Design of a Tap-Amplitude-Based Block Proportional Adaptive Filtering Algorithm

Sheng Zhang, *Senior Member, IEEE*, Hongyang Chen, *Senior Member, IEEE*, and Ali H. Sayed, *Fellow, IEEE*

***Abstract*—Proportional-type algorithms have attracted much attention because of their fast convergence ability for sparse system identification. To overcome the drawbacks of existing block proportional methods stemming from inadequate block partitioning, this article develops tap-amplitude-based block partitioning methods. In the procedure, we present two block proportional normalized least-mean-square (PNLMS) algorithms named (i) ABx-PNLMS and (ii) ABy-PNLMS. In the first algorithm, the proportional gain depends on the rank-based tap-weight block on the x-axis. In comparison, the second algorithm determines the proportional gain by dynamically partitioning the y-axis in a non-uniform manner. Nevertheless, both algorithms have a common feature, namely, weights of similar magnitude are grouped together. By further incorporating an adaptive decorrelation mechanism into the ABx-PNLMS and ABy-PNLMS, two improved decorrelation proportional alogrithms are developed to speed up the convergence for colored inputs. Computer simulation results on system identification and acoustic echo cancellation demonstrate the effectiveness of the proposed approaches.**

***Index Terms*—Adaptive filter, sparse model, block proportional, decorrelation.**

I. Introduction

SPARSE system identification is useful in many applications, such as digital network and acoustic echo cancellation and underwater acoustic communications (see [1], [2], [3], [4] and the references therein). In sparse models, there are a few components in the impulse response with significant magnitude values, while all other components have zero or small values. In traditional adaptive filtering, such as methods based on the least-mean-square (LMS) algorithm or its normalized version (NLMS), the same step-size gain is usually used for all tap weights, which can slow down convergence for sparse models.

Manuscript received 18 October 2022; revised 7 March 2023, 10 April 2023, and 2 June 2023; accepted 20 June 2023. Date of publication 29 June 2023; date of current version 11 July 2023. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Behtash Babadi. This work was supported in part by the National Natural Science Foundation of China under Grants 61801401 and 62271452 and in part by the key research project of Zhejiang Lab under Grant 2022PI0AC01. *(Corresponding author: Hongyang Chen.)*

Sheng Zhang is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China (e-mail: dr.s.zhang@ieee.org).

Hongyang Chen is with the Research Center for Graph Computing, Zhejiang Lab, Hangzhou 311121, China (e-mail: dr.h.chen@ieee.org).

Ali H. Sayed is with the School of Engineering, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: ali.sayed@epfl.ch).

Digital Object Identifier 10.1109/TSP.2023.3290659

To tackle this problem, many adaptive algorithms using *the prior* sparse features of the impulse response have been developed in the literature, including proportionate NLMS (PNLMS) [5], zero-attracting LMS (ZA-LMS) [6], feature LMS (F-LMS) [7], [8], and sparsity-promoting LMS [9] algorithms. The principle behind the PNLMS algorithm proposed is to assign large gains to large tap entries based on their current estimated magnitudes. Nevertheless, the algorithm suffers from some drawbacks. For instance, over the course of the adaptation process, PNLMS does not maintain its initial fast convergence behavior. Also, as sparsity decreases, convergence speed further degrades. In response to these issues, several improved variations have been proposed, such as PNLMS++ [10], improved PNLMS (IPNLMS) [11], μ-law PNLMS (MPNLMS) [12], gradient-controlled [13], sparseness controlled [14], and the individual-activation-factor PNLMS (IAF-PNLMS) [15], [16]. Unlike proportional-type algorithms, the ZA-LMS and F-LMS algorithms were obtained, respectively, by inserting ℓ_1-norm penalties in the cost function, resulting in low steady-state error. The performance of zero-attraction type algorithms, including the ZA-LMS, reweighted ZA-LMS, ℓ_0-LMS [17] and its improved versions [18], [19], depends on the selection of the zero-attractor controller, whose optimal value is related to the step-size, to the sparsity of the unknown channel, and to the input and noise powers. Distribtuted versions of proportional and zero-attraction algorithms are discussed in [20], [21], [22]. A variable step-size LMS algorithm is developed using an optimized differential step-size criterion related to the filter coefficients in [23], but it results in high complexity.

Different from the aforementioned algorithms for general sparse systems, block-sparse models are considered in [24], where block-sparse PNLMS (B-PNLMS) and IPNLMS (B-IPNLMS) are developed. Likewise, block-sparse LMS methods exploiting a block zero attraction strategy (BZA-LMS) is designed in [25]. Using both proportional and zero-attraction mechanisms for block-sparse system identification, improved B-IPNLMS algorithms are presented in [26], [27]. However, all these methods share a common problem, namely, improper block segmentation will lead to blocks that may contain both zero and large non-zero taps. This generates a conflict when the same proportional gain should be assigned to each block (with the block having entries that differ in the scale of their magnitudes).

