

# Sparse Adaptive Filtering by Iterative Hard Thresholding

Rajib Lochan Das<sup>1</sup> and Mrityunjoy Chakraborty<sup>2</sup>

Department of Electronics and Electrical Communication Engineering

Indian Institute of Technology, Kharagpur, INDIA

E.Mail : <sup>1</sup>rajib.das.iit@gmail.com, <sup>2</sup>mrityun@ece.iitkgp.ernet.in

**Abstract**— In this paper, we present a new algorithm for sparse adaptive filtering, drawing from the ideas of a greedy compressed sensing recovery technique called the iterative hard thresholding (IHT) and the concepts of affine projection. While usage of affine projection makes it robust against colored input, the use of IHT provides a remarkable improvement in convergence speed over the existing sparse adaptive algorithms. Further, the gains in performance are achieved with very little increase in computational complexity.

**Index terms**— Sparse Adaptive Filter, Compressed Sensing, Iterative Hard Thresholding, Affine Projection, PNLMS

## I. INTRODUCTION

Compressed sensing or compressive sampling (CS)[1] is a powerful technique to represent signals at a sub-Nyquist sampling rate while retaining the capacity of perfect (or near perfect) reconstruction of the signal, provided the signal is known to be sparse in some domain. In last few years, the CS technique has attracted considerable attention from across a wide array of fields like applied mathematics, statistics, and engineering, including signal processing areas like MR imaging, speech processing, analog to digital conversion etc. The framework of CS essentially leads to finding sparse solution of a set of under-determined linear equations. This makes it a potential tool for estimation of sparse signals and systems which are often encountered in many important practical applications like network and acoustic echo cancellers [2]-[4] where the acoustic echo path largely remains inactive due to bulk delay, HDTV where clusters of dominant echoes arrive after long periods of silence [5], wireless multipath channels which, on most of the occasions, have only a few clusters of significant paths [6], and acoustic channels in shallow underwater communication where the various multipath components caused by reflections off the sea surface and sea bed have long intermediate delays [7].

A potential challenge in the usage of the CS technique for sparse system identification, however, comes from the fact that the CS method is essentially an offline and batch-based procedure, while the algorithm for identifying the system needs to be adaptive and thus online, as the system parameters including its sparsity level usually vary with time. Last few years have, however, seen several efforts to combat this challenge and several sparse adaptive filters based on the CS recovery concepts have been proposed. In this paper, we present a new algorithm for sparse system identification,

drawing from the concepts of a recently proposed greedy CS recovery technique called the Iterative Hard Thresholding (IHT) algorithm [10]. A new sparse adaptive filter is presented towards this by combining the IHT concepts with the ideas of affine projection [11]. While usage of affine projection makes it robust against colored input, the use of IHT provides a remarkable improvement in convergence speed over the existing, popular sparse adaptive filters like the PNLMS [26] algorithm. Further, the above gains in performance are achieved with very little increase in computational complexity.

## II. BRIEF OVERVIEW OF COMPRESSED SENSING AND THE IHT ALGORITHM

Let a real valued, bandlimited signal be sampled following Nyquist sampling rate condition and over a finite observation interval, generating the  $N \times 1$  signal vector  $\mathbf{u} = (u_1, u_2, \dots, u_N)^T$ . The vector  $\mathbf{u}$  is known to be sparse in some transform domain. More specifically, if  $\Psi$  be the  $N \times N$  transform matrix (usually unitary) and  $\mathbf{x} \in \mathbb{R}^N$  be the transform coefficient vector, i.e.,  $\mathbf{u} = \Psi\mathbf{x}$ , then  $\mathbf{x}$  is known a priori to be  $K$  sparse, meaning a maximum of  $K$  no of terms in  $\mathbf{x}$  can be non-zero. According to the CS theory, it is then possible to replace the  $N$  samples  $u_i$ ,  $i = 1, 2, \dots, N$  by a set of  $M$  samples  $y_j$ ,  $j = 1, 2, \dots, M$ ,  $M < N$ , linearly related to  $u_i$ 's, while retaining the capacity to reconstruct  $\mathbf{u}$  correctly. Defining  $\mathbf{y} = [y_1, y_2, \dots, y_M]^T$  and a  $M \times N$  sensing matrix  $\mathbf{A}$ , this implies  $\mathbf{y} = \mathbf{A}\mathbf{u} = \Phi\mathbf{x}$ , where  $\Phi = \mathbf{A}\Psi$ . Under the  $K$ -sparsity assumption, as per the CS theory,  $\mathbf{x}$  can be reconstructed by solving the following  $l_0$  minimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \Phi\mathbf{x}. \quad (1)$$

