NEW INSIGHTS ON THE NOISE CONSTRAINED LMS ALGORITHM

José Gil F. Zipf, 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-mails: {gil, orlando, seara}@linse.ufsc.br

ABSTRACT

In this work, a stochastic model for the mean weight behavior and the learning curve of the noise constrained least-mean-square (NCLMS) algorithm is presented. The proposed model is simpler than that recently presented in the open literature. The main feature of this algorithm is that it takes into account the additive noise variance in the mean-square error (MSE) minimization process. As a result, some additional control parameters are included in the adaptive algorithm, affecting the convergence behavior of the algorithm. Then, some hints regarding these parameter settings for algorithm stability are also given. Through numerical simulations the accuracy of the proposed model is confirmed.

***Index Terms*—**Adaptive estimation, adaptive filters, adaptive signal processing, least-mean-square methods.

1. INTRODUCTION

The main advantages of stochastic gradient algorithms, such as the least-mean-square (LMS) algorithm, are their simplicity as well as the fact of not requiring previous knowledge of the process statistics [1]. This fact makes the LMS algorithm the most widely used for adaptive filtering applications. In the standard LMS implementation, the step-size parameter is fixed. However, variable step-size LMS (VSLMS) algorithms have also been proposed to improve some characteristics of the standard LMS one, such as faster convergence and smaller misadjustment [2], [3]. In a recent work, a new VSLMS algorithm was proposed aiming at improving the convergence characteristics regarding the measurement noise, giving rise to the noise-constrained LMS (NCLMS) algorithm [4]. The NCLMS algorithm is well suited for system identification problems involving linear finite impulse response (FIR) channels with additive white Gaussian noise. For instance, in wireless communications, in which a combination of intersymbolic interference (ISI) and fading may occur, an efficient identification to track fast signal variations is required [4]. One can then obtain some knowledge of the channel noise variance, by periodically estimating the noise power within intervals with no signal transmission, and to use this information to improve the adaptive algorithm performance [4], [5]. This is the basic idea behind the NCLMS algorithm.

This paper contributes with new insights concerning the NCLMS algorithm. Specifically, we present a stochastic model for the mean weight behavior and the learning curve. The proposed analytical model for the mean weight behavior is simpler than that presented in [4], since here it is just based on first order statistics. In addition, some hints regarding the stability and parameter settings for the algorithm in question are presented. Through simulation results, we also show the algorithm performance under an imperfect estimate condition of the noise power.

This work was supported in part by the Brazilian National Research Council for Scientific and Technological Development (CNPq) and Studies and Projects Funding Body (FINEP).

2. NCLMS ALGORITHM

By considering a system identification problem, the channel output is given by

$$d(n) = \mathbf{w}_\mathrm{o}^\mathrm{T}\mathbf{x}(n) + \eta(n) \tag{1}$$

where $\mathbf{x}(n) = [x(n)\ x(n-1)\cdots x(n-N+1)]^\mathrm{T}$ is the input vector, with the process $\{x(n)\}$ having a zero-mean and variance σ_x^2. Variable $\eta(n)$ is a zero-mean stationary noise process with variance σ_η^2 and uncorrelated with any other signal in the system. The mean-square error (MSE) of the process is

$$\varepsilon(\mathbf{w}) = E[e^2(n)] = E\{[d(n) - \mathbf{w}^\mathrm{T}\mathbf{x}(n)]^2\}\ . \tag{2}$$

By minimizing the MSE with respect to $\mathbf{w}$, this results in the optimum weight vector $\mathbf{w}_\mathrm{o} = \mathbf{R}^{-1}\mathbf{p}$, where $\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^\mathrm{T}(n)]$ is the autocorrelation matrix of the input vector and $\mathbf{p} = E[\mathbf{x}(n)d(n)]$ is the cross-correlation vector between the desired signal and the input vector. Now, we shall include the channel noise variance σ_η^2 in the minimization process of (2), which results in a constrained optimization problem of minimizing $\varepsilon(\mathbf{w})$ subject to $\varepsilon(\mathbf{w}) = \sigma_\eta^2$ [4]. Thus,

$$\varepsilon_1(\mathbf{w},\lambda) = \varepsilon(\mathbf{w}) + \lambda[\varepsilon(\mathbf{w}) - \sigma_\eta^2]\ . \tag{3}$$

