

# A STATISTICAL ANALYSIS OF THE $\varepsilon$ -NLMS AND NLMS ALGORITHMS FOR CORRELATED GAUSSIAN SIGNALS

José C.M. Bermudez and Márcio H. Costa

**Abstract** - The Normalized LMS adaptive algorithm is widely used in real life adaptive filtering and control applications mainly due to its simplicity and robustness to input signal power variations. This last characteristic is a consequence of its intrinsic normalization procedure. However, in practical applications it is usual the addition of a small positive constant to the normalization factor in order to avoid divisions by zero or overflow. In this case, the adaptive algorithm is named  $\varepsilon$ -NLMS. This work presents a statistical analysis of the  $\varepsilon$ -NLMS for Gaussian input signals. Deterministic recursive equations are obtained for the mean weight behavior and mean square error for a large number of adaptive coefficients. In addition, a closed expression is provided for the steady state misadjustment. Monte Carlo simulations show an excellent agreement between theoretical predictions and the algorithm's behavior in steady state. During transient, the new model is conservative and more accurate than the existing models. The developed equations can be used with  $\varepsilon = 0$  to predict the behavior of the NLMS algorithm. Simulations illustrate the better quality of the new NLMS model when compared to others already available in the literature.

**Keywords:** Adaptive filters, normalized LMS, NLMS, echo canceling, analytical modeling.

**Resumo** - O algoritmo LMS Normalizado é amplamente utilizado em aplicações reais de filtragem e controle adaptativos em decorrência de sua simplicidade e robustez a variações da potência do sinal de entrada. Esta última característica deve-se ao procedimento de normalização utilizado. Entretanto, em aplicações práticas é usual a adição de uma pequena constante positiva ao fator de normalização, de forma a evitar divisões por zero ou *overflow*. Nesse caso o algoritmo passa a se chamar  $\varepsilon$ -NLMS. Este trabalho apresenta uma análise estatística do algoritmo  $\varepsilon$ -NLMS para sinais de entrada Gaussianos. Equações determinísticas recursivas são obtidas para os comportamentos do valor médio dos coeficientes e do erro médio quadrático para um elevado número de coeficientes no filtro adaptativo. É também determinada uma expressão fechada para o desajuste em regime permanente. Simulações Monte Carlo mostram uma excelente concordância entre as predições teóricas e o comportamento real do algoritmo em regime permanente. No regime transitório o modelo é conservativo e mais preciso do que os modelos existentes na literatura. As equações desenvolvidas podem ser utilizadas com  $\varepsilon = 0$  para prever o comportamento do algoritmo NLMS convencional. Simulações demonstram uma maior precisão do novo

modelo, quando comparado com os demais modelos encontrados na literatura para o algoritmo NLMS.

**Palavras-chave:** Filtros adaptativos, LMS normalizado, NLMS, cancelamento de eco, modelagem analítica.

## 1. INTRODUCTION

Adaptive filtering techniques are widely employed in real life applications such as modeling, equalization, active noise control and echo cancellation [1-3]. Stochastic gradient based algorithms have proven to be both robust and easy to implement for real-time control and estimation applications.

The Least Mean Square (LMS) is the most popular adaptive algorithm due to its robustness and low computational complexity [4]. This simplicity, however, comes at the price of a large dependence of the optimal value of its design parameter (the adaptation step-size  $\mu$ ) on the statistics of the input signal. For a fixed-step-size implementation, changes in the input power lead to performance losses and frequently make the algorithm ineffective.

Among the various LMS family members, the normalized algorithms are attractive because of their capability of tuning the step-size to the input power. This property renders the algorithms less sensitive to input power variations at the cost of an increased computational complexity as well as higher misadjustment.

The Normalized-LMS (NLMS) algorithm, also known as the projection algorithm [5-7], is the most used normalized algorithm due to its simple form and good performance. The NLMS weight update follows the direction of the input vector  $\mathbf{x}(n)$ , and the step-size is normalized by the squared norm of the input vector, which works as an instantaneous estimate of the input power. The squared norm can be computed recursively, increasing the algorithm's cost by only two multiplications and one division per iteration, when compared to the LMS algorithm. In applications where a large number of coefficients is necessary, the advantages of using the NLMS algorithm easily overcome the extra implementation cost [7]. Typical applications are active noise control and echo cancellation.

Practical implementations of adaptive filters in fixed or floating-point digital signal processors require the adjustment of the analog to digital converters' (ADC's) dynamic range to cover the entire range of the input signal [8]. Hence, small amplitude signals are quantized to zero when their magnitudes are smaller than the least significant level of the ADC. Moreover, some applications can be characterized by periods of absence of signal, such as in speech communications. When such situations occur, the normalizing factor of the NLMS algorithm can become too small (even zero), which is undesirable. To avoid divisions

by zero or overflows, a small positive constant ( $\varepsilon$ ) is frequently added to the normalizing factor of the NLMS algorithm, yielding a step-size of the form  $\mu(n) = \beta / [\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)]$ . This modification characterizes the  $\varepsilon$ -NLMS algorithm, a generalization of the conventional NLMS algorithm.

