Improving the Performance of the PNLMS Algorithm Using l_1 Norm Regularization

Rajib Lochan Das, *Member, IEEE*, and Mrityunjoy Chakraborty, *Senior Member, IEEE*

***Abstract*—The proportionate normalized least mean square (PNLMS) algorithm and its variants are by far the most popular adaptive filters that are used to identify sparse systems. The convergence speed of the PNLMS algorithm, though very high initially, however, slows down at a later stage, even becoming worse than sparsity agnostic adaptive filters like the NLMS. In this paper, we address this problem by introducing a carefully constructed l_1 norm (of the coefficients) penalty in the PNLMS cost function which favors sparsity. This results in certain zero attracting terms in the PNLMS weight update equation which help in the shrinkage of the coefficients, especially the inactive taps, thereby arresting the slowing down of convergence and also producing lesser steady state excess mean square error (EMSE). A rigorous convergence analysis of the proposed algorithm is presented that expresses the steady state mean square deviation of both the active and the inactive taps in terms of a zero attracting coefficient of the algorithm. The analysis reveals that further reduction of the EMSE is possible by deploying a variable step size (VSS) simultaneously with a variable zero attracting coefficient in the weight update process. Simulation results confirm superior performance of the proposed VSS zero attracting PNLMS algorithm over existing algorithms, especially in terms of having both higher convergence speed and lesser steady state EMSE simultaneously.**

***Index Terms*—Mean square deviation (MSD), PNLMS algorithm, sparse systems, variable step size, zero attractors.**

I. Introduction

IN real life, there exist many examples of systems that have a sparse impulse response, having a few significant non-zero elements (called active taps) amidst several zero or insignificant elements (called inactive taps). One example of such systems is the network echo paths [1], [2], which have a total echo response of about 64–128 ms duration out of which the "active" region spans a duration of only 8–12 ms, while the remaining "inactive" part accounts for bulk delay due to network loading, encoding and jitter buffer delays. Another example is the acoustic echo channel caused by coupling between microphone and loudspeaker in hands free mobile telephony [3]. Other well known examples of sparse systems include HDTV where clusters of dominant echoes arrive after long periods of silence [4], wireless multipath channels which, on most of the occasions, have only a few clusters of significant paths [5], and underwater acoustic channels where the various multipath components caused by reflections off the sea surface and sea bed have long intermediate delays [6]. The last decade witnessed a flurry of research activities [7] that sought to develop sparsity aware adaptive filters which can exploit the *a priori* knowledge of the sparseness of the system and thus enjoy improved identification performance. The first and foremost in this category is the proportionate normalized least mean square (PNLMS) algorithm [8] which achieves faster initial convergence by deploying different step sizes for different weights, with each one made proportional to the magnitude of the corresponding weight estimate. The convergence rate of the PNLMS algorithm, however, slows down at a later stage of the iteration and becomes even worse than a sparsity agnostic algorithm like the NLMS [9]. This problem was later addressed in several of its variants like the improved PNLMS (IPNLMS) algorithm [10], composite proportionate and normalized LMS algorithm [10] and mu law PNLMS algorithm [12]. These algorithms improve the transient response (i.e., convergence speed) of the PNLMS algorithm for identifying sparse systems. However, all of them yield almost same steady-state excess mean square error (EMSE) performance as produced by the PNLMS. The need to improve both transient and steady-state performance subsequently led to several variable step size (VSS), proportionate type algorithms [13]–[15].

In this paper, drawing ideas from [16], [17], we aim to improve the performance of the PNLMS algorithm further, by introducing a carefully constructed l_1 norm (of the coefficients) penalty in the PNLMS cost function which favors sparsity. This results in a modified PNLMS update equation with a zero attractor for all the taps, named as the zero attracting PNLMS (ZA-PNLMS) algorithm. The zero attractors help in the shrinkage of the coefficients which is particularly desirable for the inactive taps, thereby giving rise to lesser steady state EMSE for sparse systems. Further, by drawing the inactive taps towards zero, the zero attractors help in arresting the sluggishness of the convergence of the PNLMS algorithm that takes place at a later stage of the iteration, caused by the diminishing effective step sizes of the inactive taps. We show this by presenting a detailed convergence analysis of the proposed algorithm, which evaluates the steady state mean square deviations (MSDs) of the active as well as the inactive taps separately, as function of a zero attracting constant associated with the zero attractors. Such an analysis is, however, a very daunting task, especially due to the presence of a so-called gain matrix and also the zero attractors in the update equation. To overcome the challenges

Manuscript received July 15, 2015; revised January 18, 2016 and March 15, 2016; accepted April 5, 2016. Date of publication April 11, 2016; date of current version May 9, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Simon Doclo.

R. L. Das is with the Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, Gujarat 382007, India (e-mail: rajib.das.iit@gmail.com).

M. Chakraborty is with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, West Bengal 721302, India (e-mail: 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/TASLP.2016.2552578

posed by them, we deploy a transform domain equivalent model of the proposed algorithm and separately, an elegant scheme of angular discretization of continuous valued random vectors proposed earlier in [18]. The analysis reveals that further reduction in the EMSE can be achieved (without affecting the convergence rate) by deploying a VSS in the weight update process. However, due to the presence of the zero attractors, this requires the zero attracting coefficient associated with the zero attractors to be updated time recursively. Detailed simulation of all the proposed schemes are carried out and their performances are compared with that of the PNLMS and other existing algorithms. The simulation results not only confirm the various conjectures made in this paper about the proposed schemes, but also validate superior performance of the proposed VSS ZA-PNLMS algorithm over several existing schemes in terms of having both higher convergence speed and lesser steady state EMSE simultaneously.[1]

II. Proposed Algorithm

Consider a PNLMS based adaptive filter that takes $x(n)$ as input and updates a L tap coefficient vector $\mathbf{w}(n) = [w_0(n), w_1(n), \ldots, w_{L-1}(n)]^T$ (superscript T indicates transposition) as [8]

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

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

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

$$\mathbf{G}(n) = \mathrm{diag}(g_0(n), g_1(n) \ldots . g_{L-1}(n)), \tag{2}$$

where

$$g_l(n) = \frac{\Gamma_l(n)}{\sum_{i=0}^{L-1}\Gamma_i(n)}, \; 0 \le l \le L-1, \tag{3}$$

with

$$\Gamma_l(n) = \max[\rho_g \max[\delta, |w_0(n)|, \ldots, |w_{L-1}(n)|], |w_l(n)|]. \tag{4}$$

The parameter δ is an initialization parameter that helps to prevent stalling of the weight updating at the initial stage when all the taps are initialized to zero. Similarly, if an individual tap weight becomes very small, to avoid stalling of the corresponding weight update recursion, the respective $\Gamma_l(n)$ is taken as a small fraction (given by the constant ρ_g) of the largest tap magnitude. By providing separate effective step size $\mu\, g_l(n)$ to each lth tap where $g_l(n)$ is broadly proportional to $|w_l(n)|$, the PNLMS algorithm achieves higher rate of convergence initially, caused primarily by the active taps. At a later stage, however, the convergence slows down considerably, being controlled primarily by the numerically dominant inactive taps that have progressively diminishing effective step sizes [10], [12].

It has recently been shown [20] that the PNLMS weight update recursion (i.e., Eq. (1)) can be obtained by minimizing the cost function $||\mathbf{w}(n+1) - \mathbf{w}(n)||^2_{\mathbf{G}^{-1}(n)}$ subject to the condition $d(n) - \mathbf{w}^T(n+1)\mathbf{x}(n) = 0$ (the notation $||\mathbf{x}||^2_{\mathbf{A}}$ indicates the generalized inner product $\mathbf{x}^T\mathbf{A}\mathbf{x}$)). In order to derive the ZA-PNLMS algorithm, following [16], we add an l_1 norm penalty $\gamma\left|\left|\mathbf{G}^{-1}(n)\mathbf{w}(n+1)\right|\right|_1$ to the above cost function, where γ is a very small constant. Note that unlike [16], we have, however, used a generalized form of l_1 norm penalty here which scales the elements of $\mathbf{w}(n+1)$ by $\mathbf{G}^{-1}(n)$ first before taking the l_1 norm (the above scaling makes the l_1 norm penalty governed primarily by the inactive taps). The above constrained optimization problem may then be stated as

$$\min_{\mathbf{w}(n+1)} \parallel \mathbf{w}(n+1) - \mathbf{w}(n) \parallel^2_{\mathbf{G}^{-1}} + \gamma \parallel \mathbf{G}^{-1}\mathbf{w}(n+1) \parallel_1 \tag{5}$$

subject to $d(n) - \mathbf{w}^T(n+1)\mathbf{x}(n) = 0$, where the short form notation "$\mathbf{G}^{-1}$" is used to indicate $\mathbf{G}^{-1}(n)$. Using Lagrange multiplier λ, this amounts to minimizing the cost function $J(n+1) = \parallel \mathbf{w}(n+1) - \mathbf{w}(n) \parallel^2_{\mathbf{G}^{-1}} + \gamma \parallel \mathbf{G}^{-1}\mathbf{w}(n+1) \parallel_1 + \lambda(d(n) - \mathbf{w}^T(n+1)\mathbf{x}(n))$. Setting $\partial J(n+1)/\partial \mathbf{w}(n+1) = 0$, one obtains

$$\mathbf{w}(n+1) = \mathbf{w}(n) - [\gamma \mathrm{sgn}(\mathbf{w}(n+1)) - \lambda\mathbf{G}(n)\mathbf{x}(n)], \tag{6}$$

where sgn(.) is the well known signum function, i.e., $\mathrm{sgn}(x) = 1\ (x > 0),\ 0\ (x = 0),\ -1\ (x < 0)$. Premultiplying both the left-hand side (LHS) and the right-hand side (RHS) of (6) by $\mathbf{x}^T(n)$ and using the condition $d(n) - \mathbf{w}^T(n+1)\mathbf{x}(n) = 0$, one obtains

$$\lambda = \frac{e(n) + \gamma\mathbf{x}^T(n)\mathrm{sgn}(\mathbf{w}(n+1))}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)}. \tag{7}$$

Substituting (7) in (6), we have

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)} - \gamma\left[\mathbf{I} - \frac{\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{G}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)}\right]\mathrm{sgn}(\mathbf{w}(n+1)), \tag{8}$$

where $\mathbf{I}$ is the $L \times L$ identity matrix. Note that the above equation does not provide the desired weight update relation, as the RHS contains the unknown term $\mathrm{sgn}(\mathbf{w}(n+1))$. In order to obtain a feasible weight update equation, we approximate $\mathrm{sgn}(\mathbf{w}(n+1))$ by an estimate, namely, $\mathrm{sgn}(\mathbf{w}(n))$ which is known. This is based on the assumption that most of the weights do not undergo change of sign as they get updated. This assumption may not, however, appear to be a very accurate one, especially for the inactive taps that fluctuate around

[1]The paper expands an earlier conference presentation [19] by the authors by providing detailed first and second order convergence analysis and new results on VSS ZA-PNLMS.

zero value in the steady state. Nevertheless, an analysis of the proposed algorithm, as given later in this paper, shows that this approximation does not have any serious effect on the convergence behavior of the proposed algorithm. Apart from this, we also observe that in (8), any (i,j)th element of the matrix $\frac{\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{G}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)}$, $i,j=0,1,\ldots,L-1$, as given by $\frac{x(n-i)x(n-j)g_j(n)}{\sum_{r=0}^{L-1} g_r(n)x^2(n-r)}$, have magnitudes much less than 1, especially for large order filters, and thus, this matrix can be neglected in comparison to $\mathbf{I}$.

