STOCHASTIC ANALYSIS OF THE TRANSFORM DOMAIN LMS ALGORITHM FOR A NON-STATIONARY ENVIRONMENT

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-mail: {javier, orlando, seara}@linse.ufsc.br

ABSTRACT

This paper presents an improved statistical model for the transform-domain LMS algorithm operating in non-stationary environment from a time-varying plant. The stationary case is also considered as a particular case of the non-stationary one. The derived model takes into account a fixed-length sliding window for estimating the transformed input signal power. Small step-size conditions and Gaussian input data are assumed. Using model expressions, algorithm parameters as optimum step size, excess error, and misadjustment are obtained. Through numerical simulations, the accuracy of the proposed model is assessed.

1. INTRODUCTION

The LMS algorithm in the transform domain (TDLMS) was proposed by Narayan et al. [1] aiming to improve the convergence characteristics of the standard LMS algorithm. The TDLMS algorithm is similar to the LMS one but having its input signal pre-processed by an N-subband orthogonal transform, followed by a normalization process equalizing the energy content in each subband. In practice, to implement the normalization process, estimates of the signal power at each subband are required. To this end, a means used is to measure the signal variance within a fixed-length sliding window (FLSW).

Regarding the modeling procedure, due to the normalization process, an important difficulty is, specifically, the computing of expected values such as

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

where $\hat{\mathbf{D}}(n)$ and $\hat{\mathbf{R}}(n)$ are matrices depending on the observed data. To solve (1), several simplifying approximations are generally considered [2]-[6], in particular the Averaging Principle (AP) used in [5] for approximating (1) to

$$E[\hat{\mathbf{D}}^{-1}(n)\hat{\mathbf{R}}(n)] \approx E[\hat{\mathbf{D}}^{-1}(n)]E[\hat{\mathbf{R}}(n)]. \tag{2}$$

The AP results in simpler model expressions, having satisfactory accuracy if large observation windows are used. However, such a case may not be found in practical situations, since most adaptive algorithms have restrictions with respect to computational burden. Thus, looking for an accurate model irrespective of observation window length, the use of AP approach is no longer recommended.

In particular, the model proposed in this work considers the following aspects:

i) The expected value (1) is calculated without invoking the AP approach.
ii) For the sake of generality, the proposed model is derived considering a non-stationary environment (time-varying plant) for two reasons: it has large practical importance, as well as in the open literature only a few and brief analyses have been found. Moreover, the stationary case can be straightforwardly obtained from the presented analysis.

From the proposed analysis, expressions for the excess error, optimum step-size parameter, and misadjustment are also derived.

This paper is organized as follows. Section 2 presents the model used for characterizing the time-varying plant. In Section 3, the algorithm model expressions are derived. Results of numerical simulations are presented in Section 4, demonstrating the validity of the proposed model. Finally in Section 5, some conclusions of this work are presented.

2. NON-STATIONARY ENVIRONMENT

A fundamental feature of adaptive filters is their ability to track signal variations. To make the filter analysis tractable, it is usual to assume that the data statistics (in our case, time-varying plant) vary according to a given rule. In this way, the desired signal is modeled as follows:

$$d(n) = \mathbf{x}_{\mathrm{T}}^{\mathrm{T}}(n)\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n) + z(n) \tag{3}$$

with the following rule for describing the plant evolution:

$$\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n+1) = a\,\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n) + \mathbf{g}(n) \tag{4}$$

where $\mathbf{x}_{\mathrm{T}}(n) = \mathbf{T}\mathbf{x}(n) = [x_{\mathrm{T},0}(n)\ \ x_{\mathrm{T},1}(n)\ \cdots\ x_{\mathrm{T},N-1}(n)]^{\mathrm{T}}$ is the input signal vector in the transform domain, with $\mathbf{T}$

Orlando J. Tobias is also with the Electrical Engineering and Telecom. Dept., Regional University of Blumenau (FURB), Blumenau, SC, Brazil. This work was supported in part by the National Council for Scientific and Technological Development (CNPq) and the Committee for Postgraduate Courses in Higher Education (CAPES).