The  $\varepsilon$ -NLMS algorithm was analyzed by Bershada [9] for white Gaussian inputs. He concluded that, under this input signal condition, the algorithm has neither a behavior which is independent of input data power nor a performance which is significantly better than the LMS algorithm. Later on, Douglas and Meng [10] studied normalized data nonlinearities for LMS adaptation in order to improve the algorithm's performance. The proposed generalized algorithm had an equation similar to the LMS, but with the updating term multiplied by a general scalar memoryless nonlinear function of  $\mathbf{x}^T(n)\mathbf{x}(n)$ . As a result, they derived an optimum form of nonlinearity for independent input data. This nonlinearity is optimum for independent input data samples with any even probability density function. The resulting adaptive algorithm is equivalent to the  $\varepsilon$ -NLMS with  $\beta = 1$ . In this algorithm, the convergence performance is controlled by the parameter  $\varepsilon$ . The authors of [10] have shown that up to 3.6 dB improvement can be obtained over the LMS steady state mean square error (MSE) performance when using this algorithm. However, [10] did not provide any analysis for the correlated Gaussian input case. In [11], the fast affine projection algorithm was studied for hands-free telephone applications as a generalization of the  $\varepsilon$ -NLMS algorithm. The factor  $\varepsilon$  was again used to control the performance of the adaptive filter. Recently, a generalized normalized gradient descent adaptive algorithm was proposed [12]. Improved stability and faster convergence characteristics were obtained in [12] by a continuous change of the regularization parameter  $\varepsilon$ . These results revived the interest in the  $\varepsilon$ -NLMS, now with  $\varepsilon$  being more than just a regularization constant.

This work presents a statistical analysis of the  $\varepsilon$ -NLMS algorithm for correlated Gaussian input signals. To the authors' best knowledge, there is no analytical model available in the literature for this case. The new analysis does not require any numerical procedure to determine the model parameters. For the special case of  $\varepsilon = 0$ , the derived expressions become a new model for the popular NLMS algorithm. This new NLMS model provides a better prediction of the algorithm's behavior than other models available in the literature [6,7].

Deterministic recursive equations are derived for the mean weight and MSE behaviors using the independence assumption [13] and the averaging principle [14]. Assuming algorithm's convergence, a closed form expression is derived for the steady state MSE misadjustment as a limit of the recursive model. Simulations are provided to verify the validity of the analytical results.

## 2. THE $\varepsilon$ -NLMS UPDATE EQUATION

Consider the system identification problem depicted in Fig. 1, where

$$e(n) = \mathbf{w}^{oT} \mathbf{x}(n) + z(n) - \mathbf{w}^T(n) \mathbf{x}(n) \quad (1)$$

is the error signal,  $\mathbf{w}(n) = [w_0^o w_1^o \dots w_{N-1}^o]^T$  is the unknown impulse response,  $z(n)$  is a stationary, white, zero-mean Gaussian measurement noise with variance  $r_z$  and uncorrelated with any other signal.  $\mathbf{w}(n) = [w_0(n) w_1(n) \dots w_{N-1}(n)]^T$  is the adaptive weight vector. Vectors  $\mathbf{w}^o$  and  $\mathbf{w}(n)$  are considered to have the same dimension. Different dimensions can be accommodated by zero-padding the shorter response. The input signal  $x(n)$  is stationary, zero-mean and Gaussian with variance  $r_0$ , and  $\mathbf{x}(n) = [x(n) x(n-1) \dots x(n-N+1)]^T$  is the observed data vector.

The update equation of the  $\varepsilon$ -NLMS is given by [9]:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \beta \frac{e(n) \mathbf{x}(n)}{\varepsilon + \mathbf{x}^T(n) \mathbf{x}(n)} \quad (2)$$

In (2),  $\beta$  is the fixed step-size and  $\varepsilon$  is a small positive constant, usually termed regularization constant. For  $\varepsilon = 0$  the algorithm becomes the popular NLMS algorithm [7]. The choice of  $\beta = 1$  leads to the optimal normalized data vector nonlinearity [10].

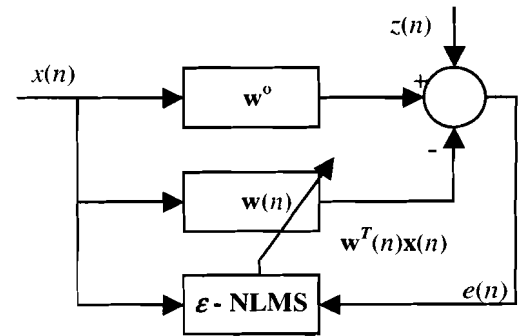


Figure 1. Block diagram of the system analyzed.