Critical values of (3) are $\mathbf{w} = \mathbf{w}_\mathrm{o}$ for any λ. According to [4], this condition could result in some convergence problems. To avoid such problems, an additional penalty term $\gamma\lambda^2$ (with $\gamma > 0$) is subtracted in (3), resulting in the augmented expression

$$\varepsilon_2(\mathbf{w},\lambda) = \varepsilon(\mathbf{w}) + \lambda[\varepsilon(\mathbf{w}) - \sigma_\eta^2] - \gamma\lambda^2. \tag{4}$$

Now, the unique critical value of (4) is $\mathbf{w} = \mathbf{w}_\mathrm{o}$ for $\lambda = 0$. The stochastic approximation given by the Robbins-Monro algorithm [6] is used to find the roots in (4), resulting in the following recursive expressions:

$$\alpha(n) = \alpha_\mathrm{f}[1 + \gamma\lambda(n)] \tag{5}$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \alpha(n)e(n)\mathbf{x}(n) \tag{6}$$

and

$$\lambda(n+1) = \lambda(n) + \beta\{\frac{1}{2}[e^2(n) - \sigma_\eta^2] - \lambda(n)\}\ . \tag{7}$$

Factor $\alpha(n)$ is the variable step-size parameter, composed of two terms, a constant α_f and a time-varying term $\lambda(n)$. Variables β and γ are positive constants.

3. NCLMS MODEL FOR MEAN WEIGHT

Expression (7) can be rewritten as

$$\lambda(n)=\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}\frac{\beta}{2}[e^2(k)-\sigma_\eta^2]. \qquad (8)$$

By substituting (8) into (5), we obtain

$$\alpha(n)=\alpha_f+\frac{\alpha_f\beta\gamma}{2}\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}[e^2(k)-\sigma_\eta^2]. \qquad (9)$$

Then, defining $\delta=0.5\alpha_f\beta\gamma$, we get

$$\alpha(n)=\alpha_f+\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}[e^2(k)-\sigma_\eta^2]. \qquad (10)$$

Now, substituting (10) into (6), we have

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\{\alpha_f+\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}[e^2(k)-\sigma_\eta^2]\}e(n)\mathbf{x}(n) \qquad (11)$$

and by redefining the time-varying term of the step size as

$$\mu(n)=\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}[e^2(k)-\sigma_\eta^2] \qquad (12)$$

we obtain the final expression for the weight update equation

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\alpha_f e(n)\mathbf{x}(n)+\mu(n)e(n)\mathbf{x}(n). \qquad (13)$$

Then, taking the expected value of both sides of (13), one has

$$E[\mathbf{w}(n+1)]=E[\mathbf{w}(n)]+\alpha_f E[e(n)\mathbf{x}(n)]+E[\mu(n)e(n)\mathbf{x}(n)] \qquad (14)$$

$$E[\mathbf{w}(n+1)]=\underbrace{E[\mathbf{w}(n)]+\alpha_f\{\mathbf{p}-\mathbf{R}E[\mathbf{w}(n)]\}}_{\text{standard LMS}}+E[\mu(n)e(n)\mathbf{x}(n)]. \qquad (15)$$

Note that (15) is identical to the mean weight value expression of the standard LMS algorithm except for the last r.h.s term, which can be written as follows:

$$\begin{aligned}E[\mu(n)e(n)\mathbf{x}(n)]&=\delta E[\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}e^2(k)e(n)\mathbf{x}(n)]\\&\quad-\delta E[\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}\sigma_\eta^2 e(n)\mathbf{x}(n)]\\&=E[\mathbf{a}(n)]-E[\mathbf{b}(n)].\end{aligned} \qquad (16)$$

By evaluating each term of (16), we get

$$\begin{aligned}E[\mathbf{b}(n)]&=\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}\sigma_\eta^2 E[e(n)\mathbf{x}(n)]\\&=\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}\sigma_\eta^2\{\mathbf{p}-\mathbf{R}E[\mathbf{w}(n)]\}\end{aligned} \qquad (17)$$

and

$$E[\mathbf{a}(n)]=\delta E[\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}e^2(k)e(n)\mathbf{x}(n)]. \qquad (18)$$