being an orthogonal transform, $\mathbf{x}(n) = [x(n)\ x(n-1)\ \cdots\ x(n-N+1)]^{\mathrm{T}}$ is the input signal vector assuming $x(n)$ as a zero-mean Gaussian stationary process with variance σ_x^2, $\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n) = \mathbf{T}\mathbf{w}^{\mathrm{o}}(n) = [w_{\mathrm{T},1}^{\mathrm{o}}(n)\ w_{\mathrm{T},2}^{\mathrm{o}}(n) \cdots w_{\mathrm{T},N-1}^{\mathrm{o}}(n)]^{\mathrm{T}}$ denotes the time-varying plant vector in the transform domain, $\mathbf{w}^{\mathrm{o}}(n) = [w_1^{\mathrm{o}}(n)\ w_2^{\mathrm{o}}(n)\ \cdots\ w_{N-1}^{\mathrm{o}}(n)]^{\mathrm{T}}$ is the time-varying plant vector, and $z(n)$ is an i.i.d. measurement noise with variance σ_z^2. Variable $a \in [0,1]$, $\mathbf{g}(n)$ is the plant perturbation vector, which is zero-mean with autocorrelation matrix $\mathbf{G}$, and $\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(0)$ is any arbitrary value.

Usually, available analyses in the literature regarding adaptive algorithms in non-stationary environment use $a=1$, and/or $\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(0) = \mathbf{0}$. Thus, for both cases one has a restricted model, since the mean value of the plant does not show any time evolution. In addition, for $a=1$, the model given in (4) is unrealistic because of the infinite variance [7]. In this work, we consider derivations taking into account a generic value of a, resulting in a less restricted model.

3. ANALYSIS

3.1 Problem Statement

In this section, we derive analytical expressions for the first and second moment of the adaptive weight vector. Then, let us start by considering the weight update equation in the transform domain, which is given by [1]

$$\mathbf{w}_{\mathrm{T}}(n+1) = \mathbf{w}_{\mathrm{T}}(n) + 2\mu\hat{\mathbf{D}}^{-1}(n)e(n)\mathbf{x}_{\mathrm{T}}(n) \tag{5}$$

where $\mathbf{w}_{\mathrm{T}}(n) = [w_{\mathrm{T},0}(n)\ w_{\mathrm{T},1}(n)\ \cdots\ w_{\mathrm{T},N-1}(n)]^{\mathrm{T}}$ denotes the adaptive filter weight vector and $\hat{\mathbf{D}}^{-1}(n) = \mathrm{diag}[\hat{\sigma}_0^{-2}(n)\ \hat{\sigma}_1^{-2}(n)\ \cdots\ \hat{\sigma}_{N-1}^{-2}(n)]$ is the step-size normalizing matrix, with $\hat{\sigma}_i^2(n)$ being the instantaneous estimate of the subband variance. In practice, a FLSW is used to estimate the variance of each subband. Thus,

$$\hat{\sigma}_i^2(n) = \frac{1}{M}\sum_{k=0}^{M-1} x_{\mathrm{T},i}^2(n-k) \tag{6}$$

where M is the window length. The error signal is obtained as

$$e(n) = d(n) - y(n) \tag{7}$$

with $y(n) = \mathbf{x}_{\mathrm{T}}^{\mathrm{T}}(n)\mathbf{w}_{\mathrm{T}}(n)$. Now, substituting (3) into (7) and the resulting expression into (5), using the weight-error vector in the transform domain, defined as $\mathbf{v}_{\mathrm{T}}(n+1) = \mathbf{w}_{\mathrm{T}}(n+1) - \mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n+1)$, we get

$$\begin{aligned}\mathbf{v}_{\mathrm{T}}(n+1) &= [\mathbf{I} - 2\mu\hat{\mathbf{D}}^{-1}(n)\mathbf{x}_{\mathrm{T}}(n)\mathbf{x}_{\mathrm{T}}^{\mathrm{T}}(n)]\mathbf{v}_{\mathrm{T}}(n) \\ &\quad + 2\mu\hat{\mathbf{D}}^{-1}(n)\mathbf{x}_{\mathrm{T}}(n)z(n) - \mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n+1) + \mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n).\end{aligned} \tag{8}$$

Then, substituting (4) into (8), the update expression in terms of $\mathbf{v}_{\mathrm{T}}(n)$ is given by

$$\begin{aligned}\mathbf{v}_{\mathrm{T}}(n+1) &= [\mathbf{I} - 2\mu\hat{\mathbf{D}}^{-1}(n)\mathbf{x}_{\mathrm{T}}(n)\mathbf{x}_{\mathrm{T}}^{\mathrm{T}}(n)]\mathbf{v}_{\mathrm{T}}(n) \\ &\quad + 2\mu\hat{\mathbf{D}}^{-1}(n)\mathbf{x}_{\mathrm{T}}(n)z(n) + (1-a)\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n) - \mathbf{g}(n).\end{aligned} \tag{9}$$

The next step is to determine the first and second moments of (9).

