A Convex Combination of NLMS and ZA-NLMS for Identifying Systems With Variable Sparsity

Bijit K. Das, *Member, IEEE*, G. Vinay Chakravarthi, and Mrityunjoy Chakraborty, *Senior Member, IEEE*

***Abstract*—This brief aims to identify and track a sparse system with time varying sparseness by a convex combination of two adaptive filters, one based on the sparsity unaware normalized least mean square (NLMS) algorithm and the other based on the sparsity aware zero-attracting NLMS (ZA-NLMS) algorithm. An analysis of the proposed combination is carried out, which reveals that while the proposed combination converges to the ZA-NLMS or the NLMS-based filter for systems that are highly sparse or highly non-sparse, respectively (i.e., better of the two under the given sparsity condition), it may, however, lead to a filter that performs better than both the constituent filters in the case of systems that lie between moderately sparse to moderately non-sparse. The same is confirmed via detailed simulation studies under different sparsity conditions.**

***Index Terms*—Adaptive filter, excess mean square error (EMSE), l_1-norm, normalized least mean square (NLMS), sparse systems.**

I. Introduction

Recently, combination schemes [1], [2] involving two or more adaptive filters have been proposed for improving filter performance in terms of convergence rate, robustness, steady state misadjustment, tracking ability, etc. In the context of sparse system identification [3], a convex combination of two adaptive filters, one employing the sparsity aware zero-attracting LMS algorithm [4] and the other deploying the conventional, sparsity unaware LMS mode of adaptation [5] was proposed in [6], following the general approach of convex combination of two adaptive filters [7]. Analysis showed that depending on the level of sparseness, from highly sparse to non-sparse, the combination switches between the two filters, always making the correct choice. The LMS, however, is a very basic form of adaptation and is often replaced by its normalized form, i.e., the so-called normalized least mean square (NLMS) algorithm [5]. In NLMS, the step size used is normalized by the Euclidean norm of the tap input vector which makes it more capable to handle fluctuations in input statistics than LMS. While it is possible to formulate a convex combination of the NLMS and its sparsity aware counterpart, the zero-attracting NLMS (ZA-NLMS) [8] along the lines of [6] for identifying and tracking a sparse system with time varying sparseness, analysis of such a combination is, however, extremely difficult due to the presence of normalized update terms in the update equation of both the filters and a nonlinear term in the update equation of the ZA-NLMS. In this brief, we consider a convex combination of NLMS and ZA-NLMS algorithms, and evaluate its behavior analytically, by using an elegant scheme of angular discretization of random vectors [9] that helps in reducing the aforesaid complexities. The analysis shows that while for systems that are highly sparse or highly non-sparse, the proposed combination converges to the ZA-NLMS or the NLMS-based filter, respectively (i.e., better of the two under the given sparsity condition), for systems that lie between moderately sparse to moderately non-sparse, the proposed combination can, however, perform better than both the constituent filters. The same is also verified via extensive simulation studies under different sparsity conditions.

Manuscript received November 28, 2016; revised January 4, 2017; accepted January 7, 2017. Date of publication January 11, 2017; date of current version August 25, 2017. This brief was recommended by Associate Editor M. Small.

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSII.2017.2651388

II. Proposed Convex Combination

We consider here the problem of identifying a system that takes a random input $x(n)$ and produces the observable output $y_d(n)$ as $y_d(n) = \mathbf{w}_0^T \mathbf{x}(n) + \eta(n)$, where $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-N+1)]^T$, $\mathbf{w}_0$ is a $N \times 1$ impulse response vector (to be identified) having variable sparsity and $\eta(n)$ is an observation noise which is assumed to be zero mean, white with variance σ_η^2 and independent of the input data vector $\mathbf{x}(m)$ for any m and n. If $\mathbf{w}_0$ is sufficiently sparse, the ZA-NLMS algorithm will be a better choice than NLMS to identify it as it will produce lesser steady state excess mean square error (EMSE) than produced by NLMS. On the other hand, if $\mathbf{w}_0$ is less sparse or non-sparse, performance of NLMS will be superior to that of ZA-NLMS. Since $\mathbf{w}_0$ is assumed to have sparsity that varies over a wide range in time, either filter alone does not make an ideal choice for identifying and tracking the system. Instead, to achieve robustness against time varying sparsity, we deploy a convex combination of the two (similar to Fig. 1 of [6]), where filter 1 uses ZA-NLMS to adapt a filter coefficient vector $\mathbf{w}_1(n)$ as [8]

$$\mathbf{w}_1(n+1) = \mathbf{w}_1(n) - \rho \,\mathrm{sgn}(\mathbf{w}_1(n)) + \mu \frac{e_1(n)\,\mathbf{x}(n)}{\delta + ||\mathbf{x}(n)||^2} \tag{1}$$

and filter 2 uses NLMS for adapting a filter coefficient vector $\mathbf{w}_2(n)$ as

$$\mathbf{w}_2(n+1) = \mathbf{w}_2(n) + \mu \frac{e_2(n)\,\mathbf{x}(n)}{\delta + ||\mathbf{x}(n)||^2} \tag{2}$$

where μ is the usual step size (assumed same for both the filters), ρ is a suitable constant (usually very very small), δ is a very small regularization parameter, and

