Mean Weight Behavior for the Generalized Subband Decomposition LMS Algorithm

Javier E. Kolodziej[1,2], Orlando J. Tobias[1,3], and Rui Seara[1]

[1]LINSE – Circuits and Signal Proc. Lab.
Department of Electrical Engineering
Fed. University of Santa Catarina, Brazil
seara@linse.ufsc.br

[2]Electronics Eng. Laboratory
National University of Misiones (UNaM)
Oberá, Misiones, Argentina
javier@linse.ufsc.br

[3]Department of Telecommunications
University of Blumenau (FURB)
Blumenau, Santa Catarina, Brazil
orlando@furb.br

***Abstract*—This paper presents an improved stochastic model for the generalized subband decomposition least-mean-square (GSD-LMS) algorithm. This algorithm is used as an alternative to the standard LMS aiming to improve the convergence speed under correlated input data. An analytical model for the first moment of the adaptive filter weights is derived considering just the independence between weight and input data vectors. Numerical simulation results confirm the accuracy of the proposed model, outperforming other models presented previously in the literature.**

***Keywords*—Averaging principle, least-mean-square (LMS) algorithm, mean weight behavior, subband adaptive filters.**

I. Introduction

One of the most popular and well-known adaptive algorithm is the least-mean-square (LMS) [1]–[3]. Due to its simple implementation and robustness, it has been used in an extensive number of applications, such as echo cancellation, channel equalization, adaptive beamforming, among others. However, the standard LMS algorithm does not work well for the case of correlated input data. In the open literature [1], [3], several adaptive algorithms have been proposed aiming to overcome this drawback. For instance, recursive least-squares (RLS) and affine projection (AP) algorithms are approaches that exhibit very good convergence characteristics for correlated input signals. As a general matter, this better performance is attained at the expense of a considerable increase in the computational load. Thus, it is interesting to allow for intermediate solutions, presenting lower computational complexity. In this context, several LMS subband algorithms have been proposed intending to improve the convergence of the standard LMS algorithm for correlated input data [2].

The generalized subband decomposition (GSD) structure has provided evidence of being an efficient implementation for FIR filters [4], [5]. This structure implements an $(N-1)$th–order FIR filter by using an M–branch parallel array with $1 \le M \le N$. Each branch, acting on a specific subband, is composed of a cascade of an interpolator and a sparse subfilter. In the adaptive GSD structure, the subfilters are adapted while the interpolators are maintained fixed.

The interpolators are implemented by using an orthogonal transform, such as discrete Fourier (DFT), discrete cosine (DCT), and Hadamard transforms, among others. The corresponding subfilters are adapted using a modified version of the LMS algorithm including a normalization stage, resulting in the GSD-LMS algorithm.

The fact of using subband filters permits increasing the convergence speed of the overall adaptive structure when correlated input signals are used. Besides, neglecting the bands which present low activity level, the computational burden is reduced.

Then, having the GSD-LMS algorithm at hand, it is possible to get a stochastic model. Such a modeling is an important tool in adaptive signal processing. Its purpose is to predict the algorithm performance under different operating conditions, as well as to derive appropriate algorithm parameters. The modeling task of adaptive algorithms leads to some mathematical challenges which can be circumvented with the use of simplifying assumptions. Thus, depending on the severity of the assumptions used, there is a trade-off between the involved mathematics and the attained model accuracy. Regarding the mean weight behavior of the GSD-LMS algorithm, the required normalizing operation is carried out from estimates of the signal power at each band [4], leading to expected values as

$$E[\hat{\mathbf{D}}_{\mathrm{a}}^{-1}(n)\hat{\mathbf{R}}_{\mathbf{u}_{\mathrm{a}}}(n)] \tag{1}$$

where $\hat{\mathbf{D}}_{\mathrm{a}}(n)$ and $\hat{\mathbf{R}}_{\mathbf{u}_{\mathrm{a}}}(n)$ are matrices depending on the input data. To compute (1), several approximations have been used in the open literature [4], [6], each of them having different degrees of accuracy. In [4], $\hat{\mathbf{D}}_{\mathrm{a}}(n)$ is considered a time-invariant matrix, leading (1) to be expressed as

$$E[\hat{\mathbf{D}}_{\mathrm{a}}^{-1}(n)\hat{\mathbf{R}}_{\mathbf{u}_{\mathrm{a}}}(n)] \cong \hat{\mathbf{D}}_{\mathrm{a}}^{-1}E[\hat{\mathbf{R}}_{\mathbf{u}_{\mathrm{a}}}(n)] \tag{2}$$

disregarding then the varying nature of $\hat{\mathbf{D}}_{\mathrm{a}}(n)$.

In [6] and [7], the Averaging Principle (AP) [8] is used to split (1) into two expected values, i.e.,

$$E[\hat{\mathbf{D}}_{\mathrm{a}}^{-1}(n)\hat{\mathbf{R}}_{\mathbf{u}_{\mathrm{a}}}(n)] \cong E[\hat{\mathbf{D}}_{\mathrm{a}}^{-1}(n)]E[\hat{\mathbf{R}}_{\mathbf{u}_{\mathrm{a}}}(n)]. \tag{3}$$

Then, $E[\hat{\mathbf{R}}_{\mathbf{u}_{\mathrm{a}}}(n)]$ is easily determined, whereas $E[\hat{\mathbf{D}}_{\mathrm{a}}^{-1}(n)]$ is obtained by assuming either independent signals [6] or (in a more accurate way) by considering the correlations between the involved signals, as in [7]. The use of the AP results in simpler model expressions, having satisfactory accuracy (mainly as applied in [7]) for large observation windows. However, when shorter windows are considered for computational complexity reasons, the use of the AP is no longer adequate. Thus, looking for an accurate model, irrespective of the observation window length, other alternative ways, different from using the AP, should be searched for.