3.2 Analysis Assumptions

To carry out the stochastic analysis, the following simplifying assumptions are considered:

i) $\mathbf{g}(n)$ and $\mathbf{g}(m)$, for $m \neq n$, are uncorrelated.

ii) $\mathbf{g}(n)$, $\mathbf{v}_{\mathrm{T}}(n)$, and $\mathbf{x}_{\mathrm{T}}(n)$ are statistically independent.

iii) $z(n)$ is uncorrelated with any other signal in the system.

3.3 First Moment of $\mathbf{v}_{\mathrm{T}}(n)$

By taking the expected value of both sides of (9) and using assumptions (i) and (ii), we obtain

$$\begin{aligned}E[\mathbf{v}_{\mathrm{T}}(n+1)] &\cong \{\mathbf{I} - 2\mu\underbrace{E[\mathbf{D}^{-1}(n)\mathbf{x}_{\mathrm{T}}(n)\mathbf{x}_{\mathrm{T}}^{\mathrm{T}}(n)]}_{\mathbf{P}}\}E[\mathbf{v}_{\mathrm{T}}(n)] \\ &\quad + (1-a)a^n\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(0)\end{aligned} \tag{10}$$

where the last r.h.s. term represents the mean of the AR(1) process given by (4). The elements of matrix $\mathbf{P}$ in (10) are determined in the Appendix.

3.4 Second Moment of $\mathbf{v}_{\mathrm{T}}(n)$ and Learning Curve

The second moment for the weight-error vector in the transform domain is obtained from $\mathbf{K}(n) = E[\mathbf{v}_{\mathrm{T}}(n)\mathbf{v}_{\mathrm{T}}^{\mathrm{T}}(n)]$. Then, transposing both sides of (9), carrying out the product $\mathbf{v}_{\mathrm{T}}(n)\mathbf{v}_{\mathrm{T}}^{\mathrm{T}}(n)$, taking the expected value of both sides of the resulting expression, and using the above assumptions (i)-(iii), we get

$$\begin{aligned}\mathbf{K}(n+1) &= \mathbf{K}(n) - 2\mu\mathbf{K}(n)\mathbf{P}^{\mathrm{T}} - 2\mu\mathbf{P}\,\mathbf{K}(n) \\ &+ (1-a)E[\mathbf{v}_{\mathrm{T}}(n)]E[\mathbf{w}_{\mathrm{T}}^{\mathrm{o}\,\mathrm{T}}(n)] + (1-a)E[\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n)]E[\mathbf{v}_{\mathrm{T}}^{\mathrm{T}}(n)] \\ &+ 4\mu^2\mathbf{\Phi}\{2\mathbf{R}_{\mathrm{T}}\mathbf{K}(n)\mathbf{R}_{\mathrm{T}} + \mathbf{R}_{\mathrm{T}}\mathrm{tr}[\mathbf{K}(n)\mathbf{R}_{\mathrm{T}}]\}\mathbf{\Phi} \\ &+ 4\mu^2\mathbf{\Phi}\mathbf{R}_{\mathrm{T}}\mathbf{\Phi}\sigma_z^2 + (1-a)^2\mathbf{K}^{\mathrm{o}}(n) \\ &- 2\mu(1-a)\mathbf{P}\,E[\mathbf{v}_{\mathrm{T}}(n)]E[\mathbf{w}_{\mathrm{T}}^{\mathrm{o}\,\mathrm{T}}(n)] \\ &- 2\mu(1-a)E[\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n)]E[\mathbf{v}_{\mathrm{T}}^{\mathrm{T}}(n)]\mathbf{P}^{\mathrm{T}} + \mathbf{G}\end{aligned} \tag{11}$$

with

$$\mathbf{R}_{\mathrm{T}} = E[\mathbf{x}_{\mathrm{T}}(n)\mathbf{x}_{\mathrm{T}}^{\mathrm{T}}(n)], \tag{12}$$

$$\mathbf{K}^{\mathrm{o}}(n) = E[\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n)\mathbf{w}_{\mathrm{T}}^{\mathrm{o}\,\mathrm{T}}(n)] = (a^2)^n\mathbf{K}^{\mathrm{o}}(0) + \frac{1-(a^2)^n}{1-a^2}\mathbf{G}, \tag{13}$$

and

$$\mathbf{\Phi} = \mathrm{diag}[\sigma_0^2\ \sigma_1^2\ \cdots\ \sigma_{N-1}^2]. \tag{14}$$