$e_i(n) = y_d(n) - y_i(n),\ i = 1, 2$ is the ith filter output error with $y_i(n) = \mathbf{w}_i^T(n)\mathbf{x}(n)$ denoting the ith filter output. The convex combination generates a combined output $y(n) = \gamma(n)y_1(n) + [1 - \gamma(n)]y_2(n)$. The variable $\gamma(n)$ 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 $\gamma(n)$ will always lie between 0 and 1. Therefore, instead of $\gamma(n)$, an equivalent variable $a(n)$ is updated which expresses $\gamma(n)$ as a sigmoidal function, i.e., $\gamma(n) = \frac{1}{1+\exp(-a(n))}$. The update equation of $a(n)$ is given by [7]

$$a(n+1) = a(n) - \frac{\mu_a}{2}\frac{\partial e^2(n)}{\partial a(n)} = a(n) + \mu_a\, e(n)\, \big[y_1(n) - y_2(n)\big]\gamma(n)\big[1 - \gamma(n)\big] \quad (3)$$

where μ_a is an appropriate step size. In practice, $\gamma(n) \approx 1$ for $a(n) >> 0$ and conversely, $\gamma(n) \approx 0$ for $a(n) << 0$. Therefore, instead of updating $a(n)$ up to $\pm\infty$, it is sufficient to restrict it to a range $[-a^+, +a^+]$ (a^+ : a large, positive number) which limits the permissible range of $\gamma(n)$ to $[1 - \gamma^+, \gamma^+]$, where $\gamma^+ = \frac{1}{1+\exp(-a^+)}$.

III. Performance Analysis

A. Some Important Definitions

Following [7], we introduce certain definitions here which will be useful in the proposed performance analysis. For $i = 1, 2$, we define the following.

1) *Weight error vectors* $\widetilde{\mathbf{w}}_i(n) = \mathbf{w}_0 - \mathbf{w}_i(n)$.
2) *Equivalent weight vector* for the combined filter $\mathbf{w}_c(n) = \gamma(n)\mathbf{w}_1(n) + [1 - \gamma(n)]\mathbf{w}_2(n)$.
3) *Equivalent weight error vector* for the combined filter $\widetilde{\mathbf{w}}_c(n) = \mathbf{w}_0 - \mathbf{w}_c(n) = \gamma(n)\widetilde{\mathbf{w}}_1(n) + [1 - \gamma(n)]\widetilde{\mathbf{w}}_2(n)$.
4) *A priori* errors $e_{a,i}(n) = \widetilde{\mathbf{w}}_i^T(n)\mathbf{x}(n)$ and $e_a(n) = \widetilde{\mathbf{w}}_c^T(n)\mathbf{x}(n)$ (clearly, $e_a(n) = \gamma(n)\, e_{a,1}(n) + [1 - \gamma(n)]\, e_{a,2}(n)$ and $e(n) = e_a(n) + \eta(n)$).
5) *EMSE* $J_{\text{ex},i}(n) = E[e_{a,i}^2(n)],\ i = 1, 2,\ J_{\text{ex}}(n) = E[e_a^2(n)]$.
6) *Cross EMSE* $J_{\text{ex},12}(n) = E[e_{a,1}(n)e_{a,2}(n)]$.

From Cauchy–Schwartz inequality, $J_{\text{ex},12}(n) \leq \sqrt{J_{\text{ex},1}(n)}\sqrt{J_{\text{ex},2}(n)}$, which means, $J_{\text{ex},12}(n)$ cannot be greater than both $J_{\text{ex},1}(n)$ and $J_{\text{ex},2}(n)$ simultaneously. Also, we define by Z and NZ the sets of filter tap indices corresponding to the inactive and active taps of $\mathbf{w}_0$, respectively.

As in [7], 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 $\eta(n)$ for all n. It is also assumed that $x(n)$ is WSS with $E[x(n)] = 0$ and $E[\mathbf{x}(n)\mathbf{x}^T(n)] = \mathbf{R}$, and that $x(n)$ and $\eta(n)$ are jointly Gaussian processes. Taking expectation on both sides of (3), one can then write

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

Substituting $e(n)$ by $e_a(n)+\eta(n)$ 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 $\eta(n)$ is zero mean, white, and assuming like [7] that in the steady state, $\gamma(n)$ is independent of the *a priori errors* $e_{a,i}(n)$, it is easy to verify [7] that for large n (theoretically, for $n \to \infty$)

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

where $\Delta J_1 = J_{\text{ex},1}(\infty) - J_{\text{ex},12}(\infty)$ and $\Delta J_2 = J_{\text{ex},2}(\infty) - J_{\text{ex},12}(\infty)$, with $J_{\text{ex},i}(\infty) = \lim_{n\to\infty} J_{\text{ex},i}(n)$ and $J_{\text{ex},12}(\infty) = \lim_{n\to\infty} J_{\text{ex},12}(n)$ denoting, respectively, the steady state EMSE and the steady state cross EMSE. 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 NLMS and the ZA-NLMS algorithms have converged. For analyzing the convergence of $E[a(n)]$, we next evaluate ΔJ_1 and ΔJ_2 for the proposed combination, which is, however, a very complicated task. To reduce the complexities, we make use of an elegant scheme of angular discretization of random vectors proposed in [9].