Due to the large eigenvalue spread of the input covariance matrix, the proportionate/zero-attracting NLMS algorithms still

experience slow convergence on highly correlated inputs, like speech signals. To enhance the convergence rate for the correlated input, several methods with decorrelation capability have been integrated into the proportional algorithm, including the affine projection algorithm [28], [29], normalized frequency-domain methods [30], [31], subband adaptive filtering schemes [32], [33], [34], [35], just to mention a few. Unfortunately, their decorrelation capabilities are limited, which are influenced by the number of data reuse, orthogonal transforms, analysis filters, and the number of subbands, respectively. Recently, an adaptive decorrelation NLMS (DNLMS) algorithm was proposed in [36]. Due to the use of an adaptive decorrelation vector in real time, the algorithm can be better adapt to colored signals. To reduce the computational complexity and steady-state error of the DNLMS method, the separated-decorrelation NLMS algorithm and its low-cost implementation have also been developed in [37], [38].

In this article, we develop new more efficient proportional gain design methods for block-sparse systems to tackle the aforementioned challenges. As a first step, a sorting operation using the magnitude of the weights is introduced into the mean-squared-error (MSE) objective function with sparse regularization, to address the shortcomings of the traditional block division method. The formulation leads to a novel rank-based block proportional NLMS method. Due to the nature of the tap-amplitude-based uniform block division along the x-axis, we refer to this method as the ABx-PNLMS algorithm. The uniform division here means that the number of tap-weights in each block is identical. Next, using the y-axis division as a benchmark, the tap-amplitude-based partitioning block proportional NLMS (ABy-PNLMS) algorithm is designed. Compared to the first algorithm, the main advantage of ABy-PNLMS is the dynamic non-uniform segmentation of the y-axis. The adaptive decorrelation mechanism incorporated into the ABx-PNLMS and ABy-PNLMS algorithms help speed up convergence for colored inputs. The main contributions of this article are summarized as follows:

- By dividing the blocks along the x- and y-axes, the ABx-PNLMS and ABy-PNLMS algorithms are derived;
- In the ABy-PNLMS algorithm, we design a dynamic partitioning mechanism to speed up convergence especially when approaching the steady-state stage;
- The mean and mean-square behavior of ABy-PNLMS is discussed. Motivated by the simulation results, we point out the limitations of the current proportional analysis method;
- Applying the recently proposed adaptive decorrelation mechanism, we also design the decorrelation ABx-PNLMS and ABy-PNLMS (named ABx-DPNLMS and ABy-DPNLMS) algorithms without significantly increasing their complexity.

The remainder of the article is organized as follows. Section II provides an overview of system model and PNLMS. Section III presents the two proportional algorithms. In Section IV, we study the mean and mean-square properties. Section V develops the accelerated proportional algorithm for colored inputs. Section VI provides simulation results, while Section VII summarizes the article.

Notation:

$\lvert\cdot\rvert$	absolute value operator.
$(\cdot)^{\mathsf{T}}$	transpose operator.
$\mathrm{diag}\{\mathbf{a}\}$	diagonal matrix constructed from vector $\mathbf{a}$
$\lVert\cdot\rVert_2$	Euclidean norm.
$\lVert\cdot\rVert_1$	ℓ_1-norm.
$\mathrm{Tr}(\cdot)$	trace of a matrix.
$\mathbf{A}\otimes\mathbf{B}$	Kronecker product of $\mathbf{A}$ and $\mathbf{B}$
$\mathbb{E}\{\cdot\}$	expectation operator.
$\mathbb{1}_L$	$L\times 1$ all one vector.
$\boldsymbol{I}_L$	$L\times L$ identity matrix.

II. Review of System Model and Proportional NLMS Algorithm

Let us consider an FIR system identification problem. The observed output signal of the unknown system at each time instant i is expressed as

$$d(i)=\mathbf{u}^{\mathsf{T}}(i)\mathbf{w}_{\mathrm{opt}}+v(i), \tag{1}$$

where $\mathbf{u}(i)=[u(i),u(i-1),\ldots,u(i-L+1)]^{\mathsf{T}}$ is the input vector, $v(i)$ is zero-mean white measurement noise, and $\mathbf{w}_{\mathrm{opt}}=[w_{\mathrm{opt},1},w_{\mathrm{opt},2},\ldots,w_{\mathrm{opt},L}]^{\mathsf{T}}$ denotes the L-length sparse impulse response to be estimated.

A. Proportionate NLMS

Define the estimate $\mathbf{w}(i)=[w_1(i),w_2(i),\ldots,w_L(i)]^{\mathsf{T}}$. The update of the standard PNLMS algorithm takes the following form [5]:

$$\mathbf{w}(i+1)=\mathbf{w}(i)+\mu\frac{\mathbf{P}(i)\mathbf{u}(i)}{\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)\mathbf{u}(i)+\epsilon}e(i), \tag{2}$$

where μ is the step-size, the error signal is calculated as $e(i)=d(i)-\mathbf{u}^{\mathsf{T}}(i)\mathbf{w}(i)$, ϵ is a small positive regularization parameter, and the diagonal proportionate matrix $\mathbf{P}(i)$ is

$$\mathbf{P}(i)=\mathrm{diag}\{p_1(i),p_2(i),\ldots,p_L(i)\}, \tag{3}$$

with the $p_\ell(i)$ denoting individual gains. In the improved PNLMS algorithm [11], each gain is computed as:

$$p_\ell(i)=\frac{1-\gamma}{2L}+\frac{(1+\gamma)|w_\ell(i)|}{2\lVert\mathbf{w}(i)\rVert_1+\epsilon},\ \ell=1,2,\ldots,L \tag{4}$$

where the parameter $-1<\gamma<1$ determines the behavior of IPNLMS between NLMS and PNLMS.

