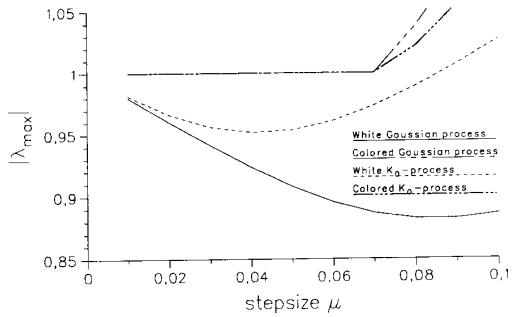

Fig. 5. Magnitude of the largest eigenvalue over step-size μ.

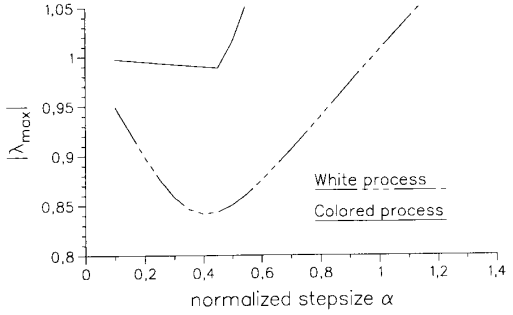

Fig. 6. Magnitude of the largest eigenvalue over normalized step-size α.

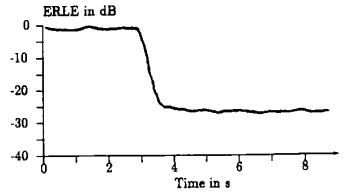

Fig. 7. ERLE of the NDLMS algorithm.

moderate step size resulted in the dynamic behavior of the algorithm as well as the steady-state mismatch being approximately proportional to the step size. Due to the given demands it is possible to choose a step size that is either optimal for the dynamic behavior of the algorithm or optimal for a small steady-state mismatch. A comparison with the results given in [7], [8] shows differences of 10 dB for the steady-state value of the excess mean-squared error.

Fig. 5 shows the magnitude of the largest eigenvalue dependent on the step-size μ for the four cases above. The investigation of the curves directly corresponds to the matrix in (23). While the cases with white excitation showed a clear optimum, the colored cases showed very flat curves. Their optimal point for fastest convergence was very close to the stability limit and within that range the convergence was very sensitive to changes in the step size. Depending on the different joint fourth-order moments the cases with white excitation had different optimal step sizes and also different optimal eigenvalues. It seemed that a larger joint moment caused a larger eigenvalue and correspondingly a lower convergence rate. When comparing with the expressions in [8], the choice for optimal step size and the bound coincided for the Gaussian process, but larger differences occured for the K_0 process. The reason for this could be found in the simplification of [8] for the joint moment $m_x^{(2,2)}$. The flat minimum of the colored processes was approximated especially poorly. The substitutions given in Section II also allow an examination of the behavior of the NDLMS algorithm.

Fig. 6 shows the magnitude of the largest eigenvalue of the matrix corresponding to (23) for the NDLMS algorithm over the normalized step-size α. As expected the probability density function does not have any influence on the dynamic behavior and the curves for Gaussian and K_0 excitation coincide. In both types of excitation, colored and white processes, the optimal eigenvalue is smaller than in the DLMS algorithm, which proves again the superiority of normalized algorithms. As in the DLMS case, the NDLMS algorithm with colored excitation reaches a much lower convergence rate. However, the curve for colored excitation is not so flat and the optimal step sizes for both types of excitation are close together. The normalization seems to equalize both the effects of the probability function as well as the effects of colored input sequences. Equation (24) delivers the stability bound $\alpha = 0.445$, which was verified for both cases. However, this bound was very restrictive for white processes. The more correlated the process is, the closer are its stability bound and the bound according to (24).

