PROPORTIONATE ADAPTIVE FILTERS BASED ON MINIMIZING DIVERSITY MEASURES FOR PROMOTING SPARSITY

Ching-Hua Lee, Bhaskar D. Rao, and Harinath Garudadri
Department of Electrical and Computer Engineering, University of California, San Diego

Contact Information:
Department of Electrical and Computer Engineering
University of California, San Diego
9500 Gilman Drive #0436, La Jolla, CA 92093

Email: chl438@ucsd.edu

Abstract

Objective:

- Propose a novel way of deriving proportionate adaptive filters that exploit sparsity in the underlying system response.

Methods:

- Diversity measure minimization using the iterative reweighting techniques [1, 2, 3] well-known in the sparse signal recovery (SSR) area.
- Affine scaling transformation (AST) [4] strategy commonly employed in the optimization literature.
- Limiting case: utilize a regularization coefficient $\lambda \to 0^+$.

Results

- Least mean square (LMS)-type and normalized LMS (NLMS)-type algorithms that can incorporate various diversity measures.
- Sparsity promoting LMS (SLMS) and Sparsity promoting NLMS (SNLMS) that realize proportionate adaptation similar to the proportionate NLMS (PNLMS) [5], but with a more systematic way of designing the step-size control factors based on SSR techniques rather than on heuristics.
- Simulation results demonstrate the flexibility of the algorithms to fit different sparsity levels of the systems.

1 Background

1.1 Adaptive Filters for System Identification in Figure. 1

Unconstrained optimization problem using instantaneous error:

$$\min_{\mathbf{h}} \quad J_n(\mathbf{h}) \triangleq e_n^2 = \left(d_n - \mathbf{u}_n^T\mathbf{h}\right)^2. \tag{1}$$

which leads to the well-known LMS and NLMS:

- LMS – apply the stochastic gradient descent:

$$\mathbf{h}_{n+1} = \mathbf{h}_n - \frac{\mu}{2}\nabla_{\mathbf{h}} J_n(\mathbf{h}_n) = \mathbf{h}_n + \mu\mathbf{u}_n e_n, \tag{2}$$

where $\mu > 0$ is the step size.

- NLMS – apply the stochastic regularized Newton's method:

$$\begin{aligned}\mathbf{h}_{n+1} &= \mathbf{h}_n - \mu\left(\nabla^2_{\mathbf{h}} J_n(\mathbf{h}_n) + 2\delta\mathbf{I}\right)^{-1}\nabla_{\mathbf{h}} J_n(\mathbf{h}_n)\\ &= \mathbf{h}_n + \frac{\mu\mathbf{u}_n e_n}{\mathbf{u}_n^T\mathbf{u}_n + \delta},\end{aligned} \tag{3}$$

where $\mu > 0$ is the step size and $\delta > 0$ is a small constant for regularization.

1.2 Diversity Measure Minimization for SSR

Finds sparse solutions to underdetermined $\mathbf{y} = \mathbf{A}\mathbf{x}$:

$$\min_{\mathbf{x}} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda G(\mathbf{x}), \qquad \lambda > 0, \tag{4}$$

where $G(\mathbf{x}) = \sum_{i=0}^{M-1} g(x_i)$ is the (separable) *general diversity measure* in which the function $g(\cdot)$ has to satisfy certain conditions [1].