[Note that uniqueness of the  $K$ -sparse solution requires that every  $2K$  columns of  $\Phi$  should be linearly independent.] The above  $l_0$  minimization problem provides the sparsest solution for  $\mathbf{x}$ . However, the  $l_0$  minimization problem is a non-convex problem and is NP-hard. In CS, this difficulty is, however, overcome by replacing the  $l_0$  norm in (1) by  $l_1$  norm and imposing certain “Restricted Isometry Property (RIP)” condition of appropriate order on  $\Phi$ . A matrix  $\Phi$  is said to satisfy the RIP of order  $K$  if there exists a “Restricted Isometry Constant”  $\delta_K \in (0, 1)$  so that

$$(1 - \delta_K) \|\mathbf{x}\|_2^2 \leq \|\Phi\mathbf{x}\|_2^2 \leq (1 + \delta_K) \|\mathbf{x}\|_2^2 \quad (2)$$

for all  $K$ -sparse  $\mathbf{x}$ . The constant  $\delta_K$  is taken as the smallest number from  $(0, 1)$  for which the RIP is satisfied. It is easy to see that if  $\Phi$  satisfies RIP of order  $K$ , then it also satisfies RIP for any order  $L$  where  $L < K$  and that  $\delta_K \geq \delta_L$ . Simple choice of a random matrix for  $\Phi$  can make it satisfy the RIP with high probability.

Given the underdetermined system  $\mathbf{y} = \Phi\mathbf{x}$  with  $\mathbf{x}$  given to be  $K$ -sparse and  $\Phi$  satisfying RIP of order  $K$ , several methods exist for the recovery of  $\mathbf{x}$ . These include convex relaxation techniques like the basis pursuit (BP) [12], the basis pursuit de-noising (BPDN) [13] and the LASSO [14] on one hand and greedy methods like the orthogonal matching pursuit (OMP) [15], the compressive sampling matching pursuit (CoSaMP) [16], the subspace pursuit (SP) [17] etc. on the other. Of these, the greedy methods have a special relevance to sparse adaptive filtering. This approach recovers the  $K$ -sparse signal by iteratively constructing the support set of the sparse signal (index of non-zero elements in the sparse vector). At each iteration, it updates its support set by appending the index of one or more columns (called atoms) of the matrix  $\Phi$  (often called dictionary) by some greedy principles based on the correlation between current residual of observation vector and the atoms.

The iterative hard thresholding (IHT) algorithm [8]-[10] is a newly proposed greedy type algorithm which uses gradient descent followed by a hard thresholding as given below

$$\mathbf{x}(n+1) = H_K (\mathbf{x}(n) + \mu \Phi^T (\mathbf{y} - \Phi \mathbf{x}(n))) \quad (3)$$

where  $\mathbf{x}(n)$  is the  $n$ -th iterate of  $\mathbf{x}$ ,  $\mu$  is the step size of the gradient descent and  $H_K$  is a hard thresholding operator that sets all but the  $K$  largest (in magnitude) elements in a vector to zero. The IHT algorithm provides guarantees of near optimal recovery under the following constraint on the RIP parameter :  $\delta_{3K} < \frac{1}{\sqrt{32}} \approx 0.177$ .

### III. OVERVIEW OF SPARSE ADAPTIVE ALGORITHMS BASED ON COMPRESSED SENSING

The CS method has influenced several researchers in recent past to develop adaptive filters for identifying sparse systems. In [18], Chen *et al*, motivated by LASSO, introduced two different sparsity constraints (the  $l_1$  norm and the log-sum penalty function) into the convex quadratic cost function of the LMS algorithm, resulting in two sparsity aware LMS algorithms, namely, the zero attracting LMS (ZA-LMS) and the reweighted zero attracting LMS (RZA-LMS) algorithms. It is shown in [18] that both the ZA-LMS and RZA-LMS algorithms outperform the standard LMS w.r.t both transient and steady state performance for sparse systems.

Separately, the LASSO has influenced the development of several RLS based sparse adaptive filters as well. In [19], a  $l_1$  norm penalty similar to [18] is introduced into the cost function of the standard RLS algorithm, which is then minimized by an expectation maximization (EM) type algorithm. In [20]-[21], Angelosante *et al* developed an algorithm for recursively generating weighted LASSO estimates using a system of normal equations or by using iterative subgradient methods. Separately, instead of using the  $l_1$  norm penalty in

the cost function (popularly called “ $l_1$  norm regularization”), an alternate approach based on projection on appropriately constructed closed convex sets and certain weighted  $l_1$  balls at each index of time was presented in [22], which is reported to have better performance than the above stated, LASSO inspired RLS algorithms, both in terms of transient convergence speed and steady state EMSE. In [23], a mixed norm  $l_{1,\infty}$  regularizer has been adopted in the recursive setting of RLS algorithm to promote group sparsity for on-line identification of group sparse system. Greedy RLS (GRLS)[24] is a recently proposed RLS based sparse adaptive algorithm which is derived from the orthogonal least squares batch algorithm, providing better performance for sparse system while requiring less complexity than the full RLS algorithm.

