A Statistical Analysis of the ε-NLMS and NLMS Algorithms for Correlated Gaussian Signals

José C. M. Bermudez and Márcio H. Costa

Resumo—**Este trabalho apresenta uma análise estatística do algoritmo ε-NLMS para sinais Gaussianos correlacionados. Equações determinísticas recursivas foram obtidas para o comportamento médio dos coeficientes e erro médio quadrático para um elevado número de coeficientes. Simulações Monte Carlo permitem verificar a concordância entre as predições do modelo e simulações.**

Palavras-Chave—**Filtros adaptativos, LMS normalizado, NLMS, modelagem analítica.**

Abstract—**This work presents a statistical analysis of the ε-NLMS algorithm for correlated Gaussian input signals. Deterministic recursive expressions are derived for the mean weight and mean square error (MSE) behaviors for a large number of adaptive weights. Monte Carlo simulations show the agreement between model predictions and simulations.**

Index Terms— **Adaptive filters, normalized LMS, NLMS, analytical modeling.**

I. INTRODUCTION

Adaptive filtering techniques are widely employed in real life applications such as modeling, equalization, active noise control and echo cancellation. Stochastic gradient based algorithms have proven to be both robust and easily-implemented for control and on-line estimation applications.

The Least Mean Square (LMS) is the most popular adaptive algorithm due to its robustness and low computational complexity [1]. 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.

The Normalized-LMS (NLMS) algorithm, also known as the projection algorithm [2,3], 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 normalizing factor is determined by the squared norm of the input vector ($\mu(n)=\beta/\left[\mathbf{X}^T(n)\mathbf{X}(n)\right]$). Albeit this procedure may seem computationally intensive, the squared norm can be computed recursively, increasing the algorithm's cost in only two multiplications and one division, when compared to the LMS algorithm. In applications where a large number of coefficients is necessary, the advantages of using the NLMS algorithm overcome the cost of implementation [3].

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. 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 this situation occurs, the normalizing factor of the NLMS algorithm can become too small (even zero), which is undesirable. To avoid divisions by zero, a small positive constant (ε) is frequently added to the normalizing factor yielding a step-size of the form $\mu(n)=\beta/\left[\varepsilon+\mathbf{X}^T(n)\mathbf{X}(n)\right]$. This modification characterizes the ε-NLMS algorithm, a generalization of the conventional NLMS.

The ε-NLMS was analyzed by Bershad [4] 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 [5] studied normalized data nonlinearities for LMS adaptation in order to improve the algorithm's performance. 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 ε-NLMS with unit step-size. In this algorithm, the convergence performance is controlled by the parameter ε. The authors of [5] have shown that up to 3.6 dB improvement can be obtained over the LMS MSE performance when using this algorithm. However, [5] did not provide any analysis for the correlated Gaussian input case.

José Bermudez is with Department of Electrical Engineering, Universidade Federal de Santa Catarina, Florianópolis, Brazil. Márcio Costa is with Lab. de Eng. Biomédica, Escola de Engenharia e Arquitetura, Universidade Católica de Pelotas, Pelotas, Brazil. E-mails: j.bermudez@ieee.org.br, m.costa@ieee.org.

Recently, the fast affine projection algorithm was studied for hands-free telephone applications as a generalization of the ε-NLMS algorithm [6]. In this paper, the ε factor was used to control the performance of the adaptive filter. These results revived the interest on the ε-NLMS, now with ε being more than just a regularization constant.

This work presents a statistical analysis of the ε-NLMS algorithm for correlated Gaussian 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 [2,3].

Deterministic recursive equations are derived for the mean weight and MSE behaviors using the independence assumption [7] and the averaging principle [8]. Assuming algorithm 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.

