Sparse Adaptive Filtering by an Adaptive Convex Combination of the LMS and the ZA-LMS Algorithms

Bijit Kumar Das and Mrityunjoy Chakraborty, *Senior Member, IEEE*

***Abstract*—In practice, one often encounters systems that have a sparse impulse response, with the degree of sparseness varying over time. This paper presents a new approach to identify such systems which adapts dynamically to the sparseness level of the system and thus works well both in sparse and non-sparse environments. The proposed scheme uses an adaptive convex combination of the LMS algorithm and the recently proposed, sparsity-aware zero-attractor LMS (ZA-LMS) algorithm. It is shown that while for non-sparse systems, the proposed combined filter always converges to the LMS algorithm (which is better of the two filters for non-sparse case in terms of lesser steady state excess mean square error (EMSE)), for semi-sparse systems, on the other hand, it actually converges to a solution that produces lesser steady state EMSE than produced by either of the component filters. For highly sparse systems, depending on the value of a proportionality constant in the ZA-LMS algorithm, the proposed combined filter may either converge to the ZA-LMS based filter or may produce a solution which, like the semi-sparse case, outperforms both the constituent filters. A simplified update formula for the mixing parameter of the adaptive convex combination is also presented. The proposed algorithm requires much less complexity than the existing algorithms and its claimed robustness against variable sparsity is well supported by simulation results.**

***Index Terms*—Convex combination, excess mean square error, sparse systems, ZA-LMS algorithm.**

I. Introduction

In practice, one often comes across systems that have a sparse impulse response. One example of such systems is the network echo channel [1], which has an "active" region of only 8–12 ms in a total echo response of about 64–128 ms duration with the "inactive" part caused by various bulk delays. Another example is the acoustic echo generated due to coupling between microphone and loudspeaker in hands free mobile telephony [2]. Other well known examples of sparse systems include HDTV where clusters of dominant echoes arrive after long periods of silence [3], wireless multipath channels which, on most of the occasions, have only a few clusters of significant paths [4] and acoustic channels in shallow underwater communication, where the various multipath components caused by reflections off the sea surface and the sea bed have long intermediate delays [5]. For most of these systems, the impulse response is not just sparse, but its degree of sparsity also varies with time and context.

The a priori information about the sparseness of the system impulse response, if exploited properly, can lead to significant improvement in the identification performance of the algorithm deployed to identify the system. Conventional system identification algorithms like the LMS and the normalized LMS (NLMS) based adaptive filters [7] are, however, sparsity agnostic, i.e., they do not make use of the a priori knowledge of the sparseness of the system and thus, fail to improve their performance both in terms of steady state excess mean square error (EMSE) and convergence speed when the system changes from a non-sparse to a highly sparse one. In recent years, several new adaptive filters have been proposed [6] that exploit the sparse nature of the system impulse response and achieve better performance, prominent amongst them being the proportionate normalized LMS (PNLMS) algorithm [8] and its variants like the μ-law PNLMS (MPNLMS) [9] and the improved PNLMS (IPNLMS) [10]. In these algorithms, each coefficient is updated with an independent step size that is made proportional (directly or indirectly) to the magnitude of the particular filter coefficient estimate, resulting in faster initial convergence for sparse systems. The performance of the PNLMS algorithm, however, starts deteriorating as the sparsity level of the system decreases, meaning it is not very effective when the system sparseness varies with time over a wide range. This problem was addressed in the IPNLMS algorithm [10] where, at the expense of some additional computations and by tuning a parameter of the algorithm suitably, the algorithm could be made to perform better for dispersive systems as well. The robustness of the IPNLMS algorithm against variable sparsity was further improved in [11] which considers a convex combination of two IPNLMS based adaptive filters that can switch between the two depending on the sparsity level. Apart from the PNLMS family, other notable approaches for sparse adaptive filtering and system identification include partial update LMS [12], [13] and the exponentiated gradient algorithm [14].

In a separate recent development, motivated by LASSO [15] and the recent progresses in compressive sensing [16], [17], an alternative approach to identify sparse systems has been proposed [18], [19], 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. This approach has a much lesser computational com-

Manuscript received February 20, 2013; revised July 27, 2013; accepted September 19, 2013. Date of publication January 09, 2014; date of current version April 24, 2014. This paper was recommended by Associate Editor T. Ogunfunmi.

The authors are with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur 721302, India (e-mail: bijitbijit@gmail.com; mrityun@ece.iitkgp.ernet.in).

Digital Object Identifier 10.1109/TCSI.2013.2289407

plexity compared to the PNLMS based adaptive filters. More importantly, for sparse systems, it produces steady state EMSE significantly less compared to the PNLMS family (thus producing more accurate estimates of system parameters) while maintaining comparable convergence rates. In order to handle the case of time varying sparsity, the ZA-LMS algorithm was modified further to a reweighted ZA-LMS (RZA-LMS) algorithm in [18] which scales down the zero attractors appropriately so that the shrinkage (i.e., zero attraction) remains confined to inactive taps only. Parameter selection for maintaining such shrinkage is, however, a tricky issue in the RZA-LMS algorithm especially for systems that have time varying sparseness with the active taps taking values over a wide range.

In this paper, we propose an alternative method to deal with variable sparseness by using an adaptive convex combination of the LMS and the ZA-LMS algorithms[1]. The proposed algorithm uses the general framework of convex combination of adaptive filters [20] and enjoys overall EMSE performance better than offered by the PNLMS family. It also requires much less complexity than required by both [11] and the RZA-LMS algorithm [18] and is free of the parameter selection problem of the RZA-LMS. The performance of the proposed scheme is first evaluated analytically by carrying out a convergence analysis. This requires evaluation of the steady state cross correlation between the output *a priori* errors generated by the LMS and the ZA-LMS based filters, and then relating it to the respective steady state EMSE of the two filters. In our treatment here, we have carried out this exercise for three different kinds of systems, namely, non-sparse, semi-sparse and sparse. The analysis shows that while for a non-sparse system, the proposed combined filter always converges to the LMS based filter (i.e., the better of the two filters in terms of lesser steady state EMSE for the non-sparse case), for semi-sparse systems, it actually converges to a solution that can outperform both the constituent filters by producing lesser EMSE than produced by either of the two constituents. For sparse systems, the proposed scheme usually converges to the ZA-LMS based filter. However, by adjusting a proportionality constant associated with the zero attractors in the ZA-LMS algorithm, it is also possible to achieve convergence to a solution which, like the semi-sparse case, outperforms both the component filters. This paper also presents a simplified update formula for the mixing parameter in the adaptive convex combination, by using certain approximations in the corresponding gradient expression. Finally, the robustness of the proposed methods against variable sparsity is verified by detailed simulation studies.

II. Problem Formulation, Proposed Algorithm and Performance Analysis

We consider the problem of identifying a system that takes $x(n)$ as input and produces the observable output $y_d(n)$ as $y_d(n) = \mathbf{w}_{\rm opt}^T \mathbf{x}(n) + v(n)$, where $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-L+1)]^T$, $\mathbf{w}_{\rm opt}$ is a $L \times 1$ impulse response vector which is known *a priori* to be sparse and $v(n)$ is an observation noise which is assumed to be zero mean, white with variance σ_v^2 and independent of the input data vector $\mathbf{x}(m)$ for any m and n. In order to identify the system, we deploy an adaptive convex combination of two adaptive filters, say, filter 1 which is ZA-LMS based and filter 2 which is LMS based as shown in Fig. 1, following the general model [20] of convex combination of adaptive filters. Filter 1 (i.e., ZA-LMS) adapts a filter coefficient vector $\mathbf{w}_1(n)$ as [18]

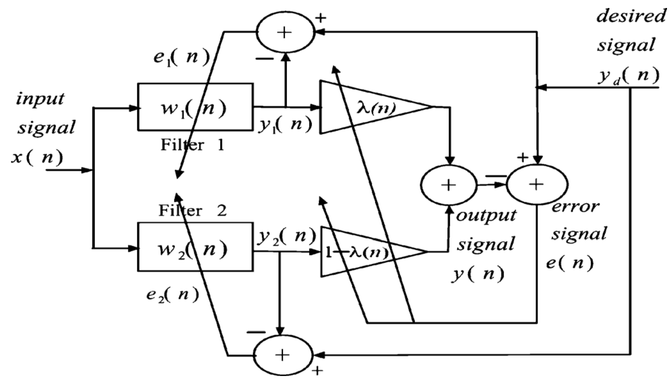

Fig. 1. The proposed adaptive convex combination of two adaptive filters : filter 1 (ZA-LMS based) and filter 2 (LMS based).

$$\mathbf{w}_1(n+1) = \mathbf{w}_1(n) - \rho\,\mathrm{sgn}(\mathbf{w}_1(n)) + \mu\, e_1(n)\,\mathbf{x}(n) \quad (1)$$

while filter 2 (i.e., LMS) adapts a filter coefficient vector $\mathbf{w}_2(n)$ as

$$\mathbf{w}_2(n+1) = \mathbf{w}_2(n) + \mu\, e_2(n)\,\mathbf{x}(n) \quad (2)$$

where μ is the usual step size, ρ is a suitable constant (usually very very small) and $e_i(n) = y_d(n) - y_i(n), i = 1,\ 2$ is the respective filter output errors with $y_i(n) = \mathbf{w}_i^T(n)\mathbf{x}(n)$ denoting the respective filter outputs. The convex combination generates a combined output $y(n) = \lambda(n)y_1(n) + [1-\lambda(n)]y_2(n)$. The variable $\lambda(n)$ is a mixing parameter that lies between 0 and 1, which is to be adapted by following a gradient descent method to minimize the quadratic error function of the overall filter, namely, $e^2(n)$ where $e(n) = y_d(n) - y(n)$. However, such adaptation does not guarantee that $\lambda(n)$ will lie between 0 and 1. Therefore, instead of $\lambda(n)$, an equivalent variable $a(n)$ is updated which expresses $\lambda(n)$ as a sigmoidal function, i.e., $\lambda(n) = 1/(1+\exp(-a(n)))$. The update equation of $a(n)$ is given by [20]

$$\begin{aligned} a(n+1) &= a(n) - \frac{\mu_a}{2}\frac{\partial e^2(n)}{\partial a(n)} \\ &= a(n) + \mu_a\, e(n)[y_1(n) - y_2(n)]\lambda(n)[1-\lambda(n)]. \end{aligned} \quad (3)$$

In practice, $\lambda(n) \approx 1$ for $a(n) \gg 0$ and conversely, $\lambda(n) \approx 0$ for $a(n) \ll 0$. Therefore, instead of updating $a(n)$ up to $\pm\infty$, it is sufficient to restrict it to a range $[-a^+, +a^+]$ (a^+ : a large, finite number) which limits the permissible range of $\lambda(n)$ to $[1-\lambda^+, \lambda^+]$, where $\lambda^+ = 1/(1+\exp(-a^+))$.