As regards to the MP, [25] proposed the so-called “Sparse Adaptive Orthogonal Matching Pursuit (SpAdOMP)” algorithm, which converts the greedy batch algorithms of [16] and [17] into equivalent online procedures, requiring linear complexity. The steady state MSE of the proposed SpAdOMP algorithm is also evaluated analytically in [25].

### IV. PROPOSED ALGORITHM

We consider an  $N$ -tap sparse system, with impulse response given as  $\mathbf{w}_{opt} = [w_{opt,1} \ w_{opt,2} \ \dots \ w_{opt,N}]^T$ , which is adaptively identified using a zero-mean white input sequence  $u(n)$  with variance  $\sigma_u^2$  and a desired response signal  $d(n) = \mathbf{u}^T(n)\mathbf{w}_{opt} + v(n)$  where  $\mathbf{u}(n) = [u(n) \ u(n-1) \ \dots \ u(n-N+1)]^T$  and  $v(n)$  is a zero-mean, observation noise with variance  $\sigma_v^2$ , independent with  $u(m)$  for all  $n, m$ . The adaptation process updates a  $N \times 1$  filter coefficient vector  $\mathbf{w}(n)$ . Defining the filter output error at index  $m, n-L+1 \leq m \leq n$ , induced by  $\mathbf{w}(n)$ , as  $e_{\mathbf{w}(n)}(m) = d(m) - \mathbf{u}^T(m)\mathbf{w}(n)$ , we consider a sliding window based cost function,

$$J_{\mathbf{w}(n)}(n) = \sum_{\tau=0}^{L-1} e_{\mathbf{w}(n)}^2(n-\tau) \equiv \|\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n)\|_2^2, \quad (4)$$

where  $\mathbf{d}_n = [d(n) \ d(n-1) \ \dots \ d(n-L+1)]^T$  and  $\mathbf{U}_n = [\mathbf{u}(n) \ \mathbf{u}(n-1) \ \dots \ \mathbf{u}(n-L+1)]^T$ . The filter weight vector  $\mathbf{w}(n)$  at index  $n$  is updated to an intermediate vector  $\mathbf{w}'(n+1)$  following a gradient descent search on  $J_{\mathbf{w}(n)}(n)$  as given below :

$$\begin{aligned} \mathbf{w}'(n+1) &= \mathbf{w}(n) - \rho \frac{\partial}{\partial \mathbf{w}(n)} (J_{\mathbf{w}(n)}(n)) \\ &= \mathbf{w}(n) + \rho \mathbf{U}_n^T (\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n)) \end{aligned} \quad (5)$$

where  $\rho$  is chosen such that norm of the a posteriori error vector for the  $L$  indices :  $n, n-1, \dots, n-L+1$  is minimized i.e.,

$$\rho = \operatorname{argmin}_{\rho} \|\mathbf{d}_n - \mathbf{U}_n \mathbf{w}'(n+1)\|_2^2 \quad (6)$$

Now, this cost function may be simplified as

$$\begin{aligned} J_{\rho} &= \|\mathbf{d}_n - \mathbf{U}_n \mathbf{w}'(n+1)\|_2^2 \\ &= \|(\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n)) - \rho \mathbf{U}_n \mathbf{U}_n^T (\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n))\|_2^2 \\ &= \|(\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n))\|_2^2 - 2\rho \|\mathbf{U}_n^T (\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n))\|_2^2 \\ &\quad + \rho^2 \|\mathbf{U}_n \mathbf{U}_n^T (\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n))\|_2^2 \end{aligned} \quad (7)$$

Therefore, solving  $\frac{\partial J_\rho}{\partial \rho} = 0$  we find

$$\rho = \frac{\|\mathbf{U}_n^T(\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n))\|_2^2}{\|\mathbf{U}_n \mathbf{U}_n^T(\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n))\|_2^2} \quad (8)$$

For smooth update, we introduce a small positive step-size  $\mu(0 < \mu \leq 1)$ , leading to

$$\mathbf{w}'(n+1) = \mathbf{w}(n) + \mu \rho \mathbf{U}_n^T(\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n)). \quad (9)$$

Next we construct the active support set  $\Lambda_n^K$  of  $\mathbf{w}'(n+1)$  by selecting the support of  $K$  largest (in magnitude) elements in  $\mathbf{w}'(n+1)$ . The final update  $\mathbf{w}(n+1)$  is then obtained by applying a hard thresholding on  $\mathbf{w}'(n+1)$ , whereby elements of  $\mathbf{w}'(n+1)$  that fall within the ambit of  $\Lambda_n^K$  are retained while others are hardlimited to zero. As the active support set gets refined over iteration, the hard thresholding progressively reduces the accumulated gradient noise due to the inactive taps. The steps of the proposed algorithm are given in TABLE I.

