A NEW ANALYTICAL MODEL FOR THE NLMS ALGORITHM

Guillaume Barrault[1], Marcio H. Costa[2], José C. M. Bermudez[2] and Arcanjo Lenzi[1]

[1] Dep. of Mechanical Engineering, Universidade Federal de Santa Catarina, Florianópolis-SC, Brazil
[2] Dep. of Electrical Engineering, Universidade Federal de Santa Catarina, Florianopólis-SC, Brazil
E-mails:guillaume@lva.ufsc.br, m.costa@ieee.org, j.bermudez@ieee.org, arcanjo@emc.ufsc.br

ABSTRACT

This paper presents a new analytical model for the Normalized Least Mean Square (NLMS) adaptive algorithm. The new model is derived using a stochastic differential equation (SDE) approach. An accurate estimate of the steady-state weight-error correlations is also derived, which leads to an improved model performance for medium and large step sizes. Numerical simulations compare the new model with existing models and show better agreement with Monte Carlo simulations.

1. INTRODUCTION

The Normalized Least Mean Square (NLMS) algorithm belongs to the family of stochastic gradient algorithms. Compared to the famous Least Mean Square (LMS) algorithm, NLMS presents the advantage of normalizing the weight vector updates with respect to the input signal power. This makes the algorithm less sensitive to variations in input power.

The NLMS algorithm behavior has been studied by several authors [1-4] for white and correlated inputs. It is well known that the statistical analysis of the algorithm behavior is complicated by the normalized weight update. This term leads to statistical expectations which are very difficult to evaluate. Accurate analysis results have been presented [1] and [2] for white inputs and for input covariance matrices with two distinct eigenvalues. However, the important case of correlated inputs could not be solved in closed form using the techniques employed in [1,2]. In [3], a new statistical model is proposed for the input signal. The input signal distribution is modeled as the product of a radial and an angular distribution. The resulting model was able to predict the algorithm behavior for correlated inputs. However, the theoretical model could not follow the knee behavior of the mean square error (MSE) of the actual algorithm. Steady-state errors could also be verified for correlated input signals. In [4], a new analytical model was proposed for Gaussian inputs and large number of adaptive coefficients. The model in [4] was more accurate than all the existing models for correlated Gaussian inputs. However, the theoretical predictions using this model still present significant errors at the knee of the MSE curve for large step sizes. Steady-state errors can also be verified for a small number of coefficients.

This paper proposes a new statistical analysis of the NLMS algorithm behavior. The analysis is based on the stochastic differential equation approach introduced in [5] to study the behavior of recursive stochastic algorithms through the associated ordinary differential equation (ODE) [6]. This approach has been successfully employed in [7] to study the behavior of finite precision implementations of adaptive IIR filters. The analysis of the NLMS algorithm behavior using the SDE approach, however, requires modifications when the step size is large. The limitations of the method for large step sizes are overcome by an accurate estimation of the algorithm's steady-state behavior.

The paper is divided in two parts. Section 2 presents the analysis of the NLMS algorithm using the SDE approach, with its limitations for large step sizes. Section 3 presents new results on the steady-state behavior of the weight-error correlation matrix, which are used to improve the model for large step sizes. Simulation results are provided comparing the new model with the model in [4].

2. NEW NLMS MODEL - THE SDE METHOD

The update equation of the NLMS algorithm is given by:

$$\mathbf{W}(n+1)=\mathbf{W}(n)+\mu\frac{e(n)\mathbf{X}(n)}{\mathbf{X}^T(n)\mathbf{X}(n)} \tag{1}$$

where $\mathbf{W}(n)=\left[w_0(n)\cdots w_{N-1}(n)\right]^T$ is the adaptive weight vector, $\mathbf{X}(n)=\left[x(n)\cdots x(n-N+1)\right]^T$ is the input data vector, μ is the step-size and $e(n)$ is the estimation error, given by:

$$e(n)=y(n)-\hat{y}(n)=\mathbf{W}_{opt}^T\mathbf{X}(n)+z(n)+\mathbf{W}^T(n)\mathbf{X}(n) \tag{2}$$

where $W_{opt} = \left[w_0 \cdots w_{N-1}\right]^T$ is the true weight vector and $z(n)$ is a zero-mean, stationary, white Gaussian noise with variance σ_z^2. Defining the weight-error vector $\mathbf{V}(n) = \mathbf{W}(n) - \mathbf{W}_{opt}$, (1) becomes:

$$\mathbf{V}(n+1) = \mathbf{V}(n) + \mu \frac{e(n)\mathbf{X}(n)}{\mathbf{X}^T(n)\mathbf{X}(n)} \tag{3}$$