A. Some Important Assumptions and Definitions

Some of the assumptions and definitions made in [20] are also useful in the present treatment and we state them below.

[1]Some preliminary results of this paper were presented at ISCAS-2011, Rio De Janeiro, Brazil, May, 2011 [21].

Firstly, as in [20], we assume that the initial conditions $\mathbf{w}_1(0)$, $\mathbf{w}_2(0)$ and $a(0)$ are independent of $\mathbf{x}(n)$, $y_d(n)$ and $v(n)$ for all n. Also, to start with, we do not make any special assumption on $x(n)$ except that it is WSS with $E[x(n)] = 0$ and $E[\mathbf{x}(n)\mathbf{x}^T(n)] = \mathbf{R}$. However, at a later stage of the analysis, it becomes necessary to assume that $x(n)$ is white, i.e., $\mathbf{R} = \sigma_x^2\mathbf{I}$, where $E[x^2(n)] = \sigma_x^2$ and $\mathbf{I}$ is the identity matrix[2]. We also assume that $x(n)$ and $v(n)$ are jointly Gaussian processes.

Next, as in [20], we define, for $i = 1, 2$ the following.

a) *Weight Error Vectors* : $\tilde{\mathbf{w}}_i(n) = \mathbf{w}_{\rm opt} - \mathbf{w}_i(n)$;
b) *Equivalent Weight Vector* for the combined filter : $\mathbf{w}_c(n) = \lambda(n)\mathbf{w}_1(n) + [1-\lambda(n)]\mathbf{w}_2(n)$;
c) *Equivalent Weight Error Vector* for the combined filter: $\tilde{\mathbf{w}}_c(n) = \mathbf{w}_{\rm opt} - \mathbf{w}_c(n) = \lambda(n)\tilde{\mathbf{w}}_1(n) + [1-\lambda(n)]\tilde{\mathbf{w}}_2(n)$;
d) *A Priori Errors* : $e_{a,i}(n) = \tilde{\mathbf{w}}_i^T(n)\mathbf{x}(n)$ and $e_a(n) = \tilde{\mathbf{w}}_c^T(n)\mathbf{x}(n)$. Clearly, $e_i(n) = e_{a,i}(n) + v(n)$, $e_a(n) = \lambda(n)\, e_{a,1}(n) + [1-\lambda(n)]\, e_{a,2}(n)$ and $e(n) = e_a(n) + v(n)$;
e) *Excess Mean Square Error (EMSE)* : $J_{{\rm ex},i}(n) = E[e_{a,i}^2(n)], i = 1, 2$ and $J_{\rm ex}(n) = E[e_a^2(n)]$;
f) *Cross EMSE* : $J_{{\rm ex},12}(n) = E[e_{a,1}(n)e_{a,2}(n)] \le \sqrt{J_{{\rm ex},1}(n)}\sqrt{J_{{\rm ex},2}(n)}$ [From Cauchy-Schwartz inequality]. This means, $J_{{\rm ex},12}(n)$ can not be greater than both $J_{{\rm ex},1}(n)$ and $J_{{\rm ex},2}(n)$ simultaneously.

Next, from (3), one can write,

$$E[a(n+1)] = E[a(n)] + \mu_a E[e(n)(y_1(n) - y_2(n)) \times\lambda(n)(1-\lambda(n))]. \quad (4)$$

As shown in [20] and also below, the convergence of $E[a(n)]$ in (4) depends on the steady state values of the EMSE and the cross EMSE, namely, $J_{{\rm ex},i}(\infty) = \lim_{n\to\infty} J_{{\rm ex},i}(n)$ and $J_{{\rm ex},12}(\infty) = \lim_{n\to\infty} J_{{\rm ex},12}(n)$ respectively. In practice, both $J_{{\rm ex},i}(n)$ and $J_{{\rm ex},12}(n)$, however, take only finite number of steps to reach their steady state values, as both the LMS and the ZA-LMS algorithms converge in finite number of iterations. Substituting $e(n)$ in (4) by $e_a(n)+v(n)$ where $e_a(n)$ is defined above, noting that $y_1(n) - y_2(n) = e_{a,2}(n) - e_{a,1}(n)$ and also that $E[v(n)] = 0$, and assuming like [20] that in the steady state, $\lambda(n)$ is independent of the *a priori errors* $e_{a,i}(n)$, it is easy to verify [20] that for large n (theoretically, for $n \to \infty$)

$$E[a(n+1)] = E[a(n)] + \mu_a\, E[\lambda(n)[1-\lambda(n)]^2]\,\Delta J_2 - \mu_a E[\lambda^2(n)[1-\lambda(n)]]\Delta J_1, \quad (5)$$

where $\Delta J_1 = J_{{\rm ex},1}(\infty) - J_{{\rm ex},12}(\infty)$ and $\Delta J_2 = J_{{\rm ex},2}(\infty) - J_{{\rm ex},12}(\infty)$. Equation (5) which provides the dynamics of the evolution of $E[a(n)]$ assumes constant ΔJ_1 and ΔJ_2, meaning, it comes in operation once both the LMS and the ZA-LMS algorithms have converged. For analyzing the convergence of $E[a(n)]$, we next evaluate ΔJ_1 and ΔJ_2 for the specific case of the ZA-LMS and the LMS algorithms.

[2]Note that the assumption of a white input is reasonable since in system identification, the input to the system is usually known and thus can be designed to be white. Also, for white input, the LMS algorithm performs at par with its higher versions like the NLMS algoritm while requiring lesser complexity which justifies our choice of LMS for filter 2.

B. Performance Analysis of the Combination

In this section, we examine the convergence behavior of the proposed convex combination for various levels of system sparsity. From [18] and using the constraint of small misadjustment, the steady state EMSE of the ZA-LMS based filter can be written as

$$J_{{\rm ex},1}(\infty) = \frac{\mu {\rm Tr}(\mathbf{R})}{2-\mu{\rm Tr}(\mathbf{R})}\sigma_v^2 + \frac{\alpha_1\rho}{\mu(2-\mu{\rm Tr}(\mathbf{R}))}\left(\rho - \frac{2\alpha_2}{\alpha_1}\right) \quad (6)$$

where

$$\alpha_1 = E[{\rm sgn}(\mathbf{w}_1(\infty))^T((\mathbf{I}-\mu\mathbf{R})^{-1}{\rm sgn}(\mathbf{w}_1(\infty))] \quad (7)$$

and

$$\alpha_2 = E[\|\, \mathbf{w}_1(\infty)\,\|_1] - \|\, \mathbf{w}_{\rm opt}\,\|_1\,. \quad (8)$$

Separately, the steady state EMSE of the LMS algorithm is given by [7],

$$J_{{\rm ex},2}(\infty) = \frac{\mu{\rm Tr}(\mathbf{R})}{2-\mu{\rm Tr}(\mathbf{R})}\sigma_v^2. \quad (9)$$