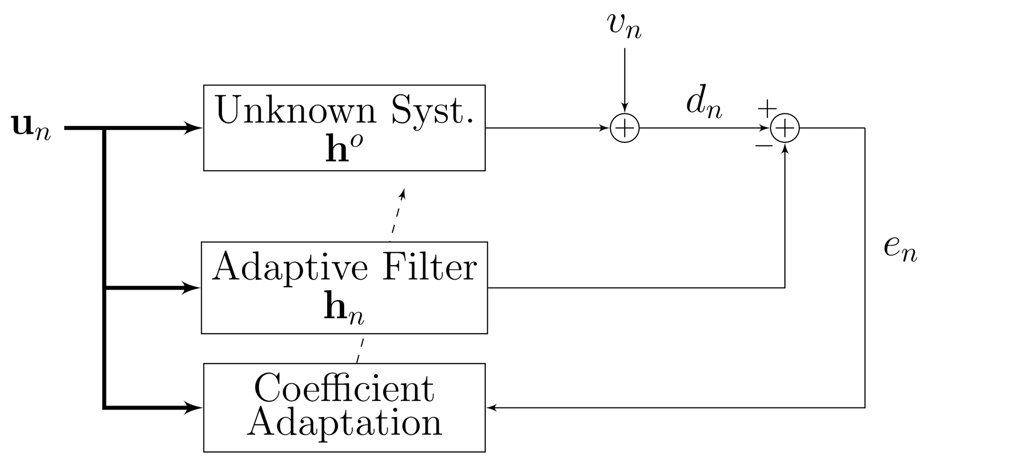

Figure 1: System identification block diagram. The adaptive filter $\mathbf{h}_n = [h_{0,n}, h_{1,n}, ..., h_{M-1,n}]^T$ is used to emulate the unknown system $\mathbf{h}^o$. $\mathbf{u}_n = [u_n, u_{n-1}, ..., u_{n-M+1}]^T$ is the input data vector. v_n is an additive noise. $e_n = d_n - \mathbf{u}_n^T\mathbf{h}_n$ is the error signal. The goal is to continuously adjust the coefficients of $\mathbf{h}_n$ such that $\mathbf{h}_n = \mathbf{h}^o$; i.e., to identify the unknown system.

Iterative reweighted ℓ_2 approach [2]: to use this approach the function $g(t)$ has to be concave in t^2; i.e., it satisfies $g(t) = f(t^2)$, where $f(z)$ is concave for $z \in \mathbb{R}_+$. It iteratively solves:

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\left\|(\mathbf{W}^{(k)})^{-1}\mathbf{x}\right\|_2^2, \tag{5}$$

where $\mathbf{W}^{(k)} = \text{diag}\{w_i^{(k)}\}$ with

$$w_i^{(k)} = \left(\left.\frac{\mathrm{d}f(z)}{\mathrm{d}z}\right|_{z=(x_i^{(k)})^2}\right)^{-\frac{1}{2}}, \tag{6}$$

and d denotes the differential operator.

2 Incorporating Sparsity into Adaptive Filters

We propose to consider the following optimization problem:

$$\min_{\mathbf{h}} \quad J_n(\mathbf{h}) + \lambda G(\mathbf{h}), \tag{7}$$

where $G(\mathbf{h}) = \sum_{i=0}^{M-1} g(h_i)$ and λ is the regularization coefficient.
As in the iterative reweighted ℓ_2 approach, we instead consider:

$$\min_{\mathbf{h}} \quad J_n(\mathbf{h}) + \lambda\left\|\mathbf{W}_n^{-1}\mathbf{h}\right\|_2^2, \tag{8}$$

where $\mathbf{W}_n = \text{diag}\{w_{i,n}\}$ and

$$w_{i,n} = \left(\left.\frac{\mathrm{d}f(z)}{\mathrm{d}z}\right|_{z=h_{i,n}^2}\right)^{-\frac{1}{2}}, \tag{9}$$

Consider the following reparameterization similar to AST [4]:

$$\mathbf{q} \triangleq \mathbf{W}_n^{-1}\mathbf{h}. \tag{10}$$

Use (10) for the objective function in (8) and perform minimization with respect to $\mathbf{q}$,:

$$\min_{\mathbf{q}} \quad J_n^{\ell_2}(\mathbf{q}) \triangleq J_n(\mathbf{W}_n\mathbf{q}) + \lambda\|\mathbf{q}\|_2^2. \tag{11}$$

Define the *a posteriori* AST variable at time n:

$$\mathbf{q}_{n|n} \triangleq \mathbf{W}_n^{-1}\mathbf{h}_n \tag{12}$$

and the *a priori* AST variable at time n:

$$\mathbf{q}_{n+1|n} \triangleq \mathbf{W}_n^{-1}\mathbf{h}_{n+1}. \tag{13}$$

Using the above equations we derive LMS-type and NLMS-type sparse adaptive filtering algorithms in the following.