## 3. MEAN WEIGHT BEHAVIOR

The following analysis assumes that the effects of the statistical dependence between  $\mathbf{x}(n)$  and  $\mathbf{w}(n)$  on the algorithm's behavior can be neglected. This corresponds to the use of the well-known independence assumption [13]. Defining the weight error vector  $\mathbf{v}(n) = \mathbf{w}(n) - \mathbf{w}^o$ , using (1) in (2) and taking the expected value yields:

$$E\{\mathbf{v}(n+1)\} = \left[ \mathbf{I} - \beta \cdot E \left\{ \frac{\mathbf{x}(n) \mathbf{x}^T(n)}{\varepsilon + \mathbf{x}^T(n) \mathbf{x}(n)} \right\} \right] \cdot E\{\mathbf{v}(n)\} \quad (3)$$

where it has been considered that  $E\{x(n)z(n)\} = 0$ , since  $z(n)$  is zero-mean and independent of  $x(n)$ .

Each element of the expectation within the square brackets in (3) has a numerator given by  $x(n-i) \cdot x(n-j)$  and a denominator given by  $\varepsilon + \sum_{k=0}^{N-1} x^2(n-k)$ . For large values of  $N$  these two random variables can be assumed weakly correlated since  $x(n-i)$  and  $x(n-j)$  affect only two of the  $N$  terms in  $\mathbf{x}^T(n)\mathbf{x}(n)$ . For ergodic  $x(n)$ , this assumption is equivalent to apply the averaging principle [14], as

$\sum_{k=0}^{N-1} x^2(n-k)$  tends to be slowly time-varying when compared to  $x(n-i)x(n-j)$  for large  $N$ . As the samples of  $x(n)$  become more correlated in time, the assumption becomes less valid. Extensive simulation results obtained by the authors have shown that this assumption holds very well for  $N$  as small as 30 and for a wide range of input eigenvalue spreads. Thus, the expected value within brackets can be approximated by

$$E \left\{ \frac{\mathbf{x}(n)\mathbf{x}^T(n)}{\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)} \right\} \cong E \left\{ \frac{1}{\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)} \right\} \cdot E \{ \mathbf{x}(n)\mathbf{x}^T(n) \} \quad (4)$$

The first expected value on the r.h.s. of (4) is difficult to evaluate because of the constant  $\varepsilon$ , which is not always very small compared to  $\mathbf{x}^T(n)\mathbf{x}(n)$  [10, 11]. To proceed with the analysis, the following approximation is used:

$$E \left\{ \frac{\mathbf{x}(n)\mathbf{x}^T(n)}{\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)} \right\} \cong \frac{E \{ \mathbf{x}(n)\mathbf{x}^T(n) \}}{E \{ \varepsilon + \mathbf{x}^T(n)\mathbf{x}(n) \}} = \frac{E \{ \mathbf{x}(n)\mathbf{x}^T(n) \}}{\varepsilon + E \{ \mathbf{x}^T(n)\mathbf{x}(n) \}} \quad (5)$$

The expected value in the numerator of (5) is the input signal correlation matrix. The expectation in the denominator is the trace of the same correlation matrix. Thus, (5) becomes [15]:

$$E \left\{ \frac{\mathbf{x}(n)\mathbf{x}^T(n)}{\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)} \right\} \cong \frac{1}{\varepsilon + Nr_0} \cdot \mathbf{R}_{\mathbf{xx}} \quad (6)$$

with  $\mathbf{R}_{\mathbf{xx}} = E \{ \mathbf{x}(n)\mathbf{x}^T(n) \}$ . Using (6) in (3) leads to the following analytical model for the mean weight error vector's behavior:

$$E \{ \mathbf{v}(n+1) \} = \left[ \mathbf{I} - \frac{\beta}{\varepsilon + Nr_0} \cdot \mathbf{R}_{\mathbf{xx}} \right] \cdot E \{ \mathbf{v}(n) \} \quad (7)$$

For  $\varepsilon = 0$ , (7) becomes the NLMS model derived in [7, Eq. (9)].

#### 4. MSE BEHAVIOR

Squaring (1) and taking the expected value leads to the well-known expression for the MSE:

$$\xi(n) = E \{ e^2(n) \} = r_\varepsilon + tr \{ \mathbf{R}_{\mathbf{xx}} \mathbf{K}(n) \} \quad (8)$$

where  $\mathbf{K}(n) = E \{ \mathbf{v}(n)\mathbf{v}^T(n) \}$  is the weight-error correlation matrix and  $tr \{ \mathbf{A} \}$  is the trace of the matrix  $\mathbf{A}$ . Post-multiplying (2) by its transpose, taking the expected value

and neglecting the statistical dependence of  $\mathbf{x}(n)$  and  $\mathbf{v}(n)$  [13] leads to the recursive expression:

$$\begin{aligned} \mathbf{K}(n+1) &= \mathbf{K}(n) \\ &- \beta \mathbf{K}(n) E \left\{ \frac{\mathbf{x}(n)\mathbf{x}^T(n)}{\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)} \right\} \\ &- \beta E \left\{ \frac{\mathbf{x}(n)\mathbf{x}^T(n)}{\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)} \right\} \mathbf{K}(n) \\ &+ \beta^2 E \left\{ \frac{\mathbf{x}(n)\mathbf{x}^T(n) \mathbf{K}(n) \mathbf{x}(n)\mathbf{x}^T(n)}{[\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)]^2} \right\} \\ &+ \beta^2 r_\varepsilon E \left\{ \frac{\mathbf{x}(n)\mathbf{x}^T(n)}{[\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)]^2} \right\} \end{aligned} \quad (9)$$