B. Partitioned Block Proportionate NLMS

Different from general sparse systems where the nonzero taps are generally positioned arbitrarily, the block-sparse channel is characterized by one or more active zones, each of which is made up of nonzero taps. When identifying block-sparse channels, the traditional PNLMS has been extended to the partitioned block PNLMS in [24]. It is described as follows.

Let us denote the number of partitions by M and number of samples per partition by P. Along the x-axis, the unknown channel is evenly divided into M partitions, illustrated in Fig. 1(a) by the vertical lines. We write

$$\mathbf{w}(i)=\left[\mathbf{w}_{[1]}^{\mathsf{T}}(i),\mathbf{w}_{[2]}^{\mathsf{T}}(i),\ldots,\mathbf{w}_{[M]}^{\mathsf{T}}(i)\right]^{\mathsf{T}}, \tag{5}$$

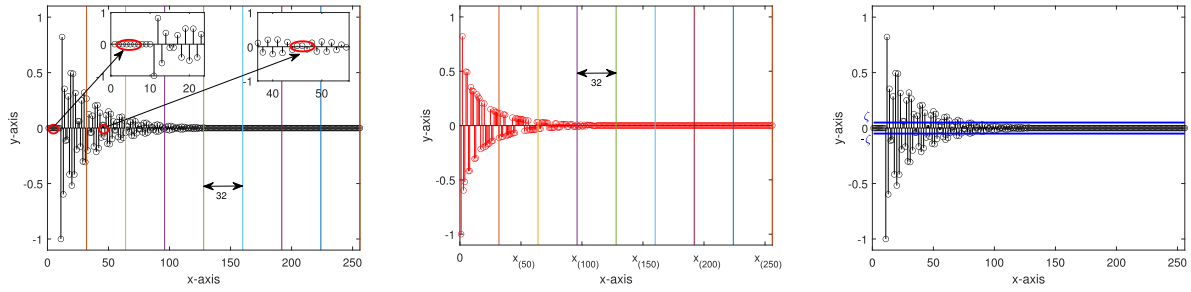


Fig. 1. Example of block partitioning, where $P = 32$. (a) Assigning partitions along the x-axis; (b) Assigning partitions along the x-axis after sorting (descending order); (c) y-axis-based partitions.

where the entries associated with the ℓ-th partition is $\mathbf{w}_{[\ell]}(i) = [w_{(\ell-1)P+1}(i), w_{(\ell-1)P+2}(i), \ldots, w_{\ell P}(i)]^{\mathsf{T}}$ and $M = L/P$.

In the B-IPNLMS, the diagonal proportionate matrix is chosen as

$$\mathbf{P}(i) = \mathsf{diag}\{p_1(i), p_2(i), ..., p_M(i)\} \otimes \boldsymbol{I}_P, \tag{6}$$

where the individual gain is now given by [24]:

$$p_m(i) = \frac{1-\gamma}{2L} + \frac{(1+\gamma)\|\mathbf{w}_{[m]}(i)\|_2}{2P\sum_{m=1}^{M}\|\mathbf{w}_{[m]}(i)\|_2+\epsilon}, \; m = 1, \ldots, M \tag{7}$$

In contrast to the standard PNLMS algorithm, the block proportional algorithm assigns an equal gain to the weights within each sub-block. However, each sub-block may contain both large and small taps. For example, in Fig. 1(a), $w_{\text{opt},2} = 0, w_{\text{opt},12} = 0.8$ in the first sub-block, and both of these entries are scaled by the same proportional gain. This fact violates the proportionate principle, where the idea is to scale entries in proportion to their sizes. Therefore, the block proportional algorithm is also not suitable for general sparse systems.

III. Proposed Tap-Amplitude-Based Partitioning

In view of the shortcomings of the block proportional algorithm, this section devises two different tap-amplitude-based partitioning block proportional methods, namely, the ABx-PNLMS and ABy-PNLMS algorithms.

A. Assigning Partitions Along the x-Axis

As a first step toward improving the concept of proportionality in the block proportional method and avoiding sub-blocks that contain both small and large taps, the weight vector $\mathbf{w}(i)$ is sorted from large to small magnitude values, as follows

$$\mathbf{w}_s(i) = [w^m_{(1)}(i), w^k_{(2)}(i), \ldots, w^h_{(L)}(i)]^{\mathsf{T}}, \tag{8}$$

where $|w^m_{(1)}(i)| \geq |w^k_{(2)}(i)| \geq \cdots \geq |w^h_{(L)}(i)|$. In (8), the subscript is the sorted order in the vector $\mathbf{w}_s(i)$ and the superscript is the original position of the element in the vector $\mathbf{w}(i)$. We introduce the mixed $\ell_{2,1}$-norm of $\mathbf{w}_s(i)$

$$\|\mathbf{w}_s(i)\|_{2,1} = \sum_{m=1}^{M}\|\mathbf{w}_{([m])}(i)\|_2, \tag{9}$$