Angular Discretization of Random Vectors: Here, the given continuous valued random vector ($\mathbf{x}(n)$ in our case) is assumed to take one of the $2N$ directions given by $\pm\mathbf{v}_i,\ i = 0, 1, \ldots, N-1$, where $\mathbf{v}_i$ is the ith normalized eigenvector of the input autocorrelation matrix $\mathbf{R}$ corresponding to the eigenvalue λ_i. More specifically, we represent $\mathbf{x}(n)$ as

$$\mathbf{x}(n) = s_n\, r_n\, \boldsymbol{v}_n \quad (6)$$

where $s_n = \pm 1$ with $Pr[s_n = \pm 1] = (1/2)$, $r_n = \|\mathbf{x}(n)\|$, i.e., r_n has the same distribution as that of $\|\mathbf{x}(n)\|$, and $\boldsymbol{v}_n \in \{\mathbf{v}_i | i = 0, 1, \ldots, N-1\}$ with $Pr[\boldsymbol{v}_n = \mathbf{v}_i] = (\lambda_i/\text{Tr}(\mathbf{R}))$. The three variables s_n, r_n, and $\boldsymbol{v}_n$ are assumed to be statistically independent of each other. It is easy to check that the RHS of (6) conforms to the following properties of $\mathbf{x}(n)$: (i) it has mean zero, (ii) its norm is given by $||\mathbf{x}(n)||$, and (iii) its autocorrelation matrix is given by $\mathbf{R}$ [9].

From now on, in addition to the above model, we will also consider the input to be white with variance σ_x^2 for convenience. Recently, for white input, the above model has been used [10] to evaluate the mean square deviation (MSD), i.e., $E[||\mathbf{w}_0 - \mathbf{w}(n)||^2]$ of a zero-attracting, proportionate NLMS (ZA-PNLMS) algorithm in the steady state (where $\mathbf{w}(n)$ is the weight vector). In ZA-PNLMS, the zero-attracting term $-\rho\ \text{sgn}(\mathbf{w}(n))$ is added to the standard PNLMS update term $\mu\frac{\mathbf{G}(n)\mathbf{x}(n)e(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)+\delta}$, where $\mathbf{G}(n)$ is a diagonal matrix, with the ith diagonal entry $g_i(n),\ i = 0, 1, \ldots, N-1$, satisfying: 1) $g_i(n) > 0$ and 2) $\sum_{i=0}^{N-1} g_i(n) = 1$. It is easy to check that for $\mathbf{G}(n) = (1/N)\mathbf{I}$ and $\rho = 0$, the ZA-PNLMS algorithm reduces to the standard NLMS algorithm. Substituting these in the MSD expression for ZA-PNLMS [10] and recalling that for white input, the EMSE for an adaptive filter equals σ_x^2 times the MSD, we obtain the following expression for $J_{\text{ex},2}(\infty)$ (which also conforms to the one derived in [9]):

$$J_{\text{ex},2}(\infty) = \frac{\mu}{2-\mu}\sigma_\eta^2 \approx \frac{\mu}{2}\sigma_\eta^2 \quad (7)$$

(i.e., assuming $\mu << 2$ for small misadjustment). We next evaluate $J_{\text{ex},1}(\infty)$ and $J_{\text{ex},12}(\infty)$. For this, we assume a large filter order (i.e., $N >> 1$) which permits us to assume that the variance of $||\mathbf{x}(n)||^2$ is small, leading to

$$E\left[\frac{1}{r_n^2}\right] \approx \frac{1}{E\big[r_n^2\big]} \equiv \frac{1}{E\big[||\mathbf{x}(n)||^2\big]} = \frac{1}{N\sigma_x^2}. \quad (8)$$

Theorem 1: For white input and large filter order (i.e., $N >> 1$) as well as small misadjustment (i.e., $\mu << 2$), the steady state EMSE of the ZA-NLMS algorithm is given by

$$J_{\mathrm{ex},1}(\infty) = J_{\mathrm{ex},2}(\infty) + \hat{J}_M \tag{9}$$

where

$$\hat{J}_M = \sigma_x^2 \left[\frac{\rho^2 N^2 M}{\mu^2} \left\{ \frac{1}{\pi} - \frac{1}{\pi}\sqrt{\frac{\pi \mu^3 \sigma_\eta^2}{N^3 \rho^2 \sigma_x^2} + 1} - 1 \right\} + \frac{\rho^2 N^3}{\mu^2} \right]$$

and M : number of inactive taps.

Proof: The result follows by making the substitution $\mathbf{G}(n) = (1/N)\mathbf{I}$ in the EMSE expression of the ZA-PNLMS [10] [as the ZA-PNLMS algorithm reduces to ZA-NLMS under $\mathbf{G}(n) = (1/N)\mathbf{I}$], then by making use of (8) (which follows from the assumption $N >> 1$) and the assumption $\mu << 2$. ■

The following may be noted from Theorem 1:

1) The term $\hat{J}_M$ is a linear function of M with negative gradient as $\{1/\pi - (1/\pi)\sqrt{(\pi\mu^3\sigma_\eta^2)/(N^3\rho^2\sigma_x^2) + 1} - 1\} < 0$.
2) For $M = 0$, i.e., for a fully non-sparse system, $\hat{J}_M > 0$. However, as M increases, $\hat{J}_M$ decreases till for some $M = M'$, it becomes zero. For $M > M'$, $\hat{J}_M$ continues to decrease linearly and we have, $\hat{J}_M < 0$. The index M' can be treated as the boundary between a non-sparse system and a sparse system, and it can be set by choosing the value of ρ appropriately (by setting $\hat{J}_M = 0$, one can verify that the value of ρ resulting in $J_{\mathrm{ex},1}(\infty) = J_{\mathrm{ex},2}(\infty)$ is given by $\rho = \sqrt{(\pi\mu^3\sigma_\eta^2)/(\sigma_x^2 N^3\{(1-\pi+(\pi N/M'))^2 - 1\})}$.
3) From 1 and 2 above, it follows that for non-sparse systems (i.e., for less values of M, in particular for $0 < M < M'$), we have $J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},2}(\infty)$, meaning ZA-NLMS performs better than NLMS, whereas, for $M' < M < N$, $J_{\mathrm{ex},1}(\infty) < J_{\mathrm{ex},2}(\infty)$, meaning ZA-NLMS performs better than NLMS. This is expected, as in ZA-NLMS, zero attraction is enforced uniformly on all taps. Thus, when number of inactive taps M is high, ZA-NLMS is beneficial, whereas, when M is less, zero-attraction on active coefficients results in enhancement of EMSE.