Since the joint probability density of  $\mathbf{w}(n)$  and  $\mathbf{x}(n)$  is not known, the third expectation in (9) can be only approximated. The following approximation is used:

$$E \left\{ \frac{\mathbf{x}(n)\mathbf{x}^T(n) \mathbf{K}(n) \mathbf{x}(n)\mathbf{x}^T(n)}{[\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)]^2} \right\} \cong \frac{E \{ \mathbf{x}(n)\mathbf{x}^T(n) \mathbf{K}(n) \mathbf{x}(n)\mathbf{x}^T(n) \}}{E \{ [\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)]^2 \}} \quad (10)$$

The expectation in the denominator of (10) can be evaluated as:

$$E \{ [\varepsilon + \mathbf{x}^T(n)\mathbf{x}(n)]^2 \} = \varepsilon^2 + 2\varepsilon Nr_0 + N^2 r_0^2 + 2 \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} r_{j-i}^2 \quad (11)$$

where  $r_{j-i} = E \{ x(n-i)x(n-j) \}$  is the element  $(i,j)$  of  $\mathbf{R}_{\mathbf{xx}}$ . Using the Gaussian moment factoring theorem [4, pp.318]:

$$\begin{aligned} E \{ \mathbf{x}(n)\mathbf{x}^T(n) \mathbf{K}(n) \mathbf{x}(n)\mathbf{x}^T(n) \} &= 2\mathbf{R}_{\mathbf{xx}} \mathbf{K}(n) \mathbf{R}_{\mathbf{xx}} \\ &+ tr \{ \mathbf{R}_{\mathbf{xx}} \mathbf{K}(n) \} \mathbf{R}_{\mathbf{xx}} \end{aligned} \quad (12)$$

Substituting (6), (10), (11) and (12) in (9) leads to the following recursion for the behavior of  $\mathbf{K}(n)$ :

$$\begin{aligned} \mathbf{K}(n+1) &= \mathbf{K}(n) \\ &- \frac{\beta}{\varepsilon + Nr_0} [\mathbf{K}(n) \mathbf{R}_{\mathbf{xx}} + \mathbf{R}_{\mathbf{xx}} \mathbf{K}(n)] \\ &+ \frac{\beta^2 \left\{ \left[ r_\varepsilon + tr \{ \mathbf{R}_{\mathbf{xx}} \mathbf{K}(n) \} \right] \mathbf{I} + 2\mathbf{R}_{\mathbf{xx}} \mathbf{K}(n) \right\}}{\varepsilon^2 + 2\varepsilon Nr_0 + N^2 r_0^2 + 2 \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} r_{j-i}^2} \mathbf{R}_{\mathbf{xx}} \end{aligned} \quad (13)$$

For the white Gaussian input case ( $\mathbf{R}_{xx} = r_0 \mathbf{I}$ ), (8) and (13) reduce to the following set of equations:

$$\left\{ \begin{array}{l} \xi(n) = r_z + r_0 \cdot T_K(n) \\ T_K(n+1) = \left\{ 1 - \frac{2\beta r_0}{\varepsilon + Nr_0} \right\} T_K(n) \\ \quad + \frac{\beta^2 (N+2)r_0^2}{\varepsilon^2 + 2\varepsilon Nr_0 + N^2 r_0^2 + 2Nr_0^2} T_K(n) \\ \quad + \frac{\beta^2 Nr_0 r_z}{\varepsilon^2 + 2\varepsilon Nr_0 + N^2 r_0^2 + 2Nr_0^2} \end{array} \right. \quad (14)$$

where  $T_K(n) = \text{tr}\{\mathbf{K}(n)\}$  is a scalar.

## 5. MISADJUSTMENT

Assuming algorithm's convergence as  $n \rightarrow \infty$ , one can use  $\lim_{n \rightarrow \infty} \mathbf{K}(n+1) = \lim_{n \rightarrow \infty} \mathbf{K}(n)$  in (13). Then, taking the trace of (13) leads, after algebraic manipulations, to:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \text{tr}\{\mathbf{K}(n)\mathbf{R}_{xx}\} \\ &= \frac{\beta[\varepsilon + Nr_0] \left\{ \text{tr}\{\mathbf{R}_{xx}\} \left( r_z + \lim_{n \rightarrow \infty} \text{tr}\{\mathbf{R}_{xx}\mathbf{K}(n)\} \right) \right\}}{2 \left[ \varepsilon^2 + 2\varepsilon Nr_0 + N^2 r_0^2 + 2 \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} r_{j-i}^2 \right]} \\ &+ \frac{\beta[\varepsilon + Nr_0] \left\{ 2 \lim_{n \rightarrow \infty} \text{tr}\{\mathbf{R}_{xx}\mathbf{K}(n)\mathbf{R}_{xx}\} \right\}}{2 \left[ \varepsilon^2 + 2\varepsilon Nr_0 + N^2 r_0^2 + 2 \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} r_{j-i}^2 \right]} \end{aligned} \quad (15)$$