From above and introducing the algorithm step size μ and a regularization parameter δ_P in (8), for a large order adaptive filter, one then obtains the following weight update equation:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)+\delta_P} - \rho\text{sgn}(\mathbf{w}(n)), \tag{9}$$

where $\rho = \mu\gamma$.

Eq. (9) provides the weight update relation for the proposed ZA-PNLMS algorithm, where the second term on the RHS is the usual PNLMS update term while the last term, i.e., $\rho\text{sgn}(\mathbf{w}(n))$ is the so-called zero attractor. The zero attractor adds $-\rho\text{sgn}(w_j(n))$ to $w_j(n)$ and thus helps in its shrinkage to zero. Ideally, the zero attraction should be confined only to the inactive taps, which means that the proposed ZA-PNLMS algorithm will perform particularly well for systems that are highly sparse, but its performance may degrade as the number of active taps increases. In such cases, Eq. (9) may be further refined by applying the reweighting concept [16] to it. For this, we replace the l_1 regularization term $\parallel \mathbf{G}^{-1}\mathbf{w}(n+1) \parallel_1$ in (5) by a log-sum penalty $\sum_{i=1}^{L} \frac{1}{g_i(n)} log(1+ \mid w_i(n+1) \mid /\epsilon)$ where $g_i(n)$ is the ith diagonal element of $\mathbf{G}(n)$ and ϵ is a small constant. Following the same steps as used above to derive the ZA-PNLMS algorithm, one can then obtain the reweighted ZA-PNLMS (RZA-PNLMS) weight update equation as given by

$$\begin{aligned} w_i(n+1) = \ & w_i(n) + \frac{\mu g_i(n)x(n-i+1)e(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)+\delta_P} \\ & - \rho\frac{\text{sgn}(w_i(n))}{1+\varepsilon \mid w_i(n) \mid}, \quad i = 0, 1, \ldots, L-1, \end{aligned} \tag{10}$$

with $\varepsilon = 1/\epsilon$ and the zero attracting constant ρ given by $\rho = \mu\gamma\varepsilon$. The last term of (10), named as reweighted zero attractor, provides a selective shrinkage to the taps. Due to this reweighted zero attractor, the inactive taps with zero magnitudes or magnitudes comparable to $1/\varepsilon$ undergo higher shrinkage compared to the active taps which enhances the performance both in terms of convergence speed and steady state EMSE.

III. Convergence Analysis of the Proposed ZA-PNLMS Algorithm

A convergence analysis of the PNLMS algorithm is known to be a daunting task, due to the presence of $\mathbf{G}(n)$ both in the numerator and the denominator of the weight update term in (1), which again depends on $\mathbf{w}(n)$. The presence of the zero attractor term makes it further complicated for the proposed ZA-PNLMS algorithm, i.e., Eq. (9). To analyze the latter, we follow here an approach adopted recently in [21] in the context of PNLMS algorithm. This involves development of an equivalent transform domain model of the proposed algorithm first. A convergence analysis of the proposed algorithm is then carried out by applying to the equivalent model a scheme of angular discretization of continuous valued random vectors proposed first by Slock [18] and used later by several other researchers [22], [23].

A. A Transform Domain Model of the Proposed Algorithm

The proposed equivalent model uses a diagonal "transform" matrix $\mathbf{G}^{\frac{1}{2}}(n)$ with $[\mathbf{G}^{\frac{1}{2}}(n)]_{i,i} = g_i^{\frac{1}{2}}(n), i = 0,1,\ldots,L-1$, to transform the input vector $\mathbf{x}(n)$ and the filter coefficient vector $\mathbf{w}(n)$ to their "transformed" versions, given respectively as $\mathbf{s}(n) = \mathbf{G}^{\frac{1}{2}}(n)\mathbf{x}(n)$ and $\mathbf{w}_N(n) = [\mathbf{G}^{\frac{1}{2}}(n)]^{-1}\mathbf{w}(n)$. It is easy to check that $\mathbf{w}_N^T(n)\mathbf{s}(n) = \mathbf{w}^T(n)\mathbf{x}(n) \equiv \mathbf{y}(n)$ (say), i.e., the filter $\mathbf{w}_N(n)$ with input vector $\mathbf{s}(n)$ produces the same output $y(n)$ as produced by $\mathbf{w}(n)$ with input vector $\mathbf{x}(n)$. To compute $\mathbf{G}^{\frac{1}{2}}(n+1)$ and $\mathbf{w}_N(n+1)$, the filter $\mathbf{w}_N(n)$ is first updated to a weight vector $\mathbf{w}_N^{'}(n+1)$ as

$$\begin{aligned} \mathbf{w}_N^{'}(n+1) = \ & \mathbf{w}_N(n) + \frac{\mu e(n)\mathbf{s}(n)}{\mathbf{s}^T(n)\mathbf{s}(n)+\delta_P} \\ & - \rho\mathbf{G}^{-\frac{1}{2}}(n)\text{sgn}(\mathbf{w}_N(n)). \end{aligned} \tag{11}$$

From (9), it is easy to check that $\mathbf{w}(n+1)$ is given by $\mathbf{w}(n+1) = \mathbf{G}^{\frac{1}{2}}(n)\mathbf{w}_N^{'}(n+1)$. The matrix $\mathbf{G}(n+1)$ follows from $\mathbf{w}(n+1)$ following its definition and $\mathbf{w}_N(n+1)$ is then evaluated as $\mathbf{w}_N(n+1) = [\mathbf{G}^{\frac{1}{2}}(n+1)]^{-1}\mathbf{w}(n+1)$. From above, it follows that $\mathbf{w}_N(n+1) = \mathbf{G}^{-\frac{1}{2}}(n+1)\mathbf{w}(n+1) = \mathbf{G}^{-\frac{1}{2}}(n+1)\mathbf{G}^{\frac{1}{2}}(n)\mathbf{w}_N^{'}(n+1)$, meaning $[\mathbf{w}_N(n+1)]_i = [\frac{g_i(n)}{g_i(n+1)}]^{\frac{1}{2}}[\mathbf{w}_N^{'}(n+1)]_i$, $i = 0,1,\ldots,L-1$. Since $\sum_{i=0}^{L-1} g_i(n) = 1$ and $0 < g_i(n) < 1, i = 0,1,\ldots,L-1$, it is reasonable to expect that $g_i(n)$ does not change significantly from index n to index $(n+1)$, especially near convergence and for large L, and thus, we can make the approximation $[g_i(n)]^{\frac{1}{2}}[\mathbf{w}_N^{'}(n+1)]_i \approx [g_i(n+1)]^{\frac{1}{2}}[\mathbf{w}_N^{'}(n+1)]_i$, which implies $\mathbf{w}_N^{'}(n+1) = \mathbf{w}_N(n+1)$.

B. Angular Discretization of a Continuous Valued Random Vector [18]

As per this, given a zero mean, $L \times 1$ random vector $\mathbf{x}$ with correlation matrix $\mathbf{R} = E[\mathbf{x}\mathbf{x}^T]$ ("E": expectation operator), it is assumed that $\mathbf{x}$ can assume only one of the $2L$ orthogonal directions, given by $\pm\mathbf{e}_i, i = 0,1,\ldots,L-1$, where $\mathbf{e}_i$ is the ith normalized eigenvector of $\mathbf{R}$ corresponding to the eigenvalue λ_i. In particular, $\mathbf{x}$ is modelled as $\mathbf{x} = s\,r\,\mathbf{v}$, where $\mathbf{v} \in \{\mathbf{e}_i | i = 0,1,\ldots,L-1\}$, with probability of $\mathbf{v} = \mathbf{e}_i$ given by p_i, $r = ||\mathbf{x}||$, i.e., r has the same distribution as that of $||\mathbf{x}||$ and $s \in \{1, -1\}$, with probability of $s = \pm 1$ given by $P(s = \pm 1) = \frac{1}{2}$. Further, the three elements s, r and $\mathbf{v}$ are assumed to be mutually independent. Note that as s is zero mean, $E[s\,r\,\mathbf{v}] = \mathbf{0}$ and thus $E[\mathbf{x}] = \mathbf{0}$ is satisfied trivially. To satisfy $E[\mathbf{x}\mathbf{x}^T] = \mathbf{R}$, the discrete probability p_i is taken as $p_i =$

$\frac{\lambda_i}{\text{Tr}[\mathbf{R}]}$ (Tr[$\mathbf{R}$]: Trace of $\mathbf{R}$), which satisfies $p_i \geq 0$, $\sum_{i=0}^{L-1} p_i = 1$ and leads to $E[\mathbf{x}\mathbf{x}^T] = E(s^2\, r^2\, \mathbf{v}\mathbf{v}^T) = E(r^2)\, E(\mathbf{v}\mathbf{v}^T) = \text{Tr}\,[\mathbf{R}] \sum_{i=0}^{L-1} p_i \mathbf{e}_i \mathbf{e}_i^T = \sum_{i=0}^{L-1} \lambda_i \mathbf{e}_i \mathbf{e}_i^T = \mathbf{R}$. Also note that if θ_i be the angle between $\mathbf{x}$ and $\mathbf{e}_i$, then $\cos(\theta_i) = \frac{\mathbf{x}^T \mathbf{e}_i}{||\mathbf{x}||}$ and $E[\cos^2(\theta_i)] \approx \frac{\lambda_i}{\text{Tr}[\mathbf{R}]}$, meaning p_i provides a measure of how far $\mathbf{x}$ is (angularly) from $\mathbf{e}_i$ on an average.

In our analysis of the proposed algorithm, we use the above model to represent the transformed input vector $\mathbf{s}(n)$ as

$$\mathbf{s}(n) = s_s(n)\, r_s(n)\, \mathbf{v}_s(n), \tag{12}$$

where, $s_s(n) \in \{+1, -1\}$ with $P(s_s(n) = \pm 1) = \frac{1}{2}$, $r_s(n) = ||\mathbf{s}(n)||$ and $\mathbf{v}_s(n) \in \{\mathbf{e}_{s,i}(n) |\, i = 0, 1, \ldots, L-1\}$ with $P(\mathbf{v}_s(n) = \mathbf{e}_{s,i}(n)) = \frac{\lambda_{s,i}(n)}{\text{Tr}(\mathbf{S}(n))}$, where, $\mathbf{S}(n) = E[\mathbf{s}(n)\mathbf{s}^T(n)]$, $\lambda_{s,i}(n)$ is the ith eigenvalue of $\mathbf{S}(n)$, and as before, the three elements $s_s(n)$, $r_s(n)$ and $\mathbf{v}_s(n)$ are mutually independent.