Finally, (11) can be used to determine the learning curve, which is given by [5]

$$E[e^2(n)] = \sigma_z^2 + \mathrm{tr}[\mathbf{R}_{\mathrm{T}}\mathbf{K}(n)]. \tag{15}$$

3.5 Excess Error

In this section, an expression for the excess error is obtained. Thus, transposing both sides of (9), carrying out the product $\hat{\mathbf{D}}(n)\mathbf{v}_{\mathrm{T}}(n+1)\mathbf{v}_{\mathrm{T}}^{\mathrm{T}}(n+1)$, taking the expected value of the resulting expression, assuming $n \to \infty$ [whereby $\mathbf{K}(n+1) \simeq \mathbf{K}(n)$, $\mathbf{K}^{\mathrm{o}}(n) \simeq (1-a^2)^{-1}\mathbf{\Phi}$, and $E[\mathbf{v}_{\mathrm{T}}(n)] \simeq 0$], and applying the trace operation, we obtain

$$\mathrm{tr}[\mathbf{K}(n)\mathbf{R}_{\mathrm{T}}] = \mu\sigma_z^2\mathrm{tr}[\mathbf{P}] + \frac{1}{2\mu(1+a)}\mathrm{tr}[\mathbf{\Phi G}] + \mu\,\underbrace{\mathrm{tr}[\mathbf{x}_{\mathrm{T}}(n)\mathbf{x}_{\mathrm{T}}^{\mathrm{T}}(n)\mathbf{v}_{\mathrm{T}}(n)\mathbf{v}_{\mathrm{T}}^{\mathrm{T}}(n)\mathbf{x}_{\mathrm{T}}(n)\mathbf{x}_{\mathrm{T}}^{\mathrm{T}}(n)\hat{\mathbf{D}}^{-1}(n)]}_{\gamma}. \tag{16}$$

Then, rearranging γ as

$$\begin{aligned}\gamma &= \mathbf{x}_{\mathrm{T}}^{\mathrm{T}}(n)\hat{\mathbf{D}}^{-1}(n)\mathbf{x}_{\mathrm{T}}(n)\mathrm{tr}[\mathbf{x}_{\mathrm{T}}^{\mathrm{T}}(n)\mathbf{v}_{\mathrm{T}}(n)\mathbf{v}_{\mathrm{T}}^{\mathrm{T}}(n)\mathbf{x}_{\mathrm{T}}(n)]\\ &= \mathrm{tr}[\mathbf{P}]\mathrm{tr}[\mathbf{R}_{\mathrm{T}}\mathbf{K}(n)]\end{aligned} \tag{17}$$

and considering that the excess error is defined as

$$\xi_{\mathrm{exc}}(n) = \mathrm{tr}[\mathbf{R}_{\mathrm{T}}\mathbf{K}(n)] \tag{18}$$

then, for $n \to \infty$, we obtain

$$\xi_{\mathrm{exc}} = \frac{1}{1-\mu\mathrm{tr}[\mathbf{P}]}\left\{\mu\sigma_z^2\mathrm{tr}[\mathbf{P}] + \frac{1}{2\mu(1+a)}\mathrm{tr}[\mathbf{\Phi G}]\right\}. \tag{19}$$

3.6 Step Size for Minimum Excess Error

The optimum step size that minimizes the excess mean-square error (MSE) is obtained by deriving (19) w.r.t. μ and making such a derivative equal to zero. Thus,

$$\frac{\partial\xi_{\mathrm{exc}}}{\partial\mu} = \frac{\sigma_z^2 c_1\mu^2 + 2c_1c_2\mu - c_2}{(1-\mu c_1)^2} = 0 \tag{20}$$

where $c_1 = \mathrm{tr}[\mathbf{P}]$ and $c_2 = [2(1+a)]^{-1}\mathrm{tr}[\mathbf{\Phi G}]$. Solving (20) for μ, we get

$$\mu_{\mathrm{opt}} = \frac{-c_1c_2 + \sqrt{c_1^2c_2^2 + \sigma_z^2 c_1 c_2}}{\sigma_z^2 c_1}. \tag{21}$$

