A Zero Attracting Proportionate Normalized Least Mean Square Algorithm

Rajib Lochan Das[1] and Mrityunjoy Chakraborty [2]
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur, INDIA
E.Mail : [1]rajib.das.iit@gmail.com, [2]mrityun@ece.iitkgp.ernet.in

Abstract— **The proportionate normalized least mean square (PNLMS) algorithm, a popular tool for sparse system identification, achieves fast initial convergence by assigning independent step sizes to the different taps, each being proportional to the magnitude of the respective tap weight. However, once the active (i.e., non-zero) taps converge, the speed of convergence slows down as the effective step sizes for the inactive (i.e., zero or near zero) taps become progressively less. In this paper, we try to improve upon both the convergence speed and the steady state excess mean square error (EMSE) of the PNLMS algorithm, by introducing a l_1 norm (of the coefficients) penalty in the cost function which introduces a so-called zero-attractor term in the PNLMS weight update recursion. The zero attractor induces further shrinkage of the coefficients, especially of those which correspond to the inactive taps and thus arrests the slowing down of the convergence of the PNLMS algorithm, apart from bringing down the steady state EMSE. We have also modified the cost function further generating a reweighted zero attractor which helps in confining the "Zero Attraction" to the inactive taps only.**

Index terms–Sparse Adaptive Filter, PNLMS Algorithm, RZA-NLMS algorithm, convergence speed, steady state performance.

I. Introduction

Sparse systems are encountered in several important applications like network echo cancelation, acoustic echo cancelation in hands free telephony, HDTV, wireless multipath channels and underwater acoustic communications. Conventional system identification algorithms like the LMS and the normalized LMS (NLMS) [1] do not, however, make use of the a priori knowledge of the sparseness of the system and thus perform poorly both in terms of steady state excess mean square error (EMSE) and convergence speed. This has resulted in a flurry of research activities [2] in the last decade for developing sparsity aware identification algorithms, prominent amongst them being the proportionate normalized LMS (PNLMS) algorithm [3] and its several variants [4]-[7]. In the PNLMS algorithm, each tap weight is updated independently by a step size that is proportional to the magnitude of the weight. This results in accelerated initial convergence of the algorithm though the rate of convergence becomes somewhat slower at a later stage. In a separate development, an alternative approach to identify sparse systems has been proposed [8], which introduces a l_1 norm (of the coefficients) penalty in the cost function which favors sparsity. This results in a modified LMS update equation with a zero attractor for all the taps, named as the Zero-Attracting LMS (ZA-LMS) algorithm. The presence of the zero attractor results in shrinkage of the coefficients, especially the inactive taps, giving rise to lesser steady state EMSE for sparse systems. In [8], the ZA-LMS algorithm was further modified to the so-called "Reweighted ZA-LMS" (RZA-LMS), where the zero attractors were reweighted by the inverse of the tap magnitude, thereby restricting the shrinkage mostly to the inactive taps only. In [9], the RZA concept was applied to the NLMS algorithm, in order to exploit the advantages that the NLMS algorithm offers vis-a-vis the LMS like faster rate of convergence, especially against colored input.

While both the PNLMS and the RZA-NLMS algorithms exploit the system sparsity to improve upon the identification performance, their working principles and objectives are, however, complementary to each other. In the case of the PNLMS algorithm, as the algorithm starts, the effective step sizes for the active taps grow rapidly which accelerates the convergence of the active taps, resulting in a very fast overall convergence rate during the initial period. Subsequently, i.e., after the active taps have converged, the rate of convergence of the algorithm gets governed primarily by the convergence rate of the inactive taps, which, however, becomes slower as the effective step sizes in this case become progressively less. The RZA-NLMS algorithm, on the other hand, aims at lesser steady state EMSE by bringing in an additional force in the update equation in the form of zero attractor which confines the inactive taps to a small range around zero value.

In this paper, we extend the RZA concept to the PNLMS algorithm in order to avail the benefits of both the PNLMS and the RZA-NLMS algorithms. In particular, once the initial period of fast convergence of the PNLMS algorithm is over, we aim to use the zero attractors as an additional force in order to "attract" the coefficients to zero. This will help in arresting the slowing down of the convergence of the PNLMS and also will lead to a lesser steady state EMSE. Towards this, we develop a ZA-PNLMS algorithm first, which is later modified to incorporate the reweighting of the zero attractors. Detailed computer simulations providing insights on the working of the proposed algorithm are also provided.