The goal of this paper is to derive an accurate statistical model, describing the GSD-LMS algorithm behavior. In particular, we derive an analytical expression for the first moment of the adaptive filter weight vector for Gaussian input signals. Through numerical simulations, the accuracy of the proposed model is verified.

II. Generalized Subband Decomposition LMS Algorithm

This section presents the basic expressions describing the GSD-LMS algorithm [4], upholding the block diagram of the adaptive subband structure illustrated in Fig. 1.

A. Standard GSD Structure

In this structure, the input data $x(n)$ is first processed by an M–point unitary transform, implemented by an $M \times M$

nonsingular matrix $\mathbf{T}$, generating the signals $u_m(n)$ with $m = 0,1,\ldots,M-1$, which are then filtered by the sparse subfilters $W_m(z)$. The samples of the transformed signals, used as inputs for the subfilters, form the (KM)–dimensional vector $\mathbf{u}_a(n)$, given by

$$\mathbf{u}_a(n) = \{\mathbf{u}^T(n)\ \ \mathbf{u}^T(n-L)\cdots\mathbf{u}^T[n-(K-1)L]\}^T \tag{4}$$

with

$$\mathbf{u}(n) = [u_0(n)\ \ u_1(n)\cdots u_{M-1}(n)]^T \tag{5}$$

where K is the number of nonzero weights in each subfilter and L, the sparseness factor.

Now, defining the generalized subband weight vector $\mathbf{w}_a(n)$ as the vector containing all weights of the subfilters, i.e.,

$$\begin{aligned}\mathbf{w}_a(n) = [&w_{0,0}(n)\ \ w_{1,0}(n)\cdots w_{M-1,0}(n)\\ &w_{0,1}(n)\ \ w_{1,1}(n)\cdots w_{M-1,1}(n)\cdots\\ &w_{0,K-1}(n)\ \ w_{1,K-1}(n)\cdots w_{M-1,K-1}(n)]^T\end{aligned} \tag{6}$$

the filter output $y(n)$ is given by

$$y(n) = \mathbf{u}_a^T(n)\mathbf{w}_a(n). \tag{7}$$

B. Adaptive GSD Structure

Now, considering that the vector $\mathbf{w}_a(n)$ is adapted by using the LMS algorithm, an adaptive version of the GSD structure can be obtained. From Fig. 1, the error signal $e(n)$ is written as

$$e(n) = d(n) - y(n) + z(n) \tag{8}$$

with $d(n)$ denoting the desired signal and $z(n)$, a measurement noise i.i.d., zero-mean with variance σ_z^2 and uncorrelated with any other signal in the system. Thus, the weight update equation is expressed as [4]

$$\mathbf{w}_a(n+1) = \mathbf{w}_a(n) + \mu\,\mathbf{D}_a^{-1}\left[d(n) + z(n) - y(n)\right]\mathbf{u}_a(n) \tag{9}$$

with

$$\mathbf{D}_a = \mathbf{I}_K \otimes \mathbf{D} \tag{10}$$

where μ is the step size, $\mathbf{I}_K$ represents a $K\times K$-dimensional identity matrix, $\mathbf{D} = \mathrm{diag}[\sigma_0^2\ \sigma_1^2\cdots\sigma_{M-1}^2]$ is a diagonal matrix having as elements the variances of each subband, and $\otimes$ denotes the Kronecker product.

Then, to implement the GSD-LMS algorithm it is necessary to determine the variances of each subband σ_m^2 for accomplishing the step-size normalization operation. Instead of using the instantaneous power, which is susceptible to fast variations of the signal during the adaptation process, it is common practice to consider the average power [2]. For such, we use here the following estimate:

$$\hat{\sigma}_m^2(n) = \frac{1}{M_w}\mathbf{u}_m^T(n)\mathbf{u}_m(n) \tag{11}$$

with

$$\mathbf{u}_m(n) = [u_m(n)\ \ u_m(n-1)\cdots u_m(n-M_w+1)]^T \tag{12}$$

where M_w is the length of the observation window. This estimation makes $\hat{\sigma}_m^2$ time varying; then, in our approach, matrix $\mathbf{D}_a$ is replaced in (9) by its time varying version $\hat{\mathbf{D}}_a(n)$, having now as diagonal elements $\hat{\sigma}_m^2(n)$. The normalization matrix $\hat{\mathbf{D}}(n)$ reads then

$$\hat{\mathbf{D}}(n) = \mathrm{diag}[\hat{\sigma}_0^2(n)\ \ldots\ \hat{\sigma}_{M-1}^2(n)\ ]. \tag{13}$$

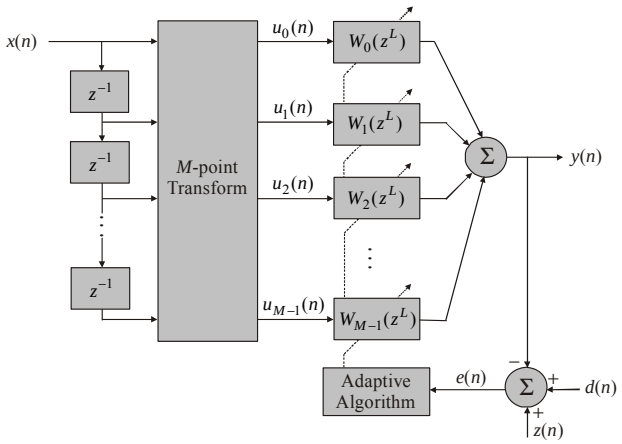

Fig. 1. Block diagram of the GSD-LMS algorithm.