where $\mathbf{w}_{([m])}(i)$ is the m-th ordered sub-block within $\mathbf{w}_s(i)$. Fig. 1(b) shows a uniform partitioning of the x-axis after sorting. Using $\boldsymbol{w}_s$, we then formulate the following regularized mean-square-error estimation problem:

$$\min \; \mathbb{E}\{(d(i) - \mathbf{u}_s^{\mathsf{T}}(i)\mathbf{w}_s)^2\} + \rho\|\mathbf{Q}^{-1}\mathbf{w}_s\|_{2,1}, \tag{10}$$

with ρ being the regularization coefficient and the matrix $\mathbf{Q} > 0$ being block diagonal, which makes $\mathbf{Q}^{-1}\mathbf{w}_s$ block sparse. We thus arrive at the ABx-PNLMS algorithm (see Appendix A):

$$\mathbf{w}_s(i+1|i) = \mathbf{w}_s(i) + \mu\frac{\mathbf{P}_s(i)\mathbf{u}_s(i)}{\mathbf{u}_s^{\mathsf{T}}(i)\mathbf{P}_s(i)\mathbf{u}_s(i)+\epsilon}e(i), \tag{11}$$

where $\mathbf{u}_s(i)$ is the original input vector with its entries reshuffled in the same manner as the entries of $\mathbf{w}(i)$ to arrive at $\mathbf{w}_s(i)$, $e(i) = d(i) - \mathbf{u}_s^{\mathsf{T}}(i)\mathbf{w}_s(i)$, and

$$\mathbf{P}_s(i) = \mathsf{diag}\{p_{[1]}(i), p_{[2]}(i), ..., p_{[M]}(i)\} \otimes \boldsymbol{I}_P, \tag{12}$$

with

$$p_{[m]}(i) = \frac{1-\gamma}{2L} + \frac{(1+\gamma)\|\mathbf{w}_{([m])}(i)\|_2}{2P\sum_{m=1}^{M}\|\mathbf{w}_{([m])}(i)\|_2+\epsilon}. \tag{13}$$

In (11), we get $\mathbf{w}_s(i+1|i)$, which is consistent with the order of $\mathbf{w}_s(i)$ and may differ from $\mathbf{w}_s(i+1)$. For continuous updating, we sort the result of (11) to generate $\mathbf{w}_s(i+1)$ at time $i+1$.

According to the original order of the taps, using (11) we find that the original weight would be rearranged and updated equivalently as

$$\mathbf{w}(i+1) = \mathbf{w}(i) + \mu\frac{\mathbf{P}'(i)\mathbf{u}(i)}{\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}'(i)\mathbf{u}(i)+\epsilon}e(i), \tag{14}$$

where $\mathbf{P}'(i) = \mathsf{diag}\{p'_1(i), p'_2(i), \ldots, p'_L(i)\}$. If the tap $w_m(i)$ belongs to the ℓ-th ordered sub-block $\mathbf{w}_{([\ell])}(i)$, we set $p'_m(i) = p_{[\ell]}(i)$.

B. Assigning Partitions Along the y-axis Dynamically

Next, we consider non-uniform partitioning along the y-axis. For simplicity, we divide all taps into two groups $\mathcal{N}_1(i)$ and $\mathcal{N}_2(i)$ according to a threshold ζ and the magnitude of the weights, i.e.,

$$|w_\ell(i)| \leq \zeta, \; \ell \in \mathcal{N}_1(i), \tag{15}$$

$$|w_\ell(i)| > \zeta, \; \ell \in \mathcal{N}_2(i). \tag{16}$$

TABLE I
SUMMARY OF PROPOSED ABx-PNLMS ALGORITHM

Initialization
$w_\ell(0)=0$, ϵ, γ, μ, P, L, M
For $i=1,2,3\cdots$
Calculate $p_\ell(i)$:
$\mathbf{w}(i)=\left[|w_1(i)|,|w_2(i)|,\cdots,|w_L(i)|\right]^{\mathsf{T}}$;
$[\mathbf{w}_s(i),\mathbf{c}]=\text{sort}(\mathbf{w}(i),\text{'descend'})$;
%Sort a vector using matlab's function sort$(\cdot)$
%$\mathbf{w}_s(i)$ lists the sorted dates and
%$\mathbf{c}$ contains the corresponding indices of $\mathbf{w}(i)$
$\mathbf{g}=$zeros$(L,1)$; $a=0$;
for $m=1:M$
$\delta=\|\mathbf{w}_{s,(m-1)P+1:mP}(i)\|_2$;
$\mathbf{g}(\mathbf{c}_{(m-1)P+1:mP})=\delta$;
$a=a+\delta$;
end
$\mathbf{p}=\frac{1-\gamma}{2L}+\frac{(1+\gamma)\mathbf{g}}{2aP+\epsilon}$;
$\mathbf{P}'(i)=\text{diag}\{\mathbf{p}\}$;
Weight update step:
$e(i)=d(i)-\mathbf{u}^{\mathsf{T}}(i)\mathbf{w}(i)$;
use (14);
end

