THE NORMALIZED LMS ALGORITHM WITH DEPENDENT NOISE

Vítor H. Nascimento[*]

Departamento de Engenharia de Sistemas Eletrônicos
Escola Politécnica, Universidade de São Paulo
São Paulo, SP
vitor@lps.usp.br

[*] This work was supported by FAPESP under award no. 2000/09569-6.

ABSTRACT

We study the behavior of the NLMS algorithm when the "noise" (the optimal estimation error) $v(k)$ is uncorrelated, but *dependent* of the regressor sequence $\{\boldsymbol{x}_k\}$. This is the most general situation in applications, arising, for example, when linear estimation is applied to systems with nonlinearities.

We show that, unlike the LMS algorithm, NLMS computes biased estimates *even* if the step-size μ is small and $\{\boldsymbol{x}_k\}$ is iid, but $v(k)$ is not independent of $\boldsymbol{x}_k$. We provide expressions for the bias, the MSD, and the MSE, which are also valid for the case of correlated regressor sequences.

1. INTRODUCTION

In this paper we analyze the behavior of the normalized-LMS (NLMS) algorithm in the general case where both the regressor and noise sequences are correlated in time. Our expressions hold even when the noise and regressor are *not* independent one from the other. This includes the case of non-Gaussian variables, as well as Gaussian variables going through some kinds of nonlinearities. In this section we provide a few motivating examples, and compare our analysis with previous results in the literature.

1.1. Optimum linear estimation and the NLMS algorithm

Given zero-mean, stationary sequences $\{y(k) \in \mathbb{R}\}$ (*desired sequence*) and $\{\boldsymbol{x}_k \in \mathbb{R}^M\}$ (*regressor sequence*), the optimum linear estimator of $y(k)$ given $\boldsymbol{x}_k$ (the Wiener filter) is the solution of [3]

$$\boldsymbol{w}_* = \arg \min_{\boldsymbol{w} \in \mathbb{R}^M} \mathrm{E}\left(y(k) - \boldsymbol{w}^T \boldsymbol{x}_k\right)^2.$$

The optimum estimation error (*noise*) is

$$v(k) \triangleq y(k) - \boldsymbol{w}_*^T \boldsymbol{x}_k.$$

From the *orthogonality principle* of linear estimation, it results that $\mathrm{E}\, v(k)\boldsymbol{x}_k = \boldsymbol{0}$ (i.e., they are uncorrelated). Therefore, it is common in the adaptive filtering literature to assume that $\{y(k)\}$ and $\{\boldsymbol{x}_k\}$ are related by a linear model $y(k) = \boldsymbol{w}_*^T \boldsymbol{x}_k + v(k)$, with $\mathrm{E}\, v(k)\boldsymbol{x}_k = \boldsymbol{0}$.

If, in addition, all variables are Gaussian, it follows from the orthogonality principle that $v(k)$ and $\boldsymbol{x}_k$ are truly independent. Given this property of Gaussian variables, many works also assume that $v(k)$ and $\boldsymbol{x}_k$ are independent (*e.g.,* [2, 5]). Although this last assumption may be true in some settings, it does not hold in general. In fact, as we shall show in an example further on, $v(k)$ and $\boldsymbol{x}_k$ are *dependent* if there is a nonlinear relation between $y(k)$ and $\boldsymbol{x}_k$.

When the statistics of $y(k)$ and $\boldsymbol{x}_k$ are known, one may compute $\boldsymbol{w}_*$ from [3]

$$\boldsymbol{w}_* = R^{-1}\boldsymbol{p},$$

where

$$R \triangleq \mathrm{E}\, \boldsymbol{x}_k \boldsymbol{x}_k^T, \qquad \boldsymbol{p} \triangleq \mathrm{E}\, y(k)\boldsymbol{x}_k.$$