The MSE is given by [11]:

$$E\left[e^2(n)\right] = \sigma_z^2 + trace\left\{\mathbf{R}_{xx}\mathbf{K}(n)\right\} \tag{4}$$

with $\mathbf{K}(n) = E\left[\mathbf{V}(n)\mathbf{V}^T(n)\right]$ and $\mathbf{R}_{xx} = E\left[\mathbf{X}(n)\mathbf{X}^T(n)\right]$. Equation (4) can be written as a function of the eigenvector matrix $\mathbf{Q}$ and the eigenvalue matrix Λ:

$$E\left[e^2(n)\right] = \sigma_z^2 + tr\left\{\mathbf{\Lambda}\tilde{\mathbf{K}}(n)\right\} = \sigma_z^2 + \boldsymbol{\lambda}^{\mathbf{T}}\tilde{\mathbf{k}}(n) \tag{5}$$

with $\mathbf{R}_{xx} = \mathbf{Q\Lambda Q^T}$, $\tilde{\mathbf{K}}(n) = \mathbf{Q^T K}(n)\mathbf{Q}$, $\boldsymbol{\lambda} = diag(\mathbf{\Lambda})$ and $\tilde{\mathbf{k}}(n) = diag\left(\tilde{\mathbf{K}}(n)\right)$. In the following, $\tilde{\mathbf{k}}(n)$ will be evaluated using the SDE method.

Following the procedure outlined in [8], the ODE corresponding to (3) is obtained as

$$\frac{d\mathbf{V}(t)}{dt} = E\left[\left.\frac{e(n)\mathbf{X}(n)}{\mathbf{X}^T(n)\mathbf{X}(n)}\right|_{\mathbf{V}(n)=\mathbf{V}(t)}\right] = -\frac{1}{2}\frac{d}{d\mathbf{V}(t)}E\left[e^2(n)\right] \tag{6}$$

where $\mathbf{V}(t)$ is the solution with $t = \mu n$ and μ is assumed sufficiently small compared to the stability limit (μ=2). It can be shown that (3) converges weakly to a stable equilibrium point $\mathbf{V}^*$.

It is shown in [5] that the process $\chi_\mu(t) = \left(\mathbf{V}(n) - \mathbf{V}(t)\right)/\sqrt{\mu}$ converges weakly as $\mu \to 0$ to the Gaussian process $\chi(t)$ which is the solution of the following linear SDE:

$$d\chi(t) = \mathbf{F}\left(\mathbf{V}(t)\right)\chi(t)dt + \mathbf{R}^{1/2}\left(\mathbf{V}(t)\right)d\mathbf{\Gamma}(t) \tag{7}$$

where $\mathbf{\Gamma}(t)$ is a Brownian motion, and $\mathbf{F}\left(\mathbf{V}(t)\right)$ and $\mathbf{R}\left(\mathbf{V}(t)\right)$ are defined as follows:

$$\mathbf{F}\left(\mathbf{V}(t)\right) = \frac{d}{d\mathbf{V}(t)}E\left[\left.\frac{e(n)\mathbf{X}(n)}{\mathbf{X}^T(n)\mathbf{X}(n)}\right|_{\mathbf{V}(n)=\mathbf{V}(t)}\right] \tag{8}$$

$$\mathbf{R}\left(\mathbf{V}(t)\right) = \sum_{n \in Z} \operatorname{cov}\left(\left.\frac{e(n)\mathbf{X}(n)}{\mathbf{X}^T(n)\mathbf{X}(n)}\right|_{\mathbf{V}(n)=\mathbf{V}(t)}, \frac{e(0)\mathbf{X}(0)}{\mathbf{X}^T(0)\mathbf{X}(0)}\right) \tag{9}$$

To analyze $\chi(t)$ near the equilibrium point $\mathbf{V}^*$, the time-variant equation (7) can be replaced by the time-invariant equation

$$d\chi(n) = \mu\mathbf{F}\left(\mathbf{V}^*\right)\chi(n)dt + \mu\mathbf{R}^{1/2}\left(\mathbf{V}^*\right)d\mathbf{\Gamma}(n) \tag{10}$$

If $\mathbf{F}(\mathbf{V})$ is independent of $\mathbf{V}$, then the SDE is linear time-invariant even away from $\mathbf{V}^*$ [7]. It will be shown that this is the case in the present analysis.