TABLE II
SUMMARY OF PROPOSED ABy-PNLMS ALGORITHM

Initialization
$\widehat{\sigma}_e^2(0)=0$, $w_\ell(0)=0$, ϵ, γ, μ, κ, β, σ_v^2, L
For $i=1,2,3\cdots$
Calculate $p_\ell(i)$:
$e(i)=d(i)-\mathbf{u}^{\mathsf{T}}(i)\mathbf{w}(i)$;
$\widehat{\sigma}_e^2(i)=\beta\widehat{\sigma}_e^2(i-1)+(1-\beta)e^2(i)$;
$\lambda(i)=\text{median}(1,\frac{\widehat{\sigma}_e^2(i)-\sigma_v^2}{\widehat{\sigma}_e^2(i)+\epsilon},0)$;
$\zeta=\kappa\lambda(i)\|\mathbf{w}(i)\|_1/L$; $a_1=0$; $N=0$;
for $\ell=1:L$
If $|w_\ell(i)|\le\zeta$
$a_1=a_1+|w_\ell(i)|$;
$N=N+1$;
end
end
$g_1=\frac{1-\gamma}{2L}+\frac{(1+\gamma)a_1/N}{2\|\mathbf{w}(i)\|_1+\epsilon}$;
$g_2=(1-g_1N)/(L-N)$;
for $\ell=1:L$
if $N=0$
$p_\ell=1/L$;
elseif $|w_\ell(i)|\le\zeta$
$p_\ell=g_1$;
else
$p_\ell=g_2$;
end
end
$\mathbf{P}(i)=\text{diag}\{p_1,p_2,\cdots,p_L\}$;
Weight update step:
use (2);
end

Since the number of elements in $\mathcal{N}_1(i)$ and $\mathcal{N}_2(i)$ may be different, we use the average of the tap amplitudes in the interval to calculate the proportional gain:

$$p_\ell(i)=\begin{cases}\frac{1-\gamma}{2L}+\frac{(1+\gamma)a_1(i)}{2(|\mathcal{N}_1(i)|a_1(i)+|\mathcal{N}_2(i)|a_2(i))+\epsilon}, & \ell\in\mathcal{N}_1(i)\\ \frac{1-\gamma}{2L}+\frac{(1+\gamma)a_2(i)}{2(|\mathcal{N}_1(i)|a_1(i)+|\mathcal{N}_2(i)|a_2(i))+\epsilon}, & \ell\in\mathcal{N}_2(i)\end{cases}\quad(17)$$

where $a_1(i)=\frac{1}{|\mathcal{N}_1(i)|}\sum_{\ell\in\mathcal{N}_1(i)}|w_\ell(i)|$ and $a_2(i)=\frac{1}{|\mathcal{N}_2(i)|}\sum_{\ell\in\mathcal{N}_2(i)}|w_\ell(i)|$. It is worth noting that (17) also uses improved proportionality principles. Recall that proportional algorithms experience slow convergence when approaching steady state [5], [12]. In light of this, we choose a time-varying threshold:

$$\zeta=\kappa\lambda(i)\frac{\|\mathbf{w}(i)\|_1}{L},\quad(18)$$

where κ is a free parameter[1] in $(0,1]$. Regarding the time-varying parameter $\lambda(i)$, we expect it to be close to 1 in the transient state, and 0 near convergence. One recommended method to update its value is:

$$\lambda(i)=\begin{cases}\frac{\widehat{\sigma}_e^2(i)-\sigma_v^2}{\widehat{\sigma}_e^2(i)+\epsilon}, & \text{if } \widehat{\sigma}_e^2(n)>\sigma_v^2\\ 0, & \text{otherwise}\end{cases}\quad(19)$$

where $\widehat{\sigma}_e^2(i)=\beta\widehat{\sigma}_e^2(i-1)+(1-\beta)e^2(i)$ with an initial value $\widehat{\sigma}_e^2(0)=0$ and $0\ll\beta<1$. Since the noise variance σ_v^2 can be estimated during periods of silence, it is assumed to be known *a priori*.[2] If $|\mathcal{N}_1(i)|=0$ occurs during the operation of the algorithm, this will result in $a_1(i)=0/0$. To avoid this problem, we set $\mathbf{P}(i)$ to $\frac{1}{L}\boldsymbol{I}_L$ in this case. Tables I and II summarize the two proposed proportional algorithms.

*Remark 1 (**ABx-PNLMS**):* By ordering the estimated weights, the ABx-PNLMS algorithm avoids zero and large non-zero weights falling within the same partition. Compared with B-IPNLMS, it enhances the concept of proportionality, while also using block information based on the magnitude of the weights.

*Remark 2 (**ABy-PNLMS**):* Similar to the two-gain NLMS (TG-NLMS) recently developed in [41], there are also two different gains in ABy-PNLMS. Taps that are smaller than the threshold receive a small gain, whereas taps that are larger than the threshold receive a large gain. The difference is that the thresholds in ABy-PNLMS follow typical dynamics varying from $\kappa\|\mathbf{w}(i)\|_1/L$ to 0. At the initial stage (i.e., $\lambda(i)\approx1$), taps larger than the average amplitude converge rapidly at first. As $\lambda(i)$ decreases, small taps attract attention and start to receive bigger gains. As such, our dynamic partitioning may be more suited for proportional matrix designs than the existing proportional methods, to attain an improved convergence performance.