The paper is organized as follows : in section II, we present a brief overview of the PNLMS and the RZA-NLMS algorithms and in section III, the proposed algorithms are derived. Section IV presents a performance evaluation of the proposed algorithms by computer simulation.

II. A Brief Review of the PNLMS and the RZA-NLMS Algorithms

A. PNLMS Algorithm

Consider a PNLMS based adaptive filter that takes $x(n)$ as input and updates a N tap coefficient vector $\mathbf{w}(n) = [w_0(n), w_1(n), \cdots, w_{N-1}(n)]^T$ as [3],

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu \mathbf{G}(n)\mathbf{x}(n)e(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) + \delta_P}, \quad (1)$$

where $\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^T$ is the input regressor vector, $\mathbf{G}(n)$ is a diagonal matrix that modifies the step size of each tap, μ is the overall step size, δ_P is a regularization parameter and $e(n) = d(n) - \mathbf{w}^T(n)\mathbf{x}(n)$ is the filter output error, with $d(n)$ denoting the so-called desired response. In the system identification problem under consideration, $d(n)$ is the observed system output, given as $d(n) = \mathbf{w}_{opt}^T \mathbf{x}(n) + v(n)$, where $\mathbf{w}_{opt}$ is the system impulse response vector (supposed to be sparse), $x(n)$ is the system input and $v(n)$ is an observation noise which is assumed to be white with variance σ_v^2 and independent of $x(m)$ for all n and m.

The matrix $\mathbf{G}(n)$ is evaluated as

$$\mathbf{G}(n) = diag(g_0(n), g_1(n)....g_{N-1}(n)), \quad (2)$$

where,

$$g_l(n) = \frac{\gamma_l(n)}{\sum_{i=0}^{N-1} \gamma_i(n)}, \; 0 \le l \le N-1, \quad (3)$$

with

$$\gamma_l(n) = \max[\rho_g \max[\delta, \mid w_0(n) \mid, .. \mid w_{N-1}(n) \mid], \mid w_l(n) \mid]. \quad (4)$$

The parameter δ is an initialization parameter that helps to prevent stalling of the weight updating at the initial stage when all the taps are initialized to zero. Similarly, if an individual tap weight becomes very small, to avoid stalling of the corresponding weight update recursion, the respective $\gamma_l(n)$ is taken as a ρ_g fraction of the largest tap magnitude. By providing separate step size to each tap, the PNLMS algorithm achieves a very fast initial convergence rate (for highly sparse systems), but this high rate is not maintained at a later stage of the adaptation process [5],[6].

B. RZA-NLMS Algorithm

The filter coefficient update equation of the RZA-NLMS algorithm is given by [9]

$$\begin{aligned} w_i(n+1) &= w_i(n) + \frac{\mu x(n-i+1)e(n)}{\mathbf{x}^T(n)\mathbf{x}(n) + \delta_N} \\ &- \rho \frac{sgn(w_i(n))}{1 + \varepsilon \mid w_i(n) \mid}, \; i = 0, 1, \cdots, N-1, \end{aligned} \quad (5)$$

where $sgn(.)$ is the well known signum function (i.e., $sgn(x) = 1 \; (x > 0), \; 0 \; (x = 0), \; -1 \; (x < 0)$ and δ_N is the so-called regularization parameter to avoid a division by zero. The last term of (5), named as reweighted zero attractor, provides a selective shrinkage to the taps. Due to this reweighted zero attractor, the inactive taps with zero magnitudes or magnitudes comparable to $1/\varepsilon$ undergo higher shrinkage compared to the active taps which enhances the performance both in terms of convergence speed and steady state EMSE.