Using Itô calculus it can be shown that [5]:

$$\begin{aligned} \frac{d}{dn}E\left[\chi(n)\chi^T(n)\right] &= \mu\mathbf{F}\left(\mathbf{V}^*\right)E\left[\chi(n)\chi^T(n)\right] \\ &+ \mu E\left[\chi(n)\chi^T(n)\right]\mathbf{F}^T\left(\mathbf{V}^*\right) + \mu^2\mathbf{R}\left(\mathbf{V}^*\right) \end{aligned} \tag{11}$$

Considering that $\mathbf{V}^* = 0$ in our problem,

$$E\left[\chi(n)\chi^T(n)\right] = E\left[\mathbf{V}(n)\mathbf{V}(n)^T\right] = \mathbf{K}(n)$$

and (11) can be rewritten as:

$$\begin{aligned} \frac{d}{dn}\tilde{\mathbf{K}}(n) &= \mu\mathbf{Q}^T\mathbf{F}\left(\mathbf{V}^*\right)\mathbf{K}(n)\mathbf{Q} \\ &+ \mu\mathbf{Q}^T\mathbf{K}(n)\mathbf{F}^T\left(\mathbf{V}^*\right)\mathbf{Q} + \mu^2\mathbf{Q}^T\mathbf{R}\left(\mathbf{V}^*\right)\mathbf{Q} \end{aligned} \tag{12}$$

To determine the solution of (12), $\mathbf{F}\left(\mathbf{V}^*\right)$ must be evaluated. From (8), it can be easily shown that

$$E\left[\left.\frac{e(n)\mathbf{X}(n)}{\mathbf{X}^T(n)\mathbf{X}(n)}\right|_{\mathbf{V}(n)=\mathbf{V}(t)}\right] = -E\left[\frac{\mathbf{X}(n)\mathbf{X}^T(n)}{\mathbf{X}^T(n)\mathbf{X}(n)}\right]\mathbf{V}(t) \tag{13}$$

where

$$e(n) = -\mathbf{X}^T(n)\mathbf{V}(n) + z(n) \tag{14}$$

To evaluate (13), approximations are necessary. A reasonable approximation is [4]:

$$\begin{aligned} E\left[\frac{\mathbf{X}(n)\mathbf{X}^T(n)}{\mathbf{X}^T(n)\mathbf{X}(n)}\right] &\approx E\left[\frac{1}{\mathbf{X}^T(n)\mathbf{X}(n)}\right]E\left[\mathbf{X}(n)\mathbf{X}^T(n)\right] \\ &= E\left[\frac{1}{\mathbf{X}^T(n)\mathbf{X}(n)}\right]\mathbf{R}_{xx} \end{aligned} \tag{15}$$

Now, for $\mathbf{X}(n)$ Gaussian, $\mathbf{X}^T(n)\mathbf{X}(n)$ is assumed to follow as a Chi-square distribution with N degrees of freedom. With these approximations (15) becomes [4]:

$$E\left[\frac{\mathbf{X}(n)\mathbf{X}^T(n)}{\mathbf{X}^T(n)\mathbf{X}(n)}\right] \approx \frac{1}{(N-2)r_0}\mathbf{R}_{xx} \tag{16}$$

with $r_0 = \mathbf{R}_{xx}(1,1)$

$\mathbf{F}$ is then approximated as follows:

$$\mathbf{F}\left(\mathbf{V}(t)\right) \approx -\frac{d}{d\mathbf{V}(t)}\left[\frac{1}{(N-2)\sigma_x^2}\mathbf{R}_{xx}\mathbf{V}(t)\right] = -\frac{\mathbf{R}_{xx}}{(N-2)r_0} \tag{17}$$