2.1 LMS-Type Sparse Adaptive Filtering Algorithm

Apply stochastic gradient descent in the $\mathbf{q}$ domain:

$$\mathbf{q}_{n+1|n} = \mathbf{q}_{n|n} - \frac{\mu}{2}\nabla_{\mathbf{q}} J_n^{\ell_2}(\mathbf{q}_{n|n}). \tag{14}$$

This leads to:

$$\mathbf{q}_{n+1|n} = (1 - \mu\lambda)\mathbf{q}_{n|n} + \mu\mathbf{W}_n\mathbf{u}_n e_n. \tag{15}$$

Multiplying both sides of (15) by $\mathbf{W}(n)$ and using the relationships (12) and (13), we will get back to the $\mathbf{h}$ domain:

$$\boxed{\mathbf{h}_{n+1} = (1 - \mu\lambda)\mathbf{h}_n + \mu\mathbf{W}_n^2\mathbf{u}_n e_n.} \tag{16}$$

This is the update rule of *the generalized LMS-type sparse adaptive filtering algorithm using reweighted ℓ_2*.

2.2 NLMS-Type Sparse Adaptive Filtering Algorithm

Apply stochastic regularized Newton's method in the $\mathbf{q}$ domain:

$$\mathbf{q}_{n+1|n} = \mathbf{q}_{n|n} - \mu\left(\nabla^2_{\mathbf{q}} J_n^{\ell_2}(\mathbf{q}_{n|n}) + 2\delta\mathbf{I}\right)^{-1}\nabla_{\mathbf{q}} J_n^{\ell_2}(\mathbf{q}_{n|n}). \tag{17}$$

This will result in:

$$\boxed{\mathbf{h}_{n+1} = (\mathbf{I} - \mu\lambda\mathbf{\Phi}_n)\mathbf{h}_n + \frac{\mu\,\mathbf{W}_n^2\mathbf{u}_n e_n}{\mathbf{u}_n^T\mathbf{W}_n^2\mathbf{u}_n + \lambda + \delta},} \tag{18}$$

where for simplicity we have combined multiple terms into a single matrix $\mathbf{\Phi}_n$. This is the update rule of *the generalized NLMS-type sparse adaptive filtering algorithm using reweighted ℓ_2*.

3 Sparsity Promoting Algorithms

Considering the limiting case of $\lambda \to 0^+$ gives rise to the following Sparsity promoting LMS (SLMS):

$$\boxed{\mathbf{h}_{n+1} = \mathbf{h}_n + \mu\mathbf{W}_n^2\mathbf{u}_n e_n,} \tag{19}$$

and Sparsity promoting NLMS (SNLMS):

$$\boxed{\mathbf{h}_{n+1} = \mathbf{h}_n + \frac{\mu\,\mathbf{W}_n^2\mathbf{u}_n e_n}{\mathbf{u}_n^T\mathbf{W}_n^2\mathbf{u}_n + \delta}.} \tag{20}$$

A diagonal matrix $\mathbf{W}_n^2$ on the gradient to leverage sparsity – realizing proportionate adaptation.

Example of $\mathbf{W}_n$ update: employing the p-norm-like diversity measure with $g(h_i) = |h_i|^p$, $0 < p \le 2$. Using (9) leads to:

$$w_{i,n} = \left(\frac{2}{p}\left(|h_{i,n}| + c\right)^{2-p}\right)^{\frac{1}{2}}, \tag{21}$$

where $c > 0$ is a small constant added for stability purposes. The parameter p plays the role for fitting different sparsity levels:

- $p \to 1$ approximates the step-size control factors of PNLMS
- $p = 2$ recovers the LMS and NLMS (sparsity-unaware)

In practice, we replace $\mathbf{W}_n^2$ in (19) and (20) with $\mathbf{S}_n$ where:

$$\boxed{\mathbf{S}_n = \frac{\mathbf{W}_n^2}{\frac{1}{M}\text{tr}\left(\mathbf{W}_n^2\right)},} \tag{22}$$

which is found to help stabilize algorithms.