III. Mean-Weight Behavior

A. Problem Statement

In this section, the mean weight expression of the adaptive weight vector is obtained by taking the expectation of both sides of (9). By using the simplifying assumption that $\mathbf{w}_a(n)$ and $\mathbf{u}_a(n)$ are statistically independent [1], we can write

$$\begin{aligned}E[\mathbf{w}_a(n+1)] = E[\mathbf{w}_a(n)] &+ \mu E[\hat{\mathbf{D}}_a^{-1}(n)\mathbf{u}_a(n)d(n)]\\ &- \mu E[\hat{\mathbf{D}}_a^{-1}(n)\mathbf{u}_a(n)\mathbf{u}_a^T(n)]E[\mathbf{w}_a(n)].\end{aligned} \tag{14}$$

To determine (14), the second term on the r.h.s., defined as the sample normalized covariance matrix

$$\mathbf{R}_{\mathbf{u}_a}^N = E[\hat{\mathbf{D}}_a^{-1}(n)\mathbf{u}_a(n)\mathbf{u}_a^T(n)] \tag{15}$$

and the third term of the r.h.s., defined as the sample normalized cross-correlation vector

$$\mathbf{p}_{\mathbf{u}_a d}^N = E[\hat{\mathbf{D}}_a^{-1}(n)\mathbf{u}_a(n)d(n)] \tag{16}$$

must be computed. The determination of (15) and (16) is notably complex, involving the calculation of high-order hyperelliptic integrals, not existing closed solutions for such integrals in the mathematical literature [9]. Thus, the available GSD-LMS models are based on some simplifying assumptions, which are briefly presented in the next section.

B. Available Models

In this section, two models available in the literature are presented. They are based on severe simplifications, resulting in models somewhat restricted to certain operating conditions.

1) Model A: In [4], expectations (15) and (16) are obtained by considering that $\hat{\mathbf{D}}_a(n)$ a time-invariant matrix, resulting in

$$\mathbf{R}_{\mathbf{u}_a\mathbf{u}_a}^N \cong \mathbf{D}_a^{-1}E[\mathbf{u}_a(n)\mathbf{u}_a^T(n)] = \mathbf{D}_a^{-1}\mathbf{R}_{\mathbf{u}_a} \tag{17}$$

and

$$\mathbf{p}_{\mathbf{u}_a d}^N \cong \mathbf{D}_a^{-1}\mathbf{p}_{\mathbf{u}_a d} \tag{18}$$

where $\mathbf{R}_{\mathbf{u}_a} = E[\mathbf{u}_a(n)\mathbf{u}_a^T(n)]$ and $\mathbf{p}_{\mathbf{u}_a d} = E[\mathbf{u}_a(n)d(n)]$ are, respectively, the autocorrelation matrix of the transformed input vector and the cross-correlation vector between the desired signal and the transformed input vector. Then, substituting (17) and (18) into (14), we get

$$E[\mathbf{w}_a(n+1)] = (\mathbf{I} - \mu\mathbf{D}_a^{-1}\mathbf{R}_{\mathbf{u}_a})E[\mathbf{w}_a(n)] + \mu\mathbf{D}_a^{-1}\mathbf{p}_{\mathbf{u}_a d}. \tag{19}$$

Note that (19) disregards the time-variant nature of $\hat{\mathbf{D}}_{\mathrm{a}}(n)$ as well as its correlation with $\mathbf{u}_{\mathrm{a}}(n)$.

2) Model B: From the approximation found in [6], expected values (15) and (16) have been computed by first splitting them into two terms, invoking the AP. Then, it is assumed that the process $\{u_m^2(n)\}$ has a chi-square distribution (even when it does not have) with M_{w} degrees of freedom [6]. In doing this, expectations (15) and (16) are, respectively, given by

$$\mathbf{R}_{\mathbf{u}_{\mathrm{a}}}^{\mathrm{N}} \cong E[\hat{\mathbf{D}}_{\mathrm{a}}^{-1}(n)]\mathbf{R}_{\mathbf{u}_{\mathrm{a}}} \tag{20}$$

and

$$\mathbf{p}_{\mathbf{u}_{\mathrm{a}}d}^{\mathrm{N}} \cong E[\hat{\mathbf{D}}_{\mathrm{a}}^{-1}(n)]\mathbf{p}_{\mathbf{u}_{\mathrm{a}}d} \tag{21}$$

where

$$E[\hat{\mathbf{D}}_{\mathrm{a}}^{-1}(n)] = \frac{M_{\mathrm{w}}}{M_{\mathrm{w}}-2}\mathbf{D}_{\mathrm{a}}^{-1}. \tag{22}$$