For echo cancellation in hybrids a transversal filter of order $M = 32$ was implemented using a fixed-point signal processor DSP 56001. Measurements have been made on a real 'PBX' with an input sequence close to a Gaussian process. The transfer function of the hybrid was determined by a 10 km wire connection to the subscriber. For several reasons the summation point of the desired and the echo replica has been built up by an analog device. A delay of $D = 3$ is inherent to the hardware, originating from the A/D and D/A converters of the system. In Fig. 7 the echo return loss enhancement (ERLE) of the NDLMS algorithm with a normalized step-size $\alpha = 0.5$ is depicted as a function of time.

IV. GENERALIZATION AND CONCLUSION

Having a filter F instead of a pure delay in the error path, the FXLMS algorithm may be applied. However, the considerations with the DLMS algorithm allow an alternative solution. Since F has to be known, the inverse filter F^{-1} can be calculated. With respect that F may have a non-minimal-phase property, the inverse filter F^{-1} includes noncausal parts. These parts can be approximated by introducing an additional delay. Linear filtering in the error path can be at least approximately compensated by the inverse filter and an additional integer delay D.

It has been shown that a filtered error—in the simplest case, a delayed error—can be used for adaptation if the LMS or NLMS algorithm is modified. If the error is only delayed, the modification is, likewise, a delay in the coefficient update. The resulting behavior of the algorithms has been analyzed and compared to the classical LMS and NLMS algorithms in the context of an application in a telephone hybrid situation. Moreover, it is possible to avoid the

drawback of a low convergence rate by adding a correction term. Indeed, its complexity is increased, but it retains, except in the case of time-variant systems, the full behavior of LMS (or NLMS) algorithms.

REFERENCES

[1] M. G. Bellanger, "Adaptive digital filters and signal analysis." New York: Marcel Dekker, 1986.

[2] H. Brehm, "Description of spherically invariant random processes by means of G-functions," *Springer Lecture Notes*, vol. 969, pp. 39–73, 1982.

[3] H. Brehm and W. Stammler, "Description and generation of spherically invariant speech-model signals," *Signal Processing*, vol. 12, no. 2, pp. 119–141, Mar. 1987.

[4] J. C. Burgess, "Active adaptive sound control in a duct: a computer simulation," *J. Acoust. Soc. Am.*, vol. 70, no. 3, pp. 715–726, Sept. 1981.

[5] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech,Signal Processing*, vol. 33, no. 1, pp. 222–230, Feb. 1985.

[6] P. Kabal, "The stability of adaptive minimum mean square error equalizers using delayed adjustment," *IEEE Trans. Commun.*, vol. 31, no. 3, pp. 430–432, Mar. 1983.

[7] G. Long, F. Ling, and J. G. Proakis, "The LMS algorithm with delayed coefficient adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 9, pp. 1397–1405, Sept. 1989.

[8] G. Long, F. Ling, and J. A. Proakis, "Corrections to 'The LMS algorithm with delayed coefficient adaptation'," *IEEE Trans. Signal Processing*, vol. 40, no. 1, pp. 230–232, Jan. 1992.

[9] D. R. Morgan, "An analysis of multiple correlation cancellation loops with a filter in the auxiliary path," *IEEE Trans. Acoust. Speech,Sigal Processing*, vol. 28, no. 4, pp. 454–467, Aug. 1980.

[10] M. Rupp, "The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 41, no. 3, pp. 1149–1160, Mar. 1993.

[11] M. Rupp, "An analog-digital echo canceller for hybrids," in *Proc. ISCAS* (Chicago, IL), Apr. 1993, pp. 1124–1127.

Comments on "A Comparison of Two Quantized State Adaptive Algorithms"

Eweda Eweda

***Abstract*— This note contains a correction to an error in a recently published ASSP paper.**

I. INTRODUCTION

A nice comparison has been carried out between the quantized error (QE) and the quantized regressor (QReg) algorithms in an interesting paper by Sethares and Johnson [1]. This note is intended for correcting an error in Theorem 1 of [1] that leads to incorrect conclusions about the performance of the QE algorithm when using a quantizer with a dead zone. An example of such an algorithm is found in [2]. The symbols of [1] are used throughout the note. Equation numbers composed of two digits, such as (3.5), refer to equations in [1], whereas numbers having a single digit, such as (3), refer to equations in this note.

