TOWARDS ANALYTICAL CONVERGENCE ANALYSIS OF PROPORTIONATE-TYPE NLMS ALGORITHMS

Kevin T. Wagner

Naval Research Laboratory
Radar Division
Washington, DC 20375, USA

Miloš I. Doroslovački

The George Washington University
Department of Electrical and Computer Engineering
Washington, DC 20052, USA

ABSTRACT

To date no theoretical results have been developed to predict the performance of the proportionate normalized least mean square (PNLMS) algorithm or any of its cousin algorithms such as the μ-law PNLMS (MPNLMS), and the ϵ-law PNLMS (EPNLMS). In this paper we develop an analytic approach to predicting the performance of the simplified PNLMS algorithm which is closely related to the PNLMS algorithm. In particular we demonstrate the ability to predict the Mean Square Output Error of the simplified PNLMS algorithm using our theory.

Index Terms— Adaptive filtering, convergence, proportionate-type normalized least mean square (PtNLMS) algorithm, sparse impulse response.

1. INTRODUCTION

We begin by assuming there is some input signal denoted as $x(k)$ for time k that excites an unknown system with impulse response $\mathbf{w}_{opt}$. Let the output of the system be $y(k) = \mathbf{w}_{opt}^T\mathbf{x}(k)$ where $\mathbf{x}(k) = [x(k), x(k-1), \ldots, x(k-L+1)]^T$. The measured output of the system, $d(k)$, contains zero-mean stationary measurement noise $v(k)$ and is equal to the sum of $y(k)$ and $v(k)$. The impulse response of the system is estimated with the adaptive filter coefficient vector, $\hat{\mathbf{w}}(k)$. The error signal $e(k)$ between the output of the adaptive filter $\hat{y}(k)$ and $d(k)$ drives the adaptive algorithm. The weight deviation (WD) vector is given by $\mathbf{z}(k) = \mathbf{w}_{opt} - \hat{\mathbf{w}}(k)$. The normalized least mean square (NLMS) algorithm for an arbitrary time-varying stepsize control matrix is shown in Table 1, as given in [1]. Here, β is the fixed stepsize parameter, $\mathbf{G}(k+1) = \textbf{diag}\ \{g_1(k+1), \ldots, g_L(k+1)\}$ is the time-varying stepsize control matrix, and L is the length of the adaptive filter. The constant δ is typically a small positive number used to avoid overflowing.

Next, we seek the representation of the Mean Square Output Error (MSE) (Learning Curve) for the proportionate-type normalized least mean square (PtNLMS) algorithm [2]. The MSE is given by $J(k) = E\{|e(k)|^2\}$. By expanding the $e(k)$ term and assuming that the input signal is white, i.e.

Table 1

NLMS Algorithm with Arbitrary Stepsize Matrix		
$\mathbf{x}(k)$	=	$[x(k)x(k-1)\ldots x(k-L+1)]^T$
$\hat{y}(k)$	=	$\mathbf{x}^T(k)\hat{\mathbf{w}}(k)$
$e(k)$	=	$d(k) - \hat{y}(k)$
$\mathbf{G}(k+1)$	=	$\textbf{diag}\{g_1(k+1), \ldots, g_L(k+1)\}$
$\hat{\mathbf{w}}(k+1)$	=	$\hat{\mathbf{w}}(k) + \frac{\beta\mathbf{G}(k+1)\mathbf{x}(k)e(k)}{\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta}$

$\mathbf{R} = \sigma_x^2\mathbf{I}$, and β is so small that the LMS coefficient estimator acts as a low pass filter, then we can rewrite the MSE in the following form [4]:

$$J(k) = J_{min} + \sigma_x^2 \sum_{i=1}^{L} E\{z_i^2(k)\}$$

where the first term J_{min} is equal to the variance of the noise, σ_v^2, and $z_i(k)$ are the elements of $\mathbf{z}(k)$. Hence in order to calculate the MSE we need to find the expected value of the square weight deviations $z_i^2(k)$.