C. Convergence Analysis of the Proposed Algorithm in Mean

First, define the weight error vector $\tilde{\mathbf{w}}(n) = \mathbf{w}_{\text{opt}} - \mathbf{w}(n)$ and the transform domain weight error vector $\tilde{\mathbf{w}}_N(n) = \mathbf{G}^{-\frac{1}{2}}(n)\tilde{\mathbf{w}}(n) \equiv \mathbf{G}^{-\frac{1}{2}}(n)\mathbf{w}_{\text{opt}} - \mathbf{w}_N(n)$. Then expressing $e(n)$ as $e(n) = \mathbf{s}^T(n)\tilde{\mathbf{w}}_N(n) + v(n)$, the recursive form of $\tilde{\mathbf{w}_N}(n)$ can be written as

$$\begin{aligned}\tilde{\mathbf{w}}_N(n+1) \approx\ & \tilde{\mathbf{w}}_N(n) - \frac{\mu \mathbf{s}(n)\mathbf{s}^T(n)\tilde{\mathbf{w}}_N(n)}{\mathbf{s}^T(n)\mathbf{s}(n) + \delta_P} \\ & - \frac{\mu \mathbf{s}(n) v(n)}{\mathbf{s}^T(n)\mathbf{s}(n) + \delta_P} + \rho \mathbf{G}^{-\frac{1}{2}}(n)\text{sgn}(\mathbf{w}_N(n)).\end{aligned} \tag{13}$$

For our analysis here, we approximate δ_P by zero in (13) as δ_P is a very small constant. Also, we define the mean weight vectors $\bar{\mathbf{w}}_N(n) = E[\mathbf{w}_N(n)]$ and $\bar{\mathbf{w}}(n) = E[\mathbf{w}(n)]$. The first order convergence of the proposed ZA-PNLMS algorithm is then provided in the following theorem.

Theorem 1: With a zero-mean input $x(n)$ of covariance matrix $\mathbf{R}$, the ZA-PNLMS algorithm produces stable $\bar{\mathbf{w}}_N(n)$ and also $\bar{\mathbf{w}}(n)$ if the step-size μ satisfies $0 < \mu < 2$ and under this condition, $\bar{\mathbf{w}}_N(n)$ and $\bar{\mathbf{w}}(n)$ converge respectively as per the following:

$$\begin{aligned}\bar{\mathbf{w}}_N(\infty) =\ & \lim_{n\to\infty} \bar{\mathbf{w}}_N(n) = E(\mathbf{G}^{-\frac{1}{2}}(n))\big|_{\infty} \mathbf{w}_{\text{opt}} \\ & - \frac{\rho}{\mu}\text{Tr}(\mathbf{S}(\infty))\mathbf{S}^{-1}(\infty) \\ & \times \lim_{n\to\infty} E(\mathbf{G}^{-\frac{1}{2}}(n)\text{sgn}(\mathbf{w}_N(n))),\end{aligned} \tag{14}$$

and

$$\begin{aligned}\bar{\mathbf{w}}(\infty) =\ & \lim_{n\to\infty} \bar{\mathbf{w}}(n) = \mathbf{w}_{\text{opt}} - \frac{\rho}{\mu}\text{Tr}(\mathbf{S}(\infty)) \\ & \times E(\mathbf{G}^{\frac{1}{2}}(n))\big|_{\infty} \mathbf{S}^{-1}(\infty) \\ & \times \lim_{n\to\infty} E(\mathbf{G}^{-\frac{1}{2}}(n)\text{sgn}(\mathbf{w}(n)),\end{aligned} \tag{15}$$

where $\mathbf{S}(n) = E(\mathbf{s}(n)\mathbf{s}^T(n)) = E(\mathbf{G}^{\frac{1}{2}}(n)\mathbf{R}\mathbf{G}^{\frac{1}{2}}(n))$.

Proof: The proof has been uploaded in [24] and is not included here for brevity. ■

Corollary 1: For white input, $\bar{w}_i(\infty)(=\lim_{n\to\infty}E(w_i(n)))$ for the ith active tap (*i.e.*, for which $w_{\text{opt},i} \neq 0$) is approximately given by

$$\bar{w}_i(\infty) = w_{\text{opt},i} - \frac{\rho}{\mu}\bar{g}_i^{-1}(\infty)\text{sgn}(w_{\text{opt},i}), \tag{16}$$

where $\bar{g}_i(\infty) = \lim_{n\to\infty}\bar{g}_i(n)$ and $\bar{g}_i(n) = [E(\mathbf{G}(n))]_{i,i}$.

Proof: For white input with variance σ_x^2, we have $\mathbf{R} = \sigma_x^2\mathbf{I}$, $\mathbf{S}(n) = \sigma_x^2 E(\mathbf{G}(n))$, $\text{Tr}(\mathbf{S}(n)) = \sigma_x^2$ and $\mathbf{S}^{-1}(n) = \frac{1}{\sigma_x^2}E(\mathbf{G}(n))^{-1}$ and then, we can have a simplified expression of $\bar{\mathbf{w}}(\infty)$ as

$$\bar{\mathbf{w}}(\infty) \approx \mathbf{w}_{\text{opt}} - \frac{\rho}{\mu}\lim_{n\to\infty} E(\mathbf{G}(n))^{-1}E(\text{sgn}(\mathbf{w}(n))), \tag{17}$$

where we have assumed that in the steady state as $n \to \infty$, $\mathbf{G}^{-\frac{1}{2}}(n)$ and $\text{sgn}(\mathbf{w}(n))$ become statistically independent and $E(\mathbf{G}(n)^{-\frac{1}{2}}) \approx E(\mathbf{G}^{\frac{1}{2}}(n))^{-1}$, which is reasonable as in the steady state, variance of each individual $g_i(n), i = 0, 1, \ldots, L-1$ is quite small (i.e., it behaves almost like a constant). Now, for an active tap with significantly large magnitude $w_{\text{opt},i}$, it is reasonable to approximate $\text{sgn}(w_i(n)) \approx \text{sgn}(w_{\text{opt},i})$ under the assumption that in the steady state, the variance of $w_i(n)$, i.e., $E((w_i(n) - w_{\text{opt},i})^2)$ is small enough compared to the magnitude of $w_{\text{opt},i}$. Then, with $E(\text{sgn}(w_i(n))) \approx E(\text{sgn}(w_{\text{opt},i})) = \text{sgn}(w_{\text{opt},i})$ for an active tap in the steady state, the result follows trivially from (17). ■

Corollary 1 shows that

$$\bar{w}_i(\infty) = \begin{cases} w_{\text{opt},i} - \dfrac{\rho}{\mu}\bar{g}_i^{-1}(\infty), & \text{if sgn}(w_{\text{opt},i}) > 0, \\ w_{\text{opt},i} + \dfrac{\rho}{\mu}\bar{g}_i^{-1}(\infty), & \text{if sgn}(w_{\text{opt},i}) < 0, \end{cases}$$

which implies that $\bar{w}_i(\infty)$ is always closer to the origin vis-a-vis $w_{\text{opt},i}$. Further, the bias (i.e., usually defined as $w_{\text{opt},i} - \bar{w}_i(\infty)$) is also proportional to $\bar{g_i}^{-1}(\infty)$, meaning active taps with comparatively smaller values will have larger bias and vice versa.

In the case of inactive taps, we have $w_{\text{opt},i} = 0$. From (15) and for $\rho = 0$ (i.e., no zero attraction), this implies $\bar{w}_i(\infty) = 0$, i.e., the tap estimates fluctuate around zero value. For $\rho > 0$, the zero attractors come into play in the update equation (9) and act as an additional force that tries to pull the coefficients to zero from either side. The effect of zero attractor is thus to confine the fluctuations in a small band around zero. On an average, one can then take $E(\text{sgn}(w_i(n)))\big|_{\infty} \approx 0$, meaning, from (17), the inactive tap estimates will largely be free of any bias.

D. Second Order Convergence Analysis of the Proposed Algorithm

In this section, we restrict the input to be white with variance σ_x^2. The EMSE in this case is given by $\sigma_x^2\eta(n)$ [9], where $\eta(n) = \sum_{i=0}^{L-1}\eta_i(n)$ is the overall MSD and $\eta_i(n) = E[(w_i(n) - w_{\text{opt},i})^2], i = 0,\ 1, \ldots, L-1$ is the MSD for the ith tap. To evaluate the EMSE, we carry out an analysis of both $\eta_i(n)$ and $\eta(n)$, and in particular, investigate the contribution of the active and the inactive taps in the overall steady state MSD $\eta(\infty)$. The analysis uses the notations NZ and Z to indicate the index sets of active and inactive taps, respectively, i.e., $w_{\text{opt},i} \neq 0$ for $i \in NZ$ and $w_{\text{opt},i} = 0$ for $i \in Z$. It is then possible to prove the following:

Theorem 2: With a zero-mean, white input $x(n)$ of variance $E(x^2(n)) = \sigma_x^2$, the ZA-PNLMS algorithm produces stable MSD performance if the step-size μ satisfies $0 < \mu < 2$ and the steady-state $MSD\ \eta_i(\infty)$ of active ($i \in NZ$) and inactive ($i \in Z$) taps are given by

1) for active taps (*i.e.*, $i \in NZ$):

$$\eta_i(\infty) = \bar{g}_i(\infty)\frac{\mu}{(2-\mu)}\frac{\sigma_v^2}{\hat{r}_s^2(\infty)} + \frac{\rho^2}{\mu^2}\left(\frac{2-\mu\bar{g}_i(\infty)}{2-\mu}\right)\frac{1}{\bar{g}_i^2(\infty)}, \quad (18)$$

where $\hat{r}_s^2(\infty) = \lim_{n\to\infty}\hat{r}_s^2(n)$, $\hat{r}_s^2(n) = [E(\frac{1}{r_s^2(n)})]^{-1}$ and $r_s^2(n) = \mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)$, and

2) for inactive taps (*i.e.*, $i \in Z$):

$$\eta_i(\infty) = \bar{g}_i(\infty)\, t^2, \quad (19)$$

where t is the positive root of the quadratic function

$$at^2 + bt - c = 0, \quad (20)$$

and

$$a = \mu(2-\mu)\rho_g \bar{g}_{\max}(\infty), \quad (21)$$

$$b = 2\sqrt{\frac{2}{\pi}}\rho\left(1 - \mu\rho_g\bar{g}_{\max}(\infty)\right)\rho_g^{-\frac{1}{2}}\bar{g}_{\max}^{-\frac{1}{2}}(\infty), \quad (22)$$

$$c = \mu^2\sigma_v^2\frac{\rho_g\bar{g}_{\max}(\infty)}{\sigma_x^2} + \rho^2\rho_g^{-1}\bar{g}_{\max}^{-1}(\infty), \quad (23)$$

where $\bar{g}_i(\infty)$ is same as defined in Corollary 1, and $\bar{g}_{\max}(n) = \max\{\bar{g}_i(n)|i = 0, 1, 2, \ldots, L-1\}$ and $\bar{g}_{\max}(\infty) = \lim_{n\to\infty}\bar{g}_{\max}(n)$.

Proof: Given in Appendix A. ■

Theorem 2 reveals several properties of the ZA-PNLMS algorithm as given in the following corollaries.

Corollary 2: For $\rho = 0$, *i.e.*, when the ZA-PNLMS algorithm corresponds to the original PNLMS algorithm and for a large order filter, the steady state MSD for both the active and inactive taps is given by

$$\eta_i(\infty) = \bar{g}_i(\infty)\frac{\mu\sigma_v^2}{(2-\mu)\sigma_x^2}, \quad (24)$$

and the total steady state MSD is given by

$$\eta(\infty) = \frac{\mu\sigma_v^2}{(2-\mu)\sigma_x^2} \quad (25)$$

[which also conform to the results obtained in [21]].

