AN IMPROVED STOCHASTIC MODEL OF THE NLMS ALGORITHM FOR CORRELATED INPUT DATA

Javier E. Kolodziej, Orlando J. Tobias, and Rui Seara

LINSE – Circuits and Signal Processing Laboratory
Department of Electrical Engineering
Federal University of Santa Catarina
88040-900 – Florianópolis – SC – Brazil
E-mails: {javier, orlando, seara}@linse.ufsc.br

ABSTRACT
This paper proposes an improved stochastic model for the normalized least-mean-square (NLMS) algorithm considering correlated input signals obtained from a spherically invariant random process (SIRP). A SIRP describes both Gaussian and a wide class of non-Gaussian processes, including the ones with Laplacian, $K_0$, and Gamma marginal density functions. Hence an approximate procedure for computing high-order hyperelliptic integrals arisen from the modeling process is developed. The resulting model outperforms other existing models discussed in the open literature. Through numerical simulations the accuracy of the proposed model is verified.

1. INTRODUCTION

In the modeling of adaptive algorithms, one initially establishes certain analysis conditions for which the sought model is expected to provide accurate predictions. In addition, the prediction quality is associated with the involved mathematical complexity in the model development. The latter is concerned with the assumed mathematical assumptions, among others. For instance, the stochastic model of the filtered-x least-mean-square (FxLMS) algorithm obtained under the light of the Independence Theory (IT) (usually applied to modeling LMS-like algorithms [1]), although mathematically simple, is very inaccurate in this case. That is so because the IT assumption disregards the correlations between input signal vectors at different time lags created by the secondary path [2], [3]. On the other hand, taking into account all correlations of the input signal vectors (even under white input signals), the obtained model is more complex but much more accurate [4]. Regarding the modeling of the normalized LMS (NLMS) algorithm discussed in [5], only one assumption, the statistical independence between the input data and the adaptive weight vectors, is considered.

Such a hypothesis is commonly used for modeling purposes of LMS-based adaptive algorithms [1]. A particularly tricky obstacle found in the modeling of the NLMS algorithm is the computation of expectations involving the normalization factor $x^T(n)x(n)$. An example is the determination of the expected value $E[\sum_{i=1}^{n} w_i x_i(n) / \sum_{i=1}^{n} x_i(n) x_i(n)]$, needed for obtaining the first and second moments of the adaptive weight vector. In [5], such an expectation is formulated by considering the input signal obtained from a spherical invariant random process (SIRP). In this formulation, a stationary multivariate Gaussian process can be taken as a particular case. As a consequence of such an approach, a high-order hyperelliptic or Abelian integral must be computed. In the literature there is no available general solution for that kind of integral; however, assuming that vector $x(n)$ contains independent samples, the computation of such an integral can be especially eased, as presented in [6]. Note that the previous assumption is not true for correlated input signals, resulting in an inaccurate model. In the open literature, the referred integral class has known solution only for order equal or less than four, in this case termed elliptic integrals [7]. In [5], comparison results between simulations and model were shown for a fourth-tap adaptive filter.

Under some considerations, the problem of computing the expected values is treated using different strategies. In [1], the approximation used is $E[\sum_{i=1}^{n} w_i x_i(n) / \sum_{i=1}^{n} x_i(n) x_i(n)] \approx E[\sum_{i=1}^{n} w_i x_i(n) x_i(n)] / E[\sum_{i=1}^{n} x_i(n) x_i(n)]$, leading to satisfactory modeling results for adaptive filters having both a large number of weights and independent input data. Recently in [6], [8], and [9], by invoking the Averaging Principle (AP) [10], the expected value $E[\sum_{i=1}^{n} w_i x_i(n) x_i(n)]$ is split as $E[\sum_{i=1}^{n} w_i x_i(n) x_i(n)] / E[\sum_{i=1}^{n} x_i(n) x_i(n)]$, simplifying notably the involved mathematics. Such an approximation is based on the ergodicity of the input process, leading to acceptable results if the dimension of vector $x(n)$ is large. Since a stationary Gaussian process is the sole one that is an ergodic SIRP [5], one cannot consider the use of AP for non-Gaussian SIRP.