III. Proposed Algorithm

We consider the following constrained optimization problem based on the principle of minimum disturbance with an l_1 norm regularization as shown below :

$$\min_{\mathbf{w}(n+1)} \| \mathbf{w}(n+1) - \mathbf{w}(n) \|_{\mathbf{G}^{-1}}^2 + \gamma \| \mathbf{G}^{-1}\mathbf{w}(n+1) \|_1 \quad (6)$$

subject to

$$d(n) - \mathbf{w}^T(n+1)\mathbf{x}(n) = 0, \quad (7)$$

where γ is a very very small constant and the short form notation "$\mathbf{G}^{-1}$" is used to indicate $\mathbf{G}^{-1}(n)$ (also, the notation $\| \mathbf{x} \|_{\mathbf{A}}^2$ indicates the generalized inner product $\mathbf{x}^T\mathbf{A}\mathbf{x}$). Note that in (6), we have introduced a l_1 norm penalty of $\mathbf{w}(n+1)$ after scaling its elements by $\mathbf{G}^{-1}(n)$ (the above scaling makes the l_1 norm penalty governed primarily by the inactive taps). Using Lagrange multiplier λ, the cost function for the above minimization problem can be defined as

$$\begin{aligned} J(n+1) = &\| \mathbf{w}(n+1) - \mathbf{w}(n) \|_{\mathbf{G}^{-1}}^2 + \gamma \| \mathbf{G}^{-1}\mathbf{w}(n+1) \|_1 \\ &+ \lambda(d(n) - \mathbf{w}^T(n+1)\mathbf{x}(n)) \end{aligned} \quad (8)$$

Setting $\partial J / \partial \mathbf{w}(n+1) = 0$ and $\partial J / \partial \lambda = 0$, we get

$$\mathbf{w}(n+1) = \mathbf{w}(n) - [\gamma sgn(\mathbf{w}(n+1)) - \lambda \mathbf{G}(n)\mathbf{x}(n)] \quad (9)$$

and

$$d(n) = \mathbf{w}^T(n+1)\mathbf{x}(n) \quad (10)$$

From (9) and (10), we can solve for the Lagrange multiplier λ as

$$\lambda = \frac{e(n) + \gamma \mathbf{x}^T(n) sgn(\mathbf{w}(n+1))}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)} \quad (11)$$

Replacing λ in (9) by above,

$$\begin{aligned} \mathbf{w}(n+1) = \mathbf{w}(n) &+ \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)} \\ &- \gamma \left[I - \frac{\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{G}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)} \right] sgn(\mathbf{w}(n+1)) \end{aligned} \quad (12)$$

The above equation does not provide the desired weight update relation, as the R.H.S. contains the unknown term $sgn(\mathbf{w}(n+1))$. In order to obtain a feasible weight update equation, we approximate $sgn(\mathbf{w}(n+1))$ by an estimate, namely, $sgn(\mathbf{w}(n))$ which is known. This is based on the assumption that most of the weights do not undergo change of sign as they get updated. This assumption may not, however, appear to be a very accurate one, especially for the inactive taps that fluctuate around zero value in the steady state. Nevertheless, an analysis of the proposed algorithm (not covered in this paper) shows that this approximation does not affect the convergence behavior of the proposed algorithm.

Making the above substitution in (12),

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)}$$
$$-\gamma \left[I - \frac{\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{G}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)} \right] sgn(\mathbf{w}(n)). \quad (13)$$

It is easy to see that the elements of the matrix $\frac{\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{G}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)}$ have magnitudes much less than 1, especially for large order filters. Neglecting this in comparison to the identity matrix, introducing algorithm step size μ and a regularization parameter δ_P, for a large order adaptive filter, the above update equation may be written as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) + \delta_P} - \rho sgn(\mathbf{w}(n)) \quad (14)$$

where $\rho = \mu\gamma$. In (14), which describes the proposed ZA-PNLMS algorithm, the second term on the R.H.S. is the usual PNLMS update term whereas the last term acts as the zero attractor.