In practice, the covariance matrix R and the correlation vector $\boldsymbol{p}$ are unknown, and one uses an iterative algorithm such as the least mean squares (LMS) algorithm, normalized-LMS (NLMS), or recursive least squares algorithm (RLS) to find approximations to $\boldsymbol{w}_*$. It is well-known that all these algorithms compute unbiased estimates for $\boldsymbol{w}_*$ if $v(k)$ is independent of $\boldsymbol{x}_k$ (even if $\{\boldsymbol{x}_k\}$ is a correlated sequence) [5]. More is known about LMS: it may be biased when $\{\boldsymbol{x}_k\}$ is a correlated sequence and $v(k)$ is uncorrelated, but dependent on $\boldsymbol{x}_k$ [7]. This bias, however, decreases to zero as the *step-size* parameter μ of LMS decreases to zero. Therefore, for slow adaptation (small μ), LMS is approximately unbiased. One of the goals of this paper is to show that, unlike LMS, NLMS may have a nonzero bias even for $\mu \approx 0$ and iid $\{\boldsymbol{x}_k\}$, if $v(k)$ is dependent on $\boldsymbol{x}_k$. We shall give a simple example of this fact below, and later proceed to find an expression for the bias and for the steady-state mean-square

error (MSE) and mean-square deviation (MSD) of NLMS with correlated input sequences.

The normalized LMS algorithm (NLMS) computes estimates of the parameter vector $\boldsymbol{w}_*$ using the recursion [5]

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \frac{\mu}{a + \|\boldsymbol{x}_k\|^2} e(k)\boldsymbol{x}_k, \qquad (1)$$

where

$$e(k) \triangleq y(k) - \boldsymbol{w}_k^T \boldsymbol{x}_k,$$

and a is a (usually small) positive constant. For comparison, the LMS algorithm uses the recursion

$$\boldsymbol{w}_{k+1}^{\text{lms}} = \boldsymbol{w}_k^{\text{lms}} + \mu e(k)\boldsymbol{x}_k,$$

where we used the superscript "lms" to differentiate the weight estimates computed by NLMS and by LMS.

To see that the NLMS recursion may compute biased estimates, consider the following simple example: Assume that $M = 1$ (i.e., all variables are scalar), that the regressor sequence is composed of Gaussian variables with variance $R = \mathrm{E}\, x_k^2 = 1$, and that $y(k)$ is obtained by $y(k) = \mathrm{sat}(x_k)$, where $\mathrm{sat}(\cdot)$ is a saturation function.[1]

Define the weight error vector $\tilde{\boldsymbol{w}}_k \triangleq \boldsymbol{w}_* - \boldsymbol{w}_k$, subtract $\boldsymbol{w}_*$ from both sides of (1) and take expectations on both sides to obtain (recall that $\{x_k\}$ is iid)

$$\begin{aligned}\mathrm{E}\,\tilde{\boldsymbol{w}}_{k+1} = &\left(1 - \mu\,\mathrm{E}\left(\frac{x_k^2}{a + x_k^2}\right)\right)\mathrm{E}\,\tilde{\boldsymbol{w}}_k - \\ &- \mu\,\mathrm{E}\left(\frac{x_k}{a + x_k^2} v(k)\right).\end{aligned} \qquad (2)$$

We computed these expectations using Matlab's symbolic toolbox, obtaining (for $a = 10^{-3}$), $\mathrm{E}\left(\frac{x_k^2}{a+x_k^2}\right) = 0.9613$, and $\mathrm{E}\left(\frac{x_k}{a+x_k^2} v(k)\right) = 0.1114$, concluding that, in steady-state, the average weight error will be $\lim_{k\to\infty} \mathrm{E}\,\tilde{\boldsymbol{w}}_k = -\big(0.9613\big)^{-1} 0.1114 = -0.1159$, which is independent of the step-size μ. In Fig. 1 we plot the average $\tilde{\boldsymbol{w}}_k$, computed from $L = 500$ simulations with $\mu = 0.01$.

LMS, on the other hand, is unbiased: $\lim_{k\to\infty} \mathrm{E}\,\tilde{\boldsymbol{w}}_k^{\text{lms}} = R^{-1}\,\mathrm{E}\,\boldsymbol{x}_k v(k) = 0$.

Another situation in which the estimation error is not independent of the regressor is the following: Assume that

[1] The function $\mathrm{sat}(x)$ used in our simulations saturates at -2.8 for $x \le -2.8$, and at +3 for $x \ge +3$. For $-2.8 < x < 3$, $\mathrm{sat}(x)$ is a polynomial such that $\mathrm{sat}(-3) = -2.8$, $\mathrm{sat}(3) = 3$, and $\mathrm{sat}(0) = \mathrm{sat}'(-3) = \mathrm{sat}'(3) = 0$ (the superscript "$'$" stands for derivative). Our saturation function may be thought of as an approximation for the output of an amplifier whose positive and negative power supply voltages are not exactly equal. Note that, with the above values, $E\big(\mathrm{sat}(x)\big) \approx 7.95 \times 10^{-3}$ if $x \sim N(0,1)$ (i.e., Gaussian with zero mean and unit variance). This mean value was subtracted from the samples prior to the application of the adaptive algorithms.