Proof: First note that for long filters, the variance of $r_s^2(n) = ||\mathbf{s}(n)||^2$ is usually small and one can write $[\hat{r}_s^2(n)]^{-1} = E(\frac{1}{r_s^2(n)}) \approx \frac{1}{E[r_s^2(n)]}$. Also, $E[r_s^2(n)] = E[||\mathbf{s}(n)||^2] = E[\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)] = \sigma_x^2$, since, from the usual "independence assumption" [9], $\mathbf{G}(n)$ and $\mathbf{x}(n)$ are statistically independent, and $\sum_{i=0}^{L-1}\bar{g}_i(\infty) = 1$. Substituting $\hat{r}_s^2(\infty)$ on the RHS of (18) by σ_x^2 and noting that for $\rho = 0$, the second term on the RHS of (18) is zero, it is easily seen that $\eta_i(\infty)$ for $i \in NZ$ satisfies (24). Again, for $i \in Z$, b in (22) and the second term of c in (23) are zero for $\rho = 0$ and thus, from (19) and (20), $\eta_i(\infty) = \frac{\bar{g}_i(\infty)c}{a} = \bar{g}_i(\infty)\frac{\mu}{(2-\mu)}\frac{\sigma_v^2}{\sigma_x^2}$. Clearly, (24) is satisfied by $\eta_i(\infty)$ for $i \in Z$, as well which proves (24). Further, since $\sum_{i=0}^{L-1}\bar{g}_i(\infty) = 1$, (25) follows trivially from (24). Hence proved. ■

Corollary 3: As ρ increases from zero, the positive square root t of (20) decreases.

Proof: For positive ρ and under the stability condition: $0 < \mu < 2$, all the three variables, a, b and c are positive as can be easily seen from (21)–(23). The positive square root of (20) is then given by $t = \frac{-b+\sqrt{b^2+4ac}}{2a}$. Defining $b' = b/\rho$ where b is given by (22) and neglecting the second term in (23) as compared to the first term as it is proportional to ρ^2 (where ρ is typically in the range of 10^{-4} in contrast to ρ_g which is of the order of 10^{-2}), one can write

$$\frac{dt}{d\rho} = -\frac{b'}{2a} - \frac{b.b'}{2a\sqrt{b^2+4ac}}.$$

Since $a > 0,\ b > 0,\ c > 0$ and $b' > 0$, we have $\frac{dt}{d\rho} < 0$. Hence proved. ■

Total MSD in Presence of Zero Attraction : Note that the overall steady state MSD can be written as $\eta(\infty) = \sum_{i\in NZ}\eta_i(\infty) + \sum_{i\in Z}\eta_i(\infty)$. For $\rho = 0$, i.e., for ordinary PNLMS, as Corollary 2 shows, $\eta_i(\infty)$ is same for both active and inactive taps, and is also proportional to $\bar{g}_i(\infty)$. Now, for a sparse system, $\bar{g}_i(\infty)$ for $i \in Z$ is very small (close to zero) while for $i \in NZ$, it is much higher (closer to one). Even then, as the number of inactive taps is very large, one may still find $\sum_{i\in Z}\eta_i(\infty)$ to be non-negligible. However, as ρ increases from zero, from Corollary 3 and Eq. (19), $\eta_i(\infty)$ decreases for $i \in Z$ and thus, $\sum_{i\in Z}\eta_i(\infty)$ may soon become negligible as compared to $\sum_{i\in NZ}\eta_i(\infty)$, which is, however, expected as the zero attractors try to pull the inactive taps towards their true value (i.e., zero) and thus help in reducing the MSD. The second term on the RHS of (18), on the other hand, increases with increasing ρ which, in fact, corresponds to error introduced by zero attraction on active taps. This term, being proportional to ρ^2, is, however, negligible as compared to the first term that is independent of ρ. Combining the two cases, for a sparse system, we then have

$$\eta(\infty) = \frac{\mu}{2-\mu}\left(\frac{\sigma_v^2}{\hat{r}_s^2(\infty)}\right)\sum_{i\in NZ}\bar{g}_i(\infty). \quad (26)$$

Note that the total MSD as per (26) takes the minimum value when the ZA-PNLMS algorithm assumes the special form of ZA-NLMS, for which, $\mathbf{G}(n) = \frac{1}{L}\mathbf{I}$ (also, $\rho_g = 1$). In a general case, however, the mean proportionate factors, $\bar{g}_i(\infty),\ i \in NZ$ takes values larger than $\frac{1}{L}$ and thus, the total MSD $\eta(\infty)$ assumes values larger than in the case of the ZA-NLMS algorithm (the higher initial convergence rate of the former in comparison to the latter, however, remains unchanged).

Given (26), one can try to reduce the total MSD $\eta(\infty)$ further, without affecting the convergence rate, by adopting a suitable VSS mechanism in the coefficient update process. However, due to the presence of the zero attractors, such VSS approach would require appropriate adjustment of ρ, as shown in the following section.

IV. A VSS ZA-PNLMS Algorithm

Before proposing the VSS ZA-PNLMS algorithm, we take a relook into the first term on the RHS of (18), which plays the dominant role in the overall steady state MSD, given approximately by (26). This term is proportional to $\frac{\mu}{2-\mu}$ which in turn is a monotonically increasing function of μ in the stable range of μ (i.e., $0<\mu<2$). In the VSS approach, we make μ directly related to $|e(n)|$ or $e^2(n)$. This ensures that as the algorithm converges and $|e(n)|$ decreases, μ and therefore $\frac{\mu}{2-\mu}$ reduces, thereby resulting in a lesser value of $\eta(\infty)$. In the proposed scheme, we follow the VSS approach of [25] which has the advantage of not being overly parametric. As per this, $\mu(n)$ is updated as

$$\mu(n)=\begin{cases}\mu^{'}(n), & \text{if } \mu^{'}(n)>0\\ 0, & \text{otherwise,}\end{cases} \tag{27}$$

where $\mu^{'}(n)=1-\frac{\sigma_v}{\hat{\sigma}_e(n)+\tilde{\epsilon}}$, and $\hat{\sigma}_e^2(n)$ is an estimate of $E(e^2(n))$ that is recursively evaluated as $\hat{\sigma}_e^2(n)=\beta\hat{\sigma}_e^2(n-1)+(1-\beta)e^2(n)$. The parameter β $(0<\beta<1,\ \beta\approx 1)$ is an exponential weighting factor, $\tilde{\epsilon}$ is a small positive constant deployed to avoid division by zero and the noise variance σ_v^2, if not known *a priori*, can be estimated as per [26].

Note that as the value of μ decreases with n, even though the first term in (18) decreases as desired, the second term, being inversely proportional to μ^2 starts increasing, meaning the contribution to the overall MSD from zero attraction on the active taps tends to become appreciable. To counter this, we also make ρ variable and update it so that as μ reduces, it also reduces in an appropriate manner. The proposed update relation for $\rho(n)$ is as follows:

$$\rho(n)=\rho_0\mu(n), \tag{28}$$

where ρ_0 is a positive constant used to set the initial value (i.e., $\rho_0\mu(0)$) of the zero attractors. Lastly, sometimes one comes across sparse systems in practice that have a non-trivial number of active taps. In such cases, the overall contribution to the MSD $\eta(\infty)$ by zero attraction on active taps, i.e., $\sum_{i\in NZ}\frac{\rho^2}{\mu^2}(\frac{2-\mu\bar{g}_i(\infty)}{2-\mu})\frac{1}{\bar{g}_i^2(\infty)}$ may not be negligible. For such cases, we restrict the zero attraction mostly to the inactive taps by using a reweighted form of zero attractors [16]. As per this, similar to RZA-PNLMS algorithm, the ith zero attractor of the VSS ZA-PNLMS update, namely, $\rho(n)\mathrm{sgn}(w_i(n))$ is replaced by $\rho(n)\frac{\mathrm{sgn}(w_i(n))}{1+\varepsilon|w_i(n)|}$, where ε is a positive constant that helps in confining the zero attraction largely to small magnitude taps with $|w_i(n)|$ in the order of $\frac{1}{\varepsilon}$. With this modification, the proposed VSS ZA-PNLMS algorithms changes to VSS RZA-PNLMS, with the update equation of the ith tap given by

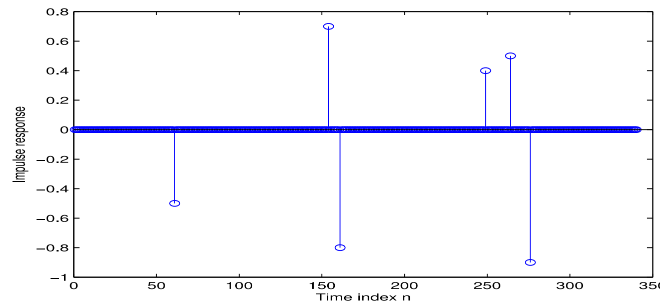

Fig. 1. Impulse response of the highly sparse system used in the first experiment.

$$\begin{aligned}w_i(n+1)=&\ w_i(n)+\frac{\mu(n)g_i(n)x(n-i+1)e(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)+\delta_P}\\ &-\rho(n)\frac{\mathrm{sgn}(w_i(n))}{1+\varepsilon\mid w_i(n)\mid},\\ i=&\ 0,1,\ldots,L-1,\end{aligned} \tag{29}$$

with $\mu(n)$ and $\rho(n)$ updated as discussed above. The VSS RZA-PNLMS algorithm requires about $6L+2$ additions, $6L+5$ multiplications and $L+3$ divisions. For the ZA-PNLMS, these figures are, respectively, $5L-1$, $5L+1$ and 2, while, for the original PNLMS, the same are $4L-1$, $5L+1$ and 2, respectively.

