A UNIFYING APPROACH TO THE DERIVATION OF THE CLASS OF PNLMS ALGORITHMS

Beth Jelfs and Danilo P. Mandic

Department of Electrical & Electronic Engineering
Imperial College London, UK
E-mail: {beth.jelfs,d.mandic}@imperial.ac.uk

Andrzej Cichocki

RIKEN, Brain Science Institute,
Saitama 351-0198, JAPAN
E-mail: cia@brain.riken.jp

ABSTRACT

A unifying approach to the derivation of the class of proportionate normalised least mean square (PNLMS) algorithms is provided. This is an important class of algorithms where the two most used algorithms are introduced empirically. It is shown that it is possible to derive PNLMS algorithms as a result of an optimisation procedure. This is achieved in a rigorous way, starting from the standard LMS through to the PNLMS with the "sparsification" factor in both the numerator and denominator of the weight update. The proposed approach is generic and also applies to other LMS types of adaptive algorithms. Simulations on benchmark sparse impulse responses support the approach.

***Index Terms*—** LMS, normalised LMS (NLMS), proportionate NLMS (PNLMS).

1. INTRODUCTION

The least mean square (LMS) family of algorithms are a de facto standard for linear adaptive filtering [1, 2]. The LMS algorithm minimises the instantaneous cost function $E(k) = \frac{1}{2}e^2(k)$, and is described by [1]

$$\begin{aligned} e(k) &= d(k) - \mathbf{x}^T(k)\mathbf{w}(k), \\ \mathbf{w}(k+1) &= \mathbf{w}(k) + \mu(k)e(k)\mathbf{x}(k), \end{aligned} \tag{1}$$

where $e(k)$ is the output error at time instant k, $d(k)$ the desired signal, and $\mathbf{x}(k) = [x(k), ..., x(k-N+1)]^T$ and $\mathbf{w}(k) = [w_1(k), ..., w_N(k)]^T$ are respectively the input signal and filter coefficient vector for a filter of length N. The parameter μ is the step–size, which is critical to the performance, and defines how fast the algorithm is converging towards the optimal solution

To facilitate the operation in a nonstationary environment that is, to allow the filter to adapt according to the time varying statistical nature of the tap input signal, normalised LMS (NLMS) [1, 2] uses an adaptive step size

$$\eta(k) = \frac{\mu}{\|\mathbf{x}(k)\|_2^2 + \varepsilon} \tag{2}$$

where $\|\cdot\|_2$ is the Euclidean norm. For practical reasons, the regularisation parameter ε is included in order to prevent the weight update becoming unstable for input vectors comprising of near to zero values which would otherwise result in a large "learning rate" $\eta(k)$.

Due to the importance and wide range of practical applications of the LMS based algorithms, research into modifications of this class of algorithms has become a key topic in statistical and adaptive signal processing [2].

1.1. Proportionate NLMS

As sparse systems occur naturally within many real-world applications (acoustics, seismics, chemical processes), investigations into sparse environments has become an increasingly large area of research [3, 4]. As both the LMS and NLMS algorithm perform in a suboptimal manner in practical settings, such as in sparse environments where the impulse response vector has a number of zero elements, one particular focus is the development of adaptive filters specifically designed for such environments.

For operation in sparse environments one such modification of the LMS, the proportionate NLMS (PNLMS) algorithm [5] has proved particularly useful. By taking advantage of the knowledge that the impulse response is sparse PNLMS develops on the existing NLMS algorithm to give an update which is proportional relative to the size of the filter coefficients. This is achieved by introducing a diagonal "tap selection matrix" $\mathbf{G}(k)$ within the coefficient update (1), giving [5]

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\frac{\mathbf{G}(k)e(k)\mathbf{x}(k)}{\|\mathbf{x}(k)\|_2^2}, \tag{3}$$

where

$$\mathbf{G}(k) = \mathrm{diag}[g_1(k), \ldots, g_N(k)]. \tag{4}$$

The diagonal elements $g_1(k), \ldots, g_N(k)$ of $\mathbf{G}(k)$ define the "proportional" amounts that each coefficient is updated by, where $g_n(k)$ are given by

$$\bar{\phi}(k) = 1/N \sum_{n=1}^{N} \phi_n(k),$$
$$\phi_n(k) = \max\left\{\rho \max\left[\delta, \|\mathbf{w}(k)\|_\infty\right], |w_n(k)|\right\},$$
$$g_n(k) = \frac{\phi_n(k)}{\bar{\phi}(k)} \qquad n = 1, \ldots, N, \tag{5}$$

The PNLMS algorithm in its original form has been introduced based on empirical evidence and subsequently most of its variants have also been designed in an ad hoc manner. Whilst some aspects of this issue have been addressed [6, 7], it would however, be beneficial if the class of PNLMS algorithms could be unified, as a result of an optimisation procedure. Our aim is to provide a unified approach to the derivation of the class of PNLMS algorithms. This is achieved based on the approach from [8], in which the focus is on minimisation of the *a posteriori* error $e(k+1)$.