©2007 EURASIP
In [6] and [8], expectation $E[1/\{x^T(n)x(n)\}]$ is obtained by assuming independent samples, failing once more for correlated signals. In [11], a model based on the stochastic differential equation (SDE) approach is discussed. Such a model provides satisfactory prediction of the mean-square error (MSE) as the adaptive algorithm is operating at the highest convergence rate. However, the model does not work well for lower convergence rates, resulting in a somewhat poor modeling process.

Looking for model accuracy, we return to the starting point presented in [5], in which just the statistical independence between the input data and the adaptive weight vectors is considered. In this way, the first and second moments of the adaptive weight vector are dependent on the computation of a high-order hyperelliptic integral. In [9], an approximate method for computing hyperelliptic integrals was considered to estimate the expected value $E[1/\{x^T(n)x(n)\}]$ for Gaussian signals; however, the accuracy of the obtained results are still dependent on the conditions for using the AP.

In this paper, we have not invoked the AP considered in [9]. However, to solve the hyperelliptic integrals the approach introduced in [9] is applied here. As a result, the accuracy of the obtained results are improved for both transient and steady-state responses as compared with previous approaches, specifically for correlated input data. In this way, the present research work confirm the accuracy of the proposed model.

2. MEAN-WEIGHT BEHAVIOR

The weight update equation of the standard NLMS algorithm is given by [1], [5]

$$w(n+1) = w(n) + \beta \frac{e(n)x(n)}{x^T(n)x(n)}$$

where $e(n) = d(n) - x^T(n)w(n) + z(n)$

$$d(n) = x(n) - x^T(n)w(n)$$

(2)

denotes the error signal, $w(n) = [w(n) w(n-1) \cdots w(n-N+1)]^T$ is the adaptive weight vector, $d(n)$ represents the desired signal, and $z(n)$ is a zero-mean i.i.d. measurement noise with variance $\sigma_z^2$ and uncorrelated with any other signal in the system. The step-size control parameter is denoted by $\beta$. The input vector is represented by $x(n) = [x(n) x(n-1) \cdots x(n-N+1)]^T$, being $\sigma_x^2$ the variance of the process $x(n)$. By substituting (1) into (2) and taking the expected value of both sides of the resulting expression, one obtains

$$E[w(n+1)] = E[w(n)] + \beta E\left[\frac{x(n)d(n)}{x^T(n)x(n)}\right]$$

$$-\beta E\left[\frac{x(n)x^T(n)}{x^T(n)x(n)}\right] + \beta E\left[\frac{x(n)z(n)}{x^T(n)x(n)}\right].$$

(3)

To determine the expectations in (3), we consider that $w(n)$ and $x(n)$ are statistically independent (slow adaptation condition [1]). Thus, we can rewrite (3) as follows:

$$E[w(n+1)] = E[w(n)] + \beta E\left[\frac{x(n)d(n)}{x^T(n)x(n)}\right]$$

$$-\beta E\left[\frac{x(n)x^T(n)}{x^T(n)x(n)}\right] E[w(n)].$$

(4)

The fourth term in the r.h.s. of (3) is zero by virtue of the nature of $z(n)$ [1]. The main point now is to determine the expected values $E_1$ and $E_2$ in (4). The procedure for computing such expected values is notably complex, since it requires that a high-order hyperelliptic integral is computed. The expressions for $E_1$ and $E_2$ can be written as [5]

$$E\left[\frac{x(n)x^T(n)}{x^T(n)x(n)}\right] = QH^T$$

(5)

and

$$E\left[\frac{x(n)d(n)}{x^T(n)x(n)}\right] = -Q A^{-1} H^T p$$

(6)

where $p = E[d(n)x(n)]$ is the cross-correlation vector between the desired signal and the input vector. $Q$ and $A$ are, respectively, the eigenvector matrix and diagonal matrix containing the eigenvalues $\lambda_j$ of $R = E[x(n)x^T(n)]$ (input autocorrelation matrix). $H$ is a diagonal matrix whose elements are given by

$$h_j = \int_0^\infty \frac{\lambda_j^d} {1 + 2\alpha \lambda_j} \prod_{j=1}^N (1 + 2\alpha \lambda_j).$$