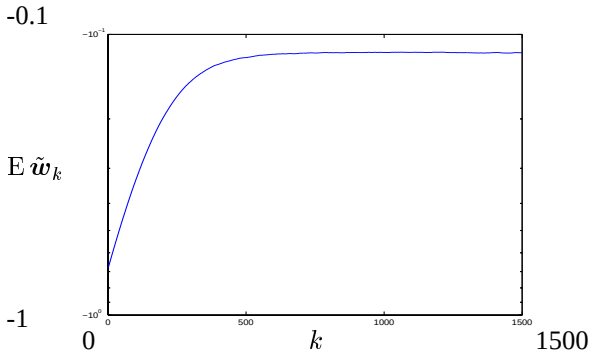

Figure 1: $\mathrm{E}\,\tilde{\boldsymbol{w}}_k$ *for the saturation example with* $\mu = 0.01$. *Average of 500 simulations.*

$y(k)$ can be described by an MA model

$$y(k) = \sum_{i=0}^{N-1} \alpha_i a(k-i) + n(k),$$

where N may be infinity, and the sequences $\{n(k) \in \mathbb{R}^M\}$ and $\{a(k) \in \mathbb{R}\}$ are random and mutually independent. Then, if we form $\boldsymbol{x}_k^T = \big[\,a(k)\ a(k-1)\ \ldots\ a(k-M+1)\,\big]$, when $M < N$ the estimation error will not, in general, be independent of $\boldsymbol{x}_k$.

1.2. Equivalence of LMS and NLMS for independent noise

There are recent results studying the behavior of NLMS when the sequence $\{\boldsymbol{x}_k\}$ is correlated, allowing in fact for very general correlation models (e.g., [2, 5]). However, these results assume that $\boldsymbol{x}_k$ is *independent* of $v(k)$, an assumption that is often not satisfied in practice if there are nonlinearities or unmodelled dynamics involved in the relation between $y(k)$ and $\boldsymbol{x}_k$, as our examples show.

We now argue that under the above assumption the normalized LMS algorithm is equivalent to the LMS algorithm applied to the bounded ("normalized") sequence

$$\boldsymbol{x}_k^{(n)} \triangleq \frac{\boldsymbol{x}_k}{\sqrt{a + \|\boldsymbol{x}_k\|^2}},$$

and with the equivalent desired sequence, noise, and estimation errors

$$\begin{aligned} y^{(n)}(k) &\triangleq \frac{y(k)}{\sqrt{a + \|\boldsymbol{x}_k\|^2}}, \\ v^{(n)}(k) &\triangleq \frac{v(k)}{\sqrt{a + \|\boldsymbol{x}_k\|^2}}, \\ e^{(n)}(k) &\triangleq y^{(n)}(k) - \boldsymbol{w}_k^T \boldsymbol{x}_k^{(n)}. \end{aligned}$$

If $v(k)$ and $\boldsymbol{x}_k$ are independent, it follows that

$$\mathrm{E}\, v^{(n)}(k)\boldsymbol{x}_k^{(n)} = \boldsymbol{0},$$

i.e., the normalized noise and regressor are uncorrelated. From the NLMS recursion, $y^{(n)}(k) = \boldsymbol{w}_*^T \boldsymbol{x}_k^{(n)} + v^{(n)}(k)$, and since $\mathrm{E}\, v^{(n)}(k)\boldsymbol{x}_k^{(n)} = \mathbf{0}$, the orthogonality principle implies that $\boldsymbol{w}_*^T \boldsymbol{x}_k^{(n)}$ is also the optimum linear estimator of $y^{(n)}(k)$ given $\boldsymbol{x}_k^{(n)}$.

In addition, if we rewrite the NLMS recursion (1) using the normalized variables, we obtain $\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \mu \boldsymbol{x}_k^{(n)} e^{(n)}(k)$, which is the LMS recursion applied to the normalized variables. This means that the theory developed to the study of the LMS algorithm with bounded regressor sequences and correlated noise can be used, without any modification, for the NLMS algorithm. This change of variables was used before to study the NLMS algorithm [10, 1, 2, 5], always assuming independence of noise and regressor. Therefore, the cited works only consider the NLMS algorithm when it is behaves in essentially the same as LMS. In this paper, we study the behavior of NLMS when an exact comparison with LMS is not possible.