2. DERIVATION OF THE CLASS OF PNLMS ALGORITHMS

Although originally Duttweiler [5] introduced the PNLMS in the form of (3) it is natural to conduct our analysis starting from LMS. In the same vein as the Duttweiler result, we shall re-write the LMS to suit sparse environments, providing us with a method to derive the class of PNLMS algorithms in a generic way.

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \mathbf{G}(k) e(k) \mathbf{x}(k). \tag{6}$$

Notice the only difference between LMS (1) and proportionate LMS (6) is the "tap selective" term $\mathbf{G}(k)$. This also has a geometric justification, since for an N-tap LMS the weight update lives in $\mathbb{R}^N$, and the direction of that update is totally dominated by the largest element of the weight vector $\mathbf{w}$. Also note that the optimisation task performed within NLMS type algorithms, is to actually minimise the *a posteriori* error $e(k+1)$, as opposed to LMS which minimises the *a priori* error $e(k)$. To arrive at the PNLMS, following the approach from [8, 9], perform the Taylor series expansion (TSE) of the *a posteriori* instantaneous output error $e(k+1)$, to give

$$e(k+1) = e(k) + \sum_{i=1}^{N} \frac{\partial e(k)}{\partial w_i(k)} \Delta w_i(k) + \sum_{i=1}^{N} \frac{\partial e(k)}{\partial x_i(k)} \Delta x_i(k) + \frac{\partial e(k)}{\partial d(k)} \Delta d(k) + \text{h.o.t.} \tag{7}$$

where h.o.t. denotes the usually neglected (due to the linearity of the filter) higher order terms of TSE (7). Unlike in the approach from [8], generally the partial derivatives with respect to $\mathbf{x}$ and d as well as higher order terms in TSE (7), cannot be neglected. Firstly from (1), we obtain the partial derivatives in (7) as

$$\begin{aligned}
\frac{\partial e(k)}{\partial w_i(k)} &= -x(k-i+1) = -x_i(k), \quad & i = 1, 2, \ldots, N \\
\frac{\partial e(k)}{\partial x_j(k)} &= -w_j(k), & j = 1, 2, \ldots, N \\
\frac{\partial e(k)}{\partial d(k)} &= 1
\end{aligned} \tag{8}$$

Next the update $\Delta w_i(k)$ (6), can be expressed as

$$\Delta w_i(k) = \mu e(k) g_i(k) x_i(k), \;\; i = 1, 2, \ldots, N \tag{9}$$

Finally substituting (8) – (9) into (7) to obtain

$$e(k+1) = e(k) - \mu e(k) \left(\sum_{i=1}^{N} x_i^2(k) g_i(k) \right) - \left(\sum_{i=1}^{N} w_i(k) \Delta x_i(k) \right) + \Delta d(k) + h.o.t. \tag{10}$$

To find the optimal learning rate η in the minimum mean squared error (MMSE) sense, differentiate $e^2(k+1)$ with respect to η and set to zero, to yield

$$\frac{\partial \frac{1}{2} e^2(k+1)}{\partial \mu} = 0 \Rightarrow e(k+1) = 0 \tag{11}$$

from which we obtain

$$\mu_{opt} = \frac{e(k) - \left(\sum_{i=1}^{N} w_i(k) \Delta x_i(k) \right) + \Delta d(k)}{e(k) \sum_{i=1}^{N} x_i^2(k) g_i(k)} \tag{12}$$

This gives an optimal learning rate providing the first order Taylor series expansion gives a good approximation of the *a posteriori* instantaneous output error $e(k+1)$. Whenever the values $\Delta x_i(k)$ and $\Delta d(k)$ can be calculated or are known (as in the case of off–line training schemes or ensemble learning), they should be included in the above analysis.