At this stage we proceed by considering the MSE for specific proportionate type NLMS algorithms. Many proportionate type NLMS algorithms, such as the PNLMS [3], MPNLMS [1], and EPNLMS [2] imply highly non-linear (threshold-based) operations. In order to simplify the derivation of analytical results we examine in this paper a simplified PNLMS algorithm. The calculation of the gain for the simplified PNLMS algorithm is given in Table 2. The simplified PNLMS algorithm avoids the usage of the maximum function which is employed in the PNLMS, MPNLMS, and EPNLMS algorithms.

Table 2

Simplified PNLMS Algorithm
$F_i(k) = \rho + \lvert\hat{w}_i(k)\rvert,\ \ i = 1, \ldots, L,\ \ \rho > 0$
$\mathbf{F}(k) = [F_1(k), \ldots, F_L(k)]^T$
$\mathbf{g}(k+1) = \frac{\mathbf{F}(k)}{1/L\sum_i F_i(k)}$

2. RECURSIVE CALCULATION OF THE MEAN WD AND MEAN SQUARE WD

We can represent the WD at time $k+1$ in terms of the prior WD at time k using the recursion for the estimated optimal coefficient vector. Using the convention that $x_i(k) = x(k-i+1)$, this recursion in component-wise form is given by

$$\begin{aligned} z_i(k+1) &= z_i(k) \\ &- \frac{\beta g_i(k+1)x_i(k)\sum_{j=1}^{L} x_j(k)z_j(k)}{\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta} \\ &- \frac{\beta g_i(k+1)x_i(k)\nu(k)}{\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta}. \end{aligned} \tag{1}$$

The component-wise form of the recursion for the square of the WD is given by

$$\begin{aligned} &z_i^2(k+1) = z_i^2(k) \\ &-\tfrac{2\beta g_i(k+1)x_i(k)\sum_{j=1}^{L} x_j(k)z_j(k)z_i(k)}{\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta} \\ &-\tfrac{2\beta g_i(k+1)x_i(k)\nu(k)z_i(k)}{\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta} \\ &-\tfrac{\beta^2 g_i^2(k+1)x_i^2(k)\sum_j\sum_m x_j(k)x_m(k)z_j(k)z_m(k)}{(\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta)^2} \\ &+\tfrac{\beta^2 g_i^2(k+1)x_i^2(k)\nu^2(k)}{(\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta)^2} \\ &+\tfrac{\beta^2 g_i^2(k+1)x_i^2(k)\sum_j x_j(k)z_j(k)\nu(k)}{(\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta)^2}. \end{aligned} \tag{2}$$

Next we take the expected value of the WD and the square WD. In order to do so we make the following set of assumptions.

Assumption I: The adaptation stepsize parameter β is sufficiently small and the LMS coefficient estimator acts as a low pass filter. Hence, $z_i(k)$ changes slowly relative to $x_i(k)$.

Assumption II: The input signal and observation noise are uncorrelated. This assumption is justified provided that the use of the linear unknown system model is applicable and the length of the Wiener optimal solution for the adaptive filter is exactly equal to the order of the unknown system.

Assumption III: The expectation of a ratio of two random variables is equal to the ratio of the expectations of each random variable. In our case the denominator of interest is typically the term $\mathbf{x}(k)^T\mathbf{G}(k+1)\mathbf{x}(k)+\delta$. This assumption holds if the denominator is nearly constant or if we have the condition that $L >> \sqrt{2\sum_{i=1}^{L} E\{g_i^2(k+1)\}}$, [5]. We can derive the expectation of the denominator term by looking at it in component-wise form and applying Assumption I, [5]:

$$\begin{aligned} &E\{\sum_{j=1}^{L} x_j^2(k)g_j(k+1)+\delta\} \\ &= E\{\sum_{j=1}^{L} E\{x_j^2(k)\}g_j(k+1)+\delta\} = \sigma_x^2 L+\delta \end{aligned} \tag{3}$$

Simulations have confirmed that this assumption holds in the situations discussed in this paper. Also, when ρ is very small ($\rho < 10^{-4}$) the experiments show that the assumption does not hold. However most real world applications use larger values for the ρ parameter and therefore this is not an issue.