Then, writing $E[\mathbf{a}(n)]=E[c(n)e(n)\mathbf{x}(n)]$ with

$$c(n)=\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}e^2(k) \qquad (19)$$

we readily note that $c(n)$ is statistically independent of $e(n)$, resulting in the following relation:

$$E[c(n)e(n)\mathbf{x}(n)]=E[c(n)]E[e(n)\mathbf{x}(n)]. \qquad (20)$$

Now, considering weight-error vector

$$\mathbf{v}(n)=\mathbf{w}(n)-\mathbf{w}_o \qquad (21)$$

and expressing $E[e^2(k)]$ as a function of **R** and $\mathbf{v}(n)$, we obtain

$$E[c(n)]=\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}(\sigma_\eta^2+\mathrm{tr}\{\mathbf{R}E[\mathbf{v}(k)\mathbf{v}^T(k)]\}). \qquad (22)$$

Since

$$E[e(n)\mathbf{x}(n)]=\mathbf{p}-\mathbf{R}E[\mathbf{w}(n)] \qquad (23)$$

and substituting (22) and (23) into (20), we get

$$\begin{aligned}E[c(n)]E[e(n)\mathbf{x}(n)]&=\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}\sigma_\eta^2\{\mathbf{p}-\mathbf{R}E[\mathbf{w}(n)]\}\\&+\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}\mathrm{tr}\{\mathbf{R}E[\mathbf{v}(k)\mathbf{v}^T(k)]\}\{\mathbf{p}-\mathbf{R}E[\mathbf{w}(n)]\}.\end{aligned} \qquad (24)$$

The first term in the r.h.s. of (24) is $E[\mathbf{b}(n)]$ given by (17). Then, substituting (24) into (16), we obtain

$$E[\mu(n)e(n)\mathbf{x}(n)]=\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}\mathrm{tr}\{\mathbf{R}E[\mathbf{v}(k)\mathbf{v}^T(k)]\}\{\mathbf{p}-\mathbf{R}E[\mathbf{w}(n)]\}. \qquad (25)$$

Finally, from (25), the model expression for the mean weight behavior is given by

$$\begin{aligned}E[\mathbf{w}(n+1)]&=E[\mathbf{w}(n)]+\alpha_f\{\mathbf{p}-\mathbf{R}E[\mathbf{w}(n)]\}\\&+\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}\mathrm{tr}\{\mathbf{R}E[\mathbf{v}(k)\mathbf{v}^T(k)]\}\{\mathbf{p}-\mathbf{R}E[\mathbf{w}(n)]\}.\end{aligned} \qquad (26)$$

Note that (26) depends on the second-order moment of $\mathbf{w}(n)$ [4]. However, by using the approximation [7]

$$E[\mathbf{v}(n)\mathbf{v}^T(n)]\approx E[\mathbf{v}(n)]E[\mathbf{v}^T(n)] \qquad (27)$$

a simplified first-order model can be obtained as

$$\begin{aligned}E[\mathbf{w}(n+1)]&=E[\mathbf{w}(n)]+\alpha_f\{\mathbf{p}-\mathbf{R}E[\mathbf{w}(n)]\}\\&+\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}\mathrm{tr}(\mathbf{R}\{E[\mathbf{w}(k)]-\mathbf{w}_o\}\{E[\mathbf{w}(k)]-\mathbf{w}_o\}^T)\\&\times\{\mathbf{p}-\mathbf{R}E[\mathbf{w}(n)]\}.\end{aligned} \qquad (28)$$

4. LEARNING CURVE MODEL

In this section, the learning curve model of the NCLMS algorithm is determined, which is given by

$$J(n)=\sigma_\eta^2+\mathrm{tr}[\mathbf{R}\mathbf{K}(n)] \qquad (29)$$

with $J(n)=E[e^2(n)]$ and $\mathbf{K}(n)=E[\mathbf{v}(n)\mathbf{v}^T(n)]$ representing the covariance matrix of the weight-error vector. By expressing (13) in terms of the weight-error vector, determining the product $\mathbf{v}(n)\mathbf{v}^T(n)$, taking the expected value of both sides of the resulting expression, and considering $\mu(n)$ independent of both $\mathbf{x}(n)$ and $\mathbf{v}(n)$, we get

$$\begin{aligned}\mathbf{K}(n+1)&=\mathbf{K}(n)-E[\alpha_f+\mu(n)][\mathbf{R}\mathbf{K}(n)+\mathbf{K}(n)\mathbf{R}]\\&+E\{[\alpha_f+\mu(n)]^2\}\{\mathbf{R}\mathrm{tr}[\mathbf{R}\mathbf{K}(n)]+2\mathbf{R}\mathbf{K}(n)\mathbf{R}\}\\&+E\{[\alpha_f+\mu(n)]^2\}\mathbf{R}\sigma_\eta^2.\end{aligned} \qquad (30)$$