II. THE ε-NLMS UPDATE EQUATION

The update equation of the ε-NLMS is given by [4]:

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \beta e(n)\mathbf{X}(n)\left[\varepsilon + \mathbf{X}^T(n)\mathbf{X}(n)\right]^{-1} \quad (1)$$

where

$$e(n) = \mathbf{W}^{oT}\mathbf{X}(n) + z(n) - \mathbf{W}^T(n)\mathbf{X}(n) \quad (2)$$

is the error signal. $\mathbf{W}^{o} = \left[w_0^o \;\; w_1^o \;\; \dots \;\; w_{N-1}^o\right]^T$ is the unknown impulse response; $z(n)$ is a stationary, white, zero-mean Gaussian measurement noise with variance σ_z^2 and uncorrelated with any other signal. β is the step-size. $\mathbf{W}(n) = \left[w_0(n) \;\; w_1(n) \;\; \dots \;\; w_{N-1}(n)\right]^T$ is the adaptive weight vector. $\mathbf{X}(n) = \left[x(n) \;\; x(n-1) \;\; \dots \;\; x(n-N+1)\right]^T$ is the observed data vector; $x(n)$ is stationary, zero-mean and Gaussian. ε is a small positive constant. For $\varepsilon = 0$ the algorithm becomes the popular NLMS algorithm [3]. The choice of $\beta = 1$ leads to the optimal normalized data vector nonlinearity [5].

III. 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 behavior can be neglected. This corresponds to the use of the well known independence assumption [7]. Defining the weight error vector $\mathbf{V}(n) = \mathbf{W}(n) - \mathbf{W}^{o}$, using (2) in (1), taking the expected value and noting that $E\{x(n)z(n)\} = 0$ yields:

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

Each element of the expectation within the square brackets has a numerator given by $x(n-i)x(n-j)$ 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 [8], 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 have shown that this assumption holds very well for N as small as 30 and for a wide range of input eigenvalue spreads. Moreover, since ε is not always very small [5,6] further approximation is required. Thus, the following approximation is used:

$$E\left\{\left[\varepsilon + \mathbf{X}^T(n)\mathbf{X}(n)\right]^{-1}\mathbf{X}(n)\mathbf{X}^T(n)\right\} \cong \left[\varepsilon + E\left\{\mathbf{X}^T(n)\mathbf{X}(n)\right\}\right]^{-1} E\left\{\mathbf{X}(n)\mathbf{X}^T(n)\right\} \quad (4)$$

The second expected value in (4) is the correlation matrix of the input signal and the first one is its trace. Using this information, (4) becomes:

$$E\left\{\left[\varepsilon + \mathbf{X}^T(n)\mathbf{X}(n)\right]^{-1}\mathbf{X}(n)\mathbf{X}^T(n)\right\} \cong \frac{1}{\varepsilon + N\sigma_x^2}\mathbf{R}_{\mathbf{xx}} \quad (5)$$

with $\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{X}(n)\mathbf{X}^T(n)\}$. Using, (5) in (3) leads to:

$$E\{\mathbf{V}(n+1)\} = \left[\mathbf{I} - \frac{\beta}{\varepsilon + N\sigma_x^2}\mathbf{R}_{\mathbf{xx}}\right] E\{\mathbf{V}(n)\} \quad (6)$$

Equation (6) determines the mean weight behavior. For $\varepsilon = 0$, (6) becomes the NLMS model derived in [3, Eq. 8].

IV. MSE BEHAVIOR

Squaring (2) and taking the expected value, results in:

$$\xi(n) = E\{e^2(n)\} = \sigma_z^2 + tr\{\mathbf{R}_{\mathbf{xx}}\mathbf{K}(n)\} \quad (7)$$

where $\mathbf{K}(n) = E\{\mathbf{V}(n)\mathbf{V}^T(n)\}$ is the weight-error correlation matrix. Postmultiplying (1) by its transpose, taking the expected value and neglecting the statistical dependence of $\mathbf{X}(n)$ and $\mathbf{V}(n)$ [7] leads to the recursive expression:

$$\begin{aligned}
\mathbf{K}(n+1) = {} & \mathbf{K}(n) + \beta^2\sigma_z^2 E\left\{\left[\varepsilon + \mathbf{X}^T(n)\mathbf{X}(n)\right]^{-2}\mathbf{X}(n)\mathbf{X}^T(n)\right\} \\
& + \beta^2 E\left\{\left[\varepsilon + \mathbf{X}^T(n)\mathbf{X}(n)\right]^{-2}\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\right\} \\
& - \beta\mathbf{K}(n) E\left\{\left[\varepsilon + \mathbf{X}^T(n)\mathbf{X}(n)\right]^{-1}\mathbf{X}(n)\mathbf{X}^T(n)\right\} \\
& - \beta E\left\{\left[\varepsilon + \mathbf{X}^T(n)\mathbf{X}(n)\right]^{-1}\mathbf{X}(n)\mathbf{X}^T(n)\right\}\mathbf{K}(n)
\end{aligned} \quad (8)$$

Since the joint probability density of $\mathbf{W}(n)$ is not known, the second expectation value in (8) can be only approximated. The following approximation is used:

$$E\left\{\left[\varepsilon + \mathbf{X}^T(n)\mathbf{X}(n)\right]^{-2}\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\right\} \cong$$

$$E\left\{\left[\varepsilon+\mathbf{X}^T(n)\mathbf{X}(n)\right]^2\right\}^{-1} E\left\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\right\} \quad (9)$$

The first expectation on the r.h.s. of (9) can be evaluated as:

$$E\left\{\left[\varepsilon+\mathbf{X}^T(n)\mathbf{X}(n)\right]^2\right\}=\varepsilon^2+2\varepsilon N\sigma_x^2+N^2\sigma_x^4+2\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} r_{j-i}^2 \quad (10)$$

where $r_{j-i}=E\{x(n-i)x(n-j)\}$ is the element (i,j) of $\mathbf{R_{xx}}$. Using the Gaussian moment factoring theorem [1, pp.318]:

$$E\left\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\right\}= 2\mathbf{R_{xx}}\mathbf{K}(n)\mathbf{R_{xx}}+tr\{\mathbf{R_{xx}}\mathbf{K}(n)\}\mathbf{R_{xx}} \quad (11)$$

Substituting (4), (5) and (9) to (11) in (8) we obtain a recursive equation for the behavior of $\mathbf{K}(n)$:

$$\mathbf{K}(n+1)=\mathbf{K}(n)-\frac{\beta}{\varepsilon+N\sigma_x^2}\left[\mathbf{K}(n)\mathbf{R_{xx}}+\mathbf{R_{xx}}\mathbf{K}(n)\right] + \frac{\beta^2\left\{\left[\sigma_z^2+tr\{\mathbf{R_{xx}}\mathbf{K}(n)\}\right]\mathbf{I}+2\mathbf{R_{xx}}\mathbf{K}(n)\right\}}{\varepsilon^2+2\varepsilon N\sigma_x^2+N^2\sigma_x^4+2\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} r_{j-i}^2}\mathbf{R_{xx}} \quad (12)$$

V. MISADJUSTMENT

Assuming algorithm convergence as $n\to\infty$, one can use $\lim_{n\to\infty}\mathbf{K}(n+1)=\lim_{n\to\infty}\mathbf{K}(n)$ in (12). Then, taking the trace of the equation leads, after algebraic manipulations, to:

$$\lim_{n\to\infty} tr\{\mathbf{K}(n)\mathbf{R_{xx}}\}=\frac{\beta\left[\varepsilon+N\sigma_x^2\right]\left[\sigma_z^2 tr\{\mathbf{R_{xx}}\}\right]}{2\left[\varepsilon^2+2\varepsilon N\sigma_x^2+N^2\sigma_x^4+2\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} r_{j-i}^2\right]} + \frac{\beta\left[\varepsilon+N\sigma_x^2\right]\left[tr\{\mathbf{R_{xx}}\}\lim_{n\to\infty} tr\{\mathbf{R_{xx}}\mathbf{K}(n)\}+2\lim_{n\to\infty} tr\{\mathbf{R_{xx}}\mathbf{K}(n)\mathbf{R_{xx}}\}\right]}{2\left[\varepsilon^2+2\varepsilon N\sigma_x^2+N^2\sigma_x^4+2\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} r_{j-i}^2\right]} \quad (13)$$