*Remark 3 (**Extensions**):* Throughout this article, we mainly use the improved proportional gain mechanism [11]. Alternatively, sparseness-controlled, μ-law and individual-activation-factor methods can also be integrated into the ABx-PNLMS or ABy-PNLMS formulations for optimization in different sparse system scenarios.

IV. STATISTICAL ANALYSIS

For the ABx-PNLMS algorithm, computing the expectation of $\mathbf{P}'(i)$ is complicated because of sorting operations. Thus, in this section, we only carry out the statistical analysis of ABy-PNLMS. In general, the mean and mean-square analysis of proportional-type algorithms is challenging, due to the magnitude of the weight entries, which determine the proportional matrices. To facilitate the convergence analysis, we introduce three customary assumptions.

[1]When $\kappa\approx0$, the ABy-PNLMS algorithm will degenerate into the NLMS, resulting in slow initial convergence. Through extensive numerical simulations, we find that $\kappa=0.2\sim0.5$ works well for block-sparse plants and $\kappa=0.8\sim1$ for general sparse systems.

[2]Online adaptive estimation schemes can also be found in [39], [40].

Assumption 1: The input signals $\{u(i)\}$ are independent zero-mean stationary white Gaussian random sequences with variance σ_u^2.

Assumption 2: The noise signals $\{v(i)\}$ are independent zero-mean stationary random variables with $\mathbb{E}\{v^2(i)\}=\sigma_v^2$, and are statistically independent of the sequence $\{u(i)\}$.

Assumption 3: $\mathbf{w}(i)$ is statistically independent of $\mathbf{u}(i)$. The independence assumption is well known and commonly used in the study of adaptive filters [1], [2], [3], [37], [42], [43], [44], [47].

Define the weight error vector $\widetilde{\mathbf{w}}(i)=\mathbf{w}_{\rm opt}-\mathbf{w}(i)$. From (2), we have the recursive formulas for $\widetilde{\mathbf{w}}(i)$ and $\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)$:

$$\widetilde{\mathbf{w}}(i+1)=\widetilde{\mathbf{w}}(i)-\mu\frac{\mathbf{P}(i)\mathbf{u}(i)\mathbf{u}^{\mathsf{T}}(i)}{\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)\mathbf{u}(i)}\widetilde{\mathbf{w}}(i)-\mu\frac{\mathbf{P}(i)\mathbf{u}(i)}{\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)\mathbf{u}(i)}v(i), \tag{20}$$

and

$$\begin{aligned}&\widetilde{\mathbf{w}}(i+1)\widetilde{\mathbf{w}}^{\mathsf{T}}(i+1)\\&=\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)-\mu\frac{\mathbf{P}(i)\mathbf{u}(i)\mathbf{u}^{\mathsf{T}}(i)\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)}{\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)\mathbf{u}(i)}\\&\quad-\mu\frac{\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)\mathbf{u}(i)\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)}{\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)\mathbf{u}(i)}\\&\quad+\mu^2\frac{\mathbf{P}(i)\mathbf{u}(i)\mathbf{u}^{\mathsf{T}}(i)\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)\mathbf{u}(i)\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)}{(\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)\mathbf{u}(i))^2}\\&\quad+\mu^2v^2(i)\frac{\mathbf{P}(i)\mathbf{u}(i)\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)}{(\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)\mathbf{u}(i))^2}+\text{other terms},\end{aligned} \tag{21}$$

where the diagonal elements of matrix $\mathbf{P}(i)$ are defined in (17).

Using the approximation $\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)\mathbf{u}(i)\approx\sigma_u^2$ for long adaptive filters[3] [43] and taking the expectation of (20) and (21) under Assumptions 1 to 3, we arrive at

$$\mathbb{E}\{\widetilde{\mathbf{w}}(i+1)\}\approx\mathbb{E}\{\widetilde{\mathbf{w}}(i)\}-\mu\mathbb{E}\left\{\mathbf{P}(i)\widetilde{\mathbf{w}}(i)\right\}, \tag{22}$$

and

$$\begin{aligned}&\mathbb{E}\{\widetilde{\mathbf{w}}(i+1)\widetilde{\mathbf{w}}^{\mathsf{T}}(i+1)\}\\&\approx\mathbb{E}\{\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)\}-\mu\mathbb{E}\{\mathbf{P}(i)\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)\}\\&\quad-\mu\mathbb{E}\{\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)\mathbf{P}(i)\}\\&\quad+\mu^2\frac{1}{\sigma_u^4}\mathbb{E}\{\mathbf{P}(i)\mathbf{u}(i)\mathbf{u}^{\mathsf{T}}(i)\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)\mathbf{u}(i)\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)\}\\&\quad+\mu^2\sigma_v^2\frac{1}{\sigma_u^2}\mathbb{E}\{\mathbf{P}^2(i)\}.\end{aligned} \tag{23}$$

Using the fourth-moment theorem, the fourth expected value on the right-hand side of (23) is calculated as

$$\begin{aligned}&\mathbb{E}\{\mathbf{P}(i)\mathbb{E}\{\mathbf{u}(i)\mathbf{u}^{\mathsf{T}}(i)\mathbf{K}(i)\mathbf{u}(i)\mathbf{u}^{\mathsf{T}}(i)\}\mathbf{P}(i)\}\\&=\mathbb{E}\{\mathbf{P}(i)(2\sigma_u^4\mathbf{K}(i)+\sigma_u^4\mathrm{Tr}(\mathbf{K}(i)))\mathbf{P}(i)\},\end{aligned} \tag{24}$$

where $\mathbf{K}(i)=\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)$.

