A Unifying Approach to the Derivation of the Class of PNLMS Algorithms

Beth Jelfs[1], Danilo Mandic[1] & Andrzej Chichocki[2]

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

[2]RIKEN, Brain Science Institute, Saitama, JAPAN
E-mail: cia@brain.riken.jp

Outline

- PNLMS Algorithm
- Minimisation of *a posteriori* error
- Taylor Series Expansion
- Derivation of PNLMS
- Higher Order Terms
- Verification of concept:- Simulation results on synthetic data
- Conclusions

Proportionate Normalised Least Mean Square

Sparse Systems occur naturally in many within many real-world applications (acoustics, seismics, chemical processes).

Standard LMS & NLMS algorithms can perform perform in a suboptimal manner in such practical settings.

For operation in sparse environments one of the most widely used modifications of the LMS is proportionate NLMS (PNLMS) [1]

$$
\begin{aligned}
\mathbf{w}(k+1) &= \mathbf{w}(k) + \mu \frac{\mathbf{G}(k)e(k)\mathbf{x}(k)}{\|\mathbf{x}(k)\|_2^2} \\
\bar{\phi}(k) &= 1/N\mathrm{sum} \sum_{n=1}^{N} \phi_n(k) \\
\phi_n(k) &= \max\{\rho \max\{\delta, \|\mathbf{w}(k)\|_\infty\}, |w_n(k)|\} \\
g_n(k) &= \frac{\phi_n(k)}{\bar{\phi}(k)} \qquad n = 1, \ldots, N
\end{aligned}
$$

PNLMS - disadvantages

Disadvantages of the PNLMS include:-

i) PNLMS empirically derived specifically for network echo cancellers

ii) Can be slower than NLMS when approaching steady state

iii) Also inherits problems from NLMS

Whilst some aspects of this issue have been addressed [2]
It would 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 based on minimisation of the *a posteriori* error $e(k+1)$.**

Proportionate LMS

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.

Thus providing us with a method to derive the class of PNLMS algorithms in a generic way.

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

Notice the only difference between LMS and proportionate LMS 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 $\mathbf{w}$.

Taylor Series Expansion

Note the difference in the optimisation tasks performed

LMS type algorithms minimise the *a priori* error $e(k)$

NLMS type algorithms minimise the *a posteriori* error $e(k+1)$

To arrive at the PNLMS, perform the Taylor series expansion (TSE) of the *a posteriori* instantaneous output error $e(k+1)$ [3], to give

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

where h.o.t. denotes the usually neglected (due to the linearity of the filter) higher order terms of TSE.

Taylor Series Expansion cont.

Firstly from, we obtain the partial derivatives as

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

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

$$\Delta w_i(k) = \mu e(k) g_i(k) x_i(k), \quad i = 1, 2, \ldots, N$$

Finally substituting the partial derivatives into the TSE to obtain

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

Optimal Learning Rate

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

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

from which we obtain

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

This gives an optimal learning rate providing the first order Taylor series expansion gives a good approximation of the *a posteriori* instantaneous output error $e(k+1)$.

First Order Terms

Consider the first order terms associated with $\Delta x_i(k)$ and $\Delta d(k)$

- In terms of off–line training schemes (or ensemble learning) where values are known or can be calculated they should be included
- For on–line adaptive filters $\Delta d(k)$ cannot be computed and 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, the PNLMS can be derived, by neglecting the first order partial derivatives with respect to $\mathbf{x}$, d and also higher order terms of TSE.

Derivation of PNLMS

The original PNLMS was introduced empirically and so uses the standard NLMS update which is a good estimate of the optimal step size.
To minimise the *a posteriori* error $e(k+1)$, we have

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

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

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

which gives us a rigorous derivation of the update in precisely the form of what is now the standard version of PNLMS [2,4].

Higher Order Terms

Considering the neglected terms from the TSE gives

$$e(k+1) = e(k) - \mu e(k) \sum_{i=1}^{N} x_i^2(k) g_i(k) + h.o.t.(k)$$

where the higher order terms can be expressed in many ways.
One development worth noting would be development of gradient adaptive step sizes for PNLMS, including

- the learning rate proposed by Mathews [5] based on $\frac{\partial E(k)}{\partial \mu}$, where the higher order terms can be expressed as

$$h.o.t.(k) = \theta(k) e(k)$$

- or alternatively such as in the GNGD algorithm [6], where the parameter ε in the step-size of NLMS is made gradient adaptive based on $\frac{\partial E(k)}{\partial \varepsilon(k-1)}$,