Determination of an analytical expression for the misadjustment from (15) requires further approximations. It can be verified that:

$$\begin{aligned} & \text{tr}\{\mathbf{R}_{xx}\} \cdot \lim_{n \rightarrow \infty} \text{tr}\{\mathbf{R}_{xx}\mathbf{K}(n)\} \\ & \gg 2 \cdot \lim_{n \rightarrow \infty} \text{tr}\{\mathbf{R}_{xx}\mathbf{K}(n)\mathbf{R}_{xx}\} \end{aligned} \quad (16)$$

For  $N$  large and for  $x(n)$  white ( $\mathbf{R}_{xx} = r_0 \mathbf{I}$ ), the l.h.s. of (16) is  $N/2$  times the r.h.s. For colored inputs, (16) has been verified by the authors through extensive simulations. Proceeding with the general correlated case and using (16), (15) reduces to:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \text{tr}\{\mathbf{R}_{xx}\mathbf{K}(n)\} \\ & \cong \frac{\beta Nr_0 r_z (\varepsilon + Nr_0)}{2 \left( \varepsilon^2 + 2\varepsilon Nr_0 + N^2 r_0^2 + 2 \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} r_{j-i}^2 \right) - \beta Nr_0 (\varepsilon + Nr_0)} \end{aligned} \quad (17)$$

Using (17) in (8) the following approximated expression is determined for the misadjustment:

$$\begin{aligned} M &= \frac{\xi_{\infty} - \xi_{\min}}{\xi_{\min}} \\ & \cong \frac{\beta Nr_0 (\varepsilon + Nr_0)}{2 \left( \varepsilon^2 + 2\varepsilon Nr_0 + N^2 r_0^2 + 2 \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} r_{j-i}^2 \right) - \beta Nr_0 (\varepsilon + Nr_0)} \end{aligned} \quad (18)$$

## 6. SIMULATIONS

This section presents six simulation results to verify the accuracy of the analytical models given by (8), (13) and (18):

*Example 1 – Low correlation input signal, medium number of taps and small step-size:* In this example  $x(n)$  is Gaussian with  $r_0 = 1$ , and generated by a second order autoregressive (AR) filter defined by the recursion  $x(n) = -a_1 x(n-1) - a_2 x(n-2) + b_0 u(n)$  with  $b_0 = 1$ ,  $a_1 = -0.3$ ,  $a_2 = 0.4$  and  $r_u = 0.8$  (input power to the model). The eigenvalue spread ( $\lambda_{\max}/\lambda_{\min}$ ) of  $\mathbf{R}_{xx}$  is equal to 8.37, yielding a low correlation input signal. The noise power is  $r_z = 10^{-6}$ . The components of  $\mathbf{w}^0$  correspond to a 50-tap Hanning window, normalized for  $\mathbf{w}^{0T} \mathbf{w}^0 = 1$ . The parameters of the algorithm are  $\beta = 0.1$ ,  $\varepsilon = 1$  and  $\mathbf{w}(0) = \mathbf{0}$ .

The results of Monte Carlo simulations (1000 runs) and the theoretical prediction of the algorithm's behavior are shown in Fig. 2. Note that the analytical model predicts very well the steady-state algorithm's behavior. In the adaptation phase, the model is conservative and thus useful for design purposes.

*Example 2 – High correlation input signal, medium number of taps and small step-size:* All parameters are the same as in Example 1, except for  $a_2 = 0.8$  and  $r_u = 0.35$ , resulting in an eigenvalue spread of 96.53. Fig. 3 presents Monte Carlo Simulation (1000 runs) and the theoretical result.

Comparing Figs. 2 and 3 it is possible to verify the model's robustness to variations in the eigenvalue spread of the input correlation matrix (from 8.37 to 96.53). The model's accuracy in the adaptation phase is comparable to the accuracy of the existing models for the NLMS algorithm with white inputs (see, for instance, Fig. 5). The mismatches between theory and simulation in Examples 1 and 2 are largely due to the use of the independence assumption. Thus, the model gets more accurate for smaller step sizes [13].

*Example 3 – High correlation input signals, reduced number of coefficients and large step-size:* In this example  $x(n)$  Gaussian with  $r_0 = 1$  and  $\lambda_{\max}/\lambda_{\min}$  equal to 82.98 (same AR model as in Example 2);  $r_z = 10^{-6}$ . The unknown system response  $\mathbf{w}^0$  has 30 taps (normalized Hanning window). The algorithm's parameters are  $\beta = 0.9$ ,  $\varepsilon = 0.1$  and  $\mathbf{w}(0) = \mathbf{0}$ . This example uses a set of parameters that is not favorable to the model. The number of taps is close to the limit where the assumptions used to derive (5) are no longer valid. Also, large step-size and eigenvalue spread make the independence assumption less valid.