From (1) and (2), and the definitions above, we have $\tilde{\mathbf{w}}_1(n+1) = \tilde{\mathbf{w}}_1(n) - \mu\, e_1(n)\,\mathbf{x}(n) + \rho\,{\rm sgn}(\mathbf{w}_1(n))$ and $\tilde{\mathbf{w}}_2(n+1) = \tilde{\mathbf{w}}_2(n) - \mu\, e_2(n)\,\mathbf{x}(n)$. Postmultiplying $\tilde{\mathbf{w}}_1^T(n+1)$ by $\tilde{\mathbf{w}}_2(n+1)$, taking expectation, and making the substitutions $\tilde{\mathbf{w}}_i^T(n)\mathbf{x}(n) = e_{a,i}(n), i = 1, 2$, we have

$$\begin{aligned} E&\left[\tilde{\mathbf{w}}_1^T(n+1)\tilde{\mathbf{w}}_2(n+1)\right] \\ &= E\left[\tilde{\mathbf{w}}_1^T(n)\tilde{\mathbf{w}}_2(n)\right] \\ &\quad - 2\,\mu J_{{\rm ex},12}(n) + \mu^2\, E[\|\mathbf{x}(n)\|^2 e_1(n)e_2(n)] \\ &\quad + \rho E[{\rm sgn}\left\{\mathbf{w}_1^T(n)\right\}\tilde{\mathbf{w}}_2(n) \\ &\quad - \mu\rho E\left[{\rm sgn}\left\{\mathbf{w}_1^T(n)\right\}\mathbf{x}(n)e_2(n)\right] \end{aligned} \quad (10)$$

where we have used the fact that $v(n)$, being zero mean, white, Gaussian, is i.i.d. and is thus independent of $\mathbf{w}_i(n)$ (as the latter depends on $v(m), 0 \le m \le n-1$) apart from being independent of $\mathbf{x}(n)$, which enables us to write $E[\tilde{\mathbf{w}}_1^T(n)\mathbf{x}(n)e_2(n)] = E[\tilde{\mathbf{w}}_1^T(n)\mathbf{x}(n)(e_{a,2}(n) + v(n))] = E[e_{a,1}(n)e_{a,2}(n)] = J_{{\rm ex},12}(n)$. Now, as both $x(n)$ and $v(n)$ are assumed to be jointly Gaussian, it is reasonable to assume that $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$ are jointly Gaussian, as both evolve from $x(m)$ and $v(m)$, $0 \le m < n$. Then, as shown in the Appendix A, we have

$$\lim_{n\to\infty}\left[E\left[\tilde{\mathbf{w}}_1^T(n+1)\tilde{\mathbf{w}}_2(n+1)\right] - E\left[\tilde{\mathbf{w}}_1^T(n)\tilde{\mathbf{w}}_2(n)\right]\right] = 0. \quad (11)$$

In other words, in the steady state, the term $E[\tilde{\mathbf{w}}_1^T(n)\tilde{\mathbf{w}}_2(n)]$ becomes time invariant.

Substituting (11) in (10) and and using the "Separation principle" [20] which implies that in the steady state, $\|\mathbf{x}(n)\|^2$ is independent of the a priori errors $e_{a,i}(n)$, we obtain

$$\begin{aligned} 2\mu J_{\text{ex},12}(n) - \mu^2\text{Tr}(\mathbf{R})J_{\text{ex},12}(n) - \mu^2\text{Tr}(\mathbf{R})\sigma_v^2 \\ = -\mu\rho E\left[\text{sgn}\left\{\mathbf{w}_1^T(n)\right\}\mathbf{x}(n)\mathbf{x}^T(n)\tilde{\mathbf{w}}_2(n)\right] \\ + \rho E\left[\text{sgn}\left\{\mathbf{w}_1^T(n)\right\}\tilde{\mathbf{w}}_2(n)\right] \end{aligned} \quad (12)$$

as $n \rightarrow \infty$, where we have made use of the fact that $E[\|\mathbf{x}(n)\|^2] = \text{Tr}(\mathbf{R})$ and $E[e_1(n)e_2(n)] = E[(e_{a,1}(n) + v(n))(e_{a,2}(n) + v(n))] = J_{\text{ex},12}(n) + \sigma_v^2$. From (12), it is then possible to write

$$J_{\text{ex},12}(\infty) = \frac{\mu\sigma_v^2\text{Tr}(\mathbf{R})}{2-\mu\text{Tr}(\mathbf{R})} + X(\infty) \equiv J_{\text{ex},2}(\infty) + X(\infty) \quad (13)$$

where

$$X(\infty) = \frac{\rho}{\mu}\left\{\frac{E\left[\text{sgn}\left\{\mathbf{w}_1^T(\infty)\right\}(\mathbf{I}-\mu\mathbf{R})\tilde{\mathbf{w}}_2(\infty)\right]}{2-\mu\text{Tr}(\mathbf{R})}\right\} \quad (14)$$

and $J_{\text{ex},2}(\infty)$ is given by (9). At this point, we invoke the assumption that $x(n)$ is white, i.e., $\mathbf{R} = \sigma_x^2\mathbf{I}$. Substituting in (14), it is then possible to express $X(\infty)$ as

$$\begin{aligned} X(\infty) &= \frac{\rho}{\mu(2-\mu\text{Tr}(\mathbf{R}))} \\ &\quad E\left[\sum_{i=0}^{L-1}\left[\text{sgn}\{w_{1,i}(\infty)\}\left(1-\mu\sigma_x^2\right)\tilde{w}_{2,i}(\infty)\right]\right] \\ &= \frac{\rho}{\mu(2-\mu\text{Tr}(\mathbf{R}))} \\ &\quad \times\left\{\sum_{i\in Z}E\left[\text{sgn}\{w_{1,i}(\infty)\}\left(1-\mu\sigma_x^2\right)\tilde{w}_{2,i}(\infty)\right]\right. \\ &\quad \left.+\sum_{i\in \text{NZ}}E\left[\text{sgn}\{w_{1,i}(\infty)\}\left(1-\mu\sigma_x^2\right)\tilde{w}_{2,i}(\infty)\right]\right\} \end{aligned} \quad (15)$$

where $w_{k,i}(\cdot) = [\mathbf{w}_k(\cdot)]_i, \tilde{w}_{k,i}(\cdot) = [\tilde{\mathbf{w}}_k(\cdot)]_i, k = 1, 2$, and Z and NZ are the sets of indices for the zero (or near zero) and the non-zero (i.e., significant) taps of $\mathbf{w}_{\text{opt}}$ respectively (note that $Z \cap \text{NZ} = \{\phi\}$ and $Z \cup \text{NZ} = \{0, 1, \ldots, L-1\}$).

Since, for $i \in \text{NZ}$, we are considering taps for which $w_{\text{opt},i}$ has values significantly larger than zero and also, considering that for the ZA-LMS algorithm, $E[w_{1,i}(\infty)] \approx w_{\text{opt},i}$, it is reasonable to assume that for $\forall i \in \text{NZ}, \text{sgn}\{w_{1,i}(\infty)\} = \text{sgn}\{w_{\text{opt},i}\}$. From this and also from the fact that $E[\tilde{w}_{2,i}(\infty)] = 0$, (15) then leads to

$$\begin{aligned} X(\infty) &= \frac{\rho}{\mu(2-\mu\text{Tr}(\mathbf{R}))} \\ &\quad \times\left\{\sum_{i\in Z}E\left[\text{sgn}\{w_{1,i}(\infty)\}\left(1-\mu\sigma_x^2\right)\tilde{w}_{2,i}(\infty)\right]\right. \\ &\quad \left.+\sum_{i\in \text{NZ}}\text{sgn}\{w_{\text{opt},i}\}\left(1-\mu\sigma_x^2\right)E[\tilde{w}_{2,i}(\infty)]\right\} \\ &= \frac{\rho}{\mu(2-\mu\text{Tr}(\mathbf{R}))}\sum_{i\in Z}E\left[\text{sgn}\{w_{1,i}(\infty)\}\right. \\ &\quad \left.\times\left(1-\mu\sigma_x^2\right)\tilde{w}_{2,i}(\infty)\right]. \end{aligned} \quad (16)$$

We next show that $X(\infty) \leq 0$ by proving the following result:

Theorem 1: $\forall i \in Z$, $E[\text{sgn}\{w_{1,i}(\infty)\}\tilde{w}_{2,i}(\infty)] < 0$.

Proof: Given in the Appendix A □.

Since, for convergence of the LMS algorithm, $0 < \mu < 2/\text{Tr}(\mathbf{R}) = 2/(L\sigma_x^2)$, Theorem 1 implies that $X(\infty) \leq 0$ (with "$= 0$" occurring when Z is an empty or near empty set). The implication of this on the convergence of $a(n)$ is examined below for various cases of sparsity level of the system.

I : Non-Sparse Systems

For non-sparse systems, number of elements in the set Z is very small and thus, $X(\infty) \simeq 0$ (since ρ is very small), meaning $J_{\text{ex},12}(\infty) \simeq J_{\text{ex},2}(\infty)$, or equivalently, $\Delta J_2 \simeq 0$. This implies $\Delta J_1 = J_{\text{ex},1}(\infty) - J_{\text{ex},12}(\infty) = J_{\text{ex},1}(\infty) - J_{\text{ex},2}(\infty) > 0$ (since, for non-sparse systems, the LMS algorithm performs better than the ZA-LMS [18]). Equation (5) then leads to

$$E[a(n+1)] = E[a(n)] - \mu_a E[\lambda^2(n)(1-\lambda(n))]\Delta J_1. \quad (17)$$

The convergence of $E[a(n)]$ and thus of $a(n)$ (to $-a^+$) can then be established using a logic similar to [20]. Note that $\forall\ \lambda(n) \in [0, 1]$, the function $f(\lambda(n)) = \lambda^2(n)(1-\lambda(n)) \geq 0$, with a maxima at $\lambda(n) = 2/3$ and with $f(\lambda(n)) = 0$ at $\lambda(n) = 0, 1$. Assume that at the n-th index, $-a^+ < E[a(n)] < a^+$, meaning that $a(n)$ has not converged to a^+ or $-a^+$ (in all trials), but is taking values from $[-a^+, a^+]$, or, equivalently, $\lambda(n)$ has not converged to λ^+ or $(1-\lambda^+)$ but is assuming values from $[(1-\lambda^+), \lambda^+]$. For $1-\lambda^+ \leq \lambda(n) \leq \lambda^+$, $f(\lambda(n)) \geq f(1-\lambda^+) = C$ (say) with $C > 0$, meaning, $E[f(\lambda(n))] \geq C$. Substituting in (17), $E[a(n+1)] \leq E[a(n)] - \mu_a C\Delta J_1$. Since $\Delta J_1 > 0$, the above implies $\lim_{n\to\infty} E[a(n)] = -a^+$ and thus $\lim_{n\to\infty} a(n) = -a^+$ almost surely. The fallout of this is that $\lim_{n\to\infty}\lambda(n) = (1-\lambda^+)$ (almost surely) ≈ 0, and therefore, the convex combination will switch to filter 2, which for non-sparse systems performs better than filter 1 (i.e., $J_{\text{ex},1}(\infty) > J_{\text{ex},2}(\infty)$).