Determination of an analytical expression for the misadjustment from (13) requires further approximations. It can be verified that:

$$tr\{\mathbf{R_{xx}}\}\cdot\lim_{n\to\infty} tr\{\mathbf{R_{xx}}\mathbf{K}(n)\} >> 2\cdot\lim_{n\to\infty} tr\{\mathbf{R_{xx}}\mathbf{K}(n)\mathbf{R_{xx}}\} \quad (14)$$

For N large the l.h.s. of (14) is $N/2$ times the r.h.s. for $x(n)$ white ($\mathbf{R_{xx}}=\sigma_x^2\cdot\mathbf{I}$). For colored inputs, (14) has been verified by extensive simulations. Using (14), (13) reduces to:

$$\lim_{n\to\infty} tr\{\mathbf{R_{xx}}\mathbf{K}(n)\}\cong \frac{\beta N\sigma_z^2\sigma_x^2\left(\varepsilon+N\sigma_x^2\right)}{2\left(\varepsilon^2+2\varepsilon N\sigma_x^2+N^2\sigma_x^4+2\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} r_{j-i}^2\right)-\beta N\sigma_x^2\left(\varepsilon+N\sigma_x^2\right)} \quad (15)$$

Using (15) in (7) the following approximated expression is determined for the misadjustment $M=\left(\xi_\infty-\xi_{\min}\right)/\xi_{\min}$:

$$M\cong\frac{\beta N\sigma_x^2\left(\varepsilon+N\sigma_x^2\right)}{2\left(\varepsilon^2+2\varepsilon N\sigma_x^2+N^2\sigma_x^4+2\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} r_{j-i}^2\right)-\beta N\sigma_x^2\left(\varepsilon+N\sigma_x^2\right)} \quad (16)$$

VI. SIMULATIONS

This section presents four simulation results to verify the accuracy of the analytical models given by (7), (12) and (16):

Example 1 – Correlated signal, large number of taps and small step-size: $x(n)$ Gaussian with $\sigma_x^2=1$, generated by an AR filter with $a_0=1$, $a_1=-0.3$, $a_2=0.8$ and $\sigma_u^2=0.35$ (input power to the model). The eigenvalue spread ($\lambda_{\max}/\lambda_{\min}$) of $\mathbf{R_{xx}}$ is equal to 96.53. The noise power is $\sigma_z^2=10^{-6}$. The components of $\mathbf{W^o}$ correspond to a 50-tap Hanning window, normalized for $\mathbf{W}^{0^T}\mathbf{W}^0=1$. $\beta=0.1$, $\varepsilon=1$ and $\mathbf{W}(0)=\mathbf{0}$. Results of Monte Carlo simulations (500 runs) and theory are shown in Fig. 1.

Example 2 – Correlated signals and large step-size: $x(n)$ Gaussian with $\sigma_x^2=1$ and $\lambda_{\max}/\lambda_{\min}$ equal to 82.98 (same AR model as in Ex. 1); $\varepsilon=0.1$; $\sigma_z^2=10^{-6}$; 30 taps (normalized Hanning window); $\beta=0.9$ and $\mathbf{W}(0)=\mathbf{0}$. Results for 1000 runs are shown in Fig. 2.

Example 3 – Comparison with Bershad's model [4] for white signals: $x(n)$ Gaussian with $\sigma_x^2=1$. $\varepsilon=1$; $\sigma_z^2=10^{-6}$; 10 coefficients (normalized Hanning window). 1000 runs, $\beta=1$ and $\mathbf{W}(0)=\mathbf{0}$. Results in Fig. 3.