3.7 Misadjustment

The misadjustment $\mathcal{M}$ is obtained from (19), which is given by

$$\mathcal{M} = \frac{\xi_{\mathrm{exc}}}{\xi_{\mathrm{min}}} = \frac{1}{1-\mu\mathrm{tr}[\mathbf{P}]}\left\{\mu\mathrm{tr}[\mathbf{P}] + \frac{\sigma_z^{-2}}{2\mu(1+a)}\mathrm{tr}[\mathbf{\Phi G}]\right\} \tag{22}$$

where $\xi_{\mathrm{min}} = \sigma_z^2$. Note that (22) [also(19)] is composed of the sum of two terms. The first is equal to the misadjustment referred to the stationary case (when $a = 1$ and $\mathbf{G}$ is a null matrix), the second is the misadjustment coming from the non-stationary characteristic. Thus, the misadjustment for the non-stationary case is larger than that observed in a stationary one.

3.8 Degree of Non-Stationarity

The degree of non-stationarity, denoted by $\alpha(n)$, is defined as [7]

$$\alpha(n) \triangleq \left\{\frac{\mathrm{E}[|y_{\mathrm{o,inc}}(n)|^2]}{\mathrm{E}[|z(n)|^2]}\right\}^{1/2} \tag{23}$$

where

$$y_{\mathrm{o,inc}}(n) = [\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n+1) - \mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n)]^{\mathrm{T}}\mathbf{x}_{\mathrm{T}}(n) \tag{24}$$

is the output due to the difference $\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n+1) - \mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n)$ (incremental filter). The numerator of (23) denotes the power introduced by the variation of the optimum filter, and the denominator is the minimum MSE (MMSE). By using (4) in (24), one obtains

$$\begin{aligned}y_{\mathrm{o,inc}}(n) &= [(a-1)\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(n) + \mathbf{g}(n)]^{\mathrm{T}}\mathbf{x}_{\mathrm{T}}(n)\\ &= \{(a-1)[a^n\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(0) + \sum_{k=0}^{n-1}a^k\mathbf{g}(n-1-k)] + \mathbf{g}(n)\}^{\mathrm{T}}\mathbf{x}_{\mathrm{T}}(n).\end{aligned} \tag{25}$$

By considering the independence between $\mathbf{g}(n)$ and $\mathbf{x}_{\mathrm{T}}(n)$, and using assumption (i), we can write

$$E[|y_{\mathrm{o,inc}}(n)|^2] = (1-a)^2a^{2n}\mathrm{tr}[\mathbf{K}^{\mathrm{o}}(0)\mathbf{R}_{\mathrm{T}}] + \frac{2-a^{2n}+a^{2n+1}}{1+a}\mathrm{tr}[\mathbf{G}\mathbf{R}_{\mathrm{T}}]. \tag{26}$$

Now, substituting (26) into (23) and recalling that σ_z^2 is the variance of $z(n)$, we have

$$\alpha(n) = \left\{\frac{(1-a)^2a^{2n}}{\sigma_z^2}\mathrm{tr}[\mathbf{K}^{\mathrm{o}}(0)\mathbf{R}_{\mathrm{T}}] + \frac{2-a^{2n}+a^{2n+1}}{\sigma_z^2(1+a)}\mathrm{tr}[\mathbf{G}\mathbf{R}_{\mathrm{T}}]\right\}^{1/2}. \tag{27}$$

From (27), we can conclude that $\alpha(n)$ presents an exponential evolution. If the degree of non-stationarity is larger than unity, the statistical variations of the optimal weight vector are too fast for the filter to be able to track them (and the misadjustment will then be large). On the other hand, if the degree of non-stationarity is much smaller than unity, the adaptive filter would be able to track the variations in the weight vector. In this work, we are interested in evaluating the tracking performance of the adaptive filter in the latter situation, i.e., when the tracking is possible.