TABLE I

**Algorithm 1:** Hard Thresholding based Adaptive Filtering (HTAF) Algorithm

**Input:**  $u(n)$ ,  $d(n)$ ,  $N$  (length of the adaptive filter),  $L$  (width of the sliding window) and  $K$  (the sparsity level).

**Initialization:** Estimated filter weight  $\mathbf{w}(0) = \mathbf{0}$  and the iteration count  $n = 0$

**Procedure:**

1) Desired data vector, regressors vector and corresponding sensing matrix:

$$\begin{aligned} \mathbf{d}_n &= [d(n) \ d(n-1) \ \dots \ d(n-L+1)]^T, \\ \mathbf{u}(n) &= [u(n) \ u(n-1) \ \dots \ u(n-N+1)]^T \text{ and} \\ \mathbf{U}_n &= [\mathbf{u}(n) \ \mathbf{u}(n-1) \ \dots \ \mathbf{u}(n-L+1)]^T. \end{aligned}$$

2) Gradient Descent Directional Update:

$$\mathbf{w}'(n+1) = \mathbf{w}(n) + \mu \frac{\|\mathbf{U}_n^T(\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n))\|_2^2}{\|\mathbf{U}_n \mathbf{U}_n^T(\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n))\|_2^2} \times \mathbf{U}_n^T(\mathbf{d}_n - \mathbf{U}_n \mathbf{w}(n))$$

3) Active Support Set:

$\Lambda_n^K$  = Support of  $K$  largest (in magnitude) elements in  $\mathbf{w}'(n+1)$

4) Hard Thresholding:

$$\begin{aligned} \mathbf{w}(n+1) &= H_K(\mathbf{w}'(n+1)) \text{ i.e.} \\ \mathbf{w}(n+1)_{\{1, \dots, N\} - \Lambda_n^K} &= 0 \text{ and} \\ \mathbf{w}(n+1)_{\Lambda_n^K} &= \mathbf{w}'(n+1)_{\Lambda_n^K} \end{aligned}$$

5) Increment  $n$ , and return to Step 1.

**Output:**

The estimated filter weight vector  $\mathbf{w}(n+1)$ .

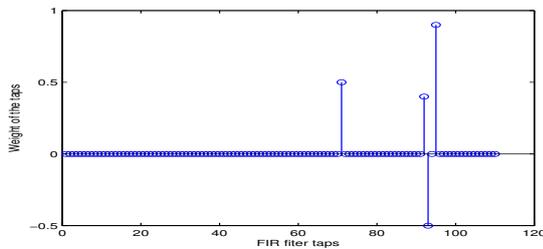


Fig. 1. Impulse response of a sparse system.

## V. SIMULATION STUDIES AND PERFORMANCE ANALYSIS

The proposed hard thresholding based adaptive filter (HTAF) was simulated for identifying a sparse system of 110 taps with the number of active taps given by  $Q = 4$ . The corresponding system impulse response is shown in Fig. 1. For this, both the input  $u(n)$  and the observation noise  $v(n)$  were taken to be zero mean, white processes with  $\sigma_u^2 = 1$  and  $\sigma_v^2 = 0.001$ ,  $L$  was taken as 15 and an initial guess of the system sparsity was taken as  $K = 5$ . Also, the performance metric chosen to evaluate the proposed algorithm was the mean square deviation (MSD) of the coefficient vector, defined by