The Monte Carlo results for 1000 runs and the theoretical prediction are shown in Fig. 4. The steady state theoretical prediction is still very good. The accuracy of the model is reduced during transient (as compared to Examples 1 and 2), but the model remains conservative and useful for design purposes.

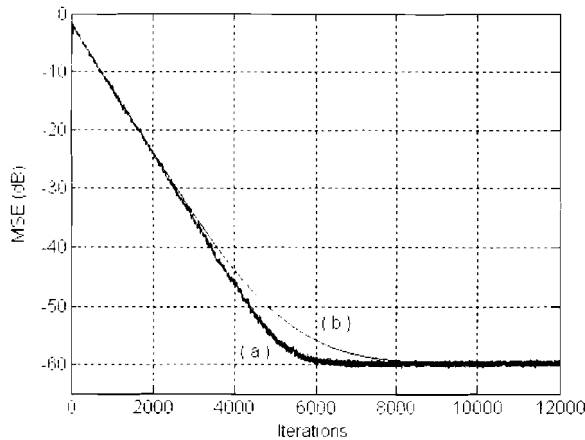


Figure 2. MSE, example 1. (a) simulations; (b) theoretical model.

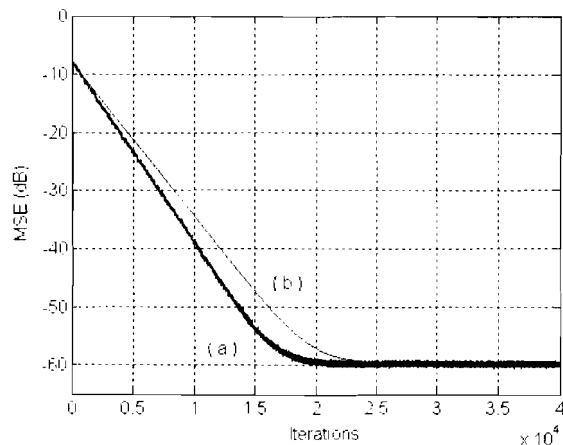


Figure 3. MSE, example 2. (a) simulations; (b) theoretical model.

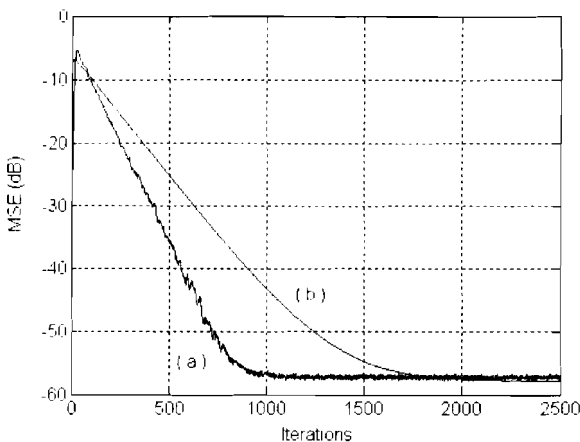


Figure 4. MSE, example 3. (a) simulations; (b) theoretical model.

*Example 4 – Comparison with Bershad's model [9] for white signals:* In this example  $x(n)$  is white and Gaussian with  $r_0 = 1$ ,  $r_z = 10^{-6}$  and  $w^0$  has 10 coefficients (normalized Hanning window). The algorithm's parameters are  $\beta = 1$ ,  $\varepsilon = 1$  and  $w(0) = 0$ . The simulation results (1000 runs) are shown in Fig. 5. Also shown in this figure are the theoretical predictions using the model presented in [9] and the new model. Note that both models present good agreement with the algorithm's behavior (even for a small number of taps).

*Example 5 – Comparison with Slock's [6] and Costa's model [7] for the NLMS case ( $\varepsilon = 0$ ) with small number of taps and correlated signals:* In this example  $x(n)$  is Gaussian with  $r_0 = 1$ . The input eigenvalue spread  $\lambda_{\max}/\lambda_{\min}$  is equal to 68.4 (same AR model as in Example 2). The noise power is  $r_z = 10^{-6}$ . The plant  $w^0$  has 20 coefficients (normalized Hanning window). The algorithm's parameters are  $\beta = 1$ ,  $\varepsilon = 0$  (NLMS) and  $w(0) = 0$ . Simulation results are shown in Fig. 6. The Monte Carlo simulation corresponds to the average of 1000 runs. The theoretical results are presented using three different analytical models. Curve (b) was obtained using the model proposed in [6], curve (c) corresponds to the model proposed in [7] and curve (d) was obtained using the new model. These plots clearly show the superiority of the new model. Nevertheless, note that this example is for a small number of taps and a large step-size, conditions that are not favorable to the new model.