Finally, we can rewrite (30) as

$$\begin{aligned}\mathbf{K}(n+1)&=\mathbf{K}(n)-\{\alpha_f+E[\mu(n)]\}[\mathbf{R}\mathbf{K}(n)+\mathbf{K}(n)\mathbf{R}]\\&+\{\alpha_f^2+2\alpha_f E[\mu(n)]+E[\mu^2(n)]\}\{2\mathbf{R}\mathbf{K}(n)\mathbf{R}+\mathbf{R}\,\mathrm{tr}[\mathbf{R}\mathbf{K}(n)]\}\\&+\{\alpha_f^2+2\alpha_f E[\mu(n)]+E[\mu^2(n)]\}\mathbf{R}\sigma_\eta^2\end{aligned} \qquad (31)$$

with

$$E[\mu(n)]=\delta\sum_{k=0}^{n-1}(1-\beta)^{n-1-k}[J(k)-\sigma_\eta^2] \qquad (32)$$

and

$$E[\mu^2(n)] = \delta^2 \sum_{k=0}^{n-1} (1-\beta)^{2(n-1-k)} [J^2(k) - 2J(k)\sigma_\eta^2 + \sigma_\eta^4] . \quad (33)$$

5. ROBUSTNESS AND STABILITY CONSIDERATIONS

Two crucial points, related to practical applications of the NCLMS algorithm, are its robustness, regarding variations of the noise variance, and the stability limits for adjusting the parameters α_f, β, and γ. The NCLMS algorithm considers a constant noise variance σ_η^2. However, in practice it can be time-varying as well as a mismatch between real and estimate noise variance may occur. In Fig. 1, the algorithm misadjustment versus noise variance mismatch in percentage is shown. From that figure, a slight linear increase of the misadjustment under large variations of $\Delta\sigma_\eta^2\%$ is observed.

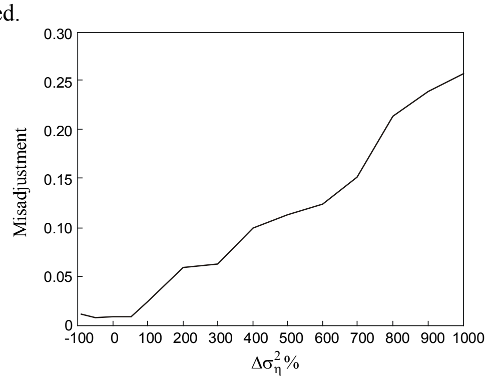

Fig. 1. NCLMS algorithm misadjustment versus $\Delta\sigma_\eta^2\%$.

Regarding the NCLMS algorithm stability, it is closely related to the dynamic behavior of $\alpha(n)$. Since the algorithm stability is determined from the recursive expression $\mathbf{K}(n)$, which also depends on $\alpha(n)$, the mathematics for such is very complex. In general, the stability of VSLMS algorithms is conditioned to a step variation within the standard LMS step-size limits [8], [9], which is given by

$$\alpha(n) < \frac{2}{3\mathrm{tr}[\mathbf{R}]} \quad (34)$$

Fig. 2 shows a typical curve of the $\alpha(n)$ evolution. The algorithm stability is ensured if the peak value α_{max} is lower than the allowed maximum step-size value. Notice that $\alpha(n)$ depends on the parameters α_f, β, and γ; thus, they must be chosen to guarantee an adequate value for α_{max}.

In this work, we heuristically investigate the behavior of α_{max} as a function of parameters α_f, β, and γ. We observe that for filters with a large number of taps, α_{max} remains constant if parameters α_f, β, and γ are adjusted considering the restriction $\delta = 0.5\alpha_f\beta\gamma$ equal to a constant. In addition, the curve describing α_{max} as a function of δ has a characteristic form independent of both input signal power σ_x^2 and filter tap number. We also observe from simulations that the upper value for δ, denoted as δ_{critic}, which guarantees algorithm stability, is a function of both input signal power and filter length. Thus,

$$\delta \ll \delta_{critic} = f(\sigma_x^2, N) . \quad (35)$$

In Section 6, some curves illustrating α_{max} as a function of δ as well as the value of δ_{critic} are presented.