V. Simulation Studies

In this section, we evaluate the performance of the proposed ZA-PNLMS, RZA-PNLMS and VSS RZA-PNLMS algorithms via simulation studies in the context of sparse system identification. For simulations, in the first experiment, we consider a highly sparse system with impulse response of length $L=340$ as shown in Fig. 1 that has only six active taps. The system is driven by zero mean white input $x(n)$ of variance $\sigma_x^2=1$ and has output observation noise $v(n)$ which is taken to be zero mean, white Gaussian with $\sigma_v^2=10^{-3}$. The objective here is to study the convergence performance of the proposed algorithms for a sparse system with white input, and also to validate some of the results from the convergence analysis. For all the algorithms, the step size μ and the regularization parameter (to avoid division by zero) are taken to be 0.5 and 0.01, respectively, while ρ_g and δ are chosen as 0.01 and 0.001, respectively. For the RZA-PNLMS algorithm, the reweighting constant ε is taken as 100. The zero attracting coefficient ρ for the ZA-PNLMS and the RZA-PNLMS are tuned to 0.00004 and 0.00009, respectively for obtaining the optimum steady state MSD. The simulations are carried out for a total of 15 000 iterations. Simulation results are demonstrated by displaying the learning curves for the MSDs for the above algorithms, obtained by averaging $||\tilde{\mathbf{w}}(n)||^2$ over 100 experiments and then plotting the average against the iteration index n. We also evaluate contributions of the active and inactive taps in the total MSDs against time, by averaging $\sum_{i\in NZ}(w_i(n)-w_{\mathrm{opt},i})^2$ and $\sum_{i\in Z}(w_i(n)-w_{\mathrm{opt},i})^2$,

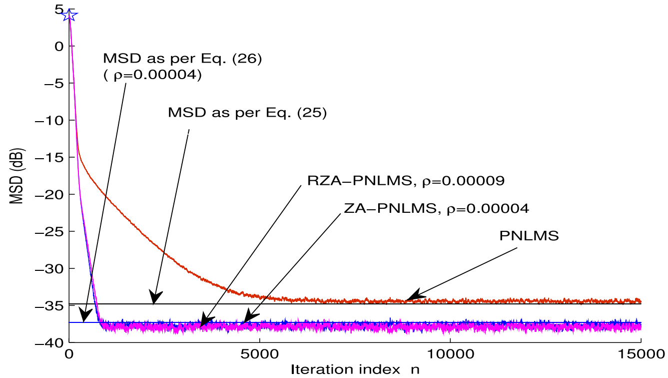

Fig. 2. Learning curves of the PNLMS, the ZA-PNLMS and the RZA-PNLMS for the system shown in Fig. 1. Also shown are the steady state MSD of the PNLMS and the ZA-PNLMS algorithms as per (25) and (26), respectively.

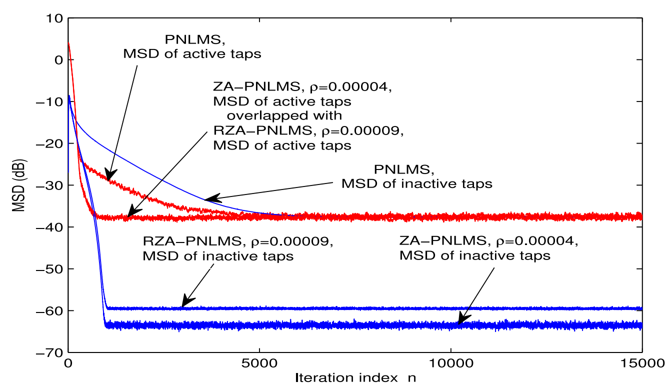

Fig. 3. Contributions of the active and the inactive taps to the total MSD of the PNLMS, the ZA-PNLMS and the RZA-PNLMS algorithms for the system shown in Fig. 1.

respectively over 100 experiments and then plotting the averages against the iteration index n for the same algorithms.

Fig. 2 displays the learning curves of the ZA-PNLMS and the RZA-PNLMS algorithms vis-a-vis the PNLMS for the system of Fig. 1. It shows that even though all the three algorithms start with identical, fast convergence rate, the convergence of the PNLMS algorithm slows down after about 300 iterations or so, whereas both the ZA-PNLMS and the RZA-PNLMS algorithms maintain their initial fast rate of convergence and also achieve lower steady state MSD. This is investigated further in Fig. 3 by identifying the contributions to the MSD from the active and the inactive taps separately in case of all the three algorithm. Fig. 3 reveals that in case of the PNLMS algorithm, the convergence of the inactive taps slows down as the effective step sizes for them become less and less progressively, and on the other hand, the active taps tend to converge very fast, in about 300 iterations and then, slow down even with the higher effective step size, which is probably due to the noisy estimate of the inactive taps. In case of the RZA-PNLMS and the ZA-PNLMS algorithms, however, the inactive taps come under the influence of an additional force, exerted by the zero attractors, which try to pull them towards their true value, i.e., zero. As a result, for both the RZA-PNLMS and the ZA-PNLMS algorithms, the inactive taps converge quickly to -60 dB or less in about 800 iterations or so (the slight difference between the steady state MSD levels of the RZA-PNLMS and the ZA-PNLMS algorithms is due to difference in the value of ρ chosen). It is also interesting to note that in this case, the active taps also converge to their steady state level in about 800 iterations, i.e., the convergence rate of the active taps also improves which is possibly due to the better estimate obtained in the inactive taps. Thereby, the total MSD for both the algorithms converges in about 800 iterations as can be seen from Fig. 2. Further, the figure also shows that both the above mentioned zero attracting algorithms attain almost 3 dB improvements in the steady state MSD vis-a-vis the PNLMS algorithm and that their learning curves are also identical. This incidentally shows that the reweighting process involved in the RZA-PNLMS algorithm has negligible effect vis-a-vis the ZA-PNLMS for a highly sparse system. Lastly, Fig. 2 is seen to validate the expressions for the steady state MSD obtained for the PNLMS and the ZA-PNLMS algorithms in (25) and (26), respectively.

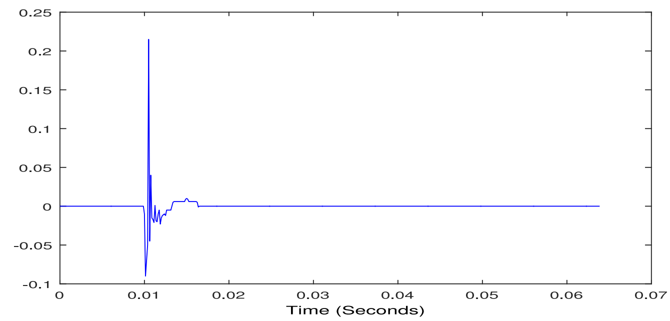

Fig. 4. Sparse network echo path impulse response (ITU-T G.168) used in the second experiment.

Next, to study the performance of the VSS RZA-PNLMS, we consider the sparse network echo path as per ITU-T G.168 that has an impulse response of duration 0.064 s (or, equivalently, length 512 at 8 KHz sampling rate), with a total of 51 active taps as shown in Fig. 4. The network echo path was identified by the proposed VSS RZA-PNLMS algorithm , taking the system input $x(n)$ to be a zero mean, unit variance white sequence and using the following choice of algorithm parameters: $\varepsilon = 100$, $\delta_P = 0.01$, $\rho_g = 0.01$, $\delta = 0.01$, $\rho_0 = 0.000002$ and $\sigma_v^2 = 0.001$. For comparative assessment of its performance, same identification exercise was also carried out by some of the well known or closely related sparse adaptive filters, like, the conventional PNLMS, VSS-PNLMS, VSS-IPNLMS (obtained by applying the VSS mechanism (27) to PNLMS and IPNLMS, respectively, with $\beta = 0.997$) and also the proposed RZA-PNLMS (with $\rho = 0.000002$). For fastest convergence, μ was set at one for both PNLMS and RZA-PNLMS. The performance of each algorithm was assessed by evaluating the normalized misalignment $\chi(n) = 10\log_{10}\frac{||\mathbf{w}(n)-\mathbf{w}_{\rm opt}||^2}{||\mathbf{w}_{\rm opt}||^2}$ and taking its ensemble average over 20 experiments. The corresponding plots vis-a-vis the iteration index n, i.e., the learning curves, are shown in Fig. 5. The figure again confirms that for sparse systems, the RZA-PNLMS algorithm enjoys both faster convergence rate as well as lesser steady state MSD as compared to the PNLMS algorithm due to zero attraction on a large number of (inactive) taps. Also, the MSD reduces considerably in comparison to PNLMS (by about 10 dB) in case of the VSS-PNLMS and the VSS-IPNLMS algorithms, though the convergence rate remains comparable to that of PNLMS. The proposed VSS RZA-

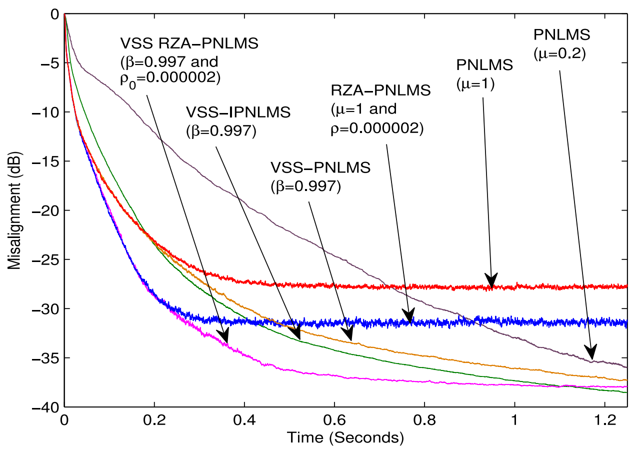

Fig. 5. Learning curve of the proposed VSS RZA-PNLMS algorithm vis-a-vis those of a few other recently proposed algorithms for the system shown in Fig. 4 with white input.

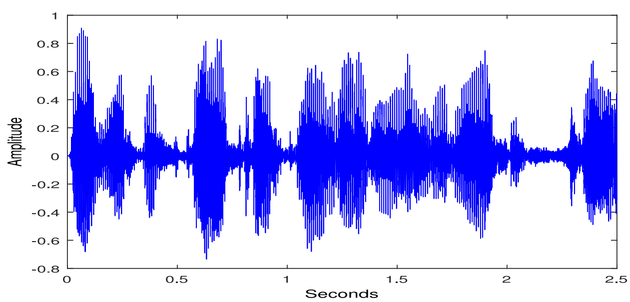

Fig. 6. Speech signal used to identify and track the system impulse response that undergoes sudden change at the 20th second in the third experiment. The actual signal used is obtained by repeating the above waveform 16 times.

PNLMS algorithm, on the other hand, enjoys *both the faster initial convergence rate of the RZA-PNLMS as well as the very low steady state MSD of the VSS-PNLMS and the VSS-IPNLMS simultaneously*, as clearly shown by Fig. 5. Lastly, for comparing convergence rates under same steady state MSD conditions, the PNLMS algorithm was also simulated with $\mu = 0.2$ that produces the same MSD as that of the VSS RZA-PNLMS. The rate of convergence of the PNLMS algorithm in this case is seen to become much slower than the VSS RZA-PNLMS as shown by Fig. 5.

Next we consider the case of colored input including speech signals. Also, we try to evaluate the tracking performance of the proposed algorithms by considering sudden changes in the system impulse response. For these, we start with the system of Fig. 4. In particular, we consider two scenarios:

A) In this, the system is fed with a colored input $x(n)$ generated by driving an AR(1) model with zero mean, unit variance white noise $u(n)$, with the model given by, $x(n) = rx(n-1) + \sqrt{1-r^2}u(n)$ $(r = 0.7)$. The system impulse response is given a sudden change at 0.625 s, with the new impulse response obtained by applying a right circular shift to the impulse response given in Fig. 4 by 30 samples (i.e., the length of the impulse response remains same with the cluster of active samples moving right by 30 samples). The system is identified and tracked by four algorithms, namely, PNLMS, VSS-PNLMS, the proposed RZA-PNLMS as well as the VSS

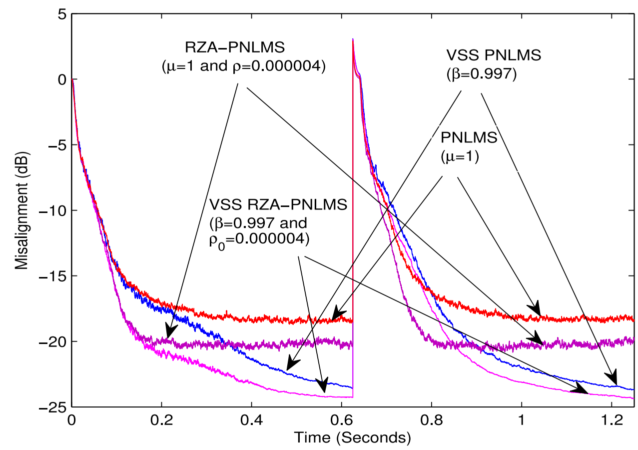

Fig. 7. Learning curve of the proposed VSS RZA-PNLMS algorithm vis-a-vis those of PNLMS, VSS-PNLMS and RZA-PNLMS (proposed) for colored input with the system undergoing sudden change at 0.625 s.