Therefore, substituting (22) into (21) and (20) and the resulting expressions into (14), we obtain

$$E[\mathbf{w}_{\mathrm{a}}(n+1)] = \left(\mathbf{I} - \mu\frac{M_{\mathrm{w}}}{M_{\mathrm{w}}-2}\mathbf{D}_{\mathrm{a}}^{-1}\mathbf{R}_{\mathbf{u}_{\mathrm{a}}}\right)E[\mathbf{w}_{\mathrm{a}}(n)] + \mu\frac{M_{\mathrm{w}}}{M_{\mathrm{w}}-2} \times \mathbf{D}_{\mathrm{a}}^{-1}\mathbf{p}_{\mathbf{u}_{\mathrm{a}}d}. \tag{23}$$

Such an approach is true for white signals, failing for this application, since the signal $u_m(n)$ is correlated to each other, i.e., it can be seen as the output of a bandpass filter. Thus, ignoring such correlations, the obtained model becomes somewhat inaccurate. However, the required mathematics for this approach is very simple, giving good results for very small step sizes and long observation windows.

C. Proposed Model

In this work, (15) and (16) are computed without using AP and not restricting the signals involved to be white. Thereby, the elements of matrix $\mathbf{R}_{\mathbf{u}_{\mathrm{a}}}^{\mathrm{N}}$ are given by

$$r_{\mathbf{u}_{\mathrm{a}}}^{\mathrm{N}}(i,j) = E\left[\frac{u_{m_1}(n-l_1)u_{m_2}(n-l_2)}{\hat{\sigma}_{m_1}^2(n)}\right] \tag{24}$$

with $l_1 = L(\lceil i/M \rceil - 1)$, $l_2 = L(\lceil j/M \rceil - 1)$, and $m_1 = \mathrm{mod}(i-1, M)$ and $m_2 = \mathrm{mod}(j-1, M)$, where $\lceil\cdot\rceil$ is the ceiling function and $\mathrm{mod}(\cdot)$ denotes the modulo function [4]. To determine (24), we define a vector given by

$$\mathbf{u}_{i,j}(n) = [u_{m_1}(n)\cdots u_{m_1}(n-M_{\mathrm{w}}+1)\ \ u_{m_1}(n-l_1)\ \ u_{m_2}(n-l_2)]^{\mathrm{T}} \tag{25}$$

such that

$$\hat{\sigma}_{m_1}^2(n) = \frac{1}{M_{\mathrm{w}}}\mathbf{u}_{i,j}^{\mathrm{T}}(n)\mathbf{I}_{\mathrm{s}}\mathbf{u}_{i,j}(n) \tag{26}$$

with $\mathbf{I}_{\mathrm{s}} = \mathrm{diag}[\overbrace{1....1}^{M_{\mathrm{w}}}0\ \ 0]$.

By considering a jointly Gaussian random process [10], (24) is then obtained by

$$r_{\mathbf{u}_{\mathrm{a}}}^{\mathrm{N}}(i,j) = \frac{M_{\mathrm{w}}}{\sqrt{(2\pi)^{M_{\mathrm{w}}+2}\det(\mathbf{R}_{i,j})}} \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{M_{\mathrm{w}}+2 \text{ fold}} \frac{u_{m_1}(n-l_1)u_{m_2}(n-l_2)}{\mathbf{u}_{i,j}^{\mathrm{T}}(n)\mathbf{I}_{\mathrm{s}}\mathbf{u}_{i,j}(n)} \times \mathrm{e}^{-\mathbf{u}_{i,j}^{\mathrm{T}}(n)\mathbf{R}_{i,j}^{-1}\mathbf{u}_{i,j}(n)/2} d\mathbf{u}_{i,j} \tag{27}$$

where $\mathbf{R}_{i,j} = E[\mathbf{u}_{i,j}(n)\mathbf{u}_{i,j}^{\mathrm{T}}(n)]$ is the autocorrelation matrix of $\mathbf{u}_{i,j}(n)$. To determine (27), we use an approach similar to that presented in [11] (see Appendix), resulting in

$$r_{\mathbf{u}_{\mathrm{a}}}^{\mathrm{N}}(i,j) = M_{\mathrm{w}}(\mathbf{r}_i^{\mathrm{T}}\mathbf{R}_{m_1}^{\mathrm{N}}\mathbf{r}_{m_1,j} - d_{m_1}\mathbf{r}_i^{\mathrm{T}}\mathbf{R}_{m_1}^{-1}\mathbf{r}_{m_1,j} + d_{m_1}r_{i,j}) \tag{28}$$

with

$$\mathbf{R}_{m_1}^{\mathrm{N}} = \mathbf{Q}_{m_1}\mathbf{H}_{m_1}\mathbf{Q}_{m_1}^{\mathrm{T}} \tag{29}$$

where $\mathbf{Q}_{m_1}$ is the eigenvector matrix of $\mathbf{R}_{m_1} = E[\mathbf{u}_{m_1}(n)\mathbf{u}_{m_1}^{\mathrm{T}}(n)]$, with $\mathbf{u}_{m_1}(n)$ defined as in (12), $\mathbf{r}_i = E[\mathbf{u}_{m_1}(n)u_{m_1}(n-l_1)]$, $\mathbf{r}_{m_1,j} = E[\mathbf{u}_{m_1}(n)u_{m_2}(n-l_2)]$, $r_{i,j} = E[u_{m_1}(n-l_1)u_{m_2}(n-l_2)]$, and $\mathbf{H}_{m_1}$ is a diagonal matrix whose elements are given by