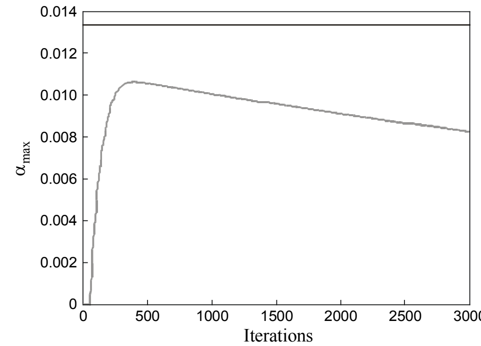

Fig. 2. Evolution of $\alpha(n)$ for a 50-tap filter.

6. SIMULATION RESULTS

In this section, we present some numerical simulations aiming to confirm the accuracy of the proposed model in describing the behavior of the NCLMS algorithm. In addition, some insights regarding algorithm stability and robustness are given.

Example 1: NCLMS modeling results

By considering a system identification problem, the used plant is $\mathbf{w}_o = [0.227\ 0.460\ 0.688\ 0.320\ 0.110]^T$. In this example, white and correlated Gaussian input signals with zero-mean and variance $\sigma_x^2 = 1$ are used. The correlated signal is obtained from an AR(1) process, given by $x(n) = ax(n-1) + u(n)$, where $a = 0.8$ and $u(n)$ is a white noise process such that $\sigma_x^2 = 1$. The variance of the additive noise is $\sigma_\eta^2 = 0.01$. In both cases, we adjust α_f, β, and γ according to [4]. Fig. 3 shows the results obtained from the Monte Carlo (MC) simulation (average of 200 independent runs) and the proposed model. From that figure, we notice a very good matching between simulation and prediction. The small mismatch in the transient phase in Fig. 3(a) is due to approximation (27). In Fig. 4, the correlated input case is depicted. Again, a very good matching between MC simulation and model prediction is verified.

Example 2: stability and robustness issues

In this example, numerical simulations verifying stability limits considering different parameter adjusting are given. In addition, the NCLMS algorithm misadjustment robustness under noise power variation is investigated. To verify the stability limit a 50-tap filter is used. From Section 5, (11), and (12) we observe that a stable operation of the NCLMS algorithm depends on the parameters α_f, β, and γ in a complex way. In particular, the convergence depends on α_f (for the fixed step size) as well as on $\delta = 0.5\alpha_f\beta\gamma$ [for the time-varying step-size term (12)]. Fig. 5 depicts a curve family of α_{max} versus δ, each of them is obtained by varying two parameters, while maintaining the remainder constant. Note from that figure that all obtained curves are tightly concentrated, meaning that for a certain δ there is an approximately single value for α_{max}. Also, for values of δ larger than δ_{critic}, no algorithm convergence is achieved.

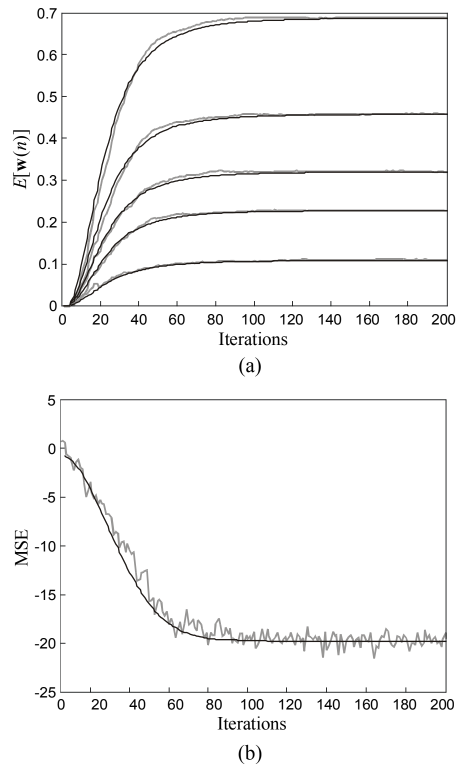

Fig. 3. Example 1. White input signal. (a) Mean weight behavior. (Gray lines) MC simulation. (Black lines) proposed model. (b) Learning curve. (Ragged line) MC simulation. (Black line) proposed model.