II. Semi-Sparse Systems

For such systems, we still have $J_{\text{ex},1}(\infty) > J_{\text{ex},2}(\infty)$. However, the set Z in such cases consists of non-negligible number of elements, meaning $X(\infty) < 0$, or, equivalently, $J_{\text{ex},2}(\infty) - J_{\text{ex},12}(\infty) > 0$. Combining, $J_{\text{ex},1}(\infty) > J_{\text{ex},2}(\infty) > J_{\text{ex},12}(\infty)$ and thus, both $\Delta J_1 > 0$, $\Delta J_2 > 0$. This is analogous to the case (3), Section III of [20]. Under this, a stationary point is obtained by setting the update term in (5) as zero as $n \rightarrow \infty$, leading to $E[\lambda(\infty)\{1-\lambda(\infty)\}^2]\Delta J_2 = E[\lambda(\infty)^2\{1-\lambda(\infty)\}]\Delta J_1$. Assuming a negligibly small variance for $\lambda(\infty)$, i.e., assuming $E[\lambda(\infty)]^2 \rightarrow 0$ which implies that $\lambda(n) \rightarrow$ a constant (almost surely) as $n \rightarrow \infty$, one can then obtain from the above $\{1 - E[\lambda(\infty)]\}\Delta J_2 = E[\lambda(\infty)]\Delta J_1$, or, equivalently,

$$E[\lambda(\infty)] = \frac{\Delta J_2}{\Delta J_1 + \Delta J_2} > 0.5. \quad (18)$$

As proved in [20], this case is not sub-optimal. Rather, it leads to

$$J_{\text{ex}}(\infty) \leq \min[J_{\text{ex},1}(\infty), J_{\text{ex},2}(\infty)] \quad (19)$$

meaning in this case, the proposed convex combination works even better than each of its component filters.

III. Sparse Systems

For sparse systems, we have $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},1}(\infty)$, as the ZA-LMS algorithm in this case outperforms the LMS algorithm [18]. Also, the set Z in this case contains significant number of elements and thus, we have $X(\infty) < 0$, implying $\Delta J_2 = J_{\mathrm{ex},2}(\infty) - J_{\mathrm{ex},12}(\infty) > 0$. In the following, we first show that for systems that are highly sparse, both the possibilities, $J_{\mathrm{ex},12}(\infty) \geq J_{\mathrm{ex},1}(\infty)$ and $J_{\mathrm{ex},12}(\infty) < J_{\mathrm{ex},1}(\infty)$ are possible, with the former occurring for ρ taking values in the range of zero to some small, positive value and the latter for higher values of ρ. For this, we first prove the following:

Theorem 2: Under small misadjustment condition and with zero-mean, Gaussian i.i.d. input, the steady state EMSE of the ZA-LMS algorithm is given approximately by $J_{\mathrm{ex},1}(\infty) = J_{\mathrm{ex},2}(\infty) + (\rho/(2\mu))[L\rho - 2\sqrt{2/\pi}\sum_{i\in Z}\sigma_{w_{1,i}}]$, where $\sigma_{w_{1,i}} = \sqrt{E[w_{1,i}^2(\infty)]}, i \in Z$ (i.e., the steady state standard deviation of $w_{1,i}(n)$).

Proof: Given in the Appendix B. □

Note that since $x(n)$ is i.i.d., $\sigma_{w_{1,i}}$ is same for all taps $w_{1,i}, i \in Z$. We then drop the index i from $\sigma_{w_{1,i}}$ and denote $\sigma_{w_{1,i}}$ by σ_{w_1}. Also, for highly sparse systems, the set NZ contains very few elements, thus allowing us to make the following approximation : $J_{\mathrm{ex},1}(\infty) \approx J_{\mathrm{ex},2}(\infty) + (\rho/(2\mu))[L\rho - 2\sqrt{2/\pi}\,L\sigma_{w_1}]$. We next evaluate $J_{\mathrm{ex},1}(\infty) - J_{\mathrm{ex},12}(\infty)$. From (13) and above, $J_{\mathrm{ex},1}(\infty) - J_{\mathrm{ex},12}(\infty) = (\rho/(2\mu))[L\rho - 2\sqrt{2/\pi}\,L\sigma_{w_1}] - X(\infty)$, where $X(\infty)$ is given by (16) which, for $\mu \ll 2/\mathrm{Tr}(\mathbf{R}) = 2/(L\sigma_x^2)$, reduces to $(\rho/(2\mu))\sum_{i\in Z} E[\mathrm{sgn}\{w_{1,i}(\infty)\}\tilde{w}_{2,i}(\infty)]$. In Appendix A, it is shown that for $i \in Z$, $E[\mathrm{sgn}\{w_{1,i}(n)\}\tilde{w}_{2,i}(n)] = -gE[w_{1,i}(n)w_{2,i}(n)] = -gE[\tilde{w}_{1,i}(n)\tilde{w}_{2,i}(n)]$, where $g = \sqrt{2/(\pi\sigma_{w_1}^2)}$. From this, (A.8) and (A.6), and using the approximation $\mu \ll 2/(L\sigma_x^2)$, we then obtain, for $i \in Z$, $E[\mathrm{sgn}\{w_{1,i}(n)\}\tilde{w}_{2,i}(n)] \approx -\mu g\sigma_v^2/2 = -(\mu\sigma_v^2/2)\sqrt{2/(\pi\sigma_{w_1}^2)}$. As for highly sparse systems, the set NZ is nearly empty, $X(\infty)$ can then be expressed approximately as $X(\infty) = (L\rho/(2\mu))E[\mathrm{sgn}\{w_{1,i}(n)\}\tilde{w}_{2,i}(n)] \approx -(L\rho\sigma_v^2/4)\sqrt{2/(\pi\sigma_{w_1}^2)}$, leading to

$$J_{\mathrm{ex},1}(\infty) - J_{\mathrm{ex},12}(\infty) = L\rho\left[\frac{\rho}{2\mu} - \sqrt{\frac{2}{\pi}}\frac{\sigma_{w_1}}{\mu} + \sqrt{\frac{2}{\pi}}\frac{\sigma_v^2}{4\sigma_{w_1}}\right]. \tag{20}$$

To analyze further, we note that for an i.i.d. input, the EMSE for the LMS algorithm is given by [22] $L\sigma_x^2\sigma_{w_2}^2$, where $\sigma_{w_2}^2$ is the weight error variance for each tap of the LMS-based filter (common for all taps as the input is i.i.d., zero-mean, white), given, under the small μ assumption, by $\mu\sigma_v^2/2$ [22]. Since, for sparse systems, $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},1}(\infty)$ and $J_{\mathrm{ex},1}(\infty) \approx L\sigma_x^2\sigma_{w_1}^2$, we have in such cases $\sigma_{w_1}^2 < \mu\sigma_v^2/2$. Now, consider the expression within parentheses on the RHS of (20), i.e., $[\rho/(2\mu) - \sqrt{2/\pi}\sigma_{w_1}/\mu + \sqrt{2/\pi}\sigma_v^2/(4\sigma_{w_1})]$. It is easy to see that this function always has a negative gradient w.r.t. σ_{w_1}. At $\rho = 0$ when the ZA-LMS algorithm becomes identical to the LMS, $\sigma_{w_1}^2$ attains its maximum value of $\mu\sigma_v^2/2$ for which this becomes maximally negative $(= -\sigma_v/(2\sqrt{\pi\mu}))$. Afterwards, as ρ increases slowly, this function becomes less and less negative, crossing zero approximately at $\sigma_{w_1}^2 = \mu\sigma_v^2/4$ and then becoming positive with further increase in ρ.[3]

Clearly, from above, two cases can occur for a highly sparse system, namely

Case I: $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},12}(\infty) \geq J_{\mathrm{ex},1}(\infty)$, (i.e., when $\mu\sigma_v^2/4 \leq \sigma_{w_1}^2 < \mu\sigma_v^2/2$) meaning $\Delta J_1 \leq 0$.

This is analogous to the case (1), Section III of [20]. Equation (5) in this case leads to

$$E[a(n+1)] = E[a(n)] + \mu_a E[g(\lambda(n))]\Delta J_2 - \mu_a E[f(\lambda(n))]\Delta J_1, \tag{21}$$

where $g(\lambda(n)) = \lambda(n)[1-\lambda(n)]^2 = f(1-\lambda(n))$. Like $f(\lambda(n)), g(\lambda(n)) \geq g(\lambda^+) = f(1-\lambda^+) = C, \forall\lambda(n) \in [1-\lambda^+, \lambda^+]$. From (17) and using arguments as used for the non-sparse case above, we then have, $E[a(n+1)] \geq E[a(n)] + \mu_a C(\Delta J_2 - \Delta J_1)$, meaning $\lim_{n\to\infty} E[a(n)] = a^+$ and thus, $\lim_{n\to\infty} a(n) = a^+$ (almost surely), or, equivalently, $\lim_{n\to\infty}\lambda(n) = \lambda^+$ (almost surely) ≈ 1. The combination filter in this case will be dominated by the ZA-LMS based filter which is the better of the two filters for sparse systems.