Example 4 – Comparison with Slock's [2] and Costa's model [3] for the NLMS case ($\varepsilon=0$ *) with correlated signals:* $x(n)$ Gaussian with $\sigma_x^2=1$. $\lambda_{\max}/\lambda_{\min}$ equal to 68.4 (same AR model as in Ex. 1); $\varepsilon=0$; $\sigma_z^2=10^{-6}$; 20 coefficients (normalized hanning window). 1000 runs, $\beta=1$ and $\mathbf{W}(0)=\mathbf{0}$. Results in Fig. 4.

Figs. 1 and 2 show that the new model provides good results for the correlated inputs and small step sizes. For large step sizes, the model leads to predictions that are good in steady-state and fair during transient. The agreement between simulations and theoretical predictions improves for larger *N*.

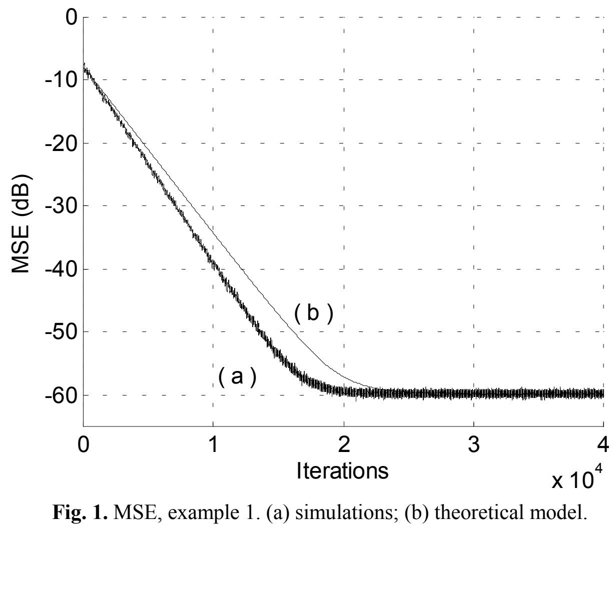

Fig. 1. MSE, example 1. (a) simulations; (b) theoretical model.

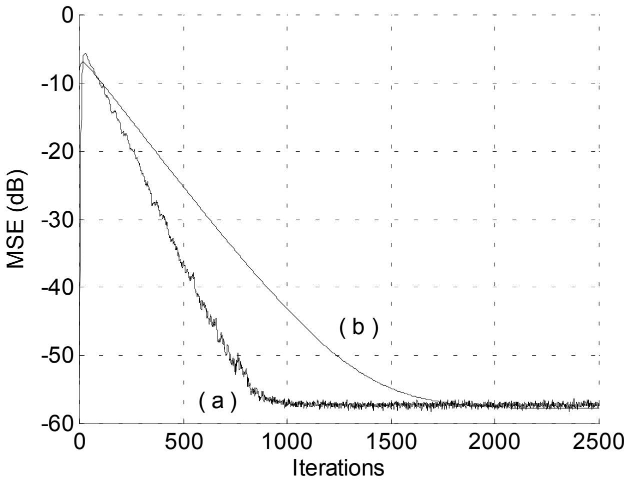

Fig. 2. MSE, example 2. (a) simulations; (b) theoretical model.

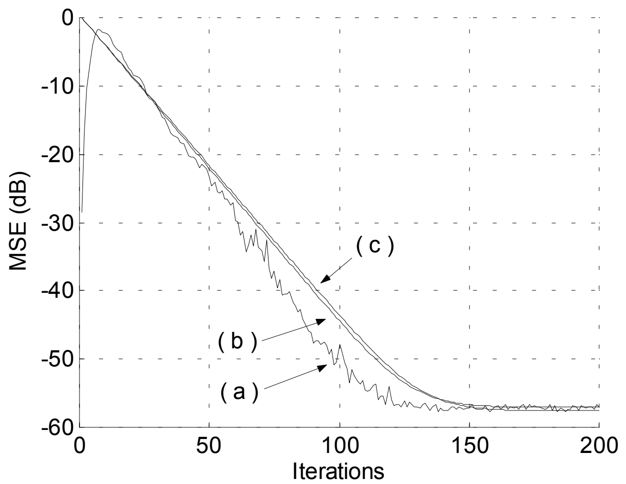

Fig. 3. MSE, example 3. (a) simulations; (b) Bershad's model [4]; (c) the new model.

