A Class of Adaptively Regularised PNLMS Algorithms

Beth Jelfs[1], Danilo Mandic[1] & Jacob Benesty[2]

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

[2]INRS-EMT, University of Quebec, Montreal, Quebec, Canada
E-mail: benesty@emt.inrs.ca

Outline

- General problem of **sparsity in adaptive filtering**
- PNLMS Algorithm
- Adaptive **step-size techniques**
- Derivation of **adaptive regularisation** for PNLMS
- Robust regularisation:- Stable in a wide range of situations and robust to parameter initialization
- Verification of concept:- Simulation results on synthetic data
- Conclusions

Why Sparsity?

- In all forms of adaptive filtering knowledge about the nature of the signal can give a useful insight, providing valuable information in many different applications
  e.g. health conditions (heart rate), radar and many others

- The theory of signals generated by linear systems, (described by Gaussian distributions or as stationary) are well understood

- All other signals are in some way non– linear, Gaussian, stationary etc. and are less clearly defined and understood

- Focusing on sparseness as a subset of nonlinearity as sparse systems occur naturally within many real-world applications (acoustics, seismics, chemical processes)

- A sparse system is usually defined as a system described by an impulse response with a relatively large number of inactive coefficients (usually zero or close to) compared to active, non-zero, ones

Proportionate Normalised Least Mean Square

One of the most widely used sparse adaptive filters is PNLMS [1]

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

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

Adaptive Step-size Techniques

Adaptive step size techniques have been developed to improve the convergence of linear adaptive filters in nonstationary environments, these include:-

- "linear" gradient adaptive learning rates based on $\partial E/\partial \mu$
  e.g. Benveniste [2], Mathews [3], Farhang [4]

- and a "nonlinear" gradient adaptive update based on $\partial E/\partial \varepsilon$
  e.g Generalised Normalised Gradient Descent (GNGD) [5]

where $E(k) = (1/2)e^2(k)$ is the cost function.

↬ **Introduction of a gradient adaptive regularisation parameter would make the existing ε–PNLMS algorithms better suited to operating in real world environments.**

PNLMS with Adaptive Regularisation

To improve the performance of the PNLMS a similar method to GNGD giving an adaptive regularisation parameter has been derived.
Evaluating the gradient $\nabla_{\varepsilon(k-1)} E(k)$ gives

$$\frac{\partial E(k)}{\partial \varepsilon(k-1)} = \frac{\partial E(k)}{\partial e(k)} \cdot \frac{\partial e(k)}{\partial y(k)} \cdot \frac{\partial y(k)}{\partial \mathbf{w}(k)} \cdot \frac{\partial \mathbf{w}(k)}{\partial \varepsilon(k-1)}$$

$$= \mu \frac{e(k)e(k-1)\mathbf{x}^T(k)\mathbf{G}(k-1)\mathbf{x}(k-1)}{\left[\|\mathbf{x}(k-1)\|_2^2 + \varepsilon(k-1)\right]^2}.$$

$$\varepsilon(k+1) = \varepsilon(k) - \beta\mu \frac{e(k)e(k-1)\mathbf{G}(k-1)\mathbf{x}^T(k)\mathbf{x}(k-1)}{\left(\|\mathbf{x}(k-1)\|_2^2 + \varepsilon(k-1)\right)^2}$$

↬ **Notice that, although the derivation starts from the original PNLMS algorithm, it could be applied equally well to any of its variants.**

Individual vs. Global Update

Including this adaptive regularisation update into PNLMS update gives us the generalised PNLMS (GPNLMS)

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

Alternatively taking into account that PNLMS favours some filter coefficients, it is natural to provide each coefficient with a corresponding regularisation parameter $\varepsilon_n(k)$. Therefore, we have

$$\varepsilon_n(k+1) = \varepsilon_n(k) - \beta\mu \frac{e(k)e(k-1)x_n(k)g_n(k-1)x_n(k-1)}{\left[\|\mathbf{x}(k-1)\|_2^2 + \varepsilon_n(k-1)\right]^2},$$

giving the generalised individually adaptive PNLMS (GIAPNLMS)

$$w_n(k+1) = w_n(k) + \mu \frac{g_n(k)e(k)x_n(k)}{\|\mathbf{x}(k)\|_2^2 + \varepsilon_n(k)}.$$