Case(II): $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},12}(\infty)$, (i.e., when $\sigma_{w_1}^2 < \mu\sigma_v^2/4$) meaning $\Delta J_1 > 0$.

This is again analogous to the case (3), Section III of [20]. Using arguments used for the semi-sparse case above, (18) and thus (19) will be satisfied in this case, meaning the proposed combined filter will perform better than both filter 1 and filter 2.

C. A New Simplified Adaptation Scheme for the Combining Parameter, $\lambda(n)$

Consider the update equation (3). Since $0 \leq \lambda(n) \leq 1$, the term $\lambda(n)(1-\lambda(n))$ is non-negative and thus, does not change the sign of the update term. A simplified update equation can then be obtained by dropping $\lambda(n)(1-\lambda(n))$ from the update term in (3), resulting in

$$a(n+1) = a(n) + \mu_a e(n)[y_1(n) - y_2(n)]. \tag{22}$$

Consequently, (5) changes to

$$E[a(n+1)] = E[a(n)] + \mu_a E[f_1(\lambda(n))]\Delta J_2 - \mu_a E[g_1(\lambda(n))]\Delta J_1, \tag{23}$$

where $f_1(\lambda(n)) = 1-\lambda(n)$ and $g_1(\lambda(n)) = \lambda(n)$. Clearly, for $1-\lambda^+ \leq \lambda(n) \leq \lambda^+, f_1(\lambda(n)) \geq f_1(\lambda^+) = 1-\lambda^+ = C_1$ (say) > 0 and also, $g_1(\lambda(n)) \geq g_1(1-\lambda^+) = C_1$. Therefore, as before, for non-sparse systems, i.e., for $\Delta J_2 = 0, \Delta J_1 > 0$, $\lim_{n\to\infty} E[a(n)] = -a^+$, meaning, $\lim_{n\to\infty}\lambda(n) = 1-\lambda^+$ (almost surely) ≈ 0 and thus the convex combination favors the filter 2. On the other hand, for sparse system, we first consider the case, $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},12}(\infty) > J_{\mathrm{ex},1}(\infty)$, i.e., $\Delta J_2 > 0$ and $\Delta J_1 < 0$. Clearly, in this case, $E[a(n+1)] \geq E[a(n)] +$

[3]Note that as ρ is increased from zero, its value, however, has to be kept very small and can not be increased indefinitely, as beyond a point, any further increase in ρ will again lead to enhancement of $J_{\mathrm{ex},1}(\infty)$ [18].

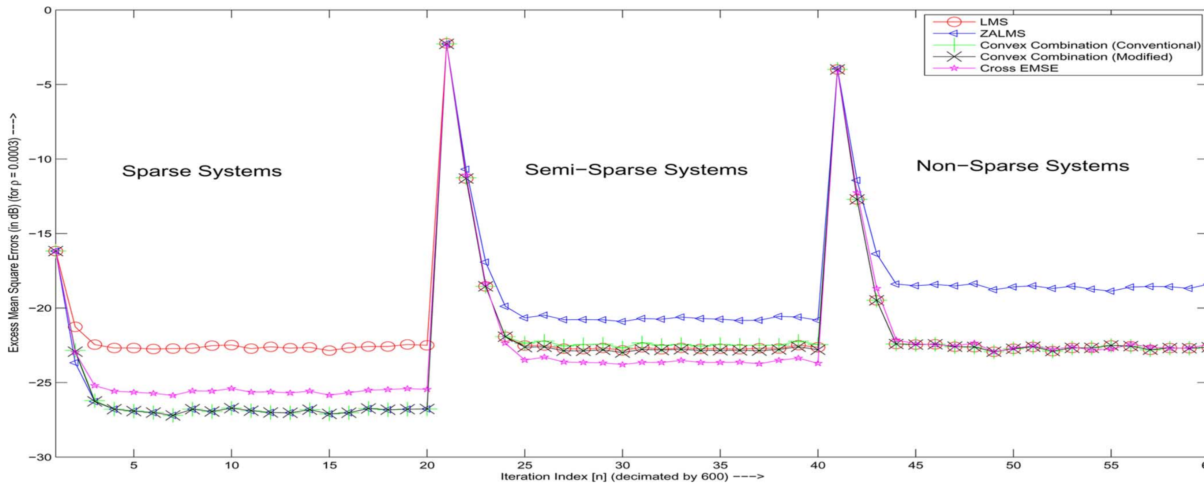

Fig. 2. The EMSE $J_{\mathrm{ex}}(n)$ for the proposed algorithm with the original update equation (3) (green line) and the modified update equation (22) (black) vis-a-vis the EMSE of the standard LMS (red) and the ZA-LMS (blue), and the cross EMSE (pink).

$\mu_a C_1[\Delta J_2 - \Delta J_1]$, meaning $\lim_{n\to\infty} E[a(n)] = a^+$ and thus $\lim_{n\to\infty} \lambda(n) = \lambda^+$ (almost surely) ≈ 1. Obviously, the convex combination in this case switches to the filter 1. For the case, $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},12}(\infty)$ and also, for $J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},12}(\infty)$ which arises in the case of semi-sparse systems, we have $\Delta J_2 > 0, \Delta J_1 > 0$ and thus a stationary point is reached, when $E[(1-\lambda(\infty))]\Delta J_2 = E[\lambda(\infty)]\Delta J_1$, which *directly leads* to (18). In this case, as discussed earlier, $J_{\mathrm{ex}}(\infty)$ satisfies (19), and thus, the convex combination works better than its individual component filters.

The update formula (22) is computationally simpler than (3) as it requires two multiplications less per time step. More importantly, for the case $\Delta J_1 > 0$, $\Delta J_2 > 0$, satisfaction of (19) (or, equivalently, of the fact that the combination works better than each component filter) does not require a strong assumption like $E[\lambda(\infty)]^2 = 0$. Instead, in this case, (18) and thus (19) follow directly from (23). In contrast, both [20] and also the analysis in part (II) above depend heavily on the assumption of $E[\lambda(\infty)]^2 = 0$ for the satisfaction of (19). Of course, by dropping the term $\lambda(n)(1-\lambda(n))$ in (3), we employ an approximate gradient in (22), or, equivalently, we introduce certain gradient noise which may lead to enhanced steady state EMSE for $a(n)$ and $\lambda(n)$ (this effect may be more pronounced near $\lambda = 0$ or 1, since, as per (3), the update term in such cases should be close to zero which is not guaranteed in (22)). A full understanding of the effect of the above approximation requires a steady state EMSE analysis for $a(n)$ or $\lambda(n)$, which seems to be intractable and is not dealt with in this paper.

III. Simulation Studies, Discussion and Conclusion

The proposed algorithm was simulated for identifying a system with 256 taps. Initially, the system was taken to be highly sparse with $\mathbf{w}_{\mathrm{opt}} = [\mathbf{0}^T_{250\times 1}, \mathbf{1}^T_{6\times 1}]^T$, where $\mathbf{0}_{m\times n}$ and $\mathbf{1}_{m\times n}$ denote $m \times n$ matrices of all zeros and all ones respectively. After 12000 time steps, the system was changed to a semi-sparse system with $\mathbf{w}_{\mathrm{opt}} = [\mathbf{0}^T_{100\times 1}, \mathbf{1}^T_{156\times 1}]^T$, and finally, after 24000 time steps, the system was changed to a non-sparse system with $\mathbf{w}_{\mathrm{opt}} = [\mathbf{1}^T_{256\times 1}]^T$. The simulation was carried out for a total of 36000 iterations, with $\mu = 0.7$, $\rho = 0.0003$, $\sigma_v^2 = 0.01$, and with the input $x(n)$ taken as a zero mean, white random process with variance 0.0039 ($\mathrm{Tr}\mathbf{R} = 1$). Filters with 256 taps were chosen both for the LMS and the ZA-LMS algorithms, with the initial tap values chosen randomly. The simulation results are shown in Fig. 2 by plotting the EMSE $J_{\mathrm{ex}}(n)$ against the iteration index n, obtained by averaging $e_a^2(n)$ over 3000 experiments, for the original update equation (3) (green line, $\mu_a = 10^4$) and the modified update equation (22) (black line, $\mu_a = 10$). Also plotted in Fig. 2 are the two EMSE's, $J_{\mathrm{ex},1}(n)$ (blue line) and $J_{\mathrm{ex},2}(n)$ (red line) and the cross EMSE $J_{\mathrm{ex},12}(n)$ (pink line). The parameter μ_a was adjusted so that $J_{\mathrm{ex}}(n)$ for both (3) and (22) maintain similar rates of convergence as that of $J_{\mathrm{ex},1}(n)$ and $J_{\mathrm{ex},2}(n)$ so that a fair comparison between them can be made based on the respective steady state EMSE. Also, to preserve clarity of visualization, the plots are shown by decimating the time interval 0–36000 by a decimation factor of 600.