Assumption IV: The expectation of the denominator term squared is equal to the square of the expectation of the denominator. This assumption leads to

$$E\{(\mathbf{x}^T(k)\mathbf{G}(k+1)\mathbf{x}(k)+\delta)^2\} = (\sigma_x^2 L+\delta)^2.$$

It holds if the denominator is nearly constant.

Therefore we can write that the expectation of the WD can be found recursively from the prior time step by

$$E\{z_i(k+1)\} = E\{z_i(k)\} - \beta_o E\{g_i(k+1)z_i(k)\} \tag{4}$$

where $\beta_o = \frac{\beta\sigma_x^2}{\sigma_x^2 L+\delta}$.

Similarly based upon our assumptions, the expected value of the square WD is given by

$$\begin{aligned} E\{z_i^2(k+1)\} &= E\{z_i^2(k)\} - 2\beta_o E\{g_i(k+1)z_i^2(k)\} \\ &+\beta_o^2 E\{g_i^2(k+1)\sum_{j=1}^{L} z_j^2(k)\} + \frac{\beta_o^2\sigma_v^2}{\sigma_x^2}E\{g_i^2(k+1)\}. \end{aligned} \tag{5}$$

At this point we have the potential to recursively estimate the expected value of the WD and the square WD vectors. One issue remaining is the calculation of terms such as

$$E\{g_i^n(k+1)z_j^m(k)\} \tag{6}$$

for $n \in \{1,2\}$, $m \in \{0,1,2\}$ and $i,j \in \{1,2,\ldots,L\}$.

We assume, that

$$E\{g_i^n(k+1)z_j^m(k)\} = E\{g_i(k+1)\}^n E\{z_j^m(k)\} \text{ if } i \neq j.$$

Now, we can take two approaches when calculating the expectation for $i = j$. In the first approach we assume that the expectation of the product of $g_i^n(k+1)$ and $z_i^m(k)$ is separable. In addition to this, we assume that the expectation of the product of the gains is equal to the product of the expectations of the gains (this assumption holds when $g_i(k+1)$ is slow varying), that is

$$E\{g_i^n(k+1)\} = E\{g_i(k+1)\}^n. \tag{7}$$

Therefore we have

$$E\{g_i^n(k+1)z_i^m(k)\} = E\{g_i(k+1)\}^n E\{z_i^m(k)\}.$$

This approach has been dubbed the 'Separable Approach'.

Alternatively, we can calculate explicitly the expectations in (6). We refer to this approach as the 'Non-Separable Approach'. In the next section we develop the needed probability distributions and expressions for the two approaches.

3. RECURSIVE CALCULATION OF EXPECTATIONS

We begin by assuming that the i^{th} component of the weight deviation at time k has a normal distribution with mean $\mu_i(k)$ and variance $\sigma_i^2(k)$ i.e.

$$z_i(k) \sim \mathcal{N}(\mu_i(k), \sigma_i^2(k)).$$

This assumption is based on a possibility of applying the central limit theorem to the recursion for the weight deviation in (1), as well as simulations. Given this assumption each component of the estimated optimal weight vector is distributed as

$$\hat{w}_i(k) = w_i - z_i(k) \sim \mathcal{N}(m_i(k), \sigma_i^2(k))$$

where $m_i(k) = w_i - \mu_i(k)$. The p.d.f. of $|\hat{w}_i(k)|$ is given by

$$f(|\hat{w}_i(k)|) = \frac{1}{\sqrt{2\pi\sigma_i^2(k)}}[e^{-\frac{(|\hat{w}_i(k)|-m_i(k))^2}{2\sigma_i^2(k)}} + e^{-\frac{(|\hat{w}_i(k)|+m_i(k))^2}{2\sigma_i^2(k)}}]U(\hat{w}_i(k)) \tag{8}$$

where $U(x)$ is the unit step function [6].