Theorem 2: For white input, large filter order and $\mu << 2$, the steady state cross-EMSE $J_{\mathrm{ex},12}(\infty)$ between the NLMS and the ZA-NLMS can be expressed as

$$J_{\mathrm{ex},12}(\infty)1 = J_{\mathrm{ex},2}(\infty)\left[1 - M \Big/ \left\{ N\left[\frac{\sqrt{2\pi}\,\mu\sigma_{w_{1,i}}}{\rho N} + 1\right]\right\}\right] \tag{10}$$

where $\sigma_{w_{1,i}}^2 = \lim_{n\to\infty} E[\tilde{w}_{1,i}^2(n)]$, $i \in Z$ (same for all i [10]).

Proof: Given in the Appendix. ■

From (10), it follows that: 1) for $M = 0$, i.e., for a fully non-sparse system, $J_{\mathrm{ex},12}(\infty) = J_{\mathrm{ex},2}(\infty)$ and 2) $J_{\mathrm{ex},12}(\infty)$ has a negative gradient with respect to M, i.e., as M increases, it falls below $J_{\mathrm{ex},2}(\infty)$ linearly. We plot $J_{\mathrm{ex},1}(\infty)$, $J_{\mathrm{ex},2}(\infty)$ and $J_{\mathrm{ex},12}(\infty)$ against M over the range $0-N$ in Fig. 1. Since NLMS is sparsity unaware, $J_{\mathrm{ex},2}(\infty)$ is constant for all values of M. Also, the two straight lines, one for $J_{\mathrm{ex},1}(\infty)$ and the other for $J_{\mathrm{ex},12}(\infty)$ are seen to intersect at $M = M''$.

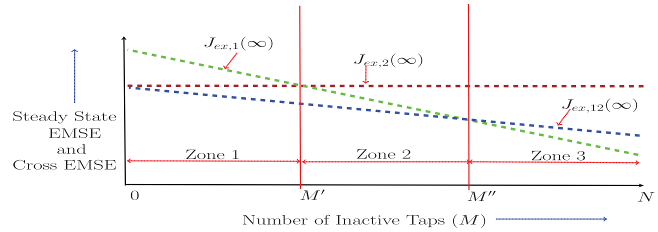

Fig. 1. Variation of $J_{\mathrm{ex},1}(\infty)$, $J_{\mathrm{ex},2}(\infty)$, and $J_{\mathrm{ex},12}(\infty)$ against the sparsity level of the system.

Fig. 1 shows the case where $M'' \le N$. In that case, the range $0-N$ can be divided into the following three zones:

1) Zone 1 ($0 \le M < M'$) for which $J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},2}(\infty) \ge J_{\mathrm{ex},12}(\infty)$.
2) Zone 2 ($M' \le M < M''$) for which $J_{\mathrm{ex},2}(\infty) \ge J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},12}(\infty)$.
3) Zone 3 ($M'' \le M \le N$) for which $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},12}(\infty) \ge J_{\mathrm{ex},1}(\infty)$.

However, if $M'' > N$, the two straight lines $J_{\mathrm{ex},1}(\infty)$ and $J_{\mathrm{ex},12}(\infty)$ do not intersect within $0-N$.[1] In that case, Fig. 1 will have only two zones, namely, zones 1 and 2.

Now, to understand the convergence behavior of the proposed convex combination, we use the following results from [7] on convex combination of any two adaptive filters:

1) If $\Delta J_2 > 0$ and $\Delta J_1 \le 0$, i.e., $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},12}(\infty) \ge J_{\mathrm{ex},1}(\infty)$, then, $\gamma(n) \to 1$ as $n \to \infty$, meaning, the proposed convex combination will converge to the filter 1 and we will have $J_{\mathrm{ex}}(\infty) \approx J_{\mathrm{ex},1}(\infty)$.
2) If $\Delta J_1 > 0$ and $\Delta J_2 \le 0$, i.e., $J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},12}(\infty) \ge J_{\mathrm{ex},2}(\infty)$, then, $\gamma(n) \to 0$ as $n \to \infty$, meaning, the proposed convex combination will converge to the filter 2, leading to $J_{\mathrm{ex}}(\infty) \approx J_{\mathrm{ex},2}(\infty)$.
3) If $\Delta J_2 > 0$ and $\Delta J_1 > 0$, i.e., $J_{\mathrm{ex},2}(\infty) > J_{\mathrm{ex},12}(\infty)$, $J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},12}(\infty)$, the convex combination may converge to a filter that produces even lesser EMSE than produced individually by filters 1 and 2, i.e., one will have $J_{\mathrm{ex}}(\infty) \le \min\{J_{\mathrm{ex},1}(\infty), J_{\mathrm{ex},2}(\infty)\}$ [note that as mentioned earlier, we cannot have the other case, namely, $J_{\mathrm{ex},2}(\infty) < J_{\mathrm{ex},12}(\infty)$, $J_{\mathrm{ex},1}(\infty) < J_{\mathrm{ex},12}(\infty)$ simultaneously].