2. NLMS WITH DEPENDENT NOISE

In this section we extend the analysis of NLMS to allow for dependent noise and regressor sequences. Our analysis is based on modifications of the averaging method used, for example, in [7], and described in [9].

2.1. Averaging results

Averaging methods provide a powerful means to analyse the performance and stability of adaptive algorithms under the assumption of sufficiently small step-sizes. There are many expositions on the subject (see, for example, [9, 4]). For this reason, we only list here to the stability and steady-state results that are needed in our derivation (see also [6]).

Consider an adaptive update (error equation) of the general form

$$\tilde{\boldsymbol{w}}_{k+1} = \tilde{\boldsymbol{w}}_k + \mu f(\boldsymbol{\xi}_k,\, \tilde{\boldsymbol{w}}_k),$$

where $\tilde{\boldsymbol{w}}_k$ is the error vector we want to minimize, and $\{\boldsymbol{\xi}_k\}$ is a stochastic sequence. Now define the averaged function f_{av} as

$$f_{av}(k,\, \tilde{\boldsymbol{w}}) = \mathrm{E}\, f(\boldsymbol{\xi}_k,\, \tilde{\boldsymbol{w}}),$$

where $\tilde{\boldsymbol{w}}$ is considered *constant* for the computation of the expected value. Define the *averaged* and the *partially averaged* systems,

$$\begin{aligned}
\tilde{\boldsymbol{w}}_{k+1}^{av} &= \tilde{\boldsymbol{w}}_k^{av} + \mu f_{av}\left(k, \tilde{\boldsymbol{w}}_k^{av}\right), \\
\tilde{\boldsymbol{w}}_{k+1}^{pav} &= \left[I + \mu \nabla_{\tilde{\boldsymbol{w}}} f_{av}(\mathbf{0})\right] \tilde{\boldsymbol{w}}_k^{pav} + \mu\big(f(k,\, \mathbf{0}) - f_{av}(k,\, \mathbf{0})\big)\,,
\end{aligned}$$

where $\nabla_{\tilde{\boldsymbol{w}}} f_{av}(\mathbf{0})$ denotes the value of the gradient of f_{av} (with respect to $\tilde{\boldsymbol{w}}$) at the origin.

The following result, proven in [9, Ch. 9], shows that if the step-size μ is sufficiently small, the original estimates $\tilde{\boldsymbol{w}}_k$ will remain close to the partially averaged estimates $\tilde{\boldsymbol{w}}_k^{pav}$, and that the steady-state covariance of $\tilde{\boldsymbol{w}}_k$ will be close to that of $\tilde{\boldsymbol{w}}_k^{pav}$. The theorem assumes that the $\{\boldsymbol{\xi}_k\}$ satisfy a *uniform mixing* property. Essentially, this condition says that the dependence of $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_j$ dies out as $|i-j|$ increases (see [8]).

Lemma 1 (Averaging result). *Consider the error equation and its averaged and partially-averaged forms as defined above, where the sequence $\{\boldsymbol{\xi}_k\}$ is uniform-mixing (see [9, p. 357]). Assume that (i) the origin, $\mathbf{0}$, is an exponentially-stable equilibrium point of the averaged system with decay rate $O(\mu)$, (ii) the gradient $\nabla_{\tilde{\boldsymbol{w}}} f_{av}(k,\, \tilde{\boldsymbol{w}})$ exists and is continuous at the origin, and (iii) there exists constants c_1 and c_2 such that, for any vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, the following Lipschitz conditions hold:*

$$\big\|f(k,\, \boldsymbol{a}) - f(k,\, \boldsymbol{b})\big\| \leq c_1 \big\|\boldsymbol{a} - \boldsymbol{b}\big\|,$$

$$\|\nabla_{\tilde{\boldsymbol{w}}} f(k,\, \boldsymbol{a}) - \nabla_{\tilde{\boldsymbol{w}}} f(k,\, \boldsymbol{b})\| \leq c_2 \big\|\boldsymbol{a} - \boldsymbol{b}\big\|.$$