(7)

A solution for this integral does not have a closed form as can be verified from the mathematics literature [7]. Here, we use the approach given in [9], replacing every two neighboring eigenvalues of matrix $R$ by its geometric mean with multiplicity two. Such a strategy eliminates the square-root operation in the denominator of (7). Then, by considering that in (7) there are only real and distinct roots (the more frequent case with correlated input signals), the partial fraction expansion of the integrand of (7) (now without the square root) leads to a closed-form solution, given by

$$h_j \approx \frac{1}{2} \sqrt{\alpha} \sum_{q=1}^{N/2} A_q \ln(\lambda^j_q) + B_{1,j} \ln(\lambda^j).$$

(8)
where 

\[ A_{q,i} = \frac{1}{\lambda_i} - \frac{1}{\lambda_q} \prod_{j=q+1}^{N/2} \frac{1}{\lambda_j} - \frac{1}{\lambda_q} \prod_{j=q+1}^{N/2} \frac{1}{\lambda_j} \]  

and \( B_{i,j} = \prod_{q=1}^{N/2} \frac{1}{\lambda_q} - \frac{1}{\lambda_q} \prod_{j=q+1}^{N/2} \frac{1}{\lambda_j} \)

\[ a_N = \prod_{i=1}^N \lambda_i \quad \text{and} \quad \lambda_i' = \sqrt{\lambda_i - \lambda_q}, \quad q = 1, 2, \ldots, N/2. \]

According to results obtained from numerical simulations, the matrix represented by \( E_2 \) has a smaller eigenvalue spread \( \chi \) than matrix \( R \) (evidencing the advantage of the NLMS algorithm over the LMS one). The proposed method to compute \( E_2 \) here preserves such a characteristic. In contrast, when \( E_2 \) is computed by using some simplifying assumption, such as the AP principle \([8, 9]\), the resulting eigenvalue spread is the same as that of \( R \), disagreeing thus with the results from simulations. For instance, considering the case \( N = 8 \), the eigenvalue spread of \( R \) is 35.11 and eigenvalue spreads of \( E_2 \) are 16.67, and 16.47 obtained from simulation and proposed model, respectively. On the other hand, the AP approach gives an eigenvalue spread equal to 35.11, verifying the previously stated.

### 3. LEARNING CURVE

By defining the weight-error vector as \( \mathbf{v}(n) = \mathbf{w}(n) - \mathbf{w}_o \), where \( \mathbf{w}_o \) is the optimum weight vector, the error signal can be expressed as

\[ e(n) = d(n) - \mathbf{x}^T(n)\mathbf{v}(n) - \mathbf{x}^T(n)\mathbf{w}_o + z(n) \]

\[ = e_o(n) - \mathbf{x}^T(n)\mathbf{w}_o + z(n) \quad (9) \]

where

\[ e_o(n) = d(n) - \mathbf{x}^T(n)\mathbf{w}_o + z(n). \quad (10) \]

By squaring both sides of (10), determining the expected value of the resulting expression, and using the Principle of Orthogonality \([1]\), we obtain

\[ E[e^2(n)] = e_{\text{min}} + E[\mathbf{x}^T(n)\mathbf{v}(n)\mathbf{x}(n)\mathbf{v}(n)] \]

\[ = e_{\text{min}} + \text{tr}(\mathbf{R} \mathbf{K}(n)) = e_{\text{min}} + \lambda^T \mathbf{k}(n) \quad (11) \]