Assuming $\mathbf{R}_{xx}$ positive define, (12) has a solution:

$$\mathbf{K}(n) = e^{\mu\mathbf{F}n}\left[\mathbf{K}(0) - \mathbf{K}(\infty)\right]e^{\mu\mathbf{F}^T n} + \mathbf{K}(\infty) \tag{18}$$

where $\mathbf{K}(\infty)$ is the solution of the Lyapunov equation:

$$\mathbf{F}\mathbf{K}(\infty)+\mathbf{K}(\infty)\mathbf{F}^T=-\mu\mathbf{R} \tag{19}$$

Using (17) and the definition of $\tilde{\mathbf{K}}(n)$, (12) can be significantly simplified to:

$$\frac{d}{dn}\tilde{k}_i(n)=-\frac{2\mu\lambda_i}{(N-2)}\tilde{k}_i(n)+\mu^2\tilde{r}_i,\quad i=1,\ldots,N \tag{20}$$

where $\tilde{\mathbf{R}}=\mathbf{Q}^{\mathrm{T}}\mathbf{R}\left(\mathbf{V}^*\right)\mathbf{Q}$ and $\tilde{\mathbf{r}}=diag\left(\tilde{\mathbf{R}}\right)$. Note that (20) is now a set of N independent equations. Its solution is given by:

$$\tilde{k}_i=e^{-\frac{2\mu\lambda_i}{(N-2)r_0}n}\left[\tilde{k}_i(0)-\tilde{k}_i(\infty)\right]+\tilde{k}_i(\infty) \tag{21}$$

with $\tilde{k}_i(\infty)$ given by (assuming nonzero eigenvalues):

$$\tilde{k}_i(\infty)=\frac{\mu(N-2)r_0}{2\lambda_i}\tilde{r}_{ii} \tag{22}$$

The next step is then to calculate $\mathbf{R}$ to obtain $\tilde{r}_i$ in (22). Using the approximations

$$\begin{aligned}&E\left\{\frac{z(0)\mathbf{X}(0)\mathbf{X}^T(0)z(0)}{\mathbf{X}^T(0)\mathbf{X}(0)\mathbf{X}^T(0)\mathbf{X}(0)}\right\}\\ &\approx E\left\{\frac{1}{\mathbf{X}^T(0)\mathbf{X}(0)\mathbf{X}^T(0)\mathbf{X}(0)}\right\}E\left\{\mathbf{X}(0)\mathbf{X}^T(0)\right\}E\left\{z(0)z(0)\right\}\\ &\approx\frac{\mathbf{R}_{xx}\sigma_z^2}{(N-2)(N-4)\mathbf{r}_0^2}\end{aligned} \tag{23}$$

and

$$\begin{aligned}&E\left\{\frac{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{V}(t)\mathbf{V}^T(t)\mathbf{X}(0)\mathbf{X}^T(0)}{\mathbf{X}^T(n)\mathbf{X}(n)\mathbf{X}^T(0)\mathbf{X}(0)}\right\}\\ &\approx E\left\{\frac{\mathbf{X}(n)\mathbf{X}^T(n)}{\mathbf{X}^T(n)\mathbf{X}(n)}\right\}\mathbf{V}(t)\mathbf{V}^T(t)E\left\{\frac{\mathbf{X}(0)\mathbf{X}^T(0)}{\mathbf{X}^T(0)\mathbf{X}(0)}\right\}\end{aligned} \tag{24}$$

One obtains:

$$\mathbf{R}\approx\frac{\mathbf{R}_{xx}\sigma_z^2}{(N-2)(N-4)\mathbf{r}_0^2} \tag{25}$$

Equations (5), (21) (22) and (25) provide the analytical model for the MSE of the NLMS algorithm. This model provides results equivalent to the model in [4] for small step sizes. Fig. 1 shows the theoretical and simulated results for $x(n)$ and unity variance autoregressive process of order 2 with coefficients –0.3 and 0.6 (eigenvalue spread of 19). N=20, μ = 0.1. The measurement noise is white Gaussian with variance equal to 10^{-6} and $\left\|\mathbf{V}(0)\right\|^2=74$.