4. SIMULATION RESULTS

The proposed model is applied to a system identification problem in which its accuracy is assessed for correlated Gaussian input data, obtained from an AR(2) process, given by

$$x(n) = a_1x(n-1) + a_2x(n-2) + v(n) \tag{28}$$

where $v(n)$ is white noise with variance σ_v^2 such that the variance of $x(n)$ is 1, a_1 and a_2 are the autoregressive coefficients, with $a_1 = 0.18$ and $a_2 = -0.85$. The resulting eigenvalue spread of the input signal for $N = 32$ is $\chi = 120$. The time-varying weights of the plant are obtained according to (4). Its initial values are selected as follows:

$$\mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(0) = \frac{\mathbf{w}_{\mathrm{aux}}}{\sqrt{\mathbf{w}_{\mathrm{aux}}^{\mathrm{T}} \mathbf{w}_{\mathrm{aux}}}} \quad (29)$$

with

$$\mathbf{w}_{\mathrm{aux}} = \mathbf{T}[\mathrm{sinc}(0) \;\; \mathrm{sinc}(1/N) \;\; \cdots \;\; \mathrm{sinc}(N-1/N)]^{\mathrm{T}}. \quad (30)$$

The elements of $\mathbf{g}(n)$ are samples from a white noise process with autocorrelation matrix $\mathbf{G} = \sigma_g^2 \mathbf{I}$. In the examples, $a = 0.99$ and two SNR values (20 and 40 dB) are used, resulting in $\alpha(0) = 0.385$ and $\alpha(0) = 0.107$, respectively. In addition, for both SNR values $\alpha(\infty) = 0.100$. The discrete cosine transform (DCT) is the orthogonal transform used. Monte Carlo (MC) simulations are obtained averaging 500 independent runs.

For the non-stationary cases, the weights are initialized as

$$\mathbf{w}_{\mathrm{T}}(0) = \mathbf{w}_{\mathrm{T}}^{\mathrm{o}}(0) \quad (31)$$

aiming to evaluate the adaptive algorithm tracking performance. In the stationary case, the weights are initialized with zero for assessing the algorithm behavior during the acquisition mode [7].

For comparison proposes, the algorithm behavior predicted by a model based on AP is presented. That model is obtained by generalizing the procedure presented in [5].

Example 1: In this example, the following parameter values are used: $N = 32$ in (29), $M = 10$, and $\mu = 0.75\mu_{\mathrm{opt}} = 0.00794$. In Figure 1, the first moment of the weight-error vector using the FLSW for power estimate with 40 dB signal-to-noise ratio (SNR) is shown. For the sake of clarity, the evolution of the expected value of only four weights is plotted. Figure 2 shows the learning curves (MSE) for two SNR values (20 and 40 dB).

Example 2: For this example, most of the parameters of Example 1 are kept the same, but M is increased $(M = 32)$ to achieve better variance estimation. For this case, $\mu = 0.75\mu_{\mathrm{op}} = 0.00689$. The obtained results for the MSE evolution are illustrated in Figure 3. Note that, in this case, the AP-based model and the proposed model present a similar behavior.

Example 3: In this example, the stationary case is considered as a particular case of the non-stationary one ($a = 1$ and degree of non-stationarity $\alpha = 0$). In this case, the parameters used are: $N = 32$, $M = 16$, and $\mu = 0.003$. The obtained results are shown in Figure 4. Note that there is a reduction in the MSE, since now the term related to the non-stationarity in (22) is equal to zero.

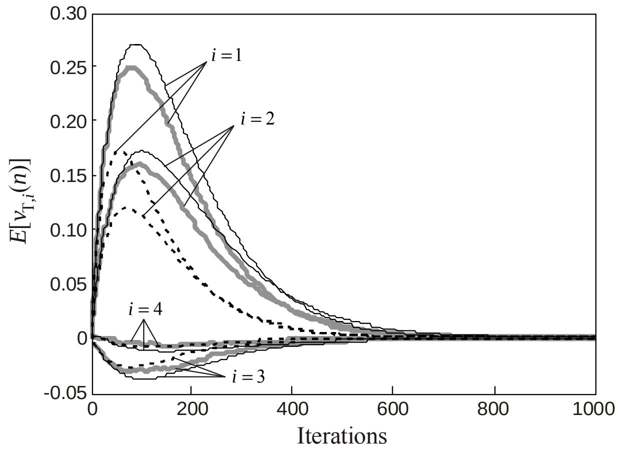

Figure 1 – Example 1. Mean weight-error behavior curves for $\mathrm{SNR} = 40\,\mathrm{dB}$. (Gray lines) MC simulations. (Dashed lines) AP-based model. (Dark solid lines) proposed model.