In practice, however, as in the case of on–line linear adaptive filters, $\Delta d(k)$ cannot be computed, and in addition, the variation of first order terms associated with $w_i(k)$, $x_i(k)$, and $d(k)$ between two successive discrete time instants might not be important. For simplicity, and using the usual independence assumptions[1], the PNLMS can be derived from (7), by neglecting the first order partial derivatives with respect to $\mathbf{x}$, d and also higher order terms of TSE (7). The original PNLMS was introduced empirically and so uses the standard NLMS update which is a good estimate of the optimal step

[1]Namely that the input signal and filter coefficient vectors are zero mean, stationary, jointly normal and with finite moments; the successive increments of tap weights are independent of one another and the error and input vector sequences are statistically independent of one another.

size, from (12), to minimise the *a posteriori* error $e(k+1)$, we have[2]

$$\mu_{opt} = \frac{1}{\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k)} \tag{13}$$

Notice that the Proportionate LMS can now be re-written as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\frac{\mathbf{G}(k)e(k)\mathbf{x}(k)}{\mathbf{x}^T(k)\mathbf{G}(k)\mathbf{x}(k)}. \tag{14}$$

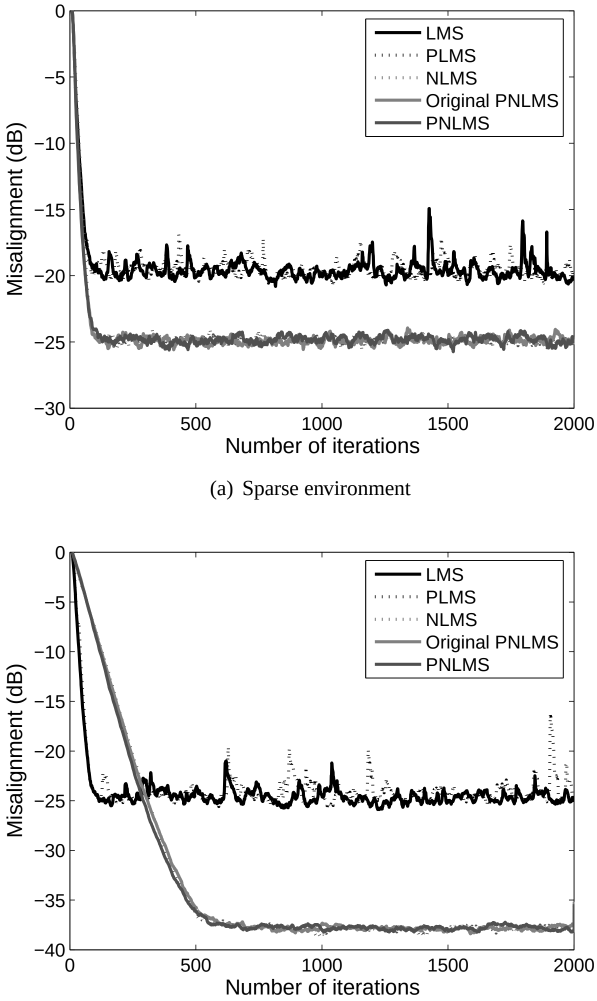
which gives us a rigorous derivation of the update in precisely the form of what is now the standard version of PNLMS [6, 10].

3. SIMULATIONS

To illustrate the performance of the algorithms, the LMS and NLMS were compared with the PLMS and PNLMS in both sparse and nonsparse environments. Learning curves were produced using the normalised misalignment in dB, given by $10\log_{10}\|\mathbf{w}_{\rm opt} - \mathbf{w}\|_2^2/\|\mathbf{w}_{\rm opt}\|_2^2$, averaged over 100 independent trials, where $\mathbf{w}_{\rm opt} = [w_{1\ \rm opt}, ..., w_{n\ \rm opt}]$ is the optimal filter coefficient vector. For the proportionate algorithms, the parameters ρ and δ were set to the recommended values of $\rho = 5/N$ and $\delta = 0.01$ [5] and for NLMS-type algorithms $\varepsilon = 0.01$. The actual sparse systems employed were the benchmark systems analysed in [11].

Figure 1 shows the performance of the algorithms for a sparse filter of length 10 and $\mu = 0.1$. Figure 1(a) shows the performance in a sparse environment, although all the filters have a similar convergence rate, the NLMS-type algorithms have approximately 5-dB smaller misalignment. In a nonsparse environment, Fig 1(b) NLMS-type algorithms still offer an improvement in misalignment this time over 10-dB but this time with a slower convergence rate. The simulation was then repeated on a filter of length 100, with $\mu = 0.01$ for the LMS-type algorithms and $\mu = 0.5$ for the NLMS-type algorithms. In this case, in the sparse environment, Fig. 2(a), the proportionate algorithms offer a faster convergence rate, with the NLMS-type algorithms again giving approximately 5-dB improvement in misalignment, meaning the PNLMS which combines both the faster convergence of the proportionate algorithms and the smaller misalignment of the NLMS-type algorithms offers the best solution. For the nonsparse environment, Fig. 2(b), the proportionate algorithms have a considerably slower speed of convergence with again the NLMS-type algorithms offering the smaller misalignment. In these cases there is little difference between the original PNLMS and the updated version with the proportionate matrix in the denominator, but it is known that although this algorithm performs well in many situations can become unstable for impulsive excitation signals [6].