$$\eta(n) = E(\|\mathbf{w}_{opt} - \mathbf{w}(n)\|_2^2). \quad (10)$$

Further, for comparison purposes, we considered some of the well known sparse adaptive filtering algorithms like the PNLMS [26], the MPNLMS [27] and a related popular adaptive filter like the NLMS [28] algorithms. Fig. 2 shows the learning curve (MSD in dB vs  $n$ ) of the proposed algorithm vis-a-vis similar learning curves for the above three algorithms, obtained by averaging  $\|\mathbf{w}_{opt} - \mathbf{w}(n)\|_2^2$  over 100 independent trials. The step size  $\mu$  was adjusted so that all the algorithms above have the same steady state MSD, which then makes it possible to have a fair comparison of them based on their respective rates of convergence. For this,  $\mu$  was taken as 0.25 for the PNLMS, the MPNLMS and the NLMS algorithms whereas it was chosen as 0.4 for the proposed method. Both the PNLMS and the MPNLMS algorithms need setting of certain parameters, such as, the initialization parameter  $\delta_g$  which is taken to be 0.01 and the parameter  $\rho$  (ratio between largest adaptation gain and the smallest gain) which is taken as 0.01. Also, there is one parameter which is common to the PNLMS, the MPNLMS and the NLMS algorithms, namely, the so-called regularization parameter  $\delta$  (used to avoid division by zero in the update term). This was taken in our simulation as 0.01. Fig. 2 confirms some of the well known aspects of the convergence behavior of the PNLMS, the MPNLMS and the NLMS algorithms, namely, that the PNLMS algorithm provides very high initial convergence rate but it slows down afterwards with convergence rate falling below that of the NLMS algorithm. The MPNLMS algorithm provides high convergence rate throughout the whole adaptation process though it has lesser initial convergence rate than that of the PNLMS algorithm. What is, however, very revealing from Fig. 2 is that *the proposed algorithm has a convergence rate that is much higher than any of the above three conventional algorithms*. Further, the performance of the proposed algorithm was compared with some recently proposed sparse adaptive algorithms as shown in Fig. 3. This figure depicts the learning curves for the proposed algorithm, Adaptive Projection-based Algorithm using Weighted  $l_1$  balls (APWL1)[22], Zero Attracting Normalized Least Mean Square (ZANLMS) and Reweighted Zero Attracting Normalized Least Mean Square (RZANLMS) algorithms[23]. The step size of all the algorithms was set to 1 in order to provide fastest convergence response by each algorithm. The figure shows that HTAF algorithm outperform other algorithms in convergence speed. It can be seen again that RZANLMS algorithm may

produce better steady-state MSD but HTAF can yield better convergence speed for the same MSD with step size  $\mu = 0.1$ .

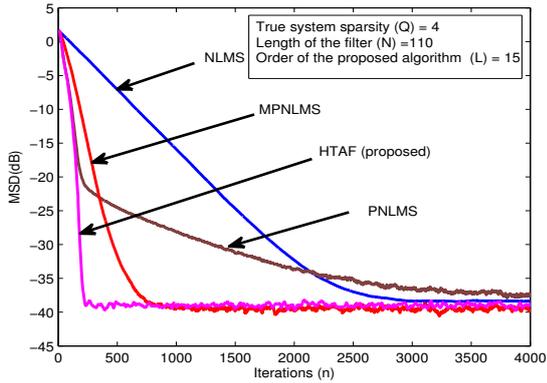


Fig. 2. Learning curves for proposed sparse adaptive algorithm.

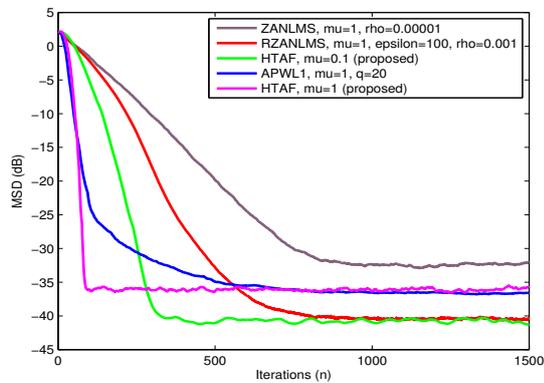


Fig. 3. Learning curves for proposed sparse adaptive algorithm with other recently proposed algorithms. Assumed sparsity  $K$  for both the algorithm APWL1 and HTAF is taken as 6 and the order  $L$  of the HTAF algorithm is 50.

We next evaluate the sensitivity of the proposed method to the initial guess of the system sparsity, by plotting the MSD for different values of (assumed)  $K$ . The results, shown in Fig. 4, suggest that as  $K$  increases from its true value in small increments, the steady state MSD deteriorates slowly, but it leaves the convergence speed almost unaffected. On the other hand, if  $K$  starts at its highest possible value (i.e., no hard thresholding) and gets progressively reduced, both the convergence rate and steady-state MSD improve, i.e., former becomes faster while the later gets lesser. Now, for sparse systems, we have broadly two types of filter taps, namely the active and the inactive taps, it is more convenient to discuss the performance of proposed algorithm by splitting the total MSD into two parts, contributed by the active and the inactive taps separately. For this, let  $NZ$  denote the index set of the active (non-zero) taps, i.e.,  $w_{opt,i} \neq 0$  for  $i \in NZ$  and let  $Z$  denote the index set of the inactive (zero) taps, i.e.,  $w_{opt,i} = 0$  for  $i \in Z$ .