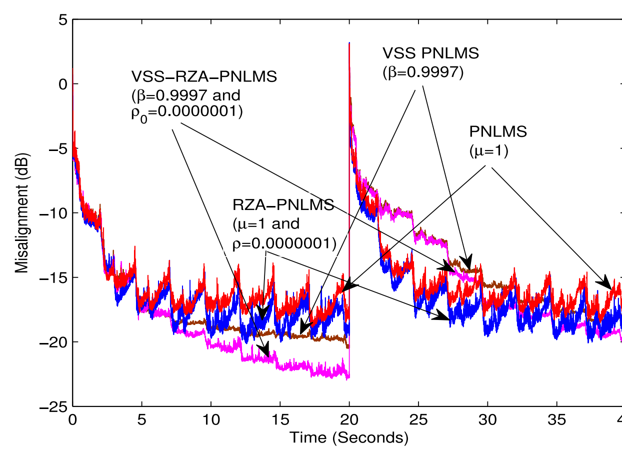

Fig. 8. Learning curve of the proposed VSS RZA-PNLMS algorithm vis-a-vis those of PNLMS, VSS PNLMS and RZA-PNLMS (proposed) for speech input with the system undergoing sudden change at the 20th second.

RZA-PNLMS, with the following set of parameters: $\sigma_v^2 = 10^{-3}$ ($\mathrm{SNR} = 30$ dB), $\rho = 4 \times 10^{-6}$, $\rho_0 = 4 \times 10^{-6}$ and $\beta = 0.997$ (other parameters remaining same as used in the context of Fig. 5).

B) Here, the system is driven by a test speech signal obtained by repeating the speech waveform of Fig. 6 16 times, with the system impulse response undergoing same sudden changes as above, at 20th second . The same four algorithms as used in (A) are used with identical set of parameters except for the following: SNR $=$ 34.68 dB, $\sigma_v^2 = 10^{-5}$, $\rho = 1 \times 10^{-6}$, $\rho_0 = 1 \times 10^{-6}$ and $\beta = 0.9997$. The simulation results for (A) and (B) are demonstrated by plotting the learning curves in Figs. 7 and 8, respectively. From both Figs. 7 and 8, it is easily seen that as before, the proposed VSS RZA-PNLMS algorithm attains steady state MSD lesser than that reached by RZA-PNLMS, PNLMS and VSS-PNLMS even when the system impulse response undergoes abrupt changes. The VSS RZA-PNLMS algorithm, like the VSS-PNLMS, is, however, somewhat slower in responding to the sudden change in the system impulse response as compared to the other two. This may be due to the dependence of $\hat{\sigma}_e(n)$ in $\mu^{'}(n)$ (i.e., (27)) on the system's past through past samples of $e(n)$.

VI. Conclusion

The PNLMS algorithm is one of the most popular algorithms for sparse adaptive filtering and sparse system identification. In this paper, an attempt has been made to overcome some of the shortcomings associated with the convergence of the PNLMS algorithm. For this, first a carefully constructed l_1 norm (of the coefficients) penalty is introduced in the PNLMS cost function which favors sparsity. This results in certain zero attractor terms in the PNLMS weight update equation which help in the shrinkage of the coefficients, especially the inactive taps, thereby arresting the slowing down of convergence and also producing lesser steady state EMSE. The paper carries out a detailed first and second order convergence analysis of the proposed zero-attracting PNLMS (ZA-PNLMS) algorithm. The analysis reveals that one can achieve further reduction in the EMSE (without affecting the convergence rate) by deploying a VSS in the weight update process. However, due to the presence of the zero attractors, this requires the zero attracting coefficient associated with the zero attractors also to be updated time recursively. Simulation results confirm superior performance of the proposed algorithms vis-a-vis the existing methods. In particular, it shows that the proposed VSS ZA-PNLMS algorithm can combine the very fast convergence rate of certain fast converging versions of the PNLMS algorithm with the very low steady state EMSE of some other categories of the PNLMS family.

Appendix
Proof of Theorem 2

Let $\mathbf{K}_s(n) = E(\tilde{\mathbf{w}}_N(n)\tilde{\mathbf{w}}_N^T(n))$ be the weight error correlation matrix for the equivalent model and let its ith diagonal entry, $i = 0, 1, \ldots, L-1$, be denoted as $\tilde{\lambda}_{s,i}(n)$, i.e., $\tilde{\lambda}_{s,i}(n) = [\mathbf{K}_s(n)]_{i,i} \equiv E[\tilde{w}_{N,i}^2(n)]$, where $\tilde{w}_{N,i}(n) = [\tilde{\mathbf{w}}_N(n)]_i$. Now, given $\tilde{\mathbf{w}}(n) = \mathbf{G}^{\frac{1}{2}}(n)\tilde{\mathbf{w}}_N(n)$, we have, $\eta_i(n) = E[\tilde{w}_i^2(n)] = E[g_i(n)\tilde{w}_{N,i}^2(n)]$. Assuming as before that $g_i(n)$ changes only marginally with time in the steady state and thus becomes largely uncorrelated with $\tilde{w}_{N,i}^2(n)$ as $n \to \infty$, we can then write, from above, $\eta_i(\infty) = \bar{g}_i(\infty)\tilde{\lambda}_{s,i}(\infty)$. To evaluate $\tilde{\lambda}_{s,i}(\infty)$, we consider the dynamics of $\mathbf{K}_s(n)$. From (13) (where we neglect the regularization parameter δ) and also the aforesaid "independence assumption" on $\mathbf{w}_N(n)$ vis-a-vis $\mathbf{s}(n)$, the recursive form of $\mathbf{K}_s(n)$ can be written as

$$\begin{aligned}
&\mathbf{K}_s(n+1) = E(\tilde{\mathbf{w}}_N(n+1)\tilde{\mathbf{w}}_N^T(n+1)) \\
&= E\left(\left(\mathbf{I} - \frac{\mu\mathbf{s}(n)\mathbf{s}^T(n)}{\mathbf{s}^T(n)\mathbf{s}(n)}\right)\mathbf{K}_s(n)\left(\mathbf{I} - \frac{\mu\mathbf{s}(n)\mathbf{s}^T(n)}{\mathbf{s}^T(n)\mathbf{s}(n)}\right)\right) \\
&\quad + \mu^2 E\left(\frac{\mathbf{s}(n)\mathbf{s}^T(n)}{\|\mathbf{s}(n)\|^4}\right)E(v^2(n)) \\
&\quad + \rho^2 E\left(\mathbf{G}^{-\frac{1}{2}}(n)\mathrm{sgn}(\mathbf{w}_N(n))\mathrm{sgn}^T(\mathbf{w}_N(n))\mathbf{G}^{-\frac{1}{2}}(n)\right) \\
&\quad + \rho E\left(\left(\mathbf{I} - \frac{\mu\mathbf{s}(n)\mathbf{s}^T(n)}{\mathbf{s}^T(n)\mathbf{s}(n)}\right)\tilde{\mathbf{w}}_N(n)\mathrm{sgn}^T(\mathbf{w}_N(n))\mathbf{G}^{-\frac{1}{2}}(n)\right) \\
&\quad + \rho E\left(\mathbf{G}^{-\frac{1}{2}}(n)\mathrm{sgn}(\mathbf{w}_N(n))\tilde{\mathbf{w}}_N^T(n)\left(\mathbf{I} - \frac{\mu\mathbf{s}(n)\mathbf{s}^T(n)}{\mathbf{s}^T(n)\mathbf{s}(n)}\right)\right)
\end{aligned} \tag{A.1}$$

where we have ignored the cross terms that become zero as $v(n)$ is zero mean and statistically independent of $\mathbf{s}(n)$. Next note that $x(n)$ is white, i.e., $E[\mathbf{x}(n)\mathbf{x}^T(n)] = \sigma_x^2\mathbf{I}$ and thus, $\mathbf{S}(n) = E[\mathbf{s}(n)\mathbf{s}^T(n)] = \sigma_x^2 E[\mathbf{G}(n)]$: a diagonal matrix, implying that the eigenvalues of $\mathbf{S}(n)$ are given by $\lambda_{s,i}(n) = [\mathbf{S}(n)]_{i,i,} = \sigma_x^2\bar{g}_i(n), i = 0, 1, \ldots, L-1$, and the corresponding eigenvectors are given by the standard basis of $\mathbb{R}^L$, i.e., $\{\mathbf{e}_i | i = 0, 1, \ldots, L-1\}$, where $\mathbf{e}_i = [0, \ldots, 0, 1, 0, \ldots, 0]^T$ with the "1" occurring at the ith position from top (note that the eigenvectors in this case are time invariant, i.e., same for all n). Also note that $\mathrm{Tr}[\mathbf{S}(n)] = \sigma_x^2$.

To simplify the evaluation of (A.1), we now invoke the angular discretization model discussed in Section III-B and replace $\mathbf{s}(n)$ by $s_s(n)\, r_s(n)\, \mathbf{v}_s(n)$ as given by (12), where $\mathbf{v}_s(n) \in \{\mathbf{e}_i | \, i = 0, 1, \ldots, L-1\}$ with $P(\mathbf{v}_s(n) = \mathbf{e}_i) = \frac{\lambda_{s,i}(n)}{\mathrm{Tr}(\mathbf{S}(n))}$. It is also interesting to observe that $\tilde{\lambda}_{s,i}(n) = [\mathbf{K}_s(n)]_{i,i}$ can be written as $\mathbf{e}_i^T\mathbf{K}_s(n)\mathbf{e}_i$. Premultiplying and postmultiplying both the LHS and the RHS of (A.1) by $\mathbf{e}_i^T$ and $\mathbf{e}_i$, respectively, we can then express $\tilde{\lambda}_{s,i}(n)$ as

$$\begin{aligned}\tilde{\lambda}_{s,i}(n+1) = \; & l_1(n) + l_2(n) \\ & + l_3(n) + l_4(n) + l_5(n),\end{aligned} \tag{A.2}$$

where $l_1(n), l_2(n), l_3(n), l_4(n)$ and $l_5(n)$ are given and also analyzed below

$$\begin{aligned}(i)\; l_1(n) = \; & \mathbf{e}_i^T E\left(\left(\mathbf{I} - \mu\frac{\mathbf{s}(n)\mathbf{s}^T(n)}{\mathbf{s}^T(n)\mathbf{s}(n)}\right)\mathbf{K}(n)\right. \\ & \left.\times \left(\mathbf{I} - \mu\frac{\mathbf{s}(n)\mathbf{s}^T(n)}{\mathbf{s}^T(n)\mathbf{s}(n)}\right)\right)\mathbf{e}_i.\end{aligned}$$