$$h.o.t.(k) = -\mu e(k) \varepsilon(k)$$

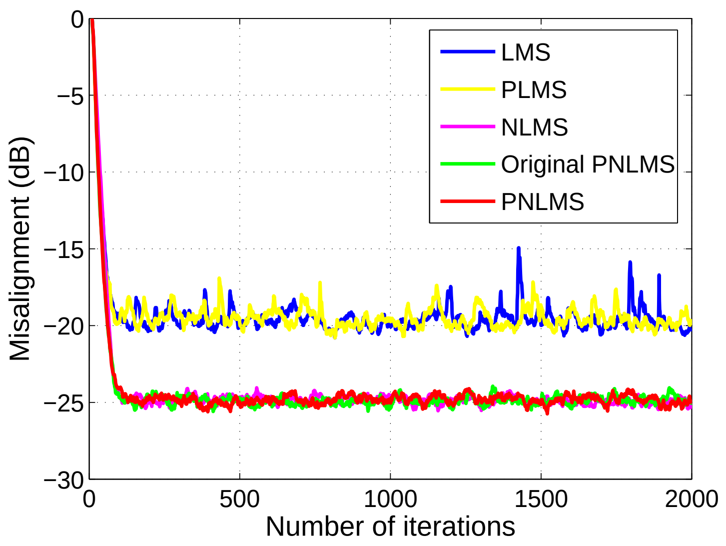
Simulations

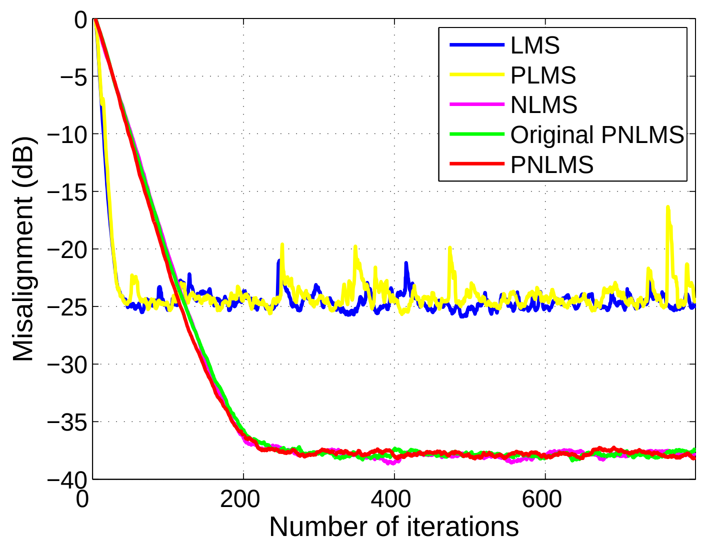
The performance of the algorithms for a benchmark sparse filter of length 10 [7] shows

- all the filters have a similar convergence rate,
- the NLMS-type algorithms have smaller misalignment.

In a nonsparse environment,

- NLMS-type algorithms still offer an improvement in misalignment
- but this time with a slower convergence rate.





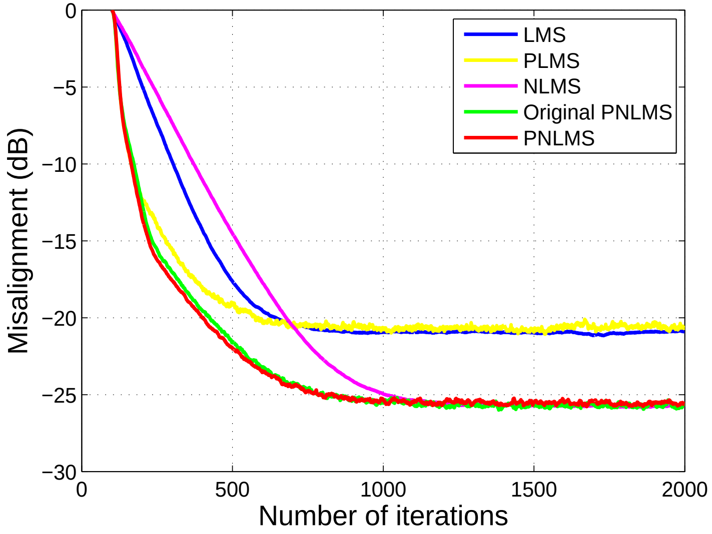
Simulations

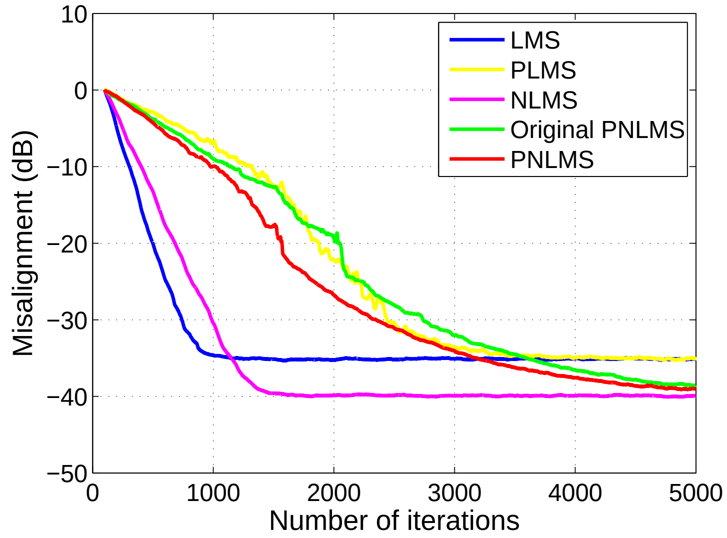
For a filter of length 100, in the sparse environment,

- the proportionate algorithms offer a faster convergence rate,
- the NLMS-type algorithms again give improvement in misalignment,

For the nonsparse environment,

- the proportionate algorithms have a considerably slower speed of convergence
- again the NLMS-type algorithms offering the smaller misalignment.





Conclusions

- We have provided a unified approach to the derivation of the class of PNLMS algorithms,
- Showing that starting from the standard LMS, 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.

References

1. D. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech, Audio Proc.*, vol. 8, no. 5, pp. 508–518, 2000.
2. 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.
3. 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.
4. 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.
5. V. Mathews & Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Proc.*, vol. 41, no. 6, pp. 2075–2087, 1993.
6. D. Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Proc. Lett.*, vol. 11, pp. 115–118, Feb. 2004.
7. R. Martin, W. Sethares, R. Williamson, & C. Johnson, Jr., "Exploiting sparsity in adaptive filters," *IEEE Trans. Signal Proc.*, vol. 50, no. 8, pp. 1883–1894, 2002.