PNLMS with Robust Regularisation

- GPNLMS algorithm is by design stable and robust to initial settings of its parameters
- A disadvantage however is it also inherits some steady state stability problems of GNGD
- One method of overcoming this is to introduce noise into the updates, helping to avoid convergence to local and spurious minima.
- We therefore propose to improve the performance of the regularised PNLMS algorithms by introducing some gradient noise into the updates of the regularisation parameter ε, based on normalising the gradient of the cost function $E(k)$ with respect to ε,

$$\frac{\nabla_{\varepsilon(k-1)}E(k)}{\left\|\nabla_{\varepsilon(k-1)}E(k)\right\|_2} = \mathrm{sgn}\Big[\nabla_{\varepsilon(k-1)}E(k)\Big].$$

PNLMS with Robust Regularisation

Considering first the individual adaptive regularisation, this results in the sign regularised GIAPNLMS (SR-GIAPNLMS) regularisation parameter update given by

$$\varepsilon_n(k+1) = \varepsilon_n(k) - \beta \operatorname{sgn} \left\{ \frac{e(k)e(k-1)x_n(k)g_n(k-1)x_n(k-1)}{\left[\|\mathbf{x}(k-1)_2^2 + \varepsilon_n(k-1)\right]^2} \right\}.$$

Notice that the denominator, which is the source of much of the computational complexity, can be omitted since it is always positive, giving

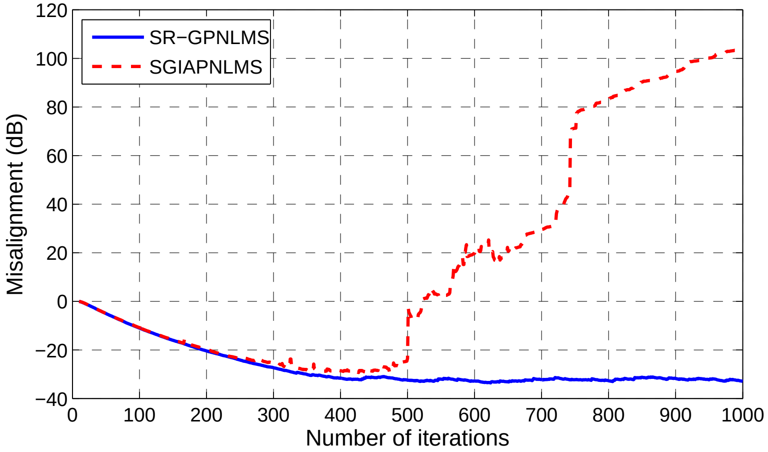
$$\varepsilon_n(k+1) = \varepsilon_n(k) - \beta \operatorname{sgn} \left[e(k)e(k-1)g_n(k-1)x_n(k)x_n(k-1) \right],$$

and for the sign regularised GPNLMS (SR-GPNLMS) this gives

$$\varepsilon(k+1) = \varepsilon(k) - \beta \operatorname{sgn} \left[e(k)e(k-1)\mathbf{x}^T(k)\mathbf{G}(k-1)\mathbf{x}(k-1) \right].$$

Analysis of SR-GIAPNLMS and SR-GPNLMS

Performance of the SR-GPNLMS and SR-GIAPNLMS, averaged over 100 independent trials, for a benchmark sparse system [6]



- Insufficient information within the individual updates of ε_n,
- reliancy on the instantaneous and possibly noisy values $x_n(k)$ and $x_n(k-1)$
- approaching steady state, changes in the sign of the error vector are also causing the value of ε_n to repeatedly fluctuate