Next we extend the reweighting concept to (14) to obtain a RZA version of the proposed ZA-PNLMS algorithm. For this, we replace the l_1 regularization term $\parallel \mathbf{G}^{-1}\mathbf{w}(n+1) \parallel_1$ in (8) by a log-sum penalty $\sum_{i=1}^{N} \frac{1}{g_i(n)} log(1+ \mid w_i(n+1) \mid /\epsilon)$ where $g_i(n)$ is the i-th diagonal element of $\mathbf{G}(n)$ and follow the same procedure resulting in the RZA-PNLMS weight update equation as given by

$$\begin{aligned} w_i(n+1) &= w_i(n) + \frac{\mu g_i(n) x(n-i+1) e(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) + \delta_P} \\ &- \rho \frac{sgn(w_i(n))}{1+\varepsilon \mid w_i(n) \mid}, \; i = 0, 1, \cdots, N-1. \end{aligned} \quad (15)$$

IV. Numerical Simulations

The proposed algorithm was simulated for identifying a sparse system with impulse response shown in Fig. 1. The system has a total of 120 taps / impulse response coefficients of which only 4 are active, i.e., non-zero and the system input $x(n)$ was taken to be a zero mean, unit variance white random process.

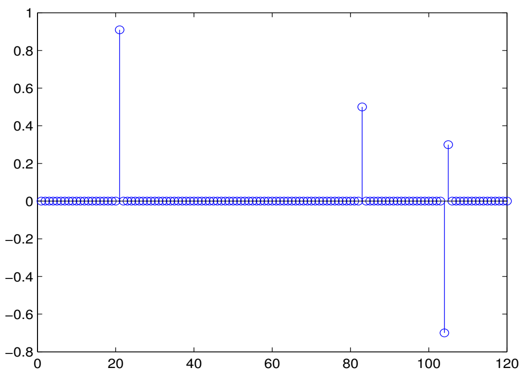

Fig. 1. Impulse response of the sparse system

Both the proposed ZA-PNLMS and the RZA-PNLMS algorithms were simulated and compared against the PNLMS algorithm. The simulation was carried out for a total of 4000 iterations, with $\mu = 0.5$, $\rho = 0.00001$, $\epsilon = 10$, $\delta_P = 0.01$, $\rho_g = 0.01$ and $\sigma_v^2 = 0.001$. For all the three algorithms, μ was chosen to be the same to maintain the same initial rate of convergence. The simulation results are shown in Fig. 2 by plotting the EMSE against the iteration index n (obtained by averaging $e^2(n)$ over 200 experiments) for the proposed RZA-PNLMS (green line) and ZA-PNLMS (red line) algorithms against the standard PNLMS (pink line) algorithm. It is easily seen that even though all the three algorithms start with identical, fast convergence rate, the convergence of the PNLMS algorithm slows down after about 250 iterations or so. This implies that in the case of the PNLMS algorithm, the active taps converge very fast, in about 250 iterations, but the convergence of the inactive taps slows down as the effective step sizes for them become less and less progressively. In the case of the RZA-PNLMS and ZA-PNLMS algorithms, however, the inactive taps come under the influence of an additional force, exerted by the zero attractors, which try to pull them towards their true value, i.e., zero. As a result, both the RZA-PNLMS and ZA-PNLMS algorithms retain much of their initial fast convergence rates even after $n = 250$. Apart from this, both the RZA-PNLMS and the ZA-PNLMS algorithms also maintain lesser steady state EMSE vis-a-vis the PNLMS algorithm, as seen easily from Fig. 2. Note that from Fig. 2, it may appear that the performance of both the RZA-PNLMS and the ZA-PNLMS algorithms are identical. This, however, happens when the active taps have sufficiently large magnitudes, as, in such cases, the effect of the zero attractors on the active taps after their convergence is negligible and as a result, one gets similar performance irrespective of whether the action of the zero attractors on the active taps is present (as in the ZA-PNLMS algorithm) or suppressed (as in the RZA-PNLMS algorithm).

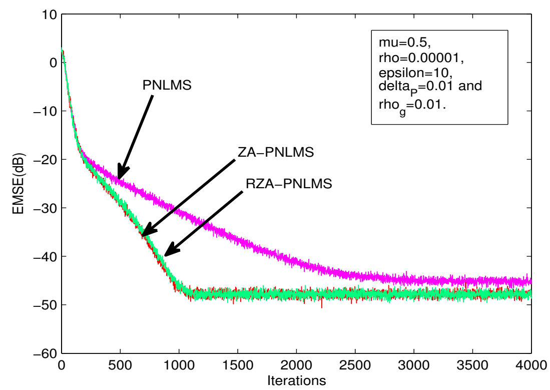

Fig. 2. The EMSE versus number of iterations for the standard PNLMS (pink), the RZA-PNLMS (green) and the ZA-PNLMS (red) algorithms.