where \( e_{\text{min}} = E[e_o^2(n)] \) is the minimum steady-state error, \( \mathbf{K}(n) = E[\mathbf{v}(n)\mathbf{v}^T(n)] \) characterizes the covariance matrix of the weight-error vector (second moment), \( \lambda \) denotes a vector containing the eigenvalues \( \lambda_j \) of \( \mathbf{R} \), and \( \mathbf{k}(n) \) is a vector containing the diagonal elements of \( \mathbf{K}'(n) = \mathbf{Q}^T \mathbf{K}(n) \mathbf{Q} \). Note that (11) is completely determined if vector \( \mathbf{k}(n) \) is known. Then, by subtracting \( \mathbf{w}_o \) from both sides of (1), determining the outer product \( \mathbf{v}(n)\mathbf{v}^T(n) \), taking the expectation of both sides of the resulting expression, and pre- and post-multiplying by \( \mathbf{Q}^T \) and \( \mathbf{Q} \), respectively, we obtain the following recursion for \( \mathbf{k}(n) \) \([5]\):

\[ \mathbf{k}(n+1) = \mathbf{Bk}(n) + \mathbf{B}^2 e_{\text{min}} \quad (12) \]

where \( \mathbf{B} \) is a matrix having the elements given by

\[ b_{i,j} = \begin{cases} \beta^2 m_{ij}, & i \neq j \\ 1 - 2\beta h_j + 3\beta^2 \ell_j, & i = j \end{cases} \quad (13) \]

where

\[ \ell_j = \int_0^\infty \frac{\lambda^2 \omega d\omega}{(1 + 2\alpha \lambda_j)^2 \prod_{k=1}^N (1 + 2\alpha \lambda_k)} \quad (14) \]

\[ m_{i,j} = \int_0^\infty \frac{\lambda_i \lambda_j \omega d\omega}{(1 + 2\alpha \lambda_i)(1 + 2\alpha \lambda_j) \prod_{k=1}^N (1 + 2\alpha \lambda_k)} \quad (15) \]

and the \( i^\text{th} \) element of vector \( \mathbf{s} \) is

\[ s_i = \int_0^\infty \frac{\lambda_i}{\prod_{k=1}^N (1 + 2\alpha \lambda_k)} d\omega \quad (16) \]

To determine (12), one requires to compute several double hyperelliptic integrals \([14, 15, 16]\). For such, we again use the approach presented in \([9]\). For illustration proposes, we consider the case of all single roots. Accordingly, expressions (14), (15), and (16), result in

\[ \ell_i \approx \frac{1}{4} \sqrt{a_N} \prod_{q=1}^{N/2} \left\{ \frac{A_{i,q} \lambda_q'}{\lambda_i - \lambda_q} \right\} \left[ 1 + \ln (\lambda_q') \right] \]

\[ + B_{i,j} \ln (\lambda_i) \quad (17) \]

\[ m_{i,j} \approx \frac{1}{4} \sqrt{a_N} \prod_{q=1}^{N/2} C_{i,j,q} \left[ 1 + \ln (\lambda_q') \right] \]

\[ + c_{i,j} \left[ 1 + \ln (\lambda_i) \right] \]

and

\[ s_i = -\frac{1}{4} \sqrt{a_N} \left\{ \sum_{q=1}^{N/2} A_{i,q}' \left[ 1 + \ln (\lambda_q') \right] + B_{i,j} \left[ 1 + \ln (\lambda_i) \right] \right\} \quad (19) \]

where

\[ B_{i,j} = \frac{1}{\lambda_i} \sum_{p=1}^{N/2} \frac{\lambda_j - \lambda_q'}{\lambda_i} \left[ \prod_{k=q+1}^{N/2} \frac{\lambda_k'}{\lambda_i' \lambda_j} \right] c_{i,j} = \frac{\lambda_j}{\lambda_j - \lambda_i} \prod_{q=1}^{N/2} \frac{\lambda_q'}{\lambda_i' \lambda_j} \]

\[ C_{i,j,q} = \frac{\lambda_j'}{\lambda_j - \lambda_i} \prod_{q=1}^{N/2} \frac{\lambda_q'}{\lambda_i' \lambda_j} \]

\[ (\lambda_i - \lambda_q')(\lambda_j - \lambda_q') \prod_{k=q+1}^{N/2} \frac{\lambda_k'}{\lambda_i' \lambda_j'} \]
4. EXPERIMENTAL RESULTS