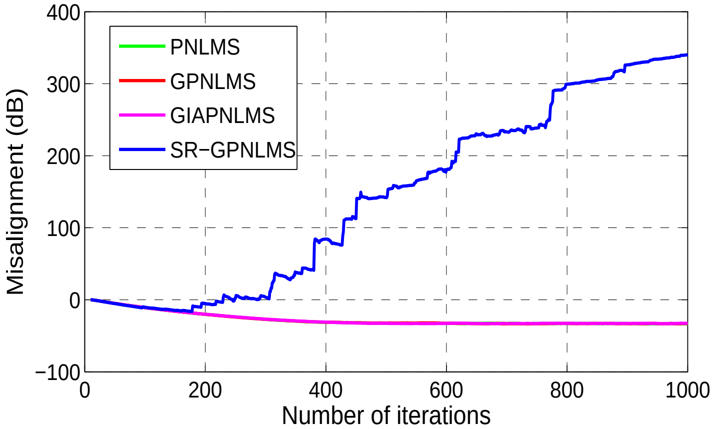
all combine, leading to instability in GIAPNLMS.

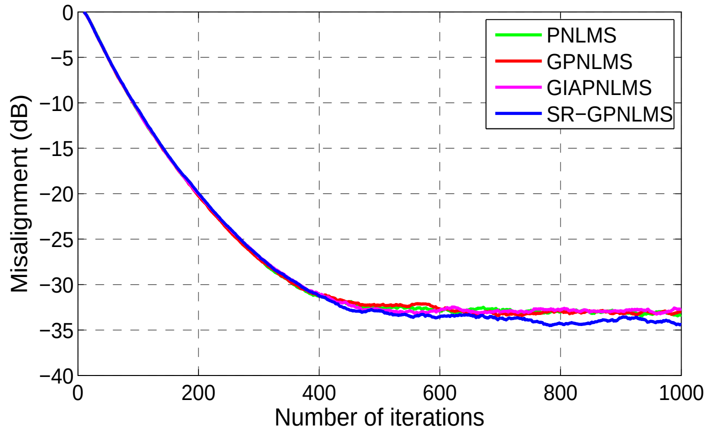
Analysis of SR-GIAPNLMS and SR-GPNLMS

- Considering the $\mathbf{G}(k-1)$ term in the GPNLMS update, omitting (as always positive) gives
$$\varepsilon(k+1) = \varepsilon(k) - \beta\,\mathrm{sgn}\left[e(k)e(k-1)\mathbf{x}^T(k)\mathbf{x}(k-1)\right].$$
- Noting the gradient vectors $\nabla_{\mathbf{w}} E(k) = -e(k)\mathbf{x}(k)$ and $\nabla_{\mathbf{w}} E(k-1) = -e(k-1)\mathbf{x}(k-1)$, and their inner product is given by
$$\nabla_{\mathbf{w}}^T E(k)\nabla_{\mathbf{w}} E(k-1) = \|\nabla_{\mathbf{w}} E(k)\| \cdot \|\nabla_{\mathbf{w}} E(k-1)\| \cdot \cos\theta.$$
- Since $\|\nabla_{\mathbf{w}} E(k)\|$ and $\|\nabla_{\mathbf{w}} E(k-1)\|$ are always positive, the update reduces to
$$\varepsilon(k+1) = \varepsilon(k) - \beta\,\mathrm{sgn}\left[\nabla_{\mathbf{w}}^T E(k)\nabla_{\mathbf{w}} E(k-1)\right] = \varepsilon(k) - \beta\,\mathrm{sgn}\left[\cos\theta\right].$$
- Therefore to cause a change in the sign of the ε update, we need $|\theta| > \pi/2$.
- $\mathbf{G}(k-1)$, has an effect on the direction of $\mathbf{x}(k-1)$ and hence the angle between the two gradients.
- The difference between $\mathbf{x}(k)$ and $\mathbf{x}(k-1)$ is generally small & slight changes to the direction of $\mathbf{x}(k-1)$ caused by $\mathbf{G}(k-1)$ are not significant.

Convergence Analysis

- $\varepsilon(k)$ has two competing unconstrained optimisation processes with different convergence speeds,
- One based on $\nabla_{\mathbf{w}}E$ and the other on $\nabla_{\varepsilon}E$
- It would benefit from constraints on the allowable values of their parameters,
- In practice for GPNLMS there was no need to include a minimum bound on ε.