*Example 6 – Comparison with Slock's [6] and Costa's model [7] for the NLMS case ( $\varepsilon = 0$ ) with large number of taps and correlated signals:* The input signal is Gaussian with  $r_0 = 1$ . The input eigenvalue spread  $\lambda_{\max}/\lambda_{\min}$  is equal to 68.4 (same AR model as in Example 1, except for  $a_2 = 0.7497$ ). The noise power is  $r_z = 10^{-6}$ . The plant  $w^0$  has 200 coefficients (normalized Hanning window). The algorithm's parameters are  $\beta = 0.1$ ,  $\varepsilon = 0$  (NLMS) and  $w(0) = 0$ . The Monte Carlo simulation corresponds to the average of 1000 runs. The results are presented in Fig. 7.

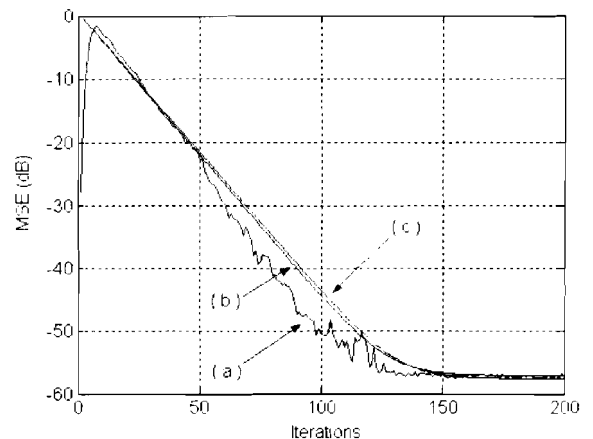


Figure 5. MSE, example 4. (a) simulations; (b) Bershad's model [9]; (c) the new model.

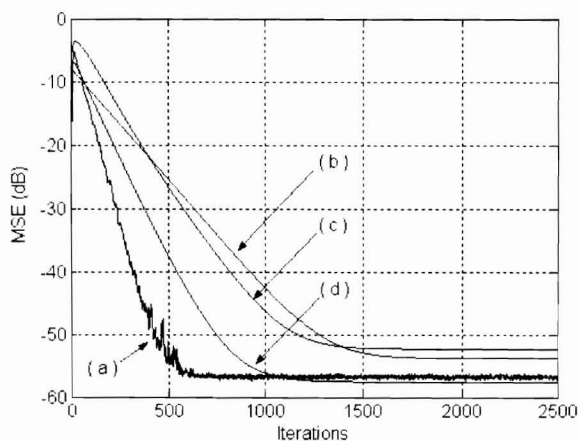


Figure 6. MSE, example 5. (a) simulations; (b) Slock's model [6]; (c) Costa's model [7]; (d) the new model.

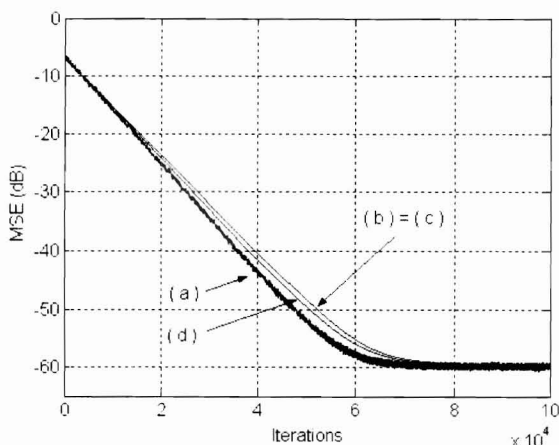


Figure 7. MSE, example 6. (a) simulations; (b) Slock's model [6]; (c) Costa's model [7]; (d) the new model.

Table 1 compares the steady-state MSE misadjustment predictions obtained by using (8) and (13), and using the closed form approximation (18). It can be verified that (18) produces very good estimates of the steady-state results obtained from running (8) and (13). Both results show a very good agreement with simulations for most design purposes.

$N$	$\beta$	$\epsilon$	$\chi$	Simulation	Eqs.(8)-(13)	Eq.(18)
5	1	1	21	-57.6	-58.2	-58.6
10	1	0.01	45	-56.4	-57.8	-58.3
15	0.1	15	59	-59.9	-59.9	-59.9
50	1	0.01	97	-57	-57.3	-57.6

Table 1. Comparisons between steady-state MSE predictions and simulations (dBs units).  $N$  is the number of taps of the plant (normalized Hanning window),  $\chi$  is the eigenvalue spread,  $r_2 = 10^{-6}$ .

## 7. CONCLUSIONS

This work presented a statistical analysis of the  $\epsilon$ -NLMS algorithm for Gaussian input signals. Deterministic recursive expressions were derived for the mean weight and MSE behaviors for slow learning and a large number of adaptive weights. The new model does not require any numerical procedure and is valid for white or correlated Gaussian input signals. For  $\epsilon=0$  the result becomes an analytical model for the behavior of the NLMS algorithm. The new model is more accurate than those previously proposed in the literature. An approximated closed expression was derived for the steady-state MSE misadjustment. Monte Carlo simulations show very good agreement between model predictions and simulation results in steady state and fair to good agreement during the acquisition phase, even for large step-sizes and small number of coefficients.