From above and also from Fig. 1, we may then conclude the following:

1) For $M = 0$, i.e., for an absolutely non-sparse system, we have $\Delta J_1 > 0$ and $\Delta J_2 = 0$, meaning, $\lim_{n\to\infty} \gamma(n) = 0$ and thus the combined filter $\mathbf{w}_c(n)$, in the steady state, will converge to $\mathbf{w}_2(n)$, i.e., the NLMS-based filter.
2) For $0 < M < M''$, we have both $\Delta J_1 > 0$ and $\Delta J_2 > 0$, meaning, $\mathbf{w}_c(n)$ in this case will converge to a filter that produces $J_{\mathrm{ex}}(\infty)$ with $J_{\mathrm{ex}}(\infty) \le \min\{J_{\mathrm{ex},1}(\infty), J_{\mathrm{ex},2}(\infty)\}$.

[1]Solving $J_{\mathrm{ex},1}(\infty) = J_{\mathrm{ex},12}(\infty)$, one obtains M'' as N/c, where c is given approximately by $c = 1 - (1/\pi) + (1/\pi)\sqrt{\pi\mu^3\sigma_\eta^2/(\rho^2 N^3 \sigma_x^2)}(1 - \sqrt{\mu\sigma_\eta^2/(8N\sigma_x^2\sigma_{w_{1,i}}^2)})$, $i \in Z$ under the assumption that ρ is very small. Now, as $\sigma_{w_{1,i}}^2$, $i \in Z$ decreases with increasing ρ [10], it is easy to verify that c decreases as ρ increases. However, as long as c remains greater than one, we have $M'' < N$. On the other hand, if c exceeds one, we have $M'' > N$ and the two straight lines $J_{\mathrm{ex},1}(\infty)$ and $J_{\mathrm{ex},12}(\infty)$ do not intersect within $0-N$.

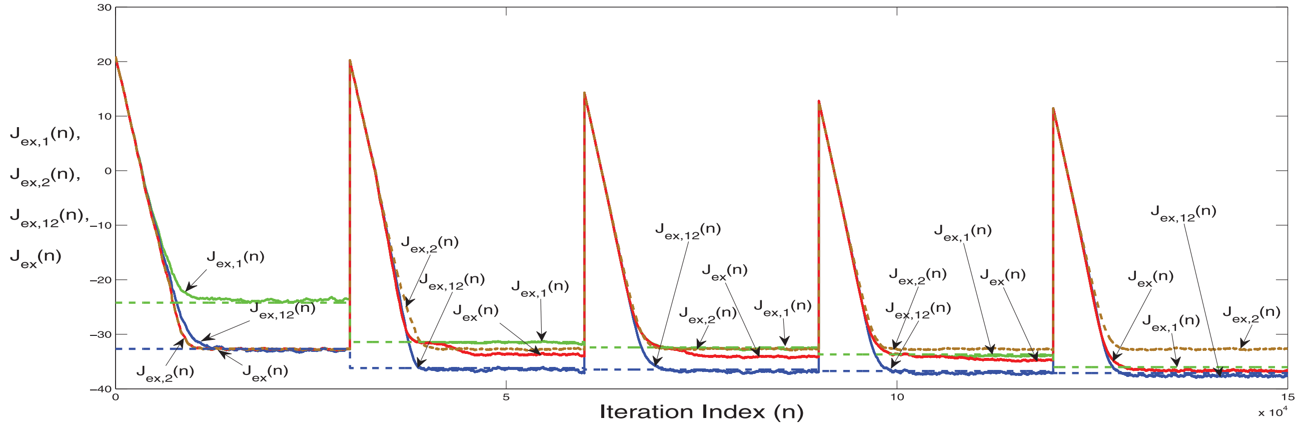

Fig. 2. EMSE and cross EMSE versus iteration index for ZA-NLMS, NLMS, and the overall combination.

3) For $M'' \leq M \leq N$, we have $\Delta J_2 > 0$ and $\Delta J_1 \leq 0$, implying, $\lim_{n\to\infty} \gamma(n) = 1$ and thus, $\mathbf{w}_c(n)$ will converge to $\mathbf{w}_1(n)$, i.e., the ZA-NLMS-based filter.

IV. Simulation Results

The proposed convex combination was simulated to identify a system with 128 coefficients that were chosen randomly from the set $\{0,\ 1\}$. Initially, the system was chosen to be highly non-sparse with the number of inactive taps (i.e., taps taking zero values) M taken to be only 4 (equivalently, number of active taps taking value 1 taken to be 124). The sparsity level of the system was gradually increased after every 30 000 time steps, leading to $M = 108,\ 113,\ 118,$ and 124 after 30 000, 60 000, 90 000 and 120 000 time steps, respectively. The simulation was carried out for a total of 150 000 iterations, with $\mu = 0.1$, $\mu_a = 15$, $\rho = 0.000004$, and with the input $x(n)$ and the observation noise $\eta(n)$ taken as zero mean, white Gaussian random processes with variances $\sigma_x^2 = 1$ and $\sigma_\eta^2 = 0.01$, respectively. Filters with 128 coefficients were chosen for both filters 1 and 2, with their initial values taken as zeros. The simulation results are displayed in Fig. 2 by plotting learning curve of the proposed convex combination, namely, the EMSE $J_{\mathrm{ex}}(n)$ against the iteration index n (red line), obtained by averaging $e_a^2(n)$ over 100 ensembles. Apart from $J_{\mathrm{ex}}(n)$, Fig. 2 also plots the individual learning curves of the two filters, namely, $J_{\mathrm{ex},1}(n)$ (green line) and $J_{\mathrm{ex},2}(n)$ (brown line), along with the cross EMSE $J_{\mathrm{ex},12}(n)$ (blue line) against n. The following can be observed from Fig. 2:

1) For $0 \leq n \leq 30\,000$, the system is almost fully non-sparse and it is seen that in the steady state, $J_{\mathrm{ex},2}(n) \approx J_{\mathrm{ex},12}(n)$ with both $J_{\mathrm{ex},2}(n)$ and $J_{\mathrm{ex},12}(n)$ less than $J_{\mathrm{ex},1}(n)$. Also seen is that $J_{\mathrm{ex}}(n) \approx J_{\mathrm{ex},2}(n)$, meaning, the convex combination switches to filter 2. This validates our conjecture above about fully non-sparse systems.
2) It is seen that for $30\,001 \leq n \leq 60\,000$, $J_{\mathrm{ex},2}(n) < J_{\mathrm{ex},1}(n)$, for $60\,001 \leq n \leq 90\,000$, $J_{\mathrm{ex},2}(n) \approx J_{\mathrm{ex},1}(n)$, while, for $90\,001 \leq n \leq 120\,000$, $J_{\mathrm{ex},2}(n) > J_{\mathrm{ex},1}(n)$ (all in the steady state). The first case corresponds to a moderately non-sparse system (zone 1 of Fig. 1), the second case corresponds to the transition point M' of Fig. 1, while the third case corresponds to a moderately sparse system (zone 2 of Fig. 1). In all the three cases, it is seen that $J_{\mathrm{ex},12}(n) < J_{\mathrm{ex},i}(n),\ i = 1, 2$ (in the steady state). More interestingly, in all the three cases above, we have $J_{\mathrm{ex}}(n) < J_{\mathrm{ex},i}(n),\ i = 1, 2$. This validates our hypothesis that in case of both zones 1 and 2 of Fig. 1, as $J_{\mathrm{ex},12}(\infty) < J_{\mathrm{ex},i}(\infty),\ i = 1, 2$, the proposed combination can converge to a filter that performs better than either of filters 1 and 2.
3) Lastly, for $120\,001 \leq n \leq 150\,000$, we have $J_{\mathrm{ex},2}(n) > J_{\mathrm{ex},12}(n) > J_{\mathrm{ex},1}(n)$ and $J_{\mathrm{ex}}(n) \approx J_{\mathrm{ex},1}(n)$ (in the steady state). This corresponds to a highly sparse system (zone 3 of Fig. 1). Again, this validates our claim that for a highly sparse system, the proposed combination will switch to filter 1.

We have also plotted in Fig. 2 the theoretical values of $J_{\mathrm{ex},1}(\infty)$ and $J_{\mathrm{ex},12}(\infty)$ as per (9) and (10) by horizontal green and blue dashed lines, respectively. It is easily seen that for each value of M, the theoretical values match the experimentally obtained values of $J_{\mathrm{ex},1}(\infty)$ and $J_{\mathrm{ex},12}(\infty)$ almost exactly.

Appendix

Evaluation of $J_{\mathrm{ex},12}(\infty)$

From Section III-A, $J_{\mathrm{ex},12}(n) = E[e_{a,1}(n)\ e_{a,2}(n)] = E[\mathbf{x}^T(n)\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)\mathbf{x}(n)] \equiv E\{\mathrm{Tr}[\mathbf{x}^T(n)\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)\ \mathbf{x}(n)]\} \equiv \mathrm{Tr}\{E[\mathbf{x}(n)\mathbf{x}^T(n)\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)]\}$. Using the statistical independence of $\mathbf{x}(n)$ vis-a-vis $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$ [5], and the fact that $x(n)$ is white with variance σ_x^2, we have, $J_{\mathrm{ex},12}(n) = \sigma_x^2 \mathrm{Tr}[\mathbf{K}_{12}(n)] = \sigma_x^2 \sum_{i=0}^{N-1} \tilde{\lambda}_{12,i}(n)$, where $\mathbf{K}_{12}(n) = E[\tilde{\mathbf{w}}_1(n)\tilde{\mathbf{w}}_2^T(n)]$ is the so-called weight error cross covariance matrix and $\tilde{\lambda}_{12,i}(n) = [\mathbf{K}_{12}(n)]_{i,i}$. To obtain $J_{\mathrm{ex},12}(\infty)$, we now consider the dynamics of the evolution of $\mathbf{K}_{12}(n)$ in time. Subtracting the LHS and RHS of (1) and (2) separately from $\mathbf{w}_0$, replacing $e_i(n)$ by $\mathbf{x}^T(n)\tilde{\mathbf{w}}_i(n) + \eta(n),\ i = 1, 2$ and neglecting δ as it is very small, we obtain

$$\tilde{\mathbf{w}}_1(n+1) = \left(\mathbf{I} - \frac{\mu \mathbf{x}(n)\mathbf{x}^T(n)}{\|\mathbf{x}(n)\|^2}\right) \times \tilde{\mathbf{w}}_1(n) - \frac{\mu \eta(n)\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2} + \rho \,\mathrm{sgn}[\mathbf{w}_1(n)] \quad (11)$$

and

$$\tilde{\mathbf{w}}_2(n+1) = \left(\mathbf{I} - \frac{\mu \mathbf{x}(n)\mathbf{x}^T(n)}{\|\mathbf{x}(n)\|^2}\right)\tilde{\mathbf{w}}_2(n) - \frac{\mu \eta(n)\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2}. \quad (12)$$