Then we can define two different performance metrics as,

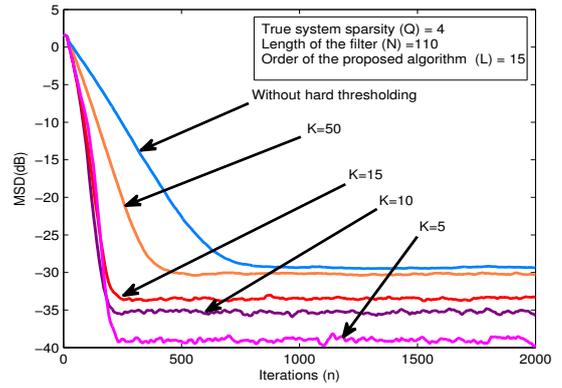


Fig. 4. Sensitivity of proposed algorithm to the choice of  $K$ .

(1) Contribution of the active taps in total MSD:

$$\eta_{NZ}(n) = E \left( \sum_{i \in NZ} (w_{opt,i} - w_i(n))^2 \right) \quad (11)$$

and

(2) Contribution of inactive taps in total MSD:

$$\eta_Z(n) = E \left( \sum_{i \in Z} (w_{opt,i} - w_i(n))^2 \right) \quad (12)$$

Fig. 5 shows the stepwise improvement of  $\eta_{NZ}(n)$  and  $\eta_Z(n)$  when  $K$  reduces from 110 to 5. To start with, we do not employ any hard thresholding (Fig. 4(a)). In this case, the proposed algorithm turns out to be just a gradient based algorithm which cannot discriminate between active and inactive taps, and thus the contribution of these two groups in the total MSD are given by  $\eta_{NZ}(\infty) = \frac{Q}{N}\eta(\infty)$  and  $\eta_Z(\infty) = \frac{N-Q}{N}\eta(\infty)$ . This is in conformity with the observation in Fig. 5(a) where it is found that  $\eta(\infty) = -29.3$  dB,  $\eta_{NZ}(\infty) = -43.69$  dB and  $\eta_Z(\infty) = -29.46$  dB. In Figs. 5(b)-(d), we gradually reduce  $K$  to 50, 15 and then 5 respectively, and in each case, compare the  $\eta_{NZ}(n)$  and  $\eta_Z(n)$  with the same for the previous choice of  $K$ . It is seen from Figs. 5(b)-(d) that both  $\eta_Z(n)$  and  $\eta_{NZ}(n)$  improve their respective convergence speeds as  $K$  decreases, the former being due to the removal of accrued gradient noise as  $K$  decreases, while the latter may be because the effective length of the filter reduces with decreasing  $K$ .

Now, it is known that the existing sparse adaptive filters like the PNLMS are not very robust against colored input. However, as we are using the concepts of affine projection algorithm in our formulation, it is expected that the proposed algorithm might enjoy robustness against colored input. To check this, the proposed HTAF was tested against a colored input, generated using an AR(1) model with pole at  $r$  using the following expression:

$$u(n) = ru(n-1) + \sqrt{1-r^2}v(n) \quad (13)$$

where  $u(n)$  is the AR process output and  $v(n)$  is a zero-mean Gaussian white noise process with unit variance. For our simulation  $r$  was set to 0.9. The results are shown in Fig. 6,

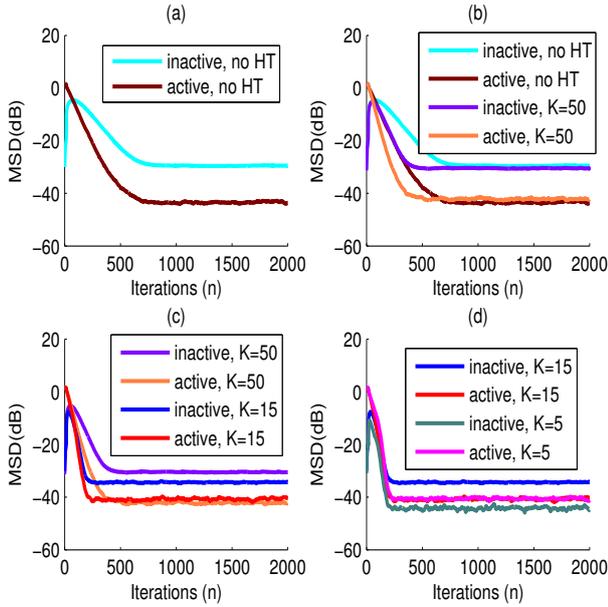


Fig. 5. Contribution of active and inactive taps in total MSD.