By considering a system identification problem, some examples are presented in this section aiming to assess the accuracy of the proposed model. The plant used is obtained from a scaled Hanning window function \( w_p \), with \( \|w_p\|=1 \). The input signal is correlated, obtained from an AR(2) process given by \( x(n) = a_1x(n-1) + a_2x(n-2) + v(n) \), where \( v(n) \) is a white noise signal with variance \( \sigma_v^2 \). The AR coefficients are \( a_1 = 0.61 \) and \( a_2 = -0.85 \).

The measurement noise \( z(n) \) has variance \( \sigma_z^2 = 10^{-4} \) (SNR = 40dB). In the presented examples, the step-size control parameter used is equal to \( 0.1 \beta_{\max} \), where \( \beta_{\max} \) is the step-size control parameter for maximum convergence speed, in this case \( \beta_{\max} = 1 \).

Fig. 1 shows the mean-weight behavior curves obtained from Monte Carlo (MC) simulation, the model using independent input data assumption (considered in [6] and [8]), the model using Averaging Principle (AP) [9], and the proposed model. The plant used is a length-8 vector, obtained from the Hanning window function, given by \( w_p = [0.0000.11620.37730.58670.58670.37730.11620.0000] \). The input signal is correlated with \( \chi = 74 \) (where \( \chi \) denotes the eigenvalue spread of the input autocorrelation matrix). The step-size control parameter is \( \beta_i = 0.1 \). From this figure, the accuracy achieved with the proposed model particularly during the transient phase is verified. For a better assessment of the model matching, Fig. 1(b) shows a detail of the transient phase for weight \( w_4(n) \), whose steady-state value is 0.5867.

To assess the effect of the number of weights on the modeling results, Fig. 2 depicts the learning curves obtained by using plants and adaptive filters having lengths 6, 8, 16, and 64, considering correlated input data with eigenvalue spreads equal to 57.3, 74, 135, and 254, respectively. In all considered cases, the results obtained by MC simulation (average of 500 independent runs), the model using independent input data assumption [8], the model based on the SDE approach [11], the model using AP [9], and the proposed model are shown. From these curves, a very good accuracy is again verified by using the proposed model, especially in the transient phase, considering all filter lengths and eigenvalue spreads used in the examples. Note that the larger the number of weights, the closer the model using AP [9] is to the proposed one, as can be seen in Fig. 2(d). This fact verifies the reasonability of using the AP for large \( N \).

5. CONCLUDING REMARKS

An improved stochastic modeling for the NLMS algorithm (first moment and learning curve) under slow adaptation condition is presented. Due to the analysis approach used here, the proposed model exhibits a very good accuracy particularly for correlated input data. The new model has been obtained taking into account a single assumption [independence between \( w(n) \) and \( x(n) \)], resulting in a model somewhat more complex than other ones (presented in the open literature), but much more accurate, especially in the transient phase. The model using the Averaging Principle allows obtaining a mathematically simpler model; however, such a model is restricted to adaptive filters having a large number of weights. In addition, an approximate method to compute high-order hyperelliptic integrals has been discussed.

REFERENCES

Figure 1 – Mean weight behavior for correlated input signal with $\chi = 74$ and plant length $N = 8$. (Gray-ragged line) MC simulation. (Dotted line) independent input data assumption [6], [8], and [11]. (Dashed line) model using AP [9]. (Dark-solid line) proposed model. (a) Curves of the adaptive filter weights. (b) Details of the transient phase of weight $w_4(n)$.

Figure 2 – Learning curves for correlated input data. (a) $N = 6$ and $\chi = 57.3$. (b) $N = 8$ and $\chi = 74$. (c) $N = 16$ and $\chi = 135$. (d) $N = 64$ and $\chi = 254$. 