Post-multiplying the RHS of (11) by the transpose of the RHS of (12), using the aforesaid statistical independence of $\mathbf{x}(n)$

with $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$, and also the fact that $\eta(n)$ is zero mean, white and independent of $x(m)$ for all n, m, we obtain

$$\begin{aligned}\mathbf{K}_{12}(n+1) = E\bigg[&\left(\mathbf{I} - \frac{\mu \mathbf{x}(n)\mathbf{x}^T(n)}{\|\mathbf{x}(n)\|^2}\right)\mathbf{K}_{12}(n)\\ &\times \left(\mathbf{I} - \frac{\mu \mathbf{x}(n)\mathbf{x}^T(n)}{\|\mathbf{x}(n)\|^2}\right)\bigg]\\ &+ \rho E\big[\mathrm{sgn}[\mathbf{w}_1(n)]\tilde{\mathbf{w}}_2^T(n)\big]\\ &\times E\left[\left(\mathbf{I} - \frac{\mu \mathbf{x}(n)\mathbf{x}^T(n)}{\|\mathbf{x}(n)\|^2}\right)\right] + \mu^2 E\Big[\eta^2(n)\Big]\\ &\times E\left[\frac{\mathbf{x}(n)\mathbf{x}^T(n)}{||\mathbf{x}(n)||^4}\right]. \end{aligned} \quad (13)$$

Using (6) to replace $\mathbf{x}(n)$ by $s_n\, r_n\, \boldsymbol{v}_n$, we obtain

$$\begin{aligned}&\mathbf{K}_{12}(n+1)\\ &= \sum_{j=0}^{N-1} \frac{\lambda_j}{\mathrm{Tr}(\mathbf{R})}\Big(\mathbf{I} - \mu \mathbf{v}_j\mathbf{v}_j^T\Big)\mathbf{K}_{12}(n)\Big(\mathbf{I} - \mu \mathbf{v}_j\mathbf{v}_j^T\Big)\\ &+ \rho E\big[\mathrm{sgn}[\mathbf{w}_1(n)]\tilde{\mathbf{w}}_2^T(n)\big]\left(\mathbf{I} - \mu \sum_{j=0}^{N-1} \frac{\lambda_j}{\mathrm{Tr}(\mathbf{R})}\mathbf{v}_j\mathbf{v}_j^T\right)\\ &+ \mu^2 E\Big[\eta^2(n)\Big]E\left[\frac{1}{r_n^2}\right]\sum_{j=0}^{N-1} \frac{\lambda_j}{\mathrm{Tr}(\mathbf{R})}\mathbf{v}_j\mathbf{v}_j^T. \end{aligned} \quad (14)$$

Pre- and post-multiplying $\mathbf{K}_{12}(n)$ by $\mathbf{v}_i^T$ and $\mathbf{v}_i$, respectively, using the orthonormality of the eigenvector $\mathbf{v}_i$'s and noting that $\tilde{\lambda}_{12,i}(n) = \mathbf{v}_i^T\mathbf{K}_{12}(n)\mathbf{v}_i$, we have

$$\begin{aligned}\tilde{\lambda}_{12,i}(n+1) = &\left(1 - \mu(2-\mu)\frac{\lambda_i}{\mathrm{Tr}(\mathbf{R})}\right)\tilde{\lambda}_{12,i}(n)\\ &+ \frac{\mu^2\sigma_\eta^2}{\mathrm{Tr}(\mathbf{R})}\lambda_i E\left[\frac{1}{r_n^2}\right] + \Omega_{12,i}(n) \end{aligned} \quad (15)$$

where

$$\begin{aligned}\Omega_{12,i}(n) = &\rho \mathbf{v}_i^T E\big[\mathrm{sgn}[\mathbf{w}_1(n)]\tilde{\mathbf{w}}_2^T(n)\big]\\ &\times \left(\mathbf{I} - \mu \sum_{j=0}^{N-1} \frac{\lambda_j}{\mathrm{Tr}(\mathbf{R})}\mathbf{v}_j\mathbf{v}_j^T\right)\mathbf{v}_i. \end{aligned} \quad (16)$$

Now, for white $x(n)$, $\mathbf{R}$ is a diagonal matrix, meaning the eigenvector $\mathbf{v}_i$ is given by the ith column of the $N \times N$ identity matrix, $\lambda_j = \sigma_x^2$ and $\mathrm{Tr}(\mathbf{R}) = N\sigma_x^2$. From (16), it then follows that:

$$\Omega_{12,i}(n) = \rho(1 - \mu/N)E\big[\mathrm{sgn}\big[w_{1,i}(n)\big]\tilde{w}_{2,i}(n)\big]. \quad (17)$$