Replacing $\mathbf{s}(n)$ by $s_s(n)\, r_s(n)\, \mathbf{v}_s(n)$ and using the angular discretization model, we can write

$$\begin{aligned}l_1(n) = \; & \sum_{j=1}^{L}\frac{\lambda_{s,j}(n)}{\mathrm{Tr}(\mathbf{S}(n))}(\mathbf{e}_i^T(\mathbf{I} - \mu\mathbf{e}_j\mathbf{e}_j^T) \\ & \times \mathbf{K}_s(n)(\mathbf{I} - \mu\mathbf{e}_j\mathbf{e}_j^T)\mathbf{e}_i).\end{aligned}$$

Using the orthonormality of the eigenvectors of $\mathbf{S}(n)$ (i.e., $\mathbf{e}_i^T\mathbf{e}_j = \delta(i-j)$) and the fact that $\frac{\lambda_{s,j}(n)}{\mathrm{Tr}[\mathbf{S}(n)]} = \bar{g}_j(n)$, we can simplify the above as $l_1(n) = \bar{g}_i(n)(1-\mu)^2\tilde{\lambda}_{s,i}(n) + \tilde{\lambda}_{s,i}(n)\sum_{j\neq i}\bar{g}_j(n)$. Noting that $\sum_{j\neq i}\bar{g}_j(n) = 1 - \bar{g}_i(n)$, we then obtain

$$l_1(n) = \; [1 - \mu(2-\mu)\bar{g}_i(n)]\,\tilde{\lambda}_{s,i}(n).$$

$$(ii)\; l_2(n) = \; \mu^2 E(v^2(n))\mathbf{e}_i^T E\left(\frac{\mathbf{s}(n)\mathbf{s}^T(n)}{\|\mathbf{s}(n)\|^4}\right)\mathbf{e}_i. \tag{A.3}$$

As before, we substitute $\mathbf{s}(n)$ by $s_s(n)r_s(n)\mathbf{v}_s(n)$. Since $s_s^2(n) = 1$ and $r_s(n)$ is statistically independent with $\mathbf{v}_s(n)$, one can then write, $E(\frac{\mathbf{s}(n)\mathbf{s}^T(n)}{\|\mathbf{s}(n)\|^4}) = E(\frac{1}{r_s^2(n)})E[\mathbf{v}_s(n)\mathbf{v}_s^T(n)]$. Writing $E[\mathbf{v}_s(n)\mathbf{v}_s^T(n)]$ as $\sum_{i=0}^{L-1}\frac{\lambda_{s,i}(n)}{\mathrm{Tr}(\mathbf{S}(n))}\mathbf{e}_i\mathbf{e}_i^T$ and noting that $\mathbf{S}(n) = \sum_{i=0}^{L-1}\lambda_{s,i}(n)\mathbf{e}_i\mathbf{e}_i^T$, this can be written as $E(\frac{\mathbf{s}(n)\mathbf{s}^T(n)}{\|\mathbf{s}(n)\|^4}) = E(\frac{1}{r_s^2(n)})\frac{\mathbf{S}(n)}{\mathrm{Tr}(\mathbf{S}(n))}$. Making substitutions in the expression for $l_2(n)$ above, noting that $\mathbf{e}_i^T\mathbf{S}(n)\mathbf{e}_i = \lambda_{s,i}(n)$

and recalling that $E(v^2(n)) = \sigma_v^2$, $\hat{r}_s^2(n) = [E(\frac{1}{r_s^2(n)})]^{-1}$ and $\frac{\lambda_{s,i}(n)}{\mathrm{Tr}[\mathbf{S}(n)]} = \bar{g}_i(n)$, we have

$$l_2(n) = \mu^2 \sigma_v^2 \left(\frac{1}{\hat{r}_s^2(n)}\right) \bar{g}_i(n). \tag{A.4}$$

(iii) $l_3(n) = \rho^2 \mathbf{e}_i^T \mathbf{C}(n)\mathbf{e}_i$, where $\mathbf{C}(n) = E(\mathbf{G}^{-\frac{1}{2}}(n)$ $\mathrm{sgn}(\mathbf{w}_N(n))\mathrm{sgn}^T(\mathbf{w}_N(n))\mathbf{G}^{-\frac{1}{2}}(n))$. Clearly,

$$\begin{aligned} l_3(n) =&\ \rho^2 [\mathbf{C}(n)]_{i,i} \\ =&\ \rho^2 [E(\mathbf{G}^{-1}(n))]_{i,i} \approx \rho^2 \bar{g}_i^{-1}(n). \end{aligned} \tag{A.5}$$

(iv) $l_4(n) = \rho \mathbf{e}_i^T \mathbf{F}(n)\mathbf{e}_i$ where $\mathbf{F}(n) = E((\mathbf{I} - \frac{\mu \mathbf{s}(n)\mathbf{s}^T(n)}{\mathbf{s}^T(n)\mathbf{s}(n)})\tilde{\mathbf{w}}_N(n)\mathrm{sgn}^T(\mathbf{w}_N(n))\mathbf{G}^{-\frac{1}{2}}(n))$. Clearly,

$$\begin{aligned} l_4(n) =&\ \rho \sum_{j=0}^{L-1} \frac{\lambda_{s,j}(n)}{\mathrm{Tr}(\mathbf{S}(n))} \Big(\mathbf{e}_i^T \left(\mathbf{I} - \mu \mathbf{e}_j \mathbf{e}_j^T\right) E(\tilde{\mathbf{w}_\mathbf{N}}(n) \\ &\times \mathrm{sgn}^T(\mathbf{w}_N(n)))E(\mathbf{G}^{-\frac{1}{2}}(n))\mathbf{e}_i\Big) \\ =&\ \rho\left(\bar{g}_i(n)(1-\mu) + 1 - \bar{g}_i(n)\right) \qquad (A.6) \\ &\times E([\tilde{\mathbf{w}_\mathbf{N}}(n)]_i \mathrm{sgn}([\mathbf{w}(n)]_i))[E(\mathbf{G}^{-\frac{1}{2}}(n))]_{i,i} \\ \approx&\ \rho\left(1 - \mu \bar{g}_i(n)\right) \bar{g}_i^{-\frac{1}{2}}(n) E([\tilde{\mathbf{w}_\mathbf{N}}(n)]_i \mathrm{sgn}([\mathbf{w}(n)]_i)), \end{aligned} \tag{A.7}$$

where, as before, we have assumed that the diagonal elements of $\mathbf{G}(n)$ have marginal variance, especially in the steady state and thus, the statistical independence between $\mathbf{w}(n)$ and $\mathbf{x}(n)$ can be extended to $\mathbf{w}_N(n)$ and $\mathbf{s}(n)$ as well.

(v) $l_5(n) = \rho \mathbf{e}_i^T \mathbf{P}(n)\mathbf{e}_i$, where $\mathbf{P}(n) = \mathbf{F}^T(n)$. Clearly, $l_5(n) = l_4(n)$.

Substituting (A.3)–(A.7) and replacing $l_5(n)$ by $l_4(n)$ in (A.2), we obtain

$$\begin{aligned} \tilde{\lambda}_{s,i}(n+1) &= [1 - \mu(2-\mu)\bar{g}_i(n)]\, \tilde{\lambda}_{s,i}(n) \\ &+ \mu^2 \sigma_v^2 \left(\frac{1}{\hat{r}_s^2(n)}\right) \bar{g}_i(n) + \rho^2 \bar{g}_i^{-1}(n) \\ &+ 2\rho\left(1 - \mu \bar{g}_i(n)\right) \bar{g}_i^{-\frac{1}{2}}(n) E([\tilde{\mathbf{w}_\mathbf{N}}(n)]_i \mathrm{sgn}([\mathbf{w}(n)]_i)). \end{aligned} \tag{A.8}$$

Note that the recursion of $\tilde{\lambda}_{s,i}(n)$ depends on E $([\tilde{\mathbf{w}_\mathbf{N}}(n)]_i \mathrm{sgn}([\mathbf{w}(n)]_i))$, which we evaluate separately for the active and the inactive taps as given below.

(a) *Active Taps* ($i \in NZ$): We assume that the active taps are having significantly large magnitudes and thus, we may approximate $\mathrm{sgn}(w_i(n)) = \mathrm{sgn}(w_{\mathrm{opt},i})\ \ \forall i \in NZ$ [especially near convergence]. Therefore, $E([\tilde{\mathbf{w}_\mathbf{N}}(n)]_i \mathrm{sgn}([\mathbf{w}(n)]_i)) \approx \mathrm{sgn}(w_{\mathrm{opt},i})E([\tilde{\mathbf{w}_\mathbf{N}}(n)]_i)$. From Theorem 1, for large n, $E([\tilde{\mathbf{w}_\mathbf{N}}(n)]_i) = \frac{\rho \mathrm{Tr}(\mathbf{S}(n))}{\mu}(\mathbf{S}^{-1}(n)E(\mathbf{G}^{-\frac{1}{2}}(n))\mathrm{sgn}(\mathbf{w}(n)))_i$, where we have used the fact that $\mathrm{sgn}(\mathbf{w}(n)) = \mathrm{sgn}(\mathbf{G}^{\frac{1}{2}}(n)$ $\mathbf{w}_N(n)) = \mathrm{sgn}(\mathbf{w}_N(n))$. Making the substitutions $\mathrm{Tr}(\mathbf{S}(n)) = \sigma_x^2$, $[\mathbf{S}^{-1}(n)]_{i,i} \approx \frac{\bar{g}_i^{-1}(n)}{\sigma_x^2}$, $[E(\mathbf{G}^{-\frac{1}{2}}(n))]_{i,i} \approx \bar{g}_i^{-\frac{1}{2}}(n)$ and using the fact that $\mathrm{sgn}(w_{\mathrm{opt},i})\mathrm{sgn}([\mathbf{w}(n)]_i) = [\mathrm{sgn}(w_{\mathrm{opt},i})]^2 = 1$, we observe that as $n \to \infty$,

$$E([\tilde{\mathbf{w}_\mathbf{N}}(n)]_i \mathrm{sgn}([\mathbf{w}(n)]_i)) \approx \frac{\rho}{\mu} \bar{g}_i^{-\frac{3}{2}}(n). \tag{A.9}$$

Substituting in (A.8), we observe that convergence condition for $\tilde{\lambda}_{s,i}(n)$ for active taps is given by $-1 < 1 - \mu(2-\mu)\bar{g}_i(n) < 1$, which after some manipulations leads to $0 < \mu < 2$. Under this, letting $n \to \infty$, we obtain

$$\begin{aligned} \tilde{\lambda}_{s,i}(\infty) =&\ \frac{\mu}{2-\mu}\sigma_v^2 \left(\frac{1}{\hat{r}_s^2(\infty)}\right) + \frac{\rho^2}{\mu(2-\mu)}\bar{g}_i^{-2}(\infty) \\ &+ 2\frac{\rho\left(1 - \mu \bar{g}_i(\infty)\right)}{\mu(2-\mu)\bar{g}_i(\infty)} \bar{g}_i^{-\frac{1}{2}}(\infty) \left(\frac{\rho}{\mu}\bar{g}_i^{-\frac{3}{2}}(\infty)\right) \\ =&\ \frac{\mu}{2-\mu}\left(\frac{\sigma_v^2}{\hat{r}_s^2(\infty)}\right) + \frac{\rho^2\left(2 - \mu \bar{g}_i(\infty)\right)}{\mu^2(2-\mu)}\bar{g}_i^{-3}(\infty). \end{aligned} \tag{A.10}$$

Substituting (A.10) in the relation $\eta_i(\infty) = \bar{g}_i(\infty)\tilde{\lambda}_{s,i}(\infty)$, the result (18) follows trivially.