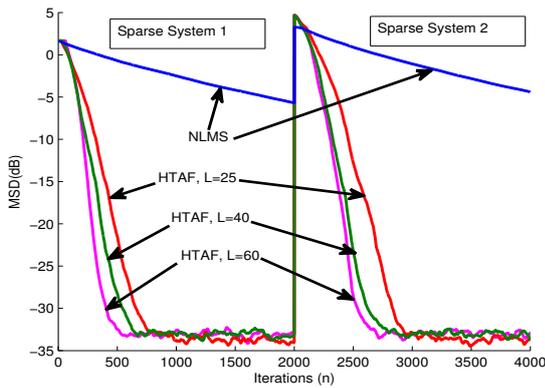


Fig. 6. Tracking performance of the proposed algorithm with increasing  $L$  and for colored input with a parametric change of the system at  $n = 2000$ .

where, taking a cue from the affine projection algorithm [11],  $L$  was increased gradually, which showed a steady increase in the convergence speed. To study the effect of increasing  $L$  on the tracking performance, the system was changed at  $n = 2000$  to a new system by shifting the 4 active taps to the left by 30 units, as shown in Fig. 7. Clearly, as confirmed by Fig. 6, the effect of increasing  $L$  on the tracking performance is marginal.

Finally, we did one more experiment to check whether HTAF algorithm could update small FIR coefficients efficiently or set all them to zeros. For this, we considered one near-sparse system response which contained a sufficiently small tap with magnitude 0.05 as shown in Fig. 8. The learning curves are plotted in Fig. 9 and it shows that proposed HTAF algorithm performs still better compared to APWL1 algorithm. By plotting the instantaneous updated steady-state weight vector  $\mathbf{w}(n)$  in Fig. 10, we show that HTAF algorithm

can update the small tap with same accuracy as obtained for other higher valued active taps.

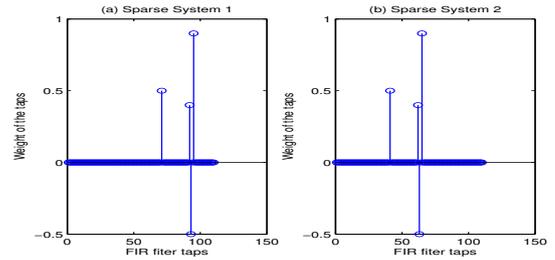


Fig. 7. Response of the time varying system. (a) Initial response (b) Response after 2000 iterations.

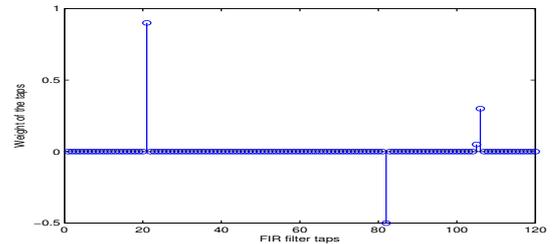


Fig. 8. Impulse response of a sparse system containing a small-magnitude tap.

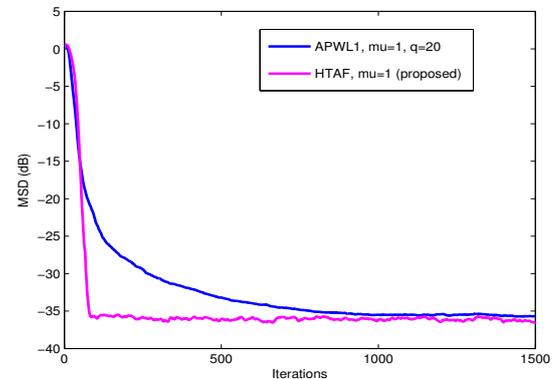


Fig. 9. Learning curves for proposed sparse adaptive algorithm.

## VI. CONCLUSIONS

In this paper, a sliding window type gradient based sparse adaptive algorithm with iterative hard thresholding is proposed. The proposed HTAF algorithm offers significant improvement over the recently proposed sparse adaptive algorithms in both convergence speed and steady-state MSD for sparse and near-sparse system. Moreover, the HTAF algorithm shows better performance for colored input, while maintaining the excellent tracking ability as well.

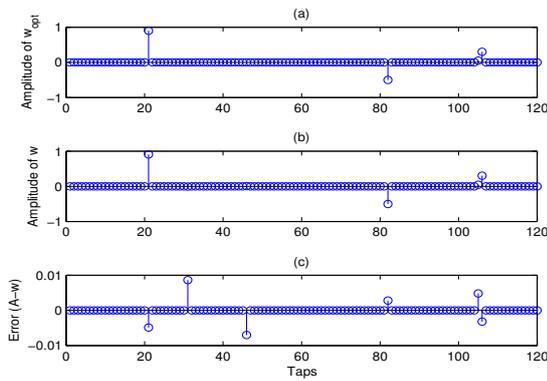


Fig. 10. Steady-state instantaneous weight vector of proposed HTAF algorithm.