Several interesting observations follow from Fig. 2. Firstly, the plots confirm our conjecture that for non-sparse systems, $J_{\mathrm{ex},2}(\infty) \approx J_{\mathrm{ex},12}(\infty)$ while for all other cases (i.e., sparse and semi-sparse systems), $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},12}(\infty)$. Secondly, for the chosen value of ρ, it is seen that $J_{\mathrm{ex},1}(\infty) < J_{\mathrm{ex},12}(\infty)$ when the system is highly sparse. As per our analysis above, the proposed combination in this case should converge to the ZA-LMS based filter, meaning we should have $J_{\mathrm{ex}}(\infty) \approx J_{\mathrm{ex},1}(\infty)$ for both the update equation (3) and (22), which is clearly satisfied in Fig. 2. Similarly, for the non-sparse system, it is seen from Fig. 2 that $J_{\mathrm{ex}}(\infty) \approx J_{\mathrm{ex},2}(\infty)$ for both (3) and (22), meaning the convex combination in this case favors filter 2, which is in conformity with our arguments above. Lastly, for the semi-sparse system, the plots confirm that $J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},12}(\infty)$. As per the discussions of the previous section, the proposed combination in this case is likely to produce a solution that performs better than both filter 1 and filter 2. A close examination of Fig. 2, however, reveals that while $J_{\mathrm{ex}}(\infty)$ for (3) remains very close to $J_{\mathrm{ex},2}(\infty)$, it is $J_{\mathrm{ex}}(\infty)$ for (22) *which actually goes below* $J_{\mathrm{ex},2}(\infty)$, i.e., attains the minimum EMSE value. This, in our opinion, is because, in the case of (22), the

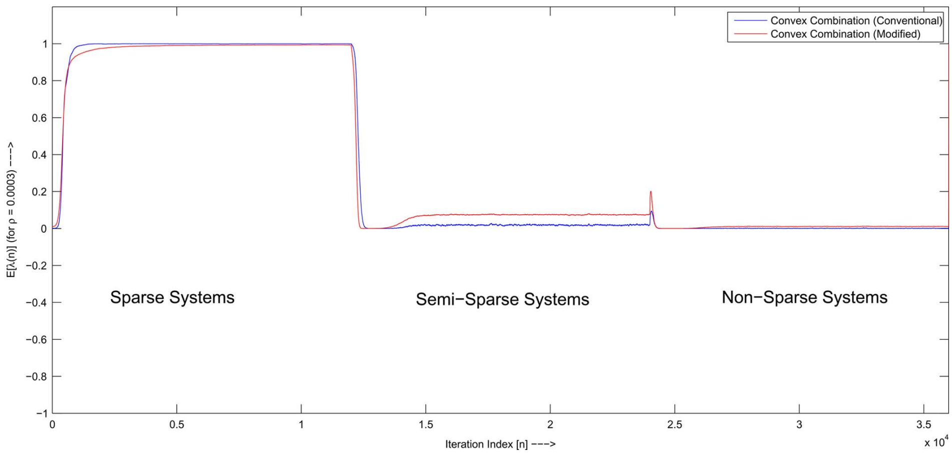

Fig. 3. Evolution of $E[\lambda(n)]$ over iterations.

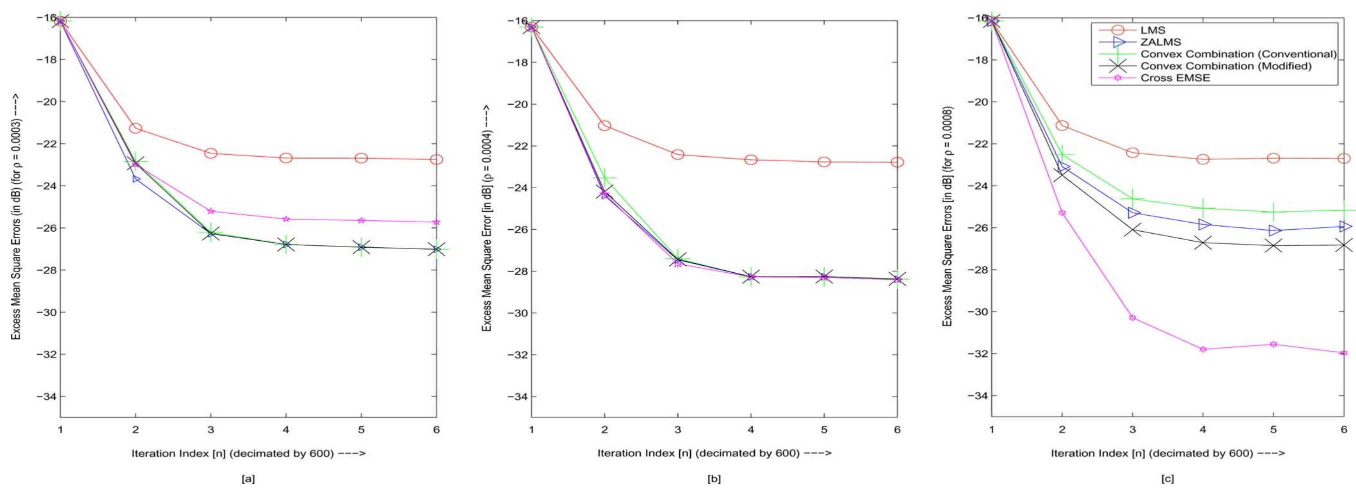

Fig. 4. Simulation results for a sparse system with (a) $\rho = 0.0003$ showing $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},12}(\infty) > J_{\mathrm{ex},1}(\infty)$, (b) $\rho = 0.0004$ showing $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},12}(\infty) \approx J_{\mathrm{ex},1}(\infty)$ and (c) $\rho = 0.0008$ showing $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},12}(\infty)$.

condition (18) follows directly, whereas in the case of (3), approximations in the form of the assumption $E[\lambda(\infty)]^2 \to 0$ are required for this to be satisfied. Also note that even though the modified update equation (22) introduces certain gradient noise which becomes more pronounced near $\lambda = 0$ and $\lambda = 1$, i.e., for sparse and non-sparse systems, no appreciable degradation in $J_{\mathrm{ex}}(\infty)$ is, however, observed for (22) due to the above. Apart from plotting the learning curves (i.e., $J_{\mathrm{ex}}(\infty)$ vs n), another useful way of studying the convergence behavior of the proposed scheme is to observe the evolution of $E[\lambda(n)]$ with time for the three systems considered. For this, we plot $E[\lambda(n)]$ vs n in Fig. 3 for both (3) and (22). As predicted, $E[\lambda(n)]$ converges to one and zero respectively for sparse and non-sparse systems, and to an intermediate value for semi-sparse systems.

Next we verify our claim that for a highly sparse system, $J_{\mathrm{ex},12}(\infty) \geq J_{\mathrm{ex},1}(\infty)$ as long as ρ remains bounded from above by a small positive constant, and that for ρ going beyond this upper limit, $J_{\mathrm{ex},12}(\infty) < J_{\mathrm{ex},1}(\infty)$. For this, we consider the sparse system used above and conduct the same simulation experiment for three different values of ρ, namely, $\rho = 0.0003, 0.0004$ and 0.0008. The corresponding results, shown in Figs. 4(a), (b) and (c) respectively, demonstrate that for $\rho = 0.0003$ (i.e., Fig. 4(a)), $J_{\mathrm{ex},12}(\infty) > J_{\mathrm{ex},1}(\infty)$, for $\rho = 0.0004$ (i.e., Fig. 4(b)), $J_{\mathrm{ex},12}(\infty) \approx J_{\mathrm{ex},1}(\infty)$, and for $\rho = 0.0008$ (i.e., Fig. 4(c)), $J_{\mathrm{ex},12}(\infty) < J_{\mathrm{ex},1}(\infty)$. Further, in both Figs. 4(a) and (b), as predicted, $J_{\mathrm{ex}}(\infty) \approx J_{\mathrm{ex},1}(\infty)$ for both (3) and (22), i.e., the proposed combination converges to filter 1. In the case of Fig. 4(c), however, $J_{\mathrm{ex}}(\infty)$ goes below both $\approx J_{\mathrm{ex},1}(\infty)$ and $\approx J_{\mathrm{ex},1}(\infty)$ only for the modified update equation (22). Reasons for this, in our opinion, are same as stated above for the case of semi-sparse systems.

Appendix A
Proof of Theorem 1

In this section, we first present a proof of (11) and later show that $E[\mathrm{sgn}\{w_{1,i}(\infty)\}\ \tilde{w}_{2,i}(\infty)] < 0$ for $i \in Z$. For this, first consider $E[\tilde{\mathbf{w}}_1(n+1)\tilde{\mathbf{w}}_2^T(n+1)]$. Substituting $\tilde{\mathbf{w}}_1(n+1) = \tilde{\mathbf{w}}_1(n) - \mu\, e_1(n)\,\mathbf{x}(n) + \rho\,\mathrm{sgn}(\mathbf{w}_1(n))$ and $\tilde{\mathbf{w}}_2(n+1) = \tilde{\mathbf{w}}_2(n) - \mu\, e_2(n)\,\mathbf{x}(n)$, replacing $e_i(n), i = 1, 2$ by $\tilde{\mathbf{w}}_i^T(n)\mathbf{x}(n) + v(n) \equiv \mathbf{x}^T(n)\tilde{\mathbf{w}}_i(n) + v(n)$, using the statistical independence between $\mathbf{w}_i(n)$ and $\mathbf{x}(n)$ (i.e., "independence assumption" [7]), and recalling that $v(n)$ is zero-mean, also independent of $\mathbf{x}(n)$ and thus of $\tilde{\mathbf{w}}_i(n)$, we have