Under these conditions, $\tilde{\boldsymbol{w}}_k$ obtained from the original error equation satisfies

$$\lim_{\mu\to 0} \sup_{k\geq 0} \frac{1}{\mu} \mathrm{E}\left\{ \big\| \tilde{\boldsymbol{w}}_k - \tilde{\boldsymbol{w}}_k^{pav} \big\|^2 \right\} = 0, \qquad \textit{and} \quad (3)$$

$$\lim_{\mu\to 0} \lim_{k\to\infty} \left(\frac{1}{\mu} \mathrm{E}\, \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^T \right) = \lim_{\mu\to 0} \lim_{k\to\infty} \left(\frac{1}{\mu} \mathrm{E}\, \tilde{\boldsymbol{w}}_k^{pav} \tilde{\boldsymbol{w}}_k^{pav,\,T} \right) . \quad (4)$$

♢

2.2. Averaging analysis of NLMS – Overview

Our goal is to use the above theorem to compute the MSE ($\mathrm{E}\, e(k)^2$) and the weight-error covariance matrix ($\mathrm{E}\, \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^T$) for small step-sizes μ. In order to use Theorem 1, we must:

1. Show that the theorem's conditions are satisfied for the problem of interest;
2. Compute the mean and mean-square behavior of $\tilde{\boldsymbol{w}}_k^{pav}$ (thus obtaining, via Theorem 1, the behavior of $\mathrm{E}\, \tilde{\boldsymbol{w}}_k$ and $\mathrm{E}\, \tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^T$ for small μ);
3. The theorem does not compare $\boldsymbol{x}_k^T \tilde{\boldsymbol{w}}_k$ and $\boldsymbol{x}_k^T \tilde{\boldsymbol{w}}_k^{pav}$, thus we must prove that

$$\begin{aligned}
&\frac{1}{\mu} \left| \mathrm{E}\, \boldsymbol{x}_k^T \tilde{\boldsymbol{w}}_k - \mathrm{E}\, \boldsymbol{x}_k^T \tilde{\boldsymbol{w}}_k^{pav} \right| \to 0 \qquad \text{and} \\
&\frac{1}{\mu} \left| \mathrm{E}\big(\boldsymbol{x}_k^T \tilde{\boldsymbol{w}}_k\big)^2 - \mathrm{E}\big(\boldsymbol{x}_k^T \tilde{\boldsymbol{w}}_k^{pav}\big)^2 \right| \to 0 \quad \text{as } \mu \to 0.
\end{aligned}$$

4. Study the mean and mean-square behavior of $\boldsymbol{x}_k^T \tilde{\boldsymbol{w}}_k^{pav}$ (finding $\mathrm{E}\, e(k)$ and $\mathrm{E}\, e(k)^2$ for small μ);

A difficulty in task 1 is that the NLMS averaged error equation (see below) may not have an equilibrium point at $\mathbf{0}$ when $v(k)$ is dependent of $\boldsymbol{x}_k$, contrary to one assumption of Theorem 1. This can be overcome by a change of variables

$$\boldsymbol{z}_k^{ave} \triangleq \tilde{\boldsymbol{w}}_k^{ave} - \tilde{\boldsymbol{w}}_\infty^{ave},$$

where

$$\tilde{\boldsymbol{w}}_\infty^{ave} = -{R^{(n)}}^{-1}\boldsymbol{b},$$

$$R^{(n)} \triangleq \left(\mathrm{E}\,\frac{\boldsymbol{x}_k\boldsymbol{x}_k^T}{a+\|\boldsymbol{x}_k\|^2}\right), \quad \text{and} \quad \boldsymbol{b} \triangleq \mathrm{E}\left(\frac{v(k)\boldsymbol{x}_k}{a+\|\boldsymbol{x}\|^2}\right), \tag{5}$$

Using this change of variables, we can show that Thm. 1 can be applied to the NLMS recursion, and prove the theorem below.