$$h_{m_1}(l,l) = \frac{1}{2\lambda_{m_1,l}^2\sqrt{a_{m_1}}}\left[\sum_{q=1}^{M_{\mathrm{w}}/2} A_{m_1,l,q}\ln(\lambda'_{m_1,q}) + B_{m_1,l}\ln(\lambda_{m_1,l})\right] \tag{30}$$

with

$$a_{m_1} = \prod_{k=1}^{M_{\mathrm{w}}}\lambda_{m_1,k}, \tag{31}$$

$$A_{m_1,l,q} = \frac{\lambda_{m_1,l}\lambda'^2_{m_1,q}}{\lambda'_{m_1,q}-\lambda_{m_1,l}}\prod_{\substack{k=1\\k\neq q}}^{M_{\mathrm{w}}/2}\frac{\lambda'_{m_1,k}}{\lambda'_{m_1,q}-\lambda'_{m_1,k}}, \tag{32}$$

$$B_{m_1,l} = \prod_{k=1}^{M_{\mathrm{w}}/2}\frac{\lambda_{m_1,l}\lambda'_{m_1,k}}{\lambda_{m_1,l}-\lambda'_{m_1,k}}, \tag{33}$$

and

$$d_{m_1} = \frac{1}{2\sqrt{a_{m_1}}}\sum_{q=1}^{M_{\mathrm{w}}/2} C_{m_1,q}\ln(\lambda'_{m_1,q}) \tag{34}$$

with

$$C_{m_1,q} = \prod_{\substack{k=1\\k\neq q}}^{M_{\mathrm{w}}/2}\frac{\lambda'_{m_1,k}\lambda'_{m_1,q}}{\lambda'_{m_1,k}-\lambda'_{m_1,q}} \tag{35}$$

being $\lambda_{m_1,k}$ the eigenvalues of $\mathbf{R}_{m_1}$. Variable $\lambda'_{m_1,k}$ is obtained as follows [11]:

$$\lambda'_{m_1,k} = \sqrt{\lambda_{m_1,2k-1}\lambda_{m_1,2k}}, \quad k = 1,2,...,M_{\mathrm{w}}/2 \text{ with } M_{\mathrm{w}} \text{ even.} \tag{36}$$

Now, the elements of vector $\mathbf{p}_{\mathbf{u}_{\mathrm{a}}d}^{\mathrm{N}}$ are given by

$$p_{\mathbf{u}_{\mathrm{a}}d}^{\mathrm{N}}(i) = E\left[\frac{u_{m_1}(n-l_1)d(n)}{\hat{\sigma}_{m_1}^2(n)}\right]. \tag{37}$$

To determine the expected value (37), a procedure similar to the used to obtain (24) is applied here. Thereby, using

$$\mathbf{u}_{\mathrm{d},i}(n) = [u_{m_1}(n)...u_{m_1}(n-M_{\mathrm{w}}+1)...u_{m_1}(n-l_1)\ d(n)]^{\mathrm{T}} \tag{38}$$

instead of the vector given by (25), we get

$$p_{\mathbf{u}_{\mathrm{a}}d}^{\mathrm{N}}(i) = M_{\mathrm{w}}(\mathbf{r}_i^{\mathrm{T}}\mathbf{R}_{m_1}^{\mathrm{N}}\mathbf{r}_{\mathrm{d},m_1} - d_{m_1}\mathbf{r}_i^{\mathrm{T}}\mathbf{R}_{m_1}^{-1}\mathbf{r}_{d,m_1} + d_{m_1}r_{d,i}) \tag{39}$$

where $\mathbf{r}_{\mathrm{d},m_1} = E[\mathbf{u}_{m_1}(n)d(n)]$ and $r_{\mathrm{d},i} = E[u_{m_1}(n-l_1)d(n)]$.

Then, substituting (28) and (39) into (14), and expressing the result in scalar form, we obtain

$$E[w_{\mathrm{a},i}(n+1)] = \sum_{j=1}^{(K\times M)} [1-2\mu M_{\mathrm{w}}(\mathbf{r}_i^{\mathrm{T}}\mathbf{R}_{m_1}^{\mathrm{N}}\mathbf{r}_{m_1,j} - d_{m_1}\mathbf{r}_i^{\mathrm{T}}\mathbf{R}_{m_1}^{-1}\mathbf{r}_{m_1,j} + d_{m_1}r_{i,j})]E[w_{\mathrm{a},j}(n)] + M_{\mathrm{w}}(\mathbf{r}_i^{\mathrm{T}}\mathbf{R}_{m_1}^{\mathrm{N}}\mathbf{r}_{\mathrm{d},m_1} - d_{m_1}\mathbf{r}_i^{\mathrm{T}}\mathbf{R}_{m_1}^{-1}\mathbf{r}_{d,m_1} + d_{m_1}r_{d,i}) \tag{40}$$

where $w_{\mathrm{a},i}(n)$ denotes the ith element of $\mathbf{w}_{\mathrm{a}}(n)$.

V. Simulation Results

To assess the accuracy of the proposed model, an example is presented considering a system identification problem with correlated input data. The goal of this example is to compare and evaluate both the available and proposed models, showing the impact on the results when using different observation window lengths.