[3]Instead, there is another approximation $\mathbf{u}^{\mathsf{T}}(i)\mathbf{P}(i)\mathbf{u}(i)\approx\sum_{\ell=1}^{L}\mathbb{E}\{t_\ell^2(i)\}$, where $t_\ell(i)=u(i)\sqrt{p_\ell(i)}$.

A vector recursion is obtained by extracting the diagonal elements from (23)

$$\begin{aligned}\mathbb{E}\{\mathbf{s}(i+1)\}\approx&\mathbb{E}\{\mathbf{s}(i)\}-2\mu\mathbb{E}\{\mathbf{P}(i)\}\mathbb{E}\{\mathbf{s}(i)\}\\&+\mu^2\left(2\mathbb{E}\{\mathbf{P}^2(i)\}\mathbb{E}\{\mathbf{s}(i)\}+\mathbb{E}\{\mathbf{p}^2(i)\}\mathbb{E}\{\mathbb{1}^{\mathsf{T}}\mathbf{s}(i)\}\right)\\&+\mu^2\frac{\sigma_v^2}{\sigma_u^2}\mathbb{E}\{\mathbf{p}^2(i)\},\end{aligned} \tag{25}$$

where $\mathbf{s}(i)$ and $\mathbf{p}^2(i)$ denote the vectors comprising the diagonal elements of $\widetilde{\mathbf{w}}(i)\widetilde{\mathbf{w}}^{\mathsf{T}}(i)$ and $\mathbf{P}^2(i)$, respectively. In the traditional proportional algorithm analysis [42], [43], [44], [45], the correlation between the proportional matrix and the weight error is ignored. Analogously, this approximation is also used in (25).

By taking the expected value of both sides of the proportionality function (17) in the ABy-PNLMS, we get

$$\begin{aligned}\mathbb{E}\{p_\ell(i)\}=&h_\ell(i)\mathbb{E}\{g_1(i)\big||w_\ell(i)|\le\zeta\}\\&+(1-h_\ell(i))\mathbb{E}\{g_2(i)\big||w_\ell(i)|>\zeta\},\end{aligned} \tag{26}$$

where $h_\ell(i)$ represents the probability of $|w_\ell(i)|\le\zeta$, and

$$g_1(i)=\frac{1-\gamma}{2L}+\frac{(1+\gamma)a_1(i)}{2\|\mathbf{w}(i)\|_1+\epsilon}, \tag{27}$$

$$g_2(i)=\frac{1-g_1(i)N(i)}{L-N(i)}, \tag{28}$$

with $N(i)$ being $|\mathcal{N}_1(i)|$ at time i. In (28), it is calculated by considering the fact that $N(i)g_1(i)+(L-N(i))g_2(i)=1$. If the weight $w_\ell(i)$ is assumed to follow a Gaussian distribution with mean $\mathbb{E}\{w_\ell(i)\}$ and variance $\mathbb{E}\{(w_\ell(i)-\mathbb{E}\{w_\ell(i)\})^2\}$ [46], at steady-state, the expected terms in (26) can be calculated as shown in Appendix B.

*Remark 4 (**Step-size**):* Ignoring the dependence between $p_\ell(i)$ and $\widetilde{\mathbf{w}}_\ell(i)$ in (22), it can be seen that the algorithm is convergent in the mean if the step-size meets $\rho(\boldsymbol{I}-\mu\mathbb{E}\{\mathbf{P}(i)\})<1$. Note that $\mathbf{P}(i)$ is positive definite and $\mathrm{Tr}(\mathbf{P}(i))=1$ as well as $0\le g_1(i)\le g_2(i)$. The step-size condition then requires $0<\mu<2/g_2(i)$. In contrast to the NLMS [47], $\mathbb{E}\{\widetilde{w}_\ell(i+1)\}=(1-\mu\frac{1}{L})\mathbb{E}\{\widetilde{w}_\ell(i)\}$, we observe that using a gain $g_\ell(i)$ larger than $1/L$, the proportional algorithm speeds up the convergence of some large weights while slowing down others.