## REFERENCES

- [1] D.L. Donoho, "Compressed sensing", *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289-1306, Apr., 2006.
- [2] J. Radecki, Z. Zilic, and K. Radecka, "Echo cancellation in IP networks", *Proceedings of the 45th Midwest Symposium on Circuits and Systems*, vol. 2, pp. 219-222, Tulsa, Okla, USA, August 2002.
- [3] V. V. Krishna, J. Rayala and B. Slade, "Algorithmic and Implementation Aspects of Echo Cancellation in Packet Voice Networks", *Proc. Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1252-1257, 2002.
- [4] E. Hansler and G. Schmidt, Eds., *Topics in Acoustic Echo and Noise Control*, Berlin, Springer-Verlag, 2006.
- [5] W. Schreiber, "Advanced Television Systems for Terrestrial Broadcasting", *Proc. IEEE*, vol. 83, no. 6, pp. 958-981, 1995.
- [6] W. Bajwa, J. Haupt, G. Raz and R. Nowak, "Compressed Channel Sensing", *Prof. IEEE CISS*, pp. 5-10, 2008.
- [7] M. Kocic, D. Brady and M. Stojanovic, "Sparse Equalization for Real-Time Digital Underwater Acoustic Communications", *Proc. IEEE OCEANS*, pp. 1417-1422, 1995.
- [8] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations", *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629-654, 2008.
- [9] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing", *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265-274, 2009.
- [10] T. Blumensath and M. E. Davies, "Normalized Iterative Hard Thresholding; guaranteed stability and performance", *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298-309, 2010.
- [11] A. H. Sayed, *Fundamentals of Adaptive Filtering*, New York, Wiley, 2003.
- [12] S. Chen and D.L. Donoho, "Basis pursuit", in *Proc. 28th Asilomar Conf. Signals, Syst. Comput.*, pp. 41-44, 1994.
- [13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit", *SIAM J. Scientif. Comput.*, vol. 20, no. 1, pp. 33-61, 1998.
- [14] R. Tibshirani, "Regression shrinkage and selection via the lasso", *J. Royal. Statist. Soc B.*, vol. 58, pp. 267-288, 1996.
- [15] J.A. Tropp, and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 4655-4666, 2007.
- [16] D. Needell and J. Tropp, "CoSaMP : Iterative Signal Recovery from Incomplete and Inaccurate Samples", *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301-321, 2009.
- [17] W. Dai and O. Milenkovic, "Subspace Pursuit for Compressive Sensing Signal Reconstruction", *IEEE Trans. Info. Theory*, vol. 55, no. 5, pp. 2230-2249, 2009.
- [18] Y. Chen, Y. Gu and A.O. Hero, "Sparse LMS for system identification", *Proc. IEEE ICASSP-2009*, pp. 3125-3128, Taipei, Taiwan, Apr. 2009.
- [19] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: The sparse RLS algorithm", *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4013-4025, 2010.
- [20] D. Angelosante and G.B. Giannakis, "RLS-Weighted LASSO for Adaptive Estimation of Sparse Signals", *Proc. IEEE ICASSP-2009*, Taipei, Taiwan, Apr. 2009.
- [21] D. Angelosante, J.A. Bazerque, and G.B. Giannakis, "Online Adaptive Estimation of Sparse Signals: Where RLS Meets the  $l_1$ -Norm", *IEEE Trans. Signal Processing*, vol. 58, no. 7, pp. 3436-3447, 2010.
- [22] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online Sparse System Identification and Signal Reconstruction using Projections onto Weighted  $l_1$  Balls", *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 936-952, Mar., 2011.
- [23] Y. Chen and A. O. Hero, "Recursive Group Lasso", *Signal Processing*, *IEEE Trans. on Signal Processing*, vol. 60, no. 8, pp. 3978-3987, 2012.
- [24] B. Dumitrescu, A. Onose, P. Helin and I. Tabus, "Greedy Sparse RLS", *IEEE Trans. on Signal Processing*, vol. 60, no. 5, pp. 2194-2207, 2012.
- [25] G. Mileounis, B. Babadi, N. Kalouptsidis, and V. Tarokh, "An Adaptive Greedy Algorithm With Application to Nonlinear Communications", *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 2998-3007, June, 2010.
- [26] 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.
- [27] 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.
- [28] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1986.