The input signal is obtained from an AR(2) process given by $x(n) = a_1 x(n-1) + a_2 x(n-2) + v(n)$, where $v(n)$ is a white noise with variance σ_v^2. The measurement noise $z(n)$ is white with variance $\sigma_z^2 = 10^{-4}$ (SNR = 40 dB). The AR coefficients are $a_1 = 0.6937$ and $a_2 = -0.8500$, resulting in an eigenvalue spread of the input autocorrelation matrix $\chi = 173.94$. Other parameters are: $\mu = 0.1\mu_{\max} = 0.002$, $M = 8$, $K = 4$, and $L = 4$. The example is presented for three values of observation window lengths, i.e., $M_{\mathrm{w}} = 8$, $M_{\mathrm{w}} = 16$, and $M_{\mathrm{w}} = 32$. The plant is the length-20 vector $(N = 20)$ defined as

$$\mathbf{p} = \frac{\mathbf{w}_{\mathrm{aux}}}{\sqrt{\mathbf{w}_{\mathrm{aux}}^{\mathrm{T}}\mathbf{w}_{\mathrm{aux}}}} \tag{41}$$

with

$$\mathbf{w}_{\mathrm{aux}} = [\mathrm{sinc}(0)\ \mathrm{sinc}(1/N)\ \cdots\ \mathrm{sinc}(N-1/N)]^{\mathrm{T}}. \tag{42}$$

The transformation used is the DCT. The curves of Fig. 2 show the first moment of three elements of vector $\mathbf{w}_{\mathrm{a}}(n)$, obtained from Monte Carlo (MC) simulations (average of 100 independent runs), for Model A (19), Model B (23), and proposed model (40). Note that there exists a very good match between numerical simulations and the proposed model for all cases assessed. Models A and B are more inaccurate mainly in the transient phase, in particular, for smaller window sizes [see Fig. 2(a)]. On the other hand, these models become more accurate as the window length increases. From a practical viewpoint, this situation is not a useful result, since short windows are generally used in practice.

VI. Concluding Remarks

In this paper, a model describing the first moment of the weight vector of the GSD-LMS algorithm was presented. The model is obtained by using only the independence assumption between the weight vector and the input signal vector, resulting in an accurate but mathematically more complex model. Comparisons with other available models from the literature showed that the proposed model provided a better prediction for both transient and steady-state responses in all assessed cases.

Appendix

Determination of (24)

To determine (24), an auxiliary function $f_{i,j}(\omega)$ defined as

$$f_{i,j}(\omega) = \frac{M_{\mathrm{w}}}{\sqrt{(2\pi)^{M_{\mathrm{w}}+2}\det(\mathbf{R}_{i,j})}} \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{M_{\mathrm{w}}+2\ \text{fold}} \frac{u_{m_1}(n-l_1)u_{m_2}(n-l_2)}{\mathbf{u}_{i,j}^{\mathrm{T}}\mathbf{I}_{\mathrm{s}}\mathbf{u}_{i,j}} \times \mathrm{e}^{-\mathbf{u}_{i,j}^{\mathrm{T}}\mathbf{L}_{i,j}^{-1}(\omega)\mathbf{u}_{i,j}/2} d\mathbf{u}_{i,j} \tag{43}$$

is used, with

$$\mathbf{L}_{i,j}(\omega) = (2\omega\mathbf{I}_{\mathrm{S}} + \mathbf{R}_{i,j}^{-1})^{-1}.$$