We now take advantage of the form of this p.d.f. and calculate several expectations which will be useful in future derivations. We begin by finding the mean of this distribution which is given by

$$E\{|\hat{w}_i(k)|\} = m_i(k)\mathrm{erf}\big(\frac{m_i(k)}{\sqrt{2\sigma_i^2(k)}}\big) + \sqrt{\frac{2}{\pi}}\sigma_i(k)e^{-\frac{m_i^2(k)}{2\sigma_i^2(k)}}. \tag{9}$$

Additionally, the second moment is given by

$$E\{|\hat{w}_i(k)|^2\} = m_i^2(k) + \sigma_i^2(k). \tag{10}$$

We can also calculate the following expectations:

$$\begin{aligned} E\{|\hat{w}_i(k)|(w_i - \hat{w}_i(k))\} &= \big(w_i\mu_i(k) - \sigma_i^2(k) - \mu_i^2(k)\big) \\ &\times\mathrm{erf}\left(\frac{m_i(k)}{\sqrt{2\sigma_i^2(k)}}\right) + \frac{2\sigma_i(k)\mu_i(k)}{\sqrt{2\pi}}e^{-\frac{m_i^2(k)}{2\sigma_i^2(k)}} \end{aligned} \tag{11}$$

$$\begin{aligned} E\{|\hat{w}_i(k)|(w_i - \hat{w}_i(k))^2\} &= \big(w_i\mu_i^2(k) + w_i\sigma_i^2(k) \\ &-3\mu_i(k)\sigma_i^2(k) - \mu_i^3(k)\big)\mathrm{erf}\left(\frac{m_i(k)}{\sqrt{2\sigma_i^2(k)}}\right) \\ &+\big(2\mu_i^2(k) + 4\sigma_i^2(k)\big)\frac{\sigma_i(k)}{\sqrt{2\pi}}e^{-\frac{m_i^2(k)}{2\sigma_i^2(k)}} \end{aligned} \tag{12}$$

$$\begin{aligned} E\{|\hat{w}_i(k)|^2(w_i - \hat{w}_i(k))^2\} &= w_i^2\big(\mu_i^2(k) + \sigma_i^2(k)\big) \\ &-2w_i\big(\mu_i^3(k) + 3\mu_i(k)\sigma_i^2(k)\big) \\ &+\mu_i^4(k) + 6\mu_i^2(k)\sigma_i^2(k) + 3\sigma_i^4(k) \end{aligned} \tag{13}$$

3.1. Separable Expectation Calculations

In the separable case the expectation of the WD and the square WD are given by

$$E\{z_i(k+1)\} = E\{z_i(k)\} - \beta_o E\{g_i(k+1)\}E\{z_i(k)\} \tag{14}$$

$$\begin{aligned} E\{z_i^2(k+1)\} &= E\{z_i^2(k)\} - 2\beta_o E\{g_i(k+1)\}E\{z_i^2(k)\} \\ &+\beta_o^2 E\{g_i(k+1)\}^2\sum_{j=1}^{L}E\{z_j^2(k)\} + \frac{\beta_o^2\sigma_v^2}{\sigma_x^2}E\{g_i(k+1)\}^2 \end{aligned} \tag{15}$$

respectively. Note $\sigma_i^2(k) = E\{z_i^2(k)\} - E^2\{z_i(k)\}$. At this point we are left to find $E\{g_i(k+1)\}$. This term can be found as

$$\begin{aligned} E\{g_i(k+1)\} &= E\{\frac{F_i(k)}{1/L\sum_j F_j(k)}\} \\ &\approx \frac{\rho + E\{|\hat{w}_i(k)|\}}{1/L\sum_j(\rho + E\{|\hat{w}_j(k)|\})}. \end{aligned} \tag{16}$$

This algorithm is initialized by setting $E\{z_i(0)\} = w_i$ and $E\{z_i^2(0)\} = w_i^2$.