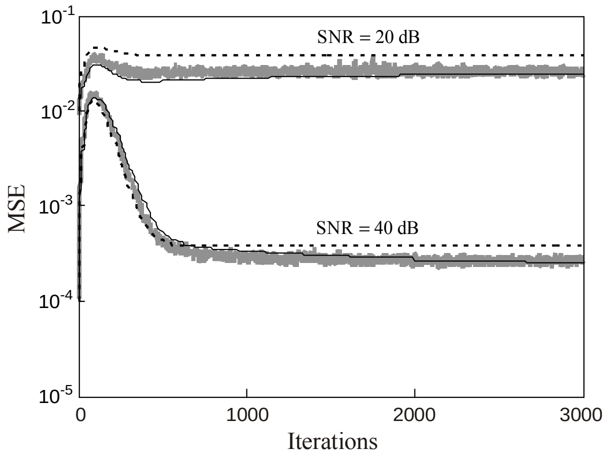

Figure 2 – Example 1. Comparison results for $M = 10$. MSE curves for $\mathrm{SNR} = 20$ and $40\,\mathrm{dB}$. (Ragged gray lines) MC simulations. (Dashed lines) AP-based model. (Dark solid lines) proposed model.

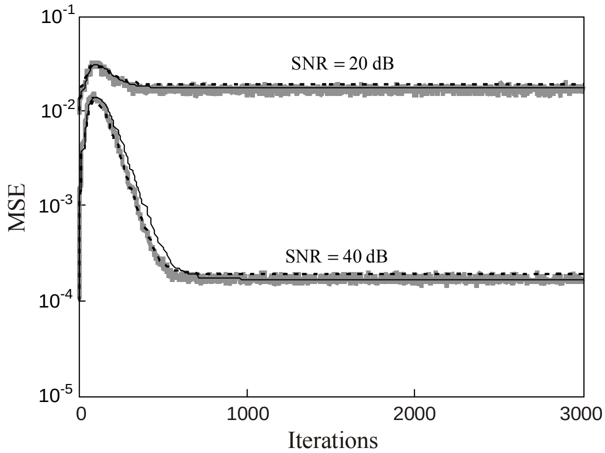

Figure 3 – Example 2. Comparison results for $M = 32$. MSE curves for $\mathrm{SNR} = 20$ and $40\,\mathrm{dB}$. (Ragged gray lines) MC simulations. (Dashed lines) AP-based model. (Dark solid lines) proposed model.

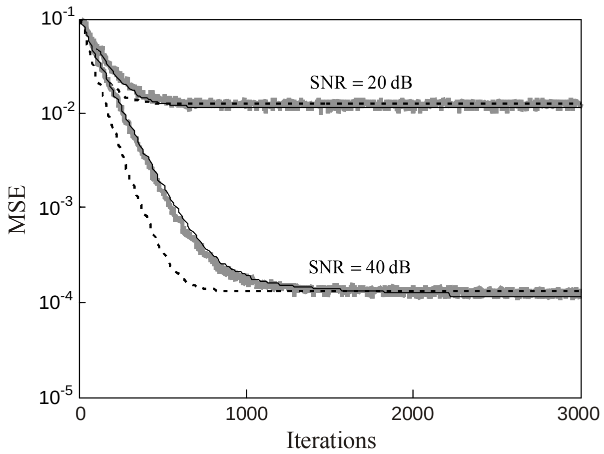

Figure 4 – Example 3. Comparison results for a stationary environment. MSE curves for $\text{SNR} = 20$ and $40\,\text{dB}$. (Ragged gray lines) MC simulations. (Dashed lines) AP-based model. (Dark solid lines) proposed model.

5. CONCLUSIONS

This paper presents a stochastic model for the TDLMS algorithm operating in non-stationary environment. This analysis is independent of the order of the filter as well as of the window length used for the subband power estimate. From the proposed model, the stationary case can also be obtained. Numerical simulations showed very good agreement between the results obtained by the MC method and the predictions using the proposed analytical model for both mean weight behavior and MSE curves.