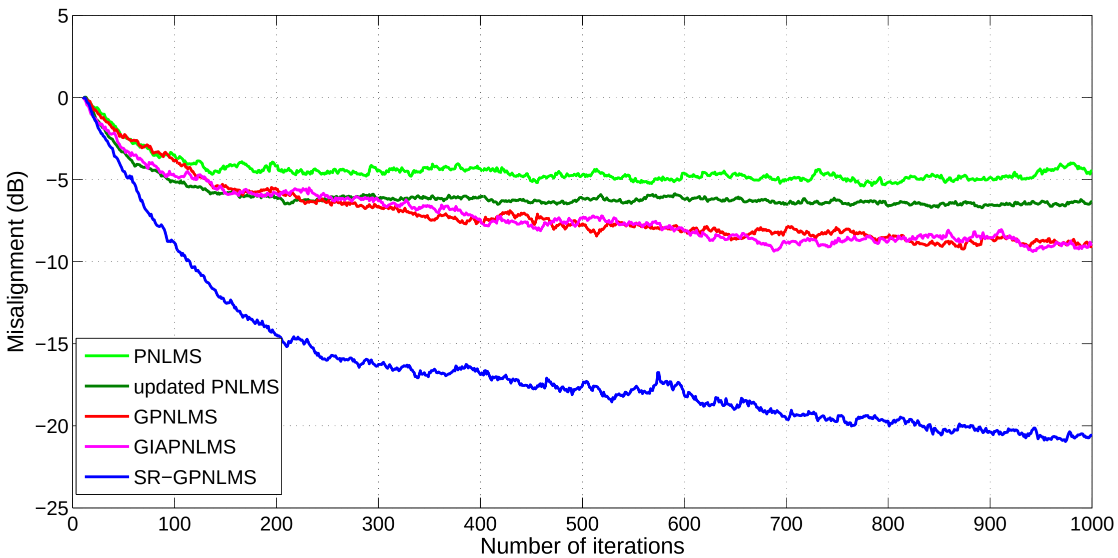


- SR-GPNLMS, has simplified (and hence noisy) updates of ε.
- In this case and due to its unconstrained nature the SR-GPNLMS diverged
- A lower bound of a lower bound of $\varepsilon_{\min} = 0$ had to be introduced

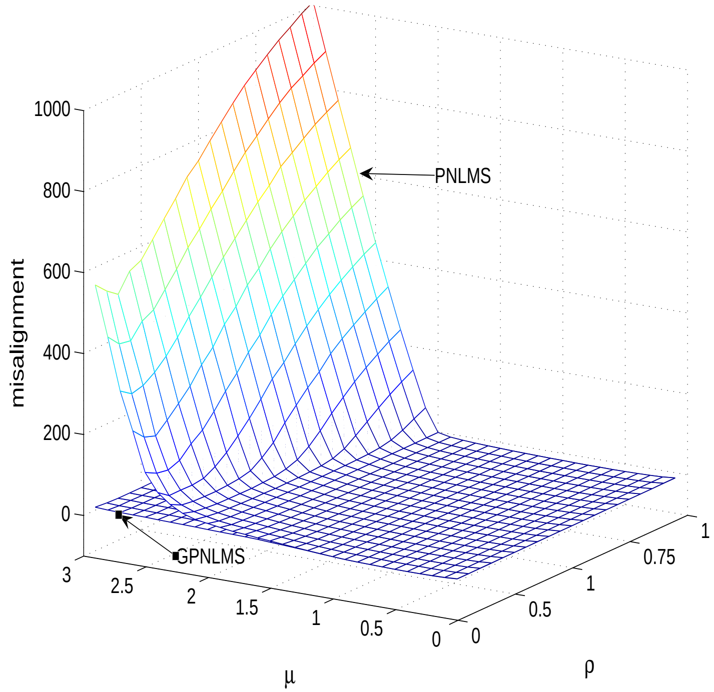
Critical Operating Conditions

- When $\|\mathbf{x}(k)\|_2^2 \longrightarrow 0$ then $\frac{\mu}{\|\mathbf{x}(k)\|_2^2+\varepsilon(k)} \longrightarrow \infty$
- To simulate this effect the value of μ is increased to $\mu = 1.95$, at which point PNLMS is on the limit of stability
- As well as comparing the algorithms with the original PNLMS the commonly used updated version was also implemented