3.2. Non-Separable Expectation Calculations

In order to calculate the mean WD and the mean square WD we find:

$$\begin{aligned} E\{g_i(k+1)z_i(k)\} &= E\{\frac{\big(\rho+|w_i-z_i(k)|\big)z_i(k)}{1/L\sum_j(\rho+|w_j-z_j(k)|)}\} \\ &\approx \frac{\rho E\{z_i(k)\}+E\{|\hat{w}_i(k)|(w_i-\hat{w}_i(k))\}}{1/L\sum_j(\rho+E\{|\hat{w}_j(k)|\})}. \end{aligned} \tag{17}$$

$$\begin{aligned} E\{g_i(k+1)z_i^2(k)\} &= E\{\frac{(\rho+|w_i-z_i(k)|)z_i^2(k)}{1/L\sum_j(\rho+|w_j-z_j(k)|)}\} \\ &\approx \frac{\rho E\{z_i^2(k)\}+E\{|\hat{w}_i(k)|(w_i-\hat{w}_i(k))^2\}}{1/L\sum_j(\rho+E\{|\hat{w}_j(k)|\})}. \end{aligned} \tag{18}$$

$$\begin{aligned} E\{g_i^2(k+1)z_i^2(k)\} &= E\{\frac{\big(\rho+|w_i-z_i(k)|\big)^2 z_i^2(k)}{\big(1/L\sum_j(\rho+|w_j-z_j(k)|)\big)^2}\} \\ &\approx \Big[\rho^2 E\{z_i^2(k)\} + 2\rho E\{|\hat{w}_i(k)|(w_i-\hat{w}_i(k))^2\} \\ &+E\{|\hat{w}_i(k)|^2(w_i-\hat{w}_i(k))^2\}\Big]/\big(\tfrac{1}{L}\textstyle\sum_j(\rho+E\{|\hat{w}_j(k)|\})\big)^2. \end{aligned} \tag{19}$$

$$\begin{aligned} E\{g_i^2(k+1)\} &= E\{\frac{\big(\rho+|w_i-z_i(k)|\big)^2}{\big(1/L\sum_j(\rho+|w_j-z_j(k)|)\big)^2}\} \\ &\approx \frac{\rho^2+2\rho E\{|\hat{w}_i(k)|\}+E\{|\hat{w}_i(k)|^2\}}{\big(1/L\sum_j(\rho+E\{|\hat{w}_j(k)|\})\big)^2}. \end{aligned} \tag{20}$$

Using equations (9)-(13) these terms can be calculated.

4. RESULTS

Now we compare the theory derived to actual results from Monte Carlo simulations. In the simulations and figures that are shown the following parameters have been chosen unless specified otherwise, $L = 512$, $\sigma_x^2 = 10^{-2}$ $\sigma_v^2 = 10^{-6}$, and $\delta = 10^{-4}$. We have developed a metric to quantitatively measure how well the theory fits the ensemble averaged results. The metric is given by

$$C = \frac{\sum_k |e_T^2(k) - e_{MC}^2(k)|}{\sum_k e_{MC}^2(k)}$$

where $e_T^2(k)$ is the squared output error generated by the theory at time k and $e_{MC}^2(k)$ is the squared output error generated by the ensemble average at time k. The term in the denominator has been added in an attempt to make the metric independent of the input signal power.

We compare the performance of the 'Separable Approach' theory versus the 'Nonseparable Approach' theory when using the echo-path impulse response presented in [7]. This impulse is sparse because very few coefficients have non-zero values. The performance of the 'Separable Approach' theory for $\rho = 10^{-2}$ is shown in Figure 1. The results when using the 'Nonseparable Approach' theory for $\rho = 10^{-2}$ is shown in Figure 2. The 'Nonseparable Approach' theory performs slightly better than the 'Separable Approach' theory. This improvement is reflected in the metric C where it has been reduced from a value of 0.14631 to 0.11011 after applying the 'Nonseparable Approach' theory.