6. APPENDIX
DETERMINATION OF P

To determine matrix $\mathbf{P}$, we define an extended vector given by $\mathbf{x}_{\text{e},i,j}(n) = [x_{\text{T},i}(n)\ x_{\text{T},i}(n-1) \ldots x_{\text{T},i}(n-M+1) \times x_{\text{T},j}(n)]^{\text{T}}$, such that

$$\sum_{k=0}^{M-1} x_{\text{T},i}^2(n-k) = \mathbf{x}_{\text{e},i,j}^{\text{T}}(n)\mathbf{I}_\text{s}\mathbf{x}_{\text{e},i,j}(n) \tag{32}$$

where $\mathbf{I}_\text{s} = \text{diag}[\underbrace{1 \cdots 1}_{M}\ 0]$. By considering jointly Gaussian random processes, $\mathbf{P}$ is calculated as

$$\{\mathbf{P}\}_{i,j} = \frac{M}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{M+1 \text{ fold}} \frac{1}{[\det(\mathbf{R}_{\text{e},i,j})]^{1/2}} \cdot \frac{x_{\text{T},i}x_{\text{T},j}}{\mathbf{x}_{\text{e},i,j}^{\text{T}}\mathbf{I}_\text{s}\mathbf{x}_{\text{e},i,j}} \times e^{-\frac{1}{2}\mathbf{x}_{\text{e},i,j}^{\text{T}}\mathbf{R}_{\text{e},i,j}^{-1}\mathbf{x}_{\text{e},i,j}} d\mathbf{x}_{\text{e},i,j} \tag{33}$$

where $\mathbf{R}_{\text{e},i,j} = E[\mathbf{x}_{\text{e},i,j}\,\mathbf{x}_{\text{e},i,j}^{\text{T}}]$ is the autocorrelation matrix of the extended vector $\mathbf{x}_{\text{e},i,j}$. In (33), we drop the time index *n* for simplicity of notation. To determine (33), we use an approach similar to that presented in [8], which results in

$$\{\mathbf{P}\}_{i,j} = M[\mathbf{Q}_{\text{s},i,j}\mathbf{H}_i\mathbf{Q}_{\text{s},i,j}^{-1}\mathbf{R}_{\text{e},i,j}]_{1,M+1} \tag{34}$$

where $\mathbf{Q}_{\text{s},i,j}$ is the matrix of eigenvectors of $\mathbf{R}_{\text{e},i,j}\mathbf{I}_\text{s}$ and $\mathbf{H}_i$ is a diagonal matrix, given by

$$\{\mathbf{H}_i\}_{l,l} \approx \frac{1}{2\sqrt{a_i}}\left[\sum_{q=1}^{N/2} A_{l,q}\ln(\lambda'_{i,q}) + B_l\ln(\lambda_l)\right] \tag{35}$$

with

$$a_i = \prod_{k=1}^{M}\lambda_{i,k} \tag{36}$$

$$A_{l,q} = \frac{1}{\left(\dfrac{1}{\lambda_{i,l}} - \dfrac{1}{\lambda'_{i,q}}\right)\displaystyle\prod_{\substack{m=1\\m\neq q}}^{N/2}\left(\dfrac{1}{\lambda'_{i,m}} - \dfrac{1}{\lambda'_{i,q}}\right)} \tag{37}$$

and

$$B_l = \frac{1}{\displaystyle\prod_{q=1}^{N/2}\left(\dfrac{1}{\lambda'_{i,q}} - \dfrac{1}{\lambda_{i,l}}\right)} \tag{38}$$

where $\lambda_{i,k}$ are the eigenvalues of the subband autocorrelation matrix $\mathbf{R}_{ii} = E[\mathbf{x}_{\text{T},i}(n)\mathbf{x}_{\text{T},i}^{\text{T}}(n)]$, with $\mathbf{x}_{\text{T},i}(n) = [x_{\text{T},i}(n)\ x_{\text{T},i}(n-1) \cdots x_{\text{T},i}(n-M+1)]^{\text{T}}$. Variable $\lambda'_{i,k}$ represents the geometric mean of adjacent pairs of the eigenvalues $\lambda_{i,k}$ [5].

REFERENCES

[1] S. S. Narayan, A. M. Peterson, and M. J. Narasimha, "Transform domain LMS algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 3, pp. 609-615, Jun. 1983.

[2] S. Hosur and A. H. Tewfik, "Wavelet transform domain adaptive FIR filtering," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 617-630, Mar. 1997.

[3] S. J. Chern, J. C. Horng, and K. M. Wong, "The performance of the hybrid LMS adaptive algorithm," *Signal Process.*, vol. 44, no. 1, pp. 67-88, Jun. 1995.

[4] D. I. Kim and P. De Wilde, "Performance analysis of the DCT-LMS adaptive filtering algorithm," *Signal Process.*, vol. 80, no. 8, pp. 1629-1654, Aug. 2000.

[5] 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.

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

[7] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*. New York: McGraw-Hill, 2000.

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