Theorem 1. *If $\{\boldsymbol{x}_k,\ v(k)\}$ is stationary and uniform-mixing, and if $\mathrm{E}\left(\frac{v(0)\boldsymbol{x}_0}{a+\|\boldsymbol{x}_0\|^2}\right) \neq 0$, the NLMS weight error $\tilde{\boldsymbol{w}}_k$ is such that, for $\mu \approx 0$,*

Average value:

$$\lim_{k\to\infty} \mathrm{E}\,\tilde{\boldsymbol{w}}_k \approx {R^{(n)}}^{-1}\,\mathrm{E}\left(\frac{v(0)\boldsymbol{x}_0}{a+\|\boldsymbol{x}_0\|^2}\right),$$

MSD:

$$\lim_{k\to\infty} \mathrm{E}\,\tilde{\boldsymbol{w}}_k^T\tilde{\boldsymbol{w}}_k \approx \left\|{R^{(n)}}^{-1}E\left(\frac{v(0)\boldsymbol{x}_0}{a+\|\boldsymbol{x}_0\|^2}\right)\right\|^2 + \mu\Omega,$$

where Ω satisfies

$$R^{(n)}\Omega + \Omega R^{(n)} = \sum_{n=0}^{\infty}\left[\mathrm{E}\left(\frac{v(0)v(n)\boldsymbol{x}_0\boldsymbol{x}_n^T}{(a+\|\boldsymbol{x}_0\|^2)(a+\|\boldsymbol{x}_n\|^2)}\right) - \right.$$
$$\left. - \mathrm{E}\left(\frac{v(0)\boldsymbol{x}_0}{a+\|\boldsymbol{x}_0\|^2}\right)\mathrm{E}\left(\frac{v(0)\boldsymbol{x}_0}{a+\|\boldsymbol{x}_0\|^2}\right)^T\right].$$

In addition, if $\sup_{k\geq 0}\|\boldsymbol{x}_k\| < \infty$,

MSE:

$$\lim_{k\to\infty} \mathrm{E}\, e^2(k) \approx \sigma_v^2 + \left\|R^{1/2}{R^{(n)}}^{-1}E\left(\frac{v(0)\boldsymbol{x}_0}{a+\|\boldsymbol{x}_0\|^2}\right)\right\|$$
$$+ \mu\big(R^{(n)}\otimes I + I\otimes R^{(n)}\big)^{-1}\,\mathrm{vec}(R)^T \times$$
$$\times \sum_{n=-\infty}^{\infty}\mathrm{vec}\left[\mathrm{E}\left(\frac{v(0)v(n)\boldsymbol{x}_n\boldsymbol{x}_0^T}{(a+\|\boldsymbol{x}_0\|^2)(a+\|\boldsymbol{x}_n\|^2)}\right) - \right.$$
$$\left. - \mathrm{E}\left(\frac{v(0)\boldsymbol{x}_0}{a+\|\boldsymbol{x}_0\|^2}\right)\mathrm{E}\left(\frac{v(0)\boldsymbol{x}_0}{a+\|\boldsymbol{x}_0\|^2}\right)^T\right],$$

where $\sigma_v^2 = \mathrm{E}\, v^2(0)$, $R^{(n)}$ is given above, $R = R^{1/2\,T}R^{1/2}$, $\mathrm{vec}(A)$ is the vector obtained by stacking the columns of A one at a time, and $A\otimes B$ is the Kronecker product of matrices A and B.

Proof: Since the proof is quite long, we only provide a brief outline of it. To obtain the above results, one starts with the partially-averaged NLMS recursion. Since this is a linear recursion on $\tilde{\boldsymbol{w}}_k^{pav}$, it is possible to write $\mathrm{E}\,\tilde{\boldsymbol{w}}_k^{pav}$ as a function of the initial condition, $\tilde{\boldsymbol{w}}_0^{pav}$ (assumed deterministic).

To compute the limits as $\mu \to 0$ and $k \to \infty$ of the MSE and MSD, we use the fact that the sequence $\{\boldsymbol{x}_k,\ v(k)\}$ is stationary and uniform-mixing, and thus for any functions $g(\cdot)$ and $h(\cdot)$,

$$\lim_{n\to\infty}\mathrm{E}\big[g(\boldsymbol{x}_0,v(0))\cdot h(\boldsymbol{x}_n,v(n))\big] =$$
$$= \mathrm{E}\,g(\boldsymbol{x}_0,v(0))\cdot\mathrm{E}\,h(\boldsymbol{x}_n,v(n)) =$$
$$= \mathrm{E}\,g(\boldsymbol{x}_0,v(0))\cdot\mathrm{E}\,h(\boldsymbol{x}_0,v(0)),$$

where the first equality follows from the uniform-mixing condition, and the second from stationarity. With this condition, we are able to find the most important terms (in an expansion around $\mu = 0$) of the MSE and MSD. ♢

3. VERIFICATION

We now verify the above results by means of a simulation example, in which the filters have dimension $M = 2$, the regression vector is

$$\boldsymbol{x}_k = \begin{bmatrix} p(k-1) & p(k)\end{bmatrix},$$

with $\{p(k)\}$ the output of a 3rd order digital Chebyshev filter with cut-off frequency at $\omega = \pi/10$, and with 3dB ripple in the pass-band whose input is an iid Gaussian random sequence with variance 16 and zero mean. The desired sequence is obtained from

$$y(k) = \mathrm{sat}(p(k)) - \mathrm{E}\,\mathrm{sat}(p(k)),$$

where $\mathrm{sat}(\cdot)$ is the saturation function defined before.