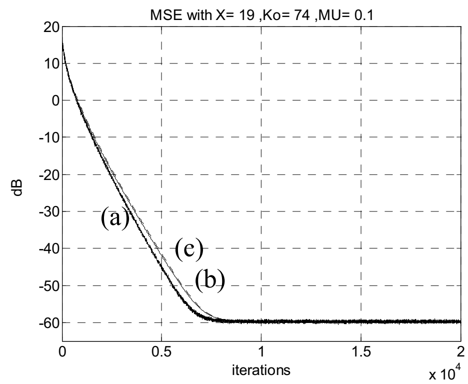

Fig. 1: Comparisons between (a) simulations, (b) model [4] (---) and (c) the new one (—) with $\mu=0.1$, $\chi=19$ and $\mathrm{K}_0=74$.

For large step sizes, $\mathbf{K}(\infty)$ becomes larger and the error in (22) becomes significant due to the errors derived from the approximations made to determine $\mathbf{R}$. In the next section, a better model is derived for the steady-state weight-error correlation matrix.

3. STEADY-STATE WEIGHT ERROR CORRELATION MATRIX

An accurate estimation of $\mathbf{K}(\infty)$ is obtained using the approach used in [4]. Post-multiplying (3) by its transpose, taking the expected value and neglecting the statistical dependence of *x(n)* and **V***(n)* leads to the recursive expression described in [4, Eq. (13)]. In evaluating the necessary expected values, a better approximation can be obtained if the following relation obtained by simple integration is used:

$$E\left\{\left[\mathbf{x}^T(n)\mathbf{x}(n)\right]^2\right\}=N^2r_0^2+2\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}r_{j-i}^2 \tag{26}$$

where $r_{j-i}=E\left\{x(n-i)x(n-j)\right\}$ is the correlation between delayed samples of the input signal. Using (26) in [4, Eq. (13)], results

$$\begin{aligned}\mathbf{K}(n+1)&=\mathbf{K}(n)-d\mu\left[\mathbf{K}(n)\mathbf{R}_{xx}+\mathbf{R}_{xx}\mathbf{K}(n)\right]\\ &+c\mu^2\sigma_z^2\mathbf{R}_{xx}+b\mu^2\left[2\mathbf{R}_{xx}\mathbf{K}(n)\mathbf{R}_{xx}+tr\left\{\mathbf{R}_{xx}\mathbf{K}(n)\right\}\mathbf{R}_{xx}\right]\end{aligned} \tag{27}$$

where

$$\begin{cases}b=\left[N^2r_0^2+2\displaystyle\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}r_{j-i}^2\right]^{-1}\\ c=\left[r_0^2(N-2)(N-4)\right]^{-1}\\ d=\left[(N-2)r_0\right]^{-1}\end{cases} \tag{28}$$

Pre and post-multiplying (27) by $\mathbf{Q}^{\mathrm{T}}$ and $\mathbf{Q}$ and taking only the main diagonal leads to:

$$\tilde{\mathbf{k}}(n+1)=\mathbf{f}\,\tilde{\mathbf{k}}(n)+g\boldsymbol{\lambda} \tag{29}$$

where

$$\begin{cases} \mathbf{f} = \mathbf{I} - 2\mu d \Lambda + 2\mu^2 b \Lambda\Lambda + \mu^2 b \lambda\lambda^{\mathrm{T}} \\ g = \mu^2 c \sigma_z^2 \end{cases} \quad (30)$$

which has the closed form solution

$$\tilde{\mathbf{k}}(n) = \mathbf{f}^n \tilde{\mathbf{k}}(0) + g \sum_{i=0}^{n-1} \mathbf{f}^i \cdot \lambda \quad (31)$$

Assuming convergence of (31) so that $\lim_{n\to\infty} \tilde{\mathbf{k}}(n) = \lim_{n\to\infty} \tilde{\mathbf{k}}(n+1)$, then

$$\lim_{n\to\infty} \tilde{\mathbf{k}}(n) = \tilde{\mathbf{k}}(\infty) = \frac{-\mu^2 c \sigma_z^2}{1 + \mu^2 b \lambda^{\mathrm{T}} \mathbf{A}^{-1} \lambda} \cdot \mathbf{A}^{-1} \lambda \quad (32)$$

where $\mathbf{A} = -2\mu d \Lambda + 2\mu^2 b \Lambda\Lambda$.

Equation (32) describes the steady state value of the main diagonal elements of the rotated weigh error correlation matrix. It is then possible to use this result in (21) to determine the algorithm behavior for large μ without depending on an accurate evaluation of **R**.