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



Sensitivity of Algorithms to Their Parameters

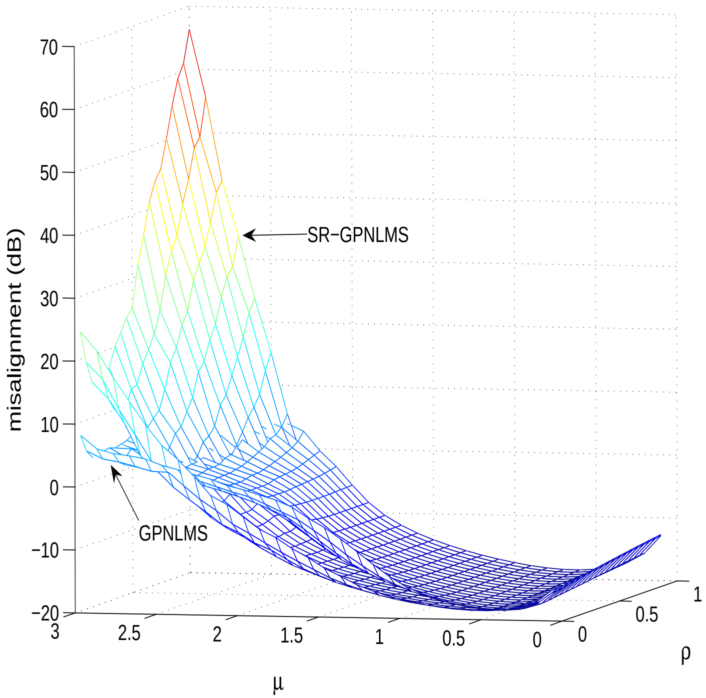


- ρ was varied in the range $[0, 1]$, for which the behaviour of the PNLMS approaches that of the NLMS.
- Changes in δ were not investigated, as only effects the algorithm initially when all filter coefficients are zero.
- The value of μ was varied in the range $[0, 3]$, with $\mu \in [2, 3)$ being outside the normal range of stability for the PNLMS.

The PNLMS and GPNLMS algorithms exhibited similar performance for $0 < \mu \leq 1.8$, but beyond this the difference between the two algorithms is clearly apparent with the PNLMS becoming extremely unstable.

Sensitivity of Algorithms to Their Parameters

Comparison of GPNLMS and SR-GPNLMS over the same ranges



- Both algorithms was clearly still under control
- For all settings with $\mu \leq 2.5$, SR-GPNLMS had similar or better performance than GPNLMS.
- For situations with $\mu > 2.5$, GPNLMS showed better stability than SR-GPNLMS, due to its more rigorous derivation,
- The SR-GPNLMS offers a viable alternative as a trade-off between performance and computational complexity

Conclusions

- A class of algorithms with gradient adaptive regularisation parameters has been proposed.
- To avoid problems experienced with PNLMS when processing inputs with large dynamics and ill conditioned tap input correlation matrices.
- The GIAPNLMS has been derived rigorously to give an individual update for each filter coefficient.
- In practice the improvement in performance offered over the GPNLMS did not compensate for the extra computational complexity.
- To overcome problems incurred in steady state & reduce computational complexity, a robust regularisation parameter has also been derived.
- Simulations show the proposed class of algorithms to be stable in a wide range of situations and robust to parameter initialization.

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. A. Benveniste, M. Metivier & P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag; 1990.
3. 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.
4. W.P. Ang WP & B. Farhang-Boroujeny "A new class of gradient adaptive step-size LMS algorithms," *IEEE Trans Signal Proc.*, vol. 49, no. 4, pp. 805–810, 2001.
5. D. Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Proc. Lett.*, vol. 11, pp. 115–118, Feb. 2004.
6. 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.
7. J. Benesty & S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Proc.(ICASSP)*, vol. 2, pp. 1881–1884, 2002.
8. M. Doroslovacki & H. Deng, "On convergence of proportionate-type NLMS adaptive algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Proc. (ICASSP)*, vol. 3, pp. 105–108, 2006.