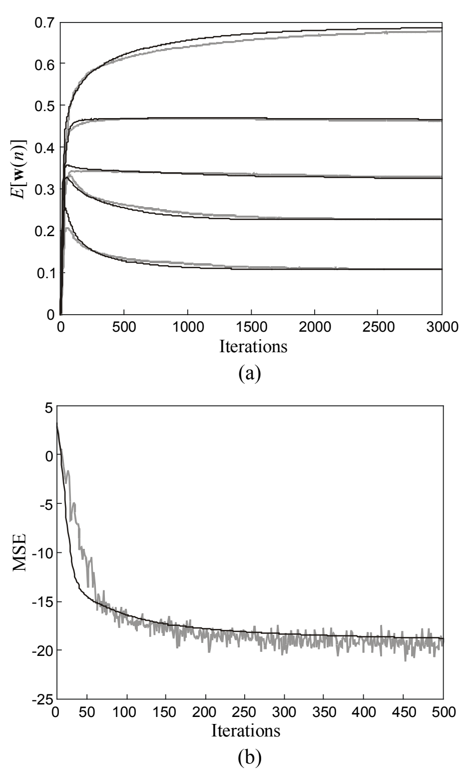

Fig. 4. Example 1. Correlated input signal. (a) Mean weight behavior. (Gray lines) MC simulation. (Black lines) proposed model. (b) Learning curve. (Ragged line) MC simulation. (Black line) proposed model.

The NCLMS robustness against variation in σ_η^2 is shown in Fig. 1. Note from that figure that for noise power variations around 50% the algorithm misadjustment is not significantly changed, as larger variations yield a slight increase. For instance, the misadjustment is increased ten times for a noise power variation of 400%, showing its robustness under an imperfect estimate of the noise power.

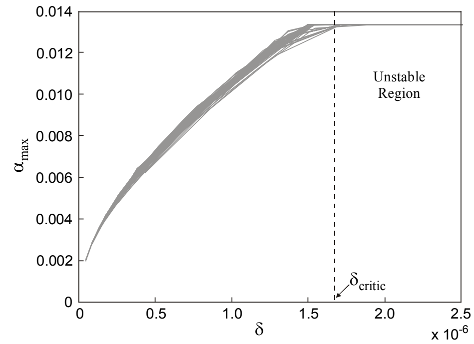

Fig. 5. $\alpha_{\max}$ versus $\delta = 0.5\alpha_f \beta\gamma$ for a 50-tap filter.

7. CONCLUSIONS

In this work, we present new model expressions for the first order mean weight model and the learning curve of the NCLMS algorithm. In addition, a brief investigation of algorithm robustness and stability is presented. It is also verified that the proposed model provides an accurate prediction for the NCLMS algorithm behavior. Since an analytical determination of the stability limits for the step size is a complex task, some heuristic investigations have been carried out, giving some insights for further analyses.

8. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 4th ed., Upper Saddle River, NJ: Prentice Hall, 2002.

[2] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1633-1642, July 1992.

[3] B. Farhang-Boroujeny, "Variable step size LMS algorithm – New developments and experiments," *IEE Proceedings – Vision, Image, Signal Process.*, vol. 141, no. 5, pp. 311-317, Oct. 1994.

[4] Y. Wei and S. B. Gelfand, "Noise-constrained least mean squares algorithm," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1961-1970, Sept. 2001.

[5] P. Kuosmanen, R. Bilcu, and C. Rusu, "A noise constrained VS-LMS algorithm," in *Proc. IEEE/AFCEA Eurocomm 2000 Information Systems for Enhanced Public Safety and Security*, Munich, Germany, May 2000, pp. 29–33.

[6] H. Robbins and S. Monro, "A stochastic approximation method," *Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400-407, Sept. 1951.

[7] N. J. Bershad, P. Celka, and J. M. Vesin, "Stochastic analysis of gradient adaptive identification of nonlinear systems with memory for Gaussian data and noisy input and output measurements," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 675-689, Mar. 1999.

[8] S. B. Gelfand, Y. Wei, and J. V. Krogmeier, "The stability of variable step-size LMS algorithms," *IEEE Trans. Signal Process.*, vol. 47, no. 12, pp. 3277-3288, Dec. 1999.

[9] T. Yang and B. Shahrrava, "Performance of variable step-size LMS algorithms for linear adaptive inverse control systems," in *Proc. IEEE Canadian Conf. on Electrical and Computer Engineering*, Saskatoon, Canada, May 2005, pp. 755-758.