$$\begin{aligned} &E\left[\tilde{\mathbf{w}}_1(n+1)\tilde{\mathbf{w}}_2^T(n+1)\right] \\ &\quad= E\left[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)\right] - \mu\mathbf{R}E\left[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)\right] \\ &\quad\quad - \mu E\left[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)\right]\mathbf{R} + \mu^2\sigma_v^2\mathbf{R} \\ &\quad\quad + \mu^2 E[\mathbf{x}(n)e_{a,1}(n)e_{a,2}(n)\mathbf{x}^T(n)] \\ &\quad\quad - \mu\rho E\left[\text{sgn}\{\mathbf{w}_1(n)\}\tilde{\mathbf{w}}_2^T(n)\right]\mathbf{R} \\ &\quad\quad + \rho E\left[\text{sgn}\{\mathbf{w}_1(n)\}\tilde{\mathbf{w}}_2^T(n)\right] \end{aligned} \quad \text{(A.1)}$$

where $\mathbf{R} = \sigma_x^2\mathbf{I}$. Next, for the term $E[\mathbf{x}(n)e_{a,1}(n) e_{a,2}(n)\mathbf{x}^T(n)] \equiv E[\mathbf{x}(n)\mathbf{x}^T(n)\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)\,\mathbf{x}(n)\mathbf{x}^T(n)]$, we recall the following result used in the second order analysis of the LMS algorithm [7]: given a set of zero-mean, uncorrelated, jointly Gaussian random variables $x_i, i = 1,2,\ldots,p$ and another set of random variables $y_i, i = 1,2,\ldots,p$ that is independent of x_i's, we have, $E[\mathbf{x}\mathbf{x}^T\mathbf{y}\mathbf{y}^T\mathbf{x}\mathbf{x}^T] = 2\mathbf{DKD} + \text{Tr}[\mathbf{DK}]\mathbf{D}$, where $\mathbf{x} = [x_1x_2\ldots x_p]^T, \mathbf{y} = [y_1y_2\ldots y_p]^T, \mathbf{D} = E[\mathbf{x}\mathbf{x}^T]$ (a diagonal matrix) and $\mathbf{K} = E[\mathbf{y}\mathbf{y}^T]$. In the present case, $\mathbf{D} = \sigma_x^2\mathbf{I}$, meaning $E[\mathbf{x}(n)e_{a,1}(n)e_{a,2}(n)\mathbf{x}^T(n)] = 2\sigma_x^4E[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)] + \sigma_x^4\text{Tr}[E[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)]] \equiv 2\sigma_x^4E[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)] + \sigma_x^4E[\tilde{\mathbf{w}}_1^T(n)\tilde{\mathbf{w}}_2(n)]]$. Substituting in (A.1), we obtain

$$\begin{aligned} &E\left[\tilde{\mathbf{w}}_1(n+1)\tilde{\mathbf{w}}_2^T(n+1)\right] = E\left[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)\right] \\ &\quad - 2\mu\sigma_x^2E\left[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)\right] + \mu^2\sigma_v^2\sigma_x^2\mathbf{I} \\ &\quad + 2\mu^2\sigma_x^4E\left[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)\right] + \mu^2\sigma_x^4E\left[\tilde{\mathbf{w}}_1^T(n)\tilde{\mathbf{w}}_2(n)\right]\mathbf{I} \\ &\quad + \rho\left(1-\mu\sigma_x^2\right)E\left[\text{sgn}\{\mathbf{w}_1(n)\}\tilde{\mathbf{w}}_2^T(n)\right]. \end{aligned} \quad \text{(A.2)}$$

Defining $\mathbf{c}(n) = \text{diag}[E[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)]$ and $\mathbf{b}(n) = \text{diag}[E[\text{sgn}[\mathbf{w}_1(n)]\tilde{\mathbf{w}}_2^T(n)]]$ ("diag[.]" denotes the vector consisting of the diagonal elements of the matrix in the argument), and collecting terms, we can then write from (A.2),

$$\begin{aligned} \mathbf{c}(n+1) &= \mathbf{c}(n)\left(1-2\mu\sigma_x^2+2\mu^2\sigma_x^4\right) + \mu^2\sigma_x^2\sigma_v^2\mathbf{1} \\ &\quad + \rho\left(1-\mu\sigma_x^2\right)\mathbf{b}(n) + \mu^2\sigma_x^4E\left[\tilde{\mathbf{w}}_1^T(n)\tilde{\mathbf{w}}_2(n)\right])\,\mathbf{1} \end{aligned} \quad \text{(A.3)}$$

where we define $\mathbf{1}$ as a column vector of appropriate dimension, whose each member is 1 (the dimension of $\mathbf{1}$ is determined by the other vectors associated with it, e.g., when we write $\mathbf{1}^T\mathbf{x}$, dimension of $\mathbf{1}$ is the same as that of $\mathbf{x}$). Next, we regroup the elements of $\mathbf{c}(n)$ into two vectors, $\mathbf{c}_Z(n)$ and $\mathbf{c}_{\text{NZ}}(n)$, consisting of $E[\tilde{w}_{1,i}(n)\tilde{w}_{2,i}(n)]$ respectively for $i \in Z$ and $i \in \text{NZ}$. Similarly, $\mathbf{b}(n)$ is regrouped into vectors $\mathbf{b}_Z(n)$ and $\mathbf{b}_{\text{NZ}}(n)$, consisting of $E[\text{sgn}\{w_{1,i}(n)\}\tilde{w}_{2,i}(n)]$ respectively for $i \in Z$ and $i \in \text{NZ}$. Let us consider first $\mathbf{b}_{\text{NZ}}(n)$. As assumed earlier, for large n (i.e., in the steady state), $\text{sgn}\{w_{1,i}(n)\} = \text{sgn}\{w_{\text{opt},i}\}$. Since $E[\tilde{w}_{2,i}(n)] \to 0$ in the steady state, we have, for large n, $\mathbf{b}_{\text{NZ}}(n) = \mathbf{0}$.

Next we consider $\mathbf{b}_Z(n)$. Note that for $i \in Z$, $\tilde{w}_{2,i}(n) = w_{\text{opt},i} - w_{2,i}(n) = -w_{2,i}(n)$. Assuming $w_{1,i}(n)$ and $w_{2,i}(n)$ to be jointly Gaussian (having mean zero in the steady state, as $w_{\text{opt},i} = 0$), we can use Price's theorem [23] to write, for large n, $E[\text{sgn}\{w_{1,i}(n)\}\tilde{w}_{2,i}(n)] = -gE[w_{1,i}(n)w_{2,i}(n)] = -gE[\tilde{w}_{1,i}(n)\tilde{w}_{2,i}(n)]$, where $g = \sqrt{2/(\pi\sigma_{w_{1,i}}^2)}$, with $\sigma_{w_{1,i}}$ denoting the steady state standard deviation of $w_{1,i}(n)$.[4] Clearly, for large n, $\mathbf{b}_Z(n) = -g\mathbf{c}_Z(n)$.

Making the substitution $\mathbf{c}(n) \to [\mathbf{c}_{\text{NZ}}^T(n), \mathbf{c}_Z^T(n)]^T$ and $\mathbf{b}(n) \to -g[\mathbf{0}^T, \mathbf{c}_Z^T(n)]^T$ in (A.3), premultiplying the L.H.S. and the R.H.S. by $\mathbf{1}^T$, and noting $\mathbf{1}^T\mathbf{c}(n) = E[\tilde{\mathbf{w}}_1^T(n)\tilde{\mathbf{w}}_2(n)] = \theta(n)$ (say), we have

$$\begin{aligned} \theta(n+1) &= \left(1-2\mu\sigma_x^2+2\mu^2\sigma_x^4\right)\mathbf{1}^T\mathbf{c}_{\text{NZ}}(n) \\ &\quad + \left(1-2\mu\sigma_x^2+2\mu^2\sigma_x^4\right. \\ &\quad \left.-\rho g\left(1-\mu\sigma_x^2\right)\right)\mathbf{1}^T\mathbf{c}_Z(n) \\ &\quad + L\mu^2\sigma_x^4\theta(n) + \mu^2\sigma_x^2\sigma_v^2L. \end{aligned} \quad \text{(A.4)}$$

Neglecting $\rho g(1-\mu\sigma_x^2)$ in comparison to $(1-2\mu\sigma_x^2+2\mu^2\sigma_x^4) = (1-\mu\sigma_x^2)^2+\mu^2\sigma_x^4$, since ρ is very very small [18], we obtain from (A.4)

$$\theta(n+1) = \left(1-2\mu\sigma_x^2+(L+2)\mu^2\sigma_x^4\right)\theta(n) + \mu^2\sigma_x^2\sigma_v^2L. \quad \text{(A.5)}$$

Clearly, $\theta(n)$ converges as $n \to \infty$, if $|1-2\mu\sigma_x^2+(L+2)\mu^2\sigma_x^4| < 1$. Let us first consider $(1-2\mu\sigma_x^2+(L+2)\mu^2\sigma_x^4) > -1$, or equivalently, $(1-\mu\sigma_x^2/2)^2+(L/2+3/4)\mu^2\sigma_x^4 > 0$, which is always satisfied. Next we consider $(1-2\mu\sigma_x^2+(L+2)\mu^2\sigma_x^4) < 1$, which is satisfied for $0 < \mu < 2/((L+2)\sigma_x^2)$. Note that this condition is almost same as the convergence (in mean) condition of the LMS algorithm : $0 < \mu < 2/\text{Tr}(\mathbf{R}) = 2/(L\sigma_x^2)$ and is thus easily satisfied. This proves (11).