Manuscript received July 8, 1992; revised March 5, 1993. The associate editor coordinating the review of this paper and approving it for publication was Dr. Hong Fan.

The author is with Military Technical College, Cairo, Egypt.

IEEE Log Number 9214636.

II. THE ERROR

Theorem 1 states that the QE algorithm satisfies (3.5) with $\delta = n\mu\beta^2 M^2/2q$. As shown below, this result holds for the QE algorithm using a quantizer without a dead zone and does not hold for the one using a quantizer with a dead zone. Neither the statement of the theorem nor its proof exclude the quantizers with dead zones. Conversely, the proof mentions that the quantizers with dead zones are included in the theorem. This appears from the condition $i \notin [j\colon Q(e_j) = 0]$ below the summation signs in (A.1) and (A.2). This condition allows the existence of j such that $Q(e_j) = 0$ i.e., quantizers with dead zones are not excluded. Hence, as it stands, Theorem 1 implies that, even with the use of a quantizer with a dead zone, the mean absolute error of the QE algorithm can be made arbitrarily small through the use of a sufficiently small value of the step size μ. This is not correct. To show this error, consider the following simple example in which

$$n = 1 \tag{1}$$

$$X_k \in \{-1, 1\} \tag{2}$$

$$Q(e_k) = 0 \text{ if and only if } |e_k| < 1/4. \tag{3}$$

In other words, (1) means that X_k and θ_k are one dimensioned, (2) means that X_k is a binary random variable, and (3) means that the quantizer has a dead zone of width 1/2. (Examples with other values of the dead zone width may be considered.) Now, assume that the initial parameter error $\tilde{\Theta}_0 = 1/8$. Then, (1.4) and (2) imply that $|e_0| = |X_0|/8 = 1/8$ and then, due to (3), $Q(e_0) = 0$. Therefore, (1.7) implies that $\hat{\Theta}_1 = \hat{\Theta}_0$ and $\tilde{\Theta}_1 = \tilde{\Theta}_0 = 1/8$. Hence, $|e_1| = 1/8, Q(e_1) = 0$, and then, (1.7) implies that $\tilde{\Theta}_2 = \tilde{\Theta}_1 = 1/8$. Continuing in this procedure, one shows that $\tilde{\Theta}_k = 1/8$ and $|e_k| = 1/8$ for all k. Hence, $\lim_{t\to\infty}(1/t)\Sigma_{i=1}^{t} |e_i| = 1/8$. Thus, the mean absolute error in this case cannot be reduced to an arbitrarily small value via the choice of a small μ as indicated by Theorem 1. Thus, Theorem 1, with the above value of δ, is not correct.

III. THE CORRECTION

In this section, we derive a correct value of δ. The derivation starts from (A.2) in [1]. Since V_t is nonnegative (sum of squares), then (A.2), with t being replaced by $t + 1$, implies that

$$0 \le \sum_{\substack{i=1 \\ i\notin[j:\, Q(e_j)=0]}}^{t} [n\mu^2\beta^2 M^2 - 2\mu\, q|e_i|] + V_0.$$

Dividing both sides of this inequality by t and then taking the limit as t tends to infinity yields

$$\begin{aligned}\lim_{t\to\infty} \frac{1}{t} \sum_{\substack{i=1 \\ i\notin[j:\, Q(e_j)=0]}}^{t} |e_i| &\le \operatorname*{Lim}_{t\to\infty} \frac{1}{t} \sum_{\substack{i=1 \\ i\notin[j:\, Q(e_j)=0]}}^{t} \frac{n\mu\beta^2 M^2}{2q} \\ &\le \operatorname*{Lim}_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \frac{n\mu\beta^2 M^2}{2q} \\ &= \frac{n\mu\beta^2 M^2}{2q}. \end{aligned} \tag{4}$$