(b) *Inactive Taps* ($i \in Z$): For the inactive taps, $w_{\mathrm{opt},i} = 0, i \in Z$, and thus, we have $\mathrm{sgn}([\mathbf{w}(n)]_i) = \mathrm{sgn}(w_i(n)) = -\mathrm{sgn}(\tilde{w}_i(n)) = -\mathrm{sgn}(\tilde{w}_{N,i}(n))$ as $\tilde{w}_i(n) = w_{\mathrm{opt},i} - w_i(n) = -w_i(n)$, $\mathrm{sgn}(\tilde{w}_{N,i}(n)) = \mathrm{sgn}(-w_{N,i}(n)) = \mathrm{sgn}(-g_i^{-\frac{1}{2}}(n)w_i(n)) = \mathrm{sgn}(g_i^{-\frac{1}{2}}(n)\tilde{w}_i(n)) = \mathrm{sgn}(\tilde{w}_i(n))$ and $g_i(n) = [\mathbf{G}(n)]_{i,i} > 0$. We now invoke the "Price's Theorem" [27], which states that if x, y are two zero mean, jointly Gaussian random variables, then, we have

$$E(x\mathrm{sgn}(y)) = \frac{1}{\sigma_y}\sqrt{\frac{2}{\pi}}E(xy). \tag{A.11}$$

Note also that when $x = y$, (A.11) implies $E(abs(y)) = \sqrt{\frac{2}{\pi}}\sigma_y$, where $\sigma_y^2 = E(y^2)$. Then, assuming that each $w_{N,i}(n)$ (or equivalently, $\tilde{w}_{N,i}(n)$) is Gaussian distributed and noting from 14 that in the steady state, $E(w_{N,i}(n)) \approx 0$ as $w_{\mathrm{opt},i} = 0$ and ρ is very small, we can write

$$\begin{aligned} &E((\tilde{\mathbf{w}}_N(n))_i \mathrm{sgn}((\mathbf{w}(n))_i) = -E((\tilde{w}_{N,i}(n))\mathrm{sgn}(\tilde{w}_{N,i}(n)) \\ &= -\sqrt{\frac{2}{\pi}}\sqrt{E(\tilde{w}_{N,i}^2(n))} = -\sqrt{\frac{2}{\pi}}\sqrt{\tilde{\lambda}_{s,i}(n)}. \end{aligned} \tag{A.12}$$

Substituting in (A.8), we obtain

$$\begin{aligned} \tilde{\lambda}_{s,i}(n+1) =&\ [1 - \mu(2-\mu)\bar{g}_i(n)]\, \tilde{\lambda}_{s,i}(n) \\ &+ \mu^2 \sigma_v^2 E\left(\frac{1}{r_s^2(n)}\right) \bar{g}_i(n) \\ &+ \rho^2 \bar{g}_i^{-1}(n) - 2\rho\left(1 - \mu \bar{g}_i(n)\right) \bar{g}_i^{-\frac{1}{2}}(n)\sqrt{\frac{2}{\pi}}\sqrt{\tilde{\lambda}_{s,i}(n)}. \end{aligned} \tag{A.13}$$

Convergence of $\tilde{\lambda}_{s,i}(n)$ then requires $-1 < [1 - \mu(2-\mu)\bar{g}_i(n)] < 1$ and also $-1 < 2\rho(1 - \mu\bar{g}_i(n))\bar{g}_i^{-\frac{1}{2}}(n)\sqrt{\frac{2}{\pi}} < 1$. It is easy to see that these are satisfied simultaneously for $0 < \mu < 2$ with small value of ρ (specifically $0 \leq \rho < \sqrt{\frac{\pi \bar{g}_i(n)}{8}}$). Under this and letting $n \to \infty$ on both sides of (A.13), we then have

$$\begin{aligned}&\mu(2-\mu)\bar{g}_i(\infty)\tilde{\lambda}_{s,i}(\infty)\\&\quad+2\rho\left(1-\mu\bar{g}_i(\infty)\right)\bar{g}_i^{-\frac{1}{2}}(\infty)\sqrt{\frac{2}{\pi}}\sqrt{\tilde{\lambda}_{s,i}(\infty)}\\&\quad-\left(\mu^2\sigma_v^2\frac{\bar{g}_i(\infty)}{\hat{r}_s^2(\infty)}+\rho^2\bar{g}_i^{-1}(\infty)\right)=0.\end{aligned}\tag{A.14}$$

Now, since $w_{\mathrm{opt},i}=0$ for $i\in Z$, in steady state, $g_i(n)=\rho_g g_{\max}(n)$, where $g_{\max}(n)=\max\{g_i(n)|i=0,1,\ldots,L-1\}$ (which follows from the definition of $\mathbf{G}(n)$). Thus, $g_i(n)$ is common for all the inactive taps, meaning $\bar{g}_i(n)$ in (A.14) can be replaced by $\rho_g\bar{g}_{\max}(n)$. It then follows from (A.14) that $\tilde{\lambda}_{s,i}(\infty)$ is given by square of the positive root of the following quadratic equation:

$$at^2+bt-c=0,\tag{A.15}$$

where

$$a=\mu(2-\mu)\rho_g\bar{g}_{\max}(\infty),\tag{A.16}$$

$$b=2\sqrt{\frac{2}{\pi}}\rho\left(1-\mu\rho_g\bar{g}_{\max}(\infty)\right)\rho_g^{-\frac{1}{2}}\bar{g}_{\max}^{-\frac{1}{2}}(\infty),\tag{A.17}$$

$$c=\mu^2\sigma_v^2\frac{\rho_g\bar{g}_{\max}(\infty)}{\sigma_x^2}+\rho^2\rho_g^{-1}\bar{g}_{\max}^{-1}(\infty).\tag{A.18}$$

Note that since a,b,c are independent of i, i.e., same for all $i\in Z$, $\tilde{\lambda}_{s,i}(\infty)$ is same for all inactive taps. From this and the above, and recalling that $\eta_i(\infty)=\bar{g}_i(\infty)\tilde{\lambda}_{s,i}(\infty)$, the relation (19) follows trivially and this completes the proof.

REFERENCES

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

[2] V. V. Krishna, J. Rayala, and B. Slade, “Algorithmic and implementation aspects of echo cancellation in packet voice networks,” in *Proc. 36th Asilomar Conf. Signals, Syst. Comput.*, 2002, vol. 2, pp. 1252–1257.

[3] E. Hansler and G. Schmidt, Eds., *Topics in Acoustic Echo and Noise Control*. Berlin, Germany, Springer-Verlag, 2006.

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

[5] W. Bajwa, J. Haupt, G. Raz, and R. Nowak, “Compressed channel sensing,” in *42nd Annu. Conf. Inform. Syst. Sci.*, 2008, pp. 5–10.

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

[7] R. L. Das and M. Chakraborty, “Sparse adaptive filters—An overview and some new results,” *Proc. IEEE Int. Symp. Circuits Syst.*, Seoul, Korea, May 2012, pp. 2745–2748.

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

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

[10] J. Benesty and S. L. Gay, “An improved PNLMS algorithm,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Orlando, FL, USA, 2002, pp. 1881–1884.

[11] S. L. Gay, “An efficient, fast converging adaptive filter for network echo cancellation,” in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 1998, pp. 394–398.

[12] 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, 2005.

[13] L. Liu, M. Fukumoto, S. Saiki, and S. Zhang, “A variable step-size proportionate affine projection algorithm for identification of sparse impulse response,” *EURASIP J. Adv. Signal Process.*, pp. 1–10, 2009.

[14] L. Liu, M. Fukumoto, S. Saiki, and S. Zhang, “A variable step-size proportionate NLMS algorithm for identification of sparse impulse response,” *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E93-A, no. 1, pp. 233–242, Jan. 2010.

[15] C. Paleologu, J. Benesty, F. Albu, and S. Ciochină, “An efficient variable step-size proportionate affine projection algorithm,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, May 2011, pp. 77–80.

[16] Y. Chen, Y. Gu, and A. O. Hero, “Sparse LMS for system identification,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 3125–3128.

[17] Y. Chen, Y. Gu, and A. O. Hero, “Regularized least-mean-square algorithms,” (Dec., 2010) [Online]. Available: http://arxiv.org/abs/1012.5066

[18] 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.

[19] R. L. Das and M. Chakraborty, “A variable step-size zero attracting proportionate normalized least mean square algorithm,” in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Melbourne, Australia, Jun. 1–5 2014, pp. 1187–1190.

[20] M. Yukawa and I. Yamada, “A unified view of adaptive variable—Metric projection algorithms,” *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–13, 2009.

[21] R. L. Das and M. Chakraborty, “On convergence of proportionate-type normalized least mean square algorithms,” *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 62, no. 5, pp. 491–495, May 2015.

[22] S. G. Sankaran and A. A. (Louis) Beex, “Convergence behavior of affine projection algorithms,” *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 1086–1096, Apr. 2000.

[23] T. K. Paul and T. Ogunfunmi, “On the convergence behavior of the affine projection algorithm for adaptive filters,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 8, pp. 1813–1826, Aug. 2011.

[24] R. L. Das and M. Chakraborty, “Zero attracting PNLMS algorithm and its convergence in mean,” (Jul., 2015). [Online]. Available: http://arxiv.org/abs/1507.02921

[25] J. Benesty, H. Rey, L. R. Vega, and S. Tressens, “A nonparametric VSS NLMS algorithm,” *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 581–584, Oct. 2006.

[26] C. Paleologu, J. Benesty, and S. Ciochină, “A variable step-size affine projection algorithm designed for acoustic echo cancellation”, *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1466–1478, 2008.

[27] R. Price, “A useful theorem for nonlinear devices using Gaussian inputs,” *IRE Trans. Inform. Theory*, vol. 4, no. 2, pp. 69–72, Jun. 1958.

Rajib Lochan Das (M’15) received the Bachelor’s degree in electronics and communication engineering from the National Institute of Technology, Durgapur, West Bengal, India, in 2000, the Master’s degree in control system engineering from Jadavpur University, Kolkata, West Bengal, in 2003, and the Ph.D. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, West Bengal, in 2015. He is currently an Assistant Professor at the Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India. His research interests include adaptive signal processing, image processing, and speech processing.

Mrityunjoy Chakraborty (M’94–SM’99) is currently a Professor in the Department of Electronics and Electrical Communication Engineering, Kharagpur, West Bengal, India, with teaching and research interests in digital and adaptive signal processing, compressive sensing, VLSI signal processing, and applied linear algebra. Details of the biography of Prof. Chakraborty can be found at http://www.ecdept.iitkgp.ernet.in/faculty/MChakraborty/.