Next, it follows easily from (A.3) that

$$\theta(\infty) = \lim_{n\to\infty}\theta(n) = \frac{\mu^2\sigma_x^2\sigma_v^2L}{\mu\sigma_x^2\left(2-(L+2)\mu\sigma_x^2\right)} > 0 \quad \text{(A.6)}$$

since $\mu < 2/((L+2)\sigma_x^2)$. From (A.3), using the same approximation as above, we can write, for $i \in Z$,

$$\begin{aligned} c_i(n+1) &= c_i(n)\left(1-2\mu\sigma_x^2+2\mu^2\sigma_x^4\right) \\ &\quad + \mu^2\sigma_x^2\left(\sigma_v^2+\sigma_x^2\theta(n)\right). \end{aligned} \quad \text{(A.7)}$$

It is easy to see that $(1-2\mu\sigma_x^2+2\mu^2\sigma_x^4) = (1-\mu\sigma_x^2)^2+\mu^2\sigma_x^4 < 1$, for $0 < \mu\sigma_x^2 < 1$, which is satisfied with the standard LMS convergence condition: $0 < \mu < 2/\text{Tr}(\mathbf{R}) = 2/(L\sigma_x^2)$ (it is assumed that $L \geq 2$). This implies that $c_i(n)$ converges as $n \to \infty$.

From (A.6) and (A.7), we can write

$$c_i(\infty) = \frac{\mu\left(\sigma_v^2+\sigma_x^2\theta(\infty)\right)}{2\left(1-\mu\sigma_x^2\right)} > 0. \quad \text{(A.8)}$$

Clearly, for $i \in Z$,

$$\begin{aligned} E[\text{sgn}\{w_{1,i}(\infty)\}\tilde{w}_{2,i}(\infty)] &= -E[\text{sgn}\{w_{1,i}(\infty)\}w_{2,i}(\infty)] \\ &= -gE[w_{1,i}(\infty)w_{2,i}(\infty)] = -gc_i(\infty) < 0. \end{aligned}$$

[4]As explained earlier, since $x(n)$ is i.i.d., for $i \in Z$, we have $\sigma_{w_{1,i}}^2 \equiv \sigma_{w_1}^2$: same for all i and thus g is taken as independent of i above, though it can be easily verified that the proof of this theorem does not require g to be same for all i.

APPENDIX B
PROOF OF THEOREM 2

First, from (6) and (9), it is easy to observe that $J_{\text{ex},1}(\infty) - J_{\text{ex},2}(\infty)$ is given by $\rho/(\mu(2-\mu\text{Tr}(\mathbf{R})))(\alpha_1\rho - 2\alpha_2)$, which under small misadjustment condition (i.e., for $\mu \ll 2/\text{Tr}(\mathbf{R}) = 2/(L\sigma_x^2)$) reduces to

$$\frac{\rho}{2\mu}(\alpha_1\rho - 2\alpha_2). \tag{B.1}$$

Next consider α_1 as given by (7). For i.i.d. input and with $\mu \ll 1/\sigma_x^2$, we have $\mathbf{I} - \mu\mathbf{R} = (1-\mu\sigma_x^2)\mathbf{I} \approx \mathbf{I}$. Since $\text{sgn}(\mathbf{w}_1(\infty))^T \text{sgn}(\mathbf{w}_1(\infty)) = L$, we then have in such case $\alpha_1 \approx L$. Next consider α_2 as given by (8) which can be equivalently expressed as

$$\alpha_2 = \sum_{i=0}^{L-1} E[|w_{1,i}(\infty)|] - |w_{\text{opt},i}|. \tag{B.2}$$

As explained earlier, $\mathbf{w}_1(\infty)$ can be assumed to be Gaussian. Then, recalling that $E[w_{1,i}(\infty)] \approx w_{\text{opt},i}$, we can write using the definition of folded normal distribution [23]

$$E[|w_{1,i}(\infty)|] = w_{\text{opt},i}\left[1 - 2\,\text{erf}\left(-\frac{w_{\text{opt},i}}{\sigma_{w_{1,i}}}\right)\right] + \sqrt{\frac{2}{\pi}}\sigma_{w_{1,i}} \exp\left(-\frac{w_{\text{opt},i}^2}{2\sigma_{w_{1,i}}^2}\right) \tag{B.3}$$

where $\text{erf}(\cdot)$ is the well known error function [23] (i.e., $\text{erf}(z) = \int_{-\infty}^{z}(1/\sqrt{2\pi})\exp(-x^2/2)dx$). First consider $w_{\text{opt},i}$ to be an active tap. Since the active taps have been assumed to have significant magnitudes and since, under small misadjustment condition, $\sigma_{w_{1,i}}$ is very small, the following observations on the R.H.S. of (B.3) can be made : for $w_{\text{opt},i} > 0$, $\text{erf}(-(w_{\text{opt},i})/(\sigma_{w_{1,i}})) \approx 0$, for $w_{\text{opt},i} < 0$, $\text{erf}(-(w_{\text{opt},i})/(\sigma_{w_{1,i}})) \approx 1$ and for both positive as well as negative $w_{\text{opt},i}$, $\exp(-(w_{\text{opt},i}^2)/(2\sigma_{w_{1,i}}^2)) \approx 0$. From these, it is easy to see that for $i \in \text{NZ}$, $E[|w_{1,i}(\infty)|] \approx |w_{\text{opt},i}|$. For $i \in Z$, on the other hand, $w_{\text{opt},i} = 0$ and thus, we have, $E[|w_{1,i}(\infty)|] = \sqrt{2/\pi}\sigma_{w_{1,i}}$. Making these substitutions in (B.2), we observe that $\alpha_2 = \sqrt{2/\pi}\sum_{i\in Z}\sigma_{w_{1,i}}$. Substituting α_2 in (B.1) by this and α_1 by L, the result is easily proved.

REFERENCES

[1] J. Radecki, Z. Zilic, and K. Radecka, "Echo cancellation in IP networks," in *Proc.45th Midwest Symp. Circuits Syst.*, 2002, vol. 2, pp. 219–222.

[2] E. Hansler, "The hands-free telephone problem—An annotated bibliography," *Signal Process.*, vol. 27, no. 3, pp. 259–271, Jun. 1992.

[3] W. Schreiber, "Advanced television systems for terrestrial broadcasting," *Proc. IEEE*, vol. 83, no. 6, pp. 958–981, 1995.

[4] W. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. IEEE CISS*, 2008, pp. 5–10.

[5] M. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in *Proc. IEEE OCEANS*, 1995, pp. 1417–1422.

[6] R. L. Das and M. Chakraborty, "Sparse adaptive filters—An overview and some new results," in *Proc. ISCAS-2012*, Seoul, South Korea, May 2012, pp. 2745–2748.

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

[8] D. L. Duttweiler, "Proportionate normalized least mean square adaptation in echo cancellers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.

[9] H. Deng and M. Doroslovacki, "Improving convergence of the PNLMS algorithm for sparse impulse response identification," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 181–184, Mar. 2005.

[10] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE ICASSP*, Orlando, FL, USA, May 2002, vol. 2, pp. 1881–1884.

[11] J. Arenas-García and A. R. Figueiras-Vidal, "Adaptive combination of proportionate filters for sparse echo cancellation," *IEEE Trans. Audio, Speech Language Process.*, vol. 17, no. 6, pp. 1087–1098, Aug. 2009.

[12] J. Homer, I. Mareels, and C. Hoang, "Enhanced detection-guided NLMS estimation of sparse FIR-modeled signal channels," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 2, pp. 1783–1791, Aug. 2006.

[13] M. Godavarti and A. O. Hero, "Partial update LMS algorithms," *IEEE Trans. Signal Processing*, vol. 53, pp. 2382–2399, 2005.

[14] R. K. Martin, W. A. Sethares, R. C. Williamson, and C. R. Johnson, Jr, "Exploiting sparsity in adaptive filters," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1883–1894, Aug. 2002.

[15] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Royal. Statist. Soc B.*, vol. 58, pp. 267–288, 1996.

[16] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[17] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 25, pp. 21–30, Mar. 2007.

[18] Y. Gu, Y. Chen, and A. O. Hero, "Sparse LMS for system identification," in *Proc. IEEE ICASSP*, Taipei, Taiwan, Apr. 2009.

[19] Y. Chen, Y. Gu, and A. O. Hero, III, Regularized Least-Mean-Square Algorithms arXiv:1012.5066v2 [stat.ME], Dec. 23, 2010.

[20] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1078–1090, Mar. 2006.

[21] B. K. Das, M. Chakraborty, and S. Banerjee, "Adaptive identification of sparse systems with variable sparsity," in *Proc. ISCAS-2011*, Rio de Janeiro, Brazil, May 2011, pp. 1267–1270.

[22] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York, NY, USA: Wiley, 2003.

[23] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw Hill, 2002.

Bijit Kumar Das (S'11) received the B.Eng. degree in electronics and telecommunication engineering from Bengal Engineering and Science University, Shibpur, India, in 2007, and the M.S. (by research) degree in 2011 in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, where he is currently pursuing the Ph.D. degree in adaptive signal processing.

His current research interests include adaptive filter theory, VLSI signal processing, digital and statistical signal processing.

Mrityunjoy Chakraborty (M'94-SM'99) is currently a professor of Electronics and Electrical Comm. Engg at the Indian Institute of Technology, Kharagpur, India, with teaching and research interests in Digital and Adaptive Signal Processing, VLSI Signal Processing, Linear Algebra and Compressed Sensing.

Dr. Chakraborty served as an Associate Editor of both IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS (2004–2007, 2010-2012) and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS (2008–2009) and currently represents IEEE SPS in the editorial board of the IEEE Computing in Science and Engineering Magazine. He has been a distinguished lecturer of the Asia Pacific Signal and Information Processing Association and is a fellow of the Indian National Academy of Engineering (INAE).