4 Simulation Results

Figure 2 shows three systems with different sparsity levels (left column): quasi-sparse, sparse, and dispersive (from top to bottom), and the corresponding mean squared error (MSE) learning curves of using SNLMS with various p values (right column).

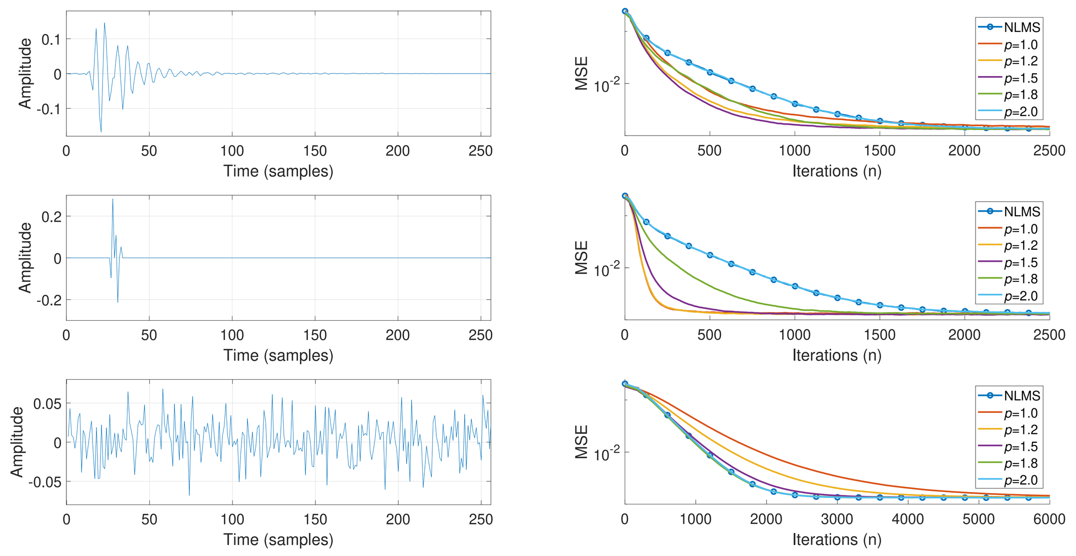

Figure 2: Learning curves of using SNLMS to identify systems with impulse responses of different sparsity levels. We see that the selection of p is crucial for obtaining optimal performance in different cases. For the quasi-sparse case, the fastest convergence is given by $p = 1.5$, which seems a reasonable value in terms of finding a balance between PNLMS ($p \to 1$) and NLMS ($p = 2$). For the sparse case, $p = 1.2$ gives the best results, which is also intuitive since the sparsity level has increased. For the dispersive case, $p = 1.8$ results in the fastest convergence and is comparable to NLMS. These results show that the algorithm exploits the underlying system structure in the way we expect. Note that since $\lambda = 0$ is utilized, the objective function in (7) exerts diminishing impact on enforcing sparsity on the solution, and the SNLMS converges toward the Wiener-Hopf solution as the NLMS. This shows that the proposed methods can leverage sparsity for speeding up convergence while not sacrificing estimation quality should sparsity be present.

5 Conclusion

We exploited the connection between sparse system identification and SSR, and utilized the iterative reweighting strategies to derive proportionate adaptive filters that incorporate sparsity. Moreover, utilizing $\lambda \to 0^+$, the proposed SLMS and SNLMS can take advantage of, though do not strictly enforce, the sparsity of the underlying system if it already exists.

6 Acknowledgements

This work was supported by NIH/NIDCD grants R01DC015436 and R33DC015046.

References

[1] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE J. Sel. Top. Signal Process.*, 2010.

[2] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. ICASSP*, 2008.

[3] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, 2008.

[4] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, 1999.

[5] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, 2000.