Note that for $\omega = 0$, (43) is by definition the desired expectation, i.e.,

$$r_{\mathbf{u}_{\mathrm{a}}}^{\mathrm{N}}(i,j) = f_{i,j}(0). \tag{44}$$

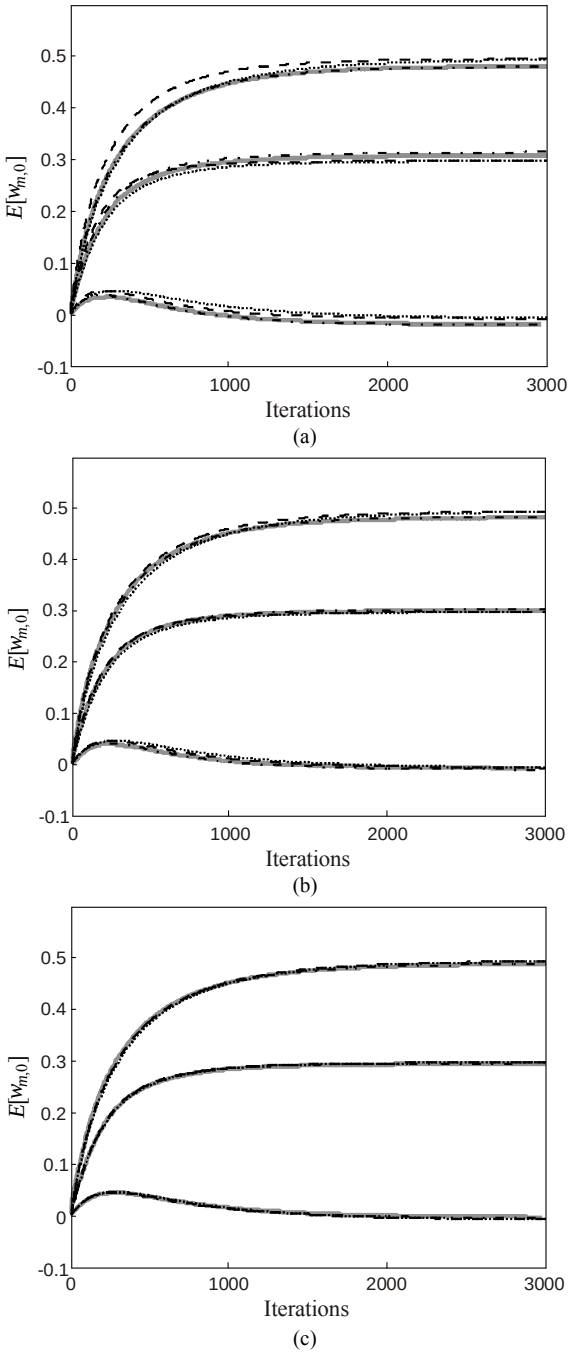

Fig. 2. Mean weight behavior curves. From top to bottom: $E[w_{1,0}(n)]$, $E[w_{2,0}(n)]$, and $E[w_{3,0}(n)]$. (Gray lines) MC simulations. (Dark-dotted lines) Model A (19). (Dark-dashed lines) Model B (23). (Dark-dot-dashed lines) proposed model (40). (a) $M_{\mathrm{w}} = 8$, (b) $M_{\mathrm{w}} = 16$, and (c) $M_{\mathrm{w}} = 64$.

Now, applying partial differentiation in (43) w.r.t ω, the term $\mathbf{u}_{i,j}^{\mathrm{T}}\mathbf{I}_{\mathrm{s}}\mathbf{u}_{i,j}$ in the denominator of the integrand (43) is eliminated, resulting in

$$\frac{\partial f_{i,j}(\omega)}{\partial \omega} = -M_{\mathrm{w}}\sqrt{\frac{\det[\mathbf{L}_{i,j}(\omega)]}{\det(\mathbf{R}_{i,j})}}\,\alpha(\omega) \tag{45}$$

with

$$\alpha(\omega) = \frac{1}{\sqrt{(2\pi)^{M_{\mathrm{w}}+2}\det[\mathbf{L}_{i,j}(\omega)]}} \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{M_{\mathrm{w}}+2 \text{ fold}} u_{m_1}(n-l_1)u_{m_2}(n-l_2) \times \mathrm{e}^{-\mathbf{u}_{i,j}^{\mathrm{T}}\mathbf{L}_{i,j}^{-1}(\omega)\mathbf{u}_{i,j}/2} d\mathbf{u}_{i,j}. \tag{46}$$

Expression (46) can be seen as the cross-correlation between $u_{m_1}(n-l_1)$ and $u_{m_2}(n-l_2)$, being these variable samples of a jointly Gaussian process with covariance matrix $\mathbf{L}_{i,j}(\omega)$. Since $u_{m_1}(n-l_1)$ and $u_{m_2}(n-l_2)$ are, respectively, the $(M_{\mathrm{w}}+1)$th and $(M_{\mathrm{w}}+2)$th elements of $\mathbf{u}_{i,j}$, and factor $\alpha(\omega)$ represents the $(M_{\mathrm{w}}+1, M_{\mathrm{w}}+2)$th element of matrix $\mathbf{L}_{i,j}(\omega)$. Therefore, we can write

$$\frac{\partial f_{i,j}(\omega)}{\partial \omega} = \frac{-M_{\mathrm{w}}[(2\omega\mathbf{R}_{i,j}\mathbf{I}_{\mathrm{s}}+\mathbf{I})^{-1}\mathbf{R}_{i,j}]_{M_{\mathrm{w}}+1,M_{\mathrm{w}}+2}}{\sqrt{\det(2\omega\mathbf{R}_{i,j}\mathbf{I}_{\mathrm{s}}+\mathbf{I})}} \tag{47}$$

where $\mathbf{I}$ is the identity matrix.

Now, considering the eigendecomposition of $\mathbf{R}_{i,j}\mathbf{I}_{\mathrm{s}} = \mathbf{Q}_{\mathrm{s},i,j}\mathbf{\Lambda}_{\mathrm{s},i,j}\mathbf{Q}_{\mathrm{s},i,j}^{-1}$, with $\mathbf{Q}_{\mathrm{s},i,j}$ and $\mathbf{\Lambda}_{\mathrm{s},i,j}$ being, respectively, the eigenvector and eigenvalue matrices of $\mathbf{R}_{i,j}\mathbf{I}_{\mathrm{s}}$, we obtain

$$\frac{\partial f_{i,j}(\omega)}{\partial \omega} = \frac{-M_{\mathrm{w}}[\mathbf{Q}_{\mathrm{s},i,j}(2\omega\mathbf{\Lambda}_{\mathrm{s},i,j}+\mathbf{I})^{-1}\mathbf{Q}_{\mathrm{s},i,j}^{-1}\mathbf{R}_{i,j}]_{M_{\mathrm{w}}+1,M_{\mathrm{w}}+2}}{\sqrt{\det(2\omega\mathbf{\Lambda}_{\mathrm{s},i,j}+\mathbf{I})}}. \tag{48}$$