## ACKNOWLEDGMENTS

Thanks to FAPERGS (Research Foundation of the State of Rio Grande do Sul - Brazil) under Grant No. 01/1173.0 and to CNPq (Brazilian Ministry of Science and Technology) under Grants Nos. 308095/2003-0 and 472762/03-6 for funding.

## REFERENCES

- [1] S. M. Kuo, D. R. Morgan, Active Noise Control Systems: Algorithms and DSP Implementations, New York: John Wiley, 1996.
- [2] C. Breining, P. Dreiscitel, E. Hansler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, J. Tilp, "Acoustic Echo Control. An Application of Very-High-Order Adaptive Filters," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 42-69, July, 1999.
- [3] W. L. B. Jeannes, P. Scalart, G. Faucon, C. Beaugeant, "Combined Noise and Echo Reduction in Hands-Free Systems: a Survey," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 808-820, November, 2001.
- [4] S. Haykin, Adaptive Filter Theory, second edition, Prentice-Hall, 1991.
- [5] J. J. Naguno, A. Noda, "A Learning Method for System Identification," *IEEE Trans. on Automatic Control*, vol. AC-12, pp. 282-287, 1967.
- [6] D. T. M. Slock, "On the Convergence Behavior of the LMS and the Normalized LMS Algorithms," *IEEE Trans. on Signal Processing*, vol. 41, no. 9, pp. 2811-2825, September, 1993.
- [7] M. H. Costa, J.C.M. Bermudez, "An Improved Model for the Normalized LMS Algorithm with Gaussian Inputs and Large Number of Coefficients," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, USA, Orlando, 2002.
- [8] P. Lapsley, J. Bier, A. Shoham, E. A. Lee, DSP Processor Fundamentals, IEEE Press, 1997.
- [9] N. J. Bershad, "Behavior of the  $\epsilon$ -Normalized LMS Algorithm with Gaussian Inputs," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-35, no. 5, pp. 636-644, May, 1987.
- [10] S. C. Douglas, T. H. Y. Meng, "Normalized Data Nonlinearities for LMS Adaptation," *IEEE Trans. Signal Processing*, vol. 42, no. 6, pp. 1352-1365, June, 1994.

- [11] V. Myllylä, G. Schmidt, "Pseudo-Optimal Regularization for Affine Projection Algorithms," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, USA, Orlando, 2002.
- [12] D. P. Mandic, "A Generalized Normalized Gradient Descent Algorithm," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 115-118, February, 2004.
- [13] J. E. Mazo, "On the Independence Theory of Equalizer Convergence," *Bell Syst. Tech. Journal*, vol. 58, pp. 963-993, May-June, 1979.
- [14] C. Samson, V. U. Reddy, "Fixed Point Error Analysis of the Normalized Ladder Algorithms," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 31, no. 5, pp. 1177-1191, October, 1983.
- [15] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, third edition, McGraw-Hill, 1991.

**José C. M. Bermudez** received the B.E.E. degree from Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, in 1978, the M.Sc. degree from COPPE/UFRJ, in 1981, and the Ph.D. degree from Concordia University, Montreal, Canada, in 1985, both in electrical engineering. He joined the Department of Electrical Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, SC, Brazil, in 1985. He is currently a Professor of electrical engineering. In the winter of 1992, he was a Visiting Researcher with the Department of Electrical Engineering, Concordia University. In 1994, he was a Visiting Researcher with the Department of Electrical Engineering and Computer Science, University of California, Irvine (UCI). His research interest have involved analog signal processing using continuous-time and sampled-data systems. His recent research interests are in digital signal processing, including linear and nonlinear adaptive filtering, active noise and vibration control, echo cancellation, image processing, and speech processing. He is a Senior Member of the IEEE and a Member of the Brazilian Telecommunications Society. He served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING in the area of adaptive filtering from 1994 to 1996 and from 1999 to 2001. He was a member of the Signal Processing Theory and Methods technical committee of the IEEE Signal Processing Society from 1998 to 2004. Prof. Bermudez is presently serving as the Signal Processing area editor for the Journal of the Brazilian Telecommunications Society.

**Márcio H. Costa** received the B.E.E. degree from Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil, in 1991, the M.Sc. degree in biomedical engineering from Universidade Federal do Rio de Janeiro (COPPE/UFRJ), Rio de Janeiro, Brazil, in 1994, and the Dr. degree in electrical engineering from Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brazil, in 2001. He joined the Department of Electrical Engineering, Universidade Federal de Santa Catarina in 2004. From 1994 to 2004 he was with the Department of Electrical Engineering, Biomedical Engineering Group, Universidade Católica de Pelotas. His present research interests are in biomedical signal processing, linear and nonlinear adaptive filters, adaptive inverse control and active noise and vibration control.