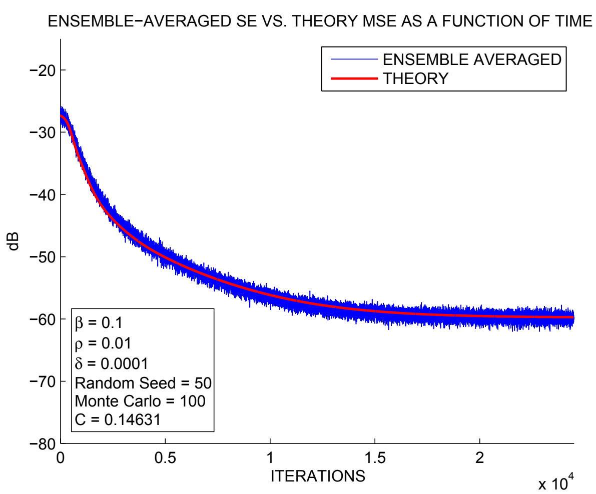

Fig. 1. Learning curve of simplified PNLMS algorithm $\rho = 10^{-2}$ using 'Separable Approach' theory

5. CONCLUSIONS

We have developed two analytical methods to predict the performance of the simplified PNLMS algorithm by developing recursions for the mean weight deviation and mean square weight deviation. The weight deviation is assumed to have a Gaussian distribution. In the first method the expectation of the product of the gain and weight deviation is considered to be separable. In the second method the expectation of the product of the gain and weight deviation is derived without assuming the separability. The second method while more computationally intensive offers some improvement in the ability to predict the performance of the simplified PNLMS algorithm. Further analysis shows that the improvement comes mainly from the direct calculation of the $E\{g_i^2(k)\}$ instead of the assumption in (7).

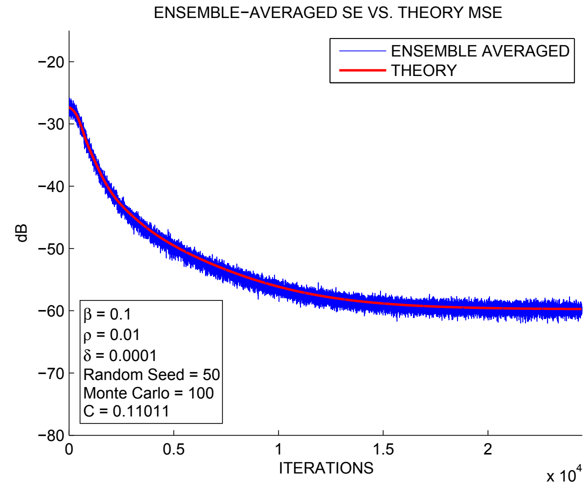

Fig. 2. Learning curve of simplified PNLMS algorithm $\rho = 10^{-2}$ using 'Nonseparable Approach' theory

6. REFERENCES

[1] H. Deng and M. Doroslovački, "Improving Convergence of the PNLMS Algorithm for Sparse Impulse Response Identification," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 181-184, Mar. 2005.

[2] K. Wagner, M. Doroslovački, and H. Deng "Convergecne of proportionate-type NLMS adaptive filters and choice of gain matrix," *Proc. 40th Asilomar on Signals, Systems, and Computers,* Pacific Grove, CA, Oct. 29 - Nov. 1, 2006.

[3] D. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancellers," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 508-518, Sept. 2000.

[4] S. Haykin, *Adaptive Filter Theory*, fourth edition, Prentice Hall 2002.

[5] H. Deng and M. Doroslovački, "On Convergence of Proportionate-Type NLMS Adaptive Algorithms," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 105-108, Toulouse, France, May 2006.

[6] A. Oppenheim and A. Willsky with S. Nawab, *Signals and Systems,* fourth edition, Prentice Hall, 1997.

[7] K. Wagner and M. Doroslovački, "Proportionate-Type Steepest Descent and NLMS Algorithms," *Proc. 41st Conference on Information Sciences and Systems,* Baltimore, MD, Mar. 14-16, 2007.