The convergence behavior of the proposed algorithms is also sensitive to the choice of the parameter ρ_g. We demonstrate this in Fig. 3 by considering the ZA-PNLMS algorithm (we do not consider the RZA-PNLMS algorithm here in order to avoid crowding, after noting from above that its performance is almost identical to that of the ZA-PNLMS algorithm) for $\rho_g = 0.01, 0.05, 0.1$. For comparison, we also plot the learning curve (EMSE-vs-iteration index n) of the PNLMS algorithm for $\rho_g = 0.01$. It is seen that the steady state EMSE of the ZA-PNLMS algorithm decreases as ρ_g increases. This can be easily explained by first noting from (4) that in the steady state, as the inactive tap weights attain values very close to zero, the corresponding $\gamma_l(n)$ is given by ρ_g times the maximum tap weight magnitude. From this and the fact that $\sum_{i=0}^{N-1} g_l(n) = 1$, it follows that as ρ_g increases, the gain $g_l(n)$ and thus the effective step sizes for the active taps decrease. As a result, their contribution to the EMSE decreases (for the inactive taps, however, the marginal increase in the EMSE that an increasing ρ_g could give rise to is offset by the zero attractors). Of course, the reduction in the effective step sizes for the active taps tries to slow down the initial fast convergence somewhat. However, as can be seen from Fig. 3, such slowing down effect is marginal.

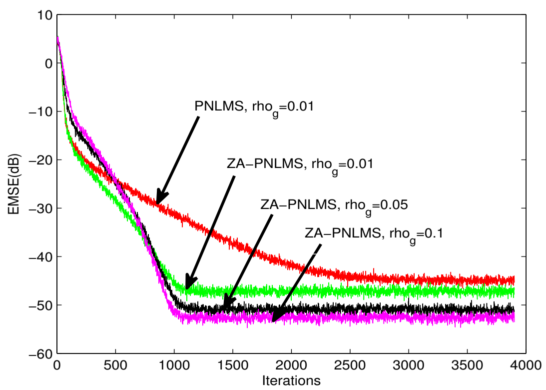

Fig. 3. The effect of ρ_g on the steady state EMSE of the ZA-PNLMS algorithm.

V. Conclusion

In this paper, we have presented a new sparse adaptive filter algorithm, namely, the zero attracting PNLMS and its reweighted version by introducing a l_1 norm penalty in the cost function of the standard PNLMS algorithm. The proposed techniques outperform the PNLMS algorithm in terms of both the convergence speed and the steady state EMSE, which is also verified by extensive simulation studies.

References

[1] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1986.

[2] R. L. Das and M. Chakraborty, "Sparse Adaptive Filters - an Overview and Some New Results", *Proc. ISCAS-2012*, pp. 2745-2748, COEX, Seoul, Korea, May 20-23, 2012.

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

[4] M. A. Mehran Nekuii, "A Fast Converging Algorithm for Network Echo Cancelation", *IEEE Signal Processing Letters*, vol. 11, no. 4, pp. 427-430, April, 2004.

[5] J. Benesty and S.L. Gay,"An improved PNLMS algorithm", *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing* , pp. 1881-1884, 2002, Orlando, Florida, USA.

[6] H. Deng and M. Doroslovacki, "Improving convergence of the PNLMS algorithm for sparse impulse response identification", *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 181-184, 2005.

[7] P. Loganathan, A. W. H. Khong, P. A. Naylor, "A Class of Sparseness-Controlled Algorithms for Echo Cancellation", *IEEE Trans. Audio, Speech Language Process.*, vol. 17, no. 11, pp. 1591-1601, Nov., 2009.

[8] Y. Chen, Y. Gu and A.O. Hero,"Sparse LMS for system identification" , *Proc. IEEE ICASSP-2009*, pp. 3125-3128, Apr. 2009, Taipei, Taiwan.

[9] Y. Chen, Y. Gu, and A.O. Hero,"Regularized Least-Mean-Square Algorithms", *Arxiv preprint arXiv:1012.5066v2[stat.ME]*, Dec. 2010.