The previous equation can be expressed in a more compact form taking into account the structure of $\mathbf{Q}_{\mathrm{s},i,j}$, resulting in

$$\frac{\partial f_{i,j}(\omega)}{\partial \omega} = \frac{-M_{\mathrm{w}}\{\mathbf{r}_i^{\mathrm{T}}\mathbf{Q}_{m_1}[(2\omega\mathbf{\Lambda}_{m_1}^2+\mathbf{\Lambda}_{m_1})^{-1}-\mathbf{\Lambda}_{m_1}]\mathbf{Q}_{m_1}^{\mathrm{T}}\mathbf{r}_{m_1,j}+r_{i,j}\}}{\sqrt{\prod_{k=1}^{M_{\mathrm{w}}}(2\omega\lambda_{m_1,k}+1)}}. \tag{49}$$

Then, considering

$$\int_0^{\infty}\frac{\partial}{\partial\omega} f_{i,j}(\omega)d\omega = f_{i,j}(\infty) - f_{i,j}(0) \tag{50}$$

and since $f_{i,j}(\infty) = 0$, we obtain

$$r_{\mathbf{u}_{\mathrm{a}}}^{\mathrm{N}}(i,j) = M_{\mathrm{w}}(\mathbf{r}_i^{\mathrm{T}}\mathbf{Q}_{m_1}\mathbf{H}_{m_1}\mathbf{Q}_{m_1}^{\mathrm{T}}\mathbf{r}_{m_1,j} - d_{m_1}\mathbf{r}_i^{\mathrm{T}}\mathbf{R}_{m_1}^{-1}\mathbf{r}_{m_1,j} + d_{m_1}r_{i,j}) \tag{51}$$

where the elements of diagonal matrix $\mathbf{H}_{m_1}$ are given by

$$h_{m_1}(l,l) = \int_0^{\infty}\frac{1}{\lambda_{m_1,l}(1+2\omega\lambda_{m_1,l})\sqrt{\prod_{k=1}^{M_{\mathrm{w}}}(2\omega\lambda_{m_1,k}+1)}}d\omega \tag{52}$$

and

$$d_{m_1} = \int_0^{\infty}\frac{1}{\sqrt{\prod_{k=1}^{M_{\mathrm{w}}}(2\omega\lambda_{m_1,k}+1)}}d\omega. \tag{53}$$

To determine (52) and (53), high-order hyperelliptic integrals must be computed. In the open literature, the referred integral has only a closed general solution for $M_{\mathrm{w}} \le 4$. However, an approximate solution of (52) and (53) can be obtained by replacing adjacent pairs of eigenvalues $\lambda_{m_1,k}$ of $\mathbf{R}_{m_1}$ by its geometric mean [denoted by (36)] with multiplicity two. Thus, (52) can be rewritten as

$$h_{m_1}(l,l) \cong \int_0^{\infty}\frac{d\omega}{\lambda_{m_1,l}(1+2\omega\lambda_{m_1,l})\prod_{q=1}^{M_{\mathrm{w}}/2}(1+2\omega\lambda'_{m_1,q})}. \tag{54}$$

Then, considering that in (54) there are only real and distinct roots (a common situation for correlated input signals), the partial fraction expansion of its integrand allows obtaining an approximate closed solution of (52). Thus,

$$h_{m_1}(l,l) \cong \frac{1}{2\lambda_{m_1,l}^2\sqrt{a_{m_1}}}[\sum_{q=1}^{M_{\mathrm{w}}/2}A_{m_1,l,q}\ln(\lambda'_{m_1,q}) + B_{m_1,l}\ln(\lambda_{m_1,l})]. \tag{55}$$

In a similar way, (53) is rewritten as

$$d_{m_1} \cong \frac{1}{2\sqrt{a_{m_1}}}\sum_{q=1}^{M_{\mathrm{w}}/2}C_{m_1,q}\ln(\lambda'_{m_1,q}). \tag{56}$$

ACKNOWLEDGMENT

The authors are thankful to the Committee for Postgraduate Courses in Higher Education (CAPES) and the National Council for Scientific and Technological Development (CNPq) by the financial supporting of this research.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 4ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[2] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*. New York: John Wiley & Sons, 1998.

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

[4] M. R. Petraglia and S. K. Mitra, "Adaptive FIR filter structure based on the generalized subband decomposition of FIR filters," *IEEE Trans. Circuits Syst. II*, vol. 40, no. 6, pp. 354-362, June 1993.

[5] E. V. Papoulis and T. Stathaki, "Extension of generalised subband decomposition-based adaptive FIR structure," *Electronics Letters*, vol. 19, no. 15, pp. 1157-1158, Jul. 2003.

[6] J. E. Kolodziej, O. J. Tobias, and R. Seara, "Stochastic model for the generalized subband decomposition εNLMS algorithm with Gaussian data," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toulouse, France, vol. 3, May 2006, pp. 764-767.

[7] E. M. Lobato, O. J. Tobias, and R. Seara, "Stochastic modeling of the transform-domain εLMS algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1840-1852, May 2008.

[8] C. G. Samson and U. Reddy, "Fixed point error analysis of the normalized ladder algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, no. 5, pp. 1177-1191, Oct. 1983.

[9] S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6.ed., Academic Press, 2000.

[10] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[11] J. E. Kolodziej, O. J. Tobias, and R. Seara, "Stochastic analysis of the transform domain LMS algorithm for a non-stationary environment," in *Proc. 17th European Signal Processing Conf.*, Glasgow, Scotland, Aug. 2009, pp. 1-4.