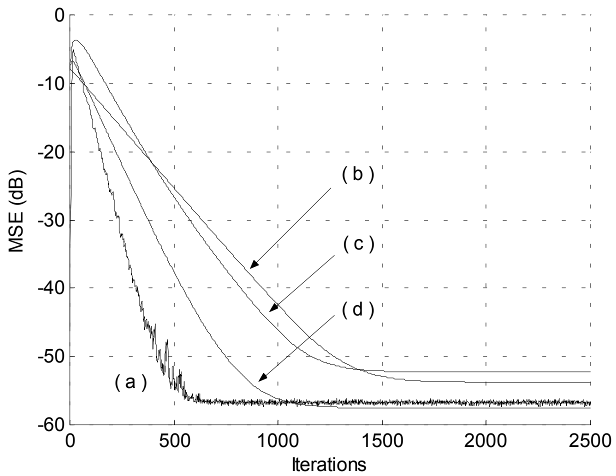

Fig. 4. MSE, example 4. (a) simulations; (b) Slock's model [2]; (c) Costa's model [3]; (d) the new model.

Fig. 3 shows that the new model produces results which are equivalent to those obtained with Bershad's model for white Gaussian inputs, even for a small number of taps.

Fig. 4 corresponds to modeling the NLMS algorithm, since $\varepsilon = 0$. The results are compared with those obtained using Slock's [2] and Costa's [3] models. The new model clearly provides a better prediction of the algorithm's behavior. Note that this example is for a small N and a large step size, conditions which are not favorable to the assumptions used in the new model.

Table 1 compares the steady-state MSE misadjustment predictions obtained running (7) and using the closed form approximation (16). It can be verified that (16) produces estimates that are very similar to those obtained from (7) after convergence. Both results have a very good agreement with simulations for most design purposes.

TABLE I

COMPARISONS BETWEEN STEADY-STATE MSE PREDICTIONS AND SIMULATIONS (DBS UNITS). N IS THE NUMBER OF TAPS, χ IS THE EIGENVALUE SPREAD, $\sigma_z^2 = 10^{-6}$.

N	β	ε	χ	Simulation	Eq. (7)	Eq. (16)
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

VII. SUMMARY

This work presented a statistical analysis of the ε-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 $\varepsilon = 0$ the result becomes an analytical model for the behavior of the NLMS algorithm. This model is more accurate and robust 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 Grant No. 300787/87-1 for funding.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, second edition, Prentice-Hall, 1991.

[2] 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.

[3] M.H. Costa and J.C.M. Bermudez, "An Improved Model for the Normalized LMS Algorithm with Gaussian Inputs and Large Number of Coefficients," *ICASSP 2002*, USA, Orlando, 2002.

[4] N.J. Bershad, "Behavior of the ε-Normalized LMS Algorithm with Gaussian Inputs," *IEEE Trans. on Acoustics, Speech and Signal Processing,* vol. ASSP-35, no. 5, pp. 636-644, May, 1987.

[5] S.C. Douglas and T.H.Y. Meng, "Normalized Data Nonlinearities for LMS Adaptation," *IEEE Trans. Signal Processing,* vol. 42, no. 6, pp. 1352-1365, June, 1994.

[6] V. Myllylä and G. Schmidt, "Pseudo-Optimal Regularization for Affine Projection Algorithms," *ICASSP 2002*, USA, Orlando, 2002.

[7] J.E. Mazo, "On the Independence Theory of Equalizer Convergence," *Bell Syst. Tech. Journal*, vol. 58, pp. 963-993, May-June, 1979.

[8] C. Samson, and 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.