Now, from [10], it follows that both $w_{1,i}(n)$ and $w_{2,i}(n)$ converge to their true values in mean. This means, for large n, $E[w_{1,i}(n)] \approx 0$ for $i \in Z$, and $E[\tilde{w}_{2,i}(n)] \approx 0$ for both $i \in Z$ and $i \in NZ$, i.e., in steady state, both $w_{1,i}(n)$ and $\tilde{w}_{2,i}(n)$, $i \in Z$ are zero mean random variables. Assuming $w_{1,i}(n)$ and $w_{2,i}(n)$ to be jointly Gaussian [which is realistic as both are generated from $x(n)$ which is assumed to be Gaussian], we can apply *Price's theorem* [11] leading to $\lim_{n\to\infty} E[\mathrm{sgn}[w_{1,i}(n)]\tilde{w}_{2,i}(n)] = -\sqrt{2/(\pi\sigma_{w_{1,i}}^2})\tilde{\lambda}_{12,i}(n)$. For large n, this implies

$$\Omega_{12,i}(n) \approx -\rho(1 - \mu/N)\sqrt{2/\Big(\pi\sigma_{w_{1,i}}^2\Big)}\tilde{\lambda}_{12,i}(n). \quad (18)$$

Substituting in (15), using the approximation $E[(1/r_n^2)] \approx (1/N\sigma_x^2)$ from (8), and letting $n \to \infty$, we have, for $i \in Z$

$$\tilde{\lambda}_{12,i}(\infty) = \frac{\mu^2\sigma_\eta^2}{N^2\sigma_x^2}\Bigg/\left\{\frac{\mu}{N}(2-\mu) + \rho\left(1 - \frac{\mu}{N}\right)\sqrt{\frac{2}{\pi\sigma_{w_{1,i}}^2}}\right\}. \quad (19)$$

Since $\sigma_{w_{1,i}}^2$ is independent of i [10], $\tilde{\lambda}_{12,i}(\infty)$ too does not depend on i. Using the approximations $\mu << 2$ and $N >> 1$ in (19), we obtain, for $i \in Z$

$$\tilde{\lambda}_{12,i}(\infty) = \mu\sigma_\eta^2\Bigg/\left\{2N\sigma_x^2\left(1 + \frac{\rho N}{2\mu}\sqrt{\frac{2}{\pi\sigma_{w_{1,i}}^2}}\right)\right\}. \quad (20)$$

Next, for $i \in NZ$ and assuming small weight error variance in the steady state, it is reasonable to assume that for large n, $\mathrm{sgn}[w_{1,i}(n)] = \mathrm{sgn}[w_{0,i}]$. Since, in the steady state, $E[\tilde{w}_{2,i}(n)] = 0$, we have, $E[\mathrm{sgn}[w_{1,i}(n)]\tilde{w}_{2,i}(n)] = E[\mathrm{sgn}[w_{0,i}]\tilde{w}_{2,i}(n)] = \mathrm{sgn}[w_{0,i}]E[\tilde{w}_{2,i}(n)] = 0$. From (17), we then have $\Omega_{12,i}(n) = 0$ for large n and for $i \in NZ$. Substituting in (15), and as before, letting $n \to \infty$ and taking $E[(1/r_n^2)] \approx (1/N\sigma_x^2)$, we obtain, for $i \in NZ$

$$\tilde{\lambda}_{12,i}(\infty) = \mu\sigma_\eta^2/\Big\{(2-\mu)N\sigma_x^2\Big\} \approx \mu\sigma_\eta^2/\Big\{2N\sigma_x^2\Big\} \quad (21)$$

as $\mu << 2$. Note that like for $i \in Z$, $\tilde{\lambda}_{12,i}(\infty)$ is independent of i for $i \in NZ$ as well. Noting that $J_{\mathrm{ex},12}(\infty) = \sigma_x^2[\sum_{i\in Z}\tilde{\lambda}_{12,i}(\infty) + \sum_{i\in NZ}\tilde{\lambda}_{12,i}(\infty)]$ where $|Z| = M$, $|NZ| = N - M$, substituting $\tilde{\lambda}_{12,i}(\infty)$ from (20) and (21), and recalling from (7) that $J_{\mathrm{ex},2}(\infty) \approx \frac{\mu}{2}\sigma_\eta^2$, the result (10) follows easily.

References

[1] J. Arenas-García, L. A. Azpicueta-Ruiz, M. T. M. Silva, V. H. Nascimento, and A. H. Sayed, "Combinations of adaptive filters: Performance and convergence properties," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 120–140, Jan. 2016.

[2] F. Huang, J. Zhang, and S. Zhang, "Combined-step-size affine projection sign algorithm for robust adaptive filtering in impulsive interference environments," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 63, no. 5, pp. 493–497, May 2016.

[3] J. W. Yoo, J. W. Shin, and P. G. Park, "An improved NLMS algorithm in sparse systems against noisy input signals," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 62, no. 3, pp. 271–275, Mar. 2015.

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

[5] S. S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[6] B. K. Das and M. Chakraborty, "Sparse adaptive filtering by an adaptive convex combination of the LMS and the ZA-LMS algorithms," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 5, pp. 1499–1507, May 2014.

[7] 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.

[8] Y. Chen, Y. Gu, and A. O. Hero, "Regularized least-mean-square algorithms," *Arxiv preprint arXiv:1012.5066v2[stat.ME]*, Dec. 2010.

[9] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2811–2825, Sep. 1993.

[10] R. L. Das and M. Chakraborty, "Improving the performance of the PNLMS algorithm using l_1 norm regularization," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 7, pp. 1280–1290, Jul. 2016.

[11] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. Boston, MA, USA: McGraw-Hill, 2002.