With these conditions, we averaged 500 runs of the normalized LMS algorithm with $\mu = 0.01$, and $a = 0.01$, obtaining the curves in fig. 2. Note that, from Thm. 1, the steady-state MSE should be 2.121. In our simulations, we obtained for the steady-state MSE the value 2.116; quite close to the predicted values.

4. CONCLUSION

In this paper, we provided expressions for the steady-state MSE and MSD for the NLMS algorithm, in conditions where

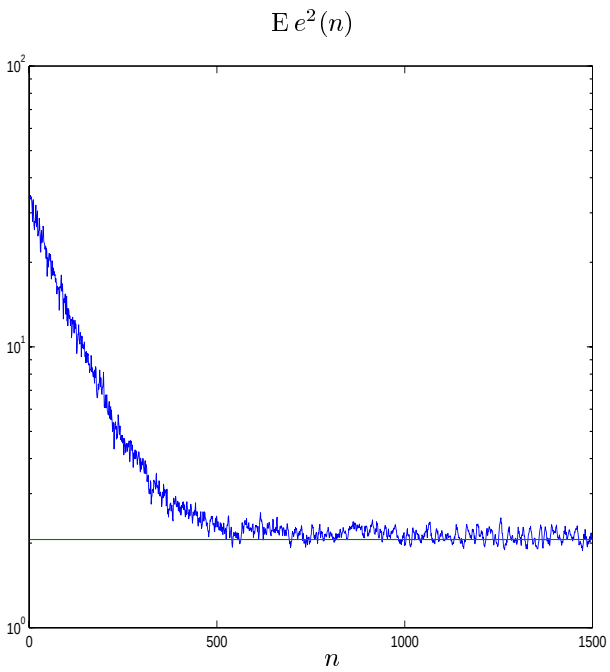

Figure 2: $\mathrm{E}\, e^2(n)$ *for the NLMS algorithm with* $\mu = 0.01$. *The plot also shows the computed level for the MSE.*

NLMS may compute biased estimates for the Wiener filter. Our expressions are dependent on knowledge of several statistics of the regressor and noise sequence, which can be estimated by performing measurements or simulations. We gave an example showing the close agreement between simulations and our expressions.

5. REFERENCES

[1] P.E. An, M. Brown, and C.J. Harris. On the convergence rate performance of the normalized least-mean-square adaptation. *IEEE Transactions on Neural Networks*, 8(5):1211–1214, September 1997.

[2] L. Guo, L. Ljung, and G.-J. Wang. Necessary and sufficient conditions for stability of LMS. *IEEE Transactions on Automatic Control*, 42(6):761–770, June 1997.

[3] T. Kailath, A.H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice-Hall, NJ, 2000.

[4] H.J. Kushner. *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic System Theory*. MIT Press, Cambridge, MA, 1984.

[5] G.V. Moustakides. Exponential convergence of products of random matrices: application to adaptive algorithms. *International Journal of Adaptive Control and Signal Processing*, 12(7):579–597, November 1998.

[6] V.H. Nascimento and A.H. Sayed. Unbiased and stable leakage-based adaptive filters. *IEEE Transactions on Signal Processing*, 47(12):3261–3276, December 1999.

[7] V. Solo. The limiting behavior of LMS. *IEEE Transactions on Signal Processing*, 37(12):1909–1922, December 1989.

[8] V. Solo. The stability of LMS. *IEEE Transactions on Signal Processing*, 45(12):3017–3026, December 1997.

[9] V. Solo and X. Kong. *Adaptive Signal Processing Algorithms*. Prentice-Hall, Englewood Cliffs, NJ, 1995.

[10] B. Widrow and M.A. Lehr. 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proceedings of the IEEE*, 78(9):1415–1441, 1990.