[2]We shall ignore the trivial solution $e(k) = 0$.



(a) Sparse environment

(b) Nonsparse environment

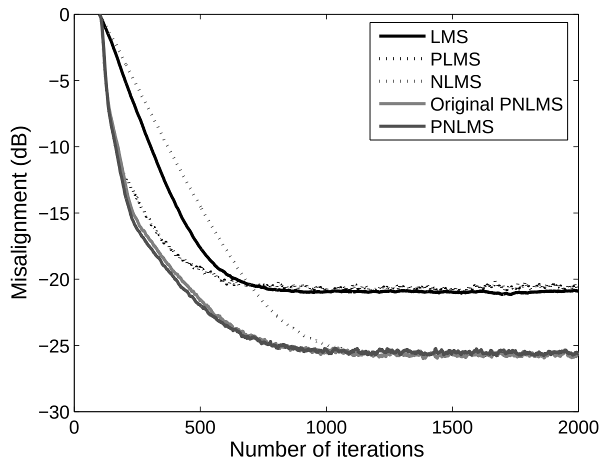
Fig. 1. Performance comparison of the algorithms for N=10

4. CONCLUSIONS

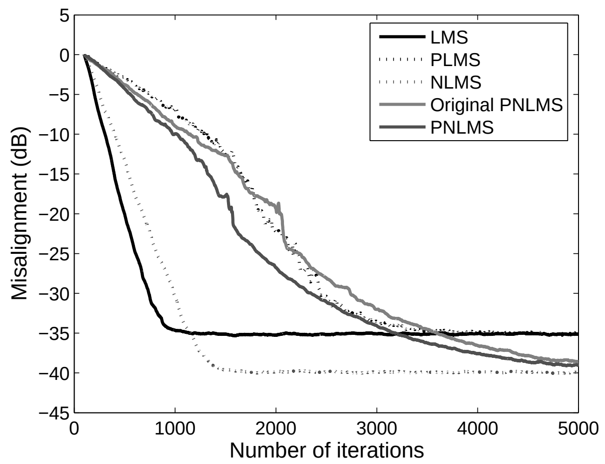
We have provided a unified approach to the derivation of the class of PNLMS algorithms, starting from the standard LMS, and showing that through the minimisation of the *a posteriori* error $e(k+1)$ it is possible to arrive at the PNLMS algorithms.. Furthermore this a generic approach which can be applied equally well to other types of LMS algorithms. Simulations on benchmark sparse and nonsparse systems support the approach, and also highlight the known existing problems with the performance of the algorithms in some situations.

5. REFERENCES

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Prentice-Hall, 1985.



(a) Sparse environment



(b) Nonsparse environment

Fig. 2. Performance comparison of the algorithms for N=100

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

[3] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing, Learning Algorithms and Applications*. John Wiley and Sons, 2003.

[4] B. Jelfs and D. Mandic, "Towards online monitoring of the changes in signal modality: The degree of sparsity," in *IMA 7th International Conference on Mathematics in Signal Processing*, 2006.

[5] D. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech, Audio Proc.*, vol. 8, no. 5, pp. 508–518, 2000.

[6] S. L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," in *Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 394–398, 1998.

[7] J. Benesty and Y. A. Huang, "The LMS, PNLMS, and exponentiated gradient algorithms," in *12th European Signal Processing Conference (EUSIPCO)*, 2004.

[8] E. Soria-Olivas, J. Calpe-Maravilla, J. F. Guerrero-Martinez, M. Martinez-Sober, and J. Espi-Lopez, "An easy demonstration of the optimum value of the adaptation constant in the LMS algorithm," *IEEE Transactions on Education*, vol. 41, no. 1, p. 81, 1998.

[9] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction*. John Wiley & Sons, LTD, 2001.

[10] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Proc.(ICASSP)*, vol. 2, pp. 1881–1884, 2002.

[11] R. Martin, W. Sethares, R. Williamson, and C. Johnson, Jr., "Exploiting sparsity in adaptive filters," *IEEE Trans. Signal Proc.*, vol. 50, no. 8, pp. 1883–1894, 2002.