Figures 2 and 3 show representative examples of the proposed model behavior for μ=1. The input signal is the same as in Fig. 1.

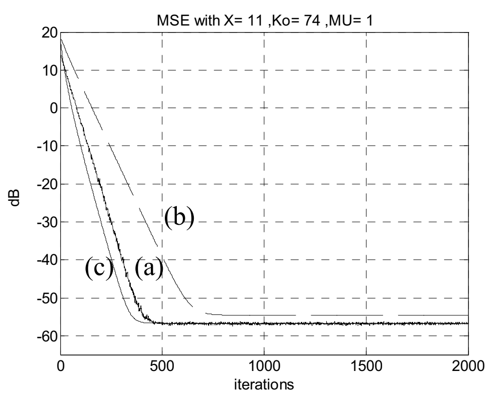

Fig. 2: Comparisons between (a) simulations, (b) model [4] (---) and (c) the new one (—) with $\mu = 1$, $\chi = 11$ and $K_0 = 74$.

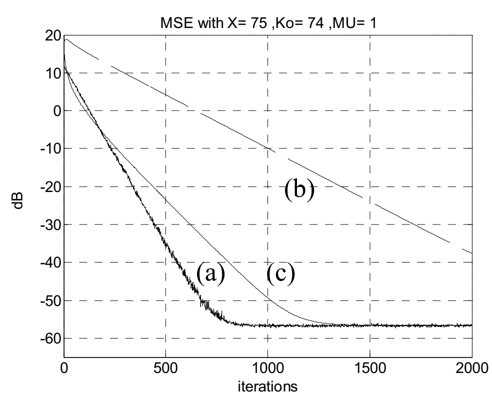

Fig. 3: Comparisons between (a) simulations, (b) model [4] (---) and (c) the new one (—) with $\mu = 1$, $\chi = 75$ and $K_0 = 74$.

4. CONCLUSION

A new statistical analysis of the NLMS algorithm has been presented. The new analysis uses the SDE to determine a closed form solution for the weight-error update equation. The solution has been corrected for the case of medium or large step sizes through an accurate estimation of the steady-state behavior of the weight-error correlation matrix. The new analytical model derived improves the results obtained by the existing models.

5. ACKNOWLEDGMENTS

Thanks to CNPq (Brazilian Ministry of Science and Technology), Embraco (Compressor Company) and RT-LAB (Real-Time Software Company) for their founding.

11. REFERENCES

[1] N.J Bershad, “Analysis of the Normalized LMS Algorithm with Gaussian inputs”, *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 34, August-1986.

[2] M. Tarrab and A. Feuer, “Convergence and Performance Analysis of the Normalized Algorithm with uncorrelated Gaussian data”, *IEEE Trans. on Information and Theory*, vol. 34, July-1988.

[3] D. Slock, "On the Convergence Behavior of the LMS and the Normalized LMS Algorithms", *IEEE Trans. on Signal Processing*, vol. 41, n. 9,setember -1993.

[4] Márcio H. Costa and José Carlos M. Bermudez, "An improved model for the Normalized LMS algorithm with Gaussian inputs and large number of coefficients", *ICASSP-2002*, vol. 2 - 2002.

[5] Albert Benveniste, “A Measure of the tracking Capability of Recursive Stochastic Algorithms with Constant Gains*”, IEEE Trans. on Automatic Control*, vol. 27 n. 3, June-1982.

[6] Lennart Ljung and T. Soderström, “Theory and Practice of Recursive Identification*”, The MIT Press, third edition*, 1986.

[7] Howard Fan, “A structural view of a asymptotic convergence speed of adapting IIR filtering algorithms-Part II finite precision”, *IEEE Trans. on Signal Processing*, vol. 45, n. 6, June-1997.

[8] Phillip A. Regalia, “Adaptive IIR Filtering in Signal Processing and Control*”, Marcel Dekker, Inc*. 1995.

[9] H.J. Kushner and H. Huang, ”Asymptotic properties of stochastic approximations with constant coefficients, ” *SIAM J. Contr. Optim.*, January-1981

[10] C. Samson and V. Reddy, “Fixed point error analysis of the normalized ladder algorithm*”, IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 31 n. 5 , October-1983.

[11] S. Haykin, “Adaptive Filter Theory*”, Prentice-Hall, fourth edition*, 2002.