*Remark 5 (**Steady-state MSD**):* From (25), it is not difficult to approximate the steady-state MSD:

$$\begin{aligned}\mathrm{MSD}&\triangleq\lim_{i\to\infty}\|\mathbb{E}\{\mathbf{s}(i)\}\|_1\\&\approx\frac{\mu\frac{\sigma_v^2}{\sigma_u^2}\sum_{\ell=1}^{L}\frac{\mathbb{E}\{p_\ell^2(\infty)\}}{\mathbb{E}\{p_\ell(\infty)\}-\mu\mathbb{E}\{p_\ell^2(\infty)\}}}{2-\mu\sum_{\ell=1}^{L}\frac{\mathbb{E}\{p_\ell^2(\infty)\}}{\mathbb{E}\{p_\ell(\infty)\}-\mu\mathbb{E}\{p_\ell^2(\infty)\}}}\\&\overset{\mu\ll1}{\approx}\frac{\mu\sigma_v^2}{2\sigma_u^2}\sum_{\ell=1}^{L}\frac{\mathbb{E}\{p_\ell^2(\infty)\}}{\mathbb{E}\{p_\ell(\infty)\}}.\end{aligned} \tag{29}$$

Note that $\mathbb{E}\{x^2\}\ge\mathbb{E}^2\{x\}$ for any random variable x. Thus, from (29), we know

$$\mathrm{MSD}\ge\frac{\mu\sigma_v^2}{2\sigma_u^2}\sum_{\ell=1}^{L}\mathbb{E}\{p_\ell(\infty)\}=\frac{\mu\sigma_v^2}{2\sigma_u^2} \tag{30}$$

TABLE III
COMPARISONS OF COMPUTATIONAL COMPLEXITY PER ITERATION

Algorithm	$\times/\div$	$+/-$	Log.	$\sqrt{\cdot}$	Comparison
NLMS	$3L$	$3L+1$	0	0	0
IPNLMS	$5L$	$5L+2$	0	0	0
IAF-PNLMS	$5L+3$	$4L$	0	0	L
EIAF-PNLMS	$5L+7$	$4L+2$	0	0	$L+1$
B-IPNLMS	$5L+M+6$	$4L+M+2$	0	M	0
MPNLMS	$6L+5$	$5L-1$	L	0	$2L-1$
TG-PNLMS	$4L+5$	$4L+1$	0	0	$3L-1$
ABx-PNLMS	$5L+M+6$	$4L+M+2$	0	M	$O(L\log_2 L)$
ABy-PNLMS	$4L+17$	$4L+\lvert\mathcal{N}_1(i)\rvert+7$	0	0	$L+2$
ABx-DPNLMS	$(5+K)L+M+4K+7$	$(4+K)L+M+4K+2$	0	M	$O(L\log_2 L)$
ABy-DPNLMS	$(4+K)L+4K+18$	$(4+K)L+\lvert\mathcal{N}_1(i)\rvert+4K+7$	0	0	$L+2$

under small step-size condition. This result indicates that the proportional NLMS algorithm may offer larger steady-state MSD than the NLMS algorithm with the same step-size. If at steady-sate $\lambda(i)\approx 0$ for small step-size in the ABy-PNLMS algorithm, it will lead to $|w_\ell(i)|>\zeta$ and $g_\ell(i)\approx\frac{1}{L}$ for almost all ℓ. In this case, the algorithm would achieve the same steady-state MSD value as the NLMS algorithm.

V. ACCELERATED CONVERGENCE METHODS ON COLORED INPUTS

It is well known that proportional NLMS suffers from slow convergence in the case of highly correlated input. To address this problem, this section designs a proportional decorrelation NLMS algorithm, using the previously designed proportional strategies.

Defining an estimated decorrelation parameter vector[4] $\widehat{\mathbf{a}}(i)=[\widehat{a}_1(i),\widehat{a}_2(i),\ldots,\widehat{a}_K(i)]^{\mathsf{T}}$ with $K>0$, the decorrelated-output $\overline{d}(i)$ and decorrelated-input vector $\overline{\mathbf{u}}(i)$ are constructed as [37], [38]

$$\overline{d}(i)=d(i)-\lambda(i)\mathbf{d}^{\mathsf{T}}(i)\widehat{\mathbf{a}}(i), \tag{32}$$

$$\overline{\mathbf{u}}(i)=\mathbf{u}(i)-\lambda(i)\mathbf{X}_u(i)\widehat{\mathbf{a}}(i), \tag{33}$$

where the separated-decorrelation parameter $\lambda(i)$ is defined as in (19), and

$$\mathbf{X}_u(i)=[\mathbf{u}(i-1),\mathbf{u}(i-2),\ldots,\mathbf{u}(i-K)], \tag{34}$$

$$\mathbf{d}(i)=[d(i-1),d(i-2),\ldots,d(i-K)]^{\mathsf{T}}. \tag{35}$$

After solving the cost function below using the Lagrange multiplier method

$$\min_{\mathbf{w}(i+1)}\|\mathbf{w}(i+1)-\mathbf{w}(i)\|^2_{\mathbf{P}(i)} \tag{36a}$$

s.t.

$$\overline{d}(i)=\overline{\mathbf{u}}^{\mathsf{T}}(i)\mathbf{w}(i+1), \tag{36b}$$

[4]Here, the vector is estimated online using the update formula:

$$\widehat{\mathbf{a}}(i+1)=\widehat{\mathbf{a}}(i)+\frac{\mu_a}{\|\mathbf{u}'(i-1)\|^2+\epsilon}\mathbf{u}'(i-1)(u(i)-\mathbf{u}'^{\mathsf{T}}(i-1)\widehat{\mathbf{a}}(i)), \tag{31}$$

where $0<\mu_a<1$ is the step-size and $\mathbf{u}'(i-1)=[u(i-1),u(i-2),\ldots,u(i-K)]^{\mathsf{T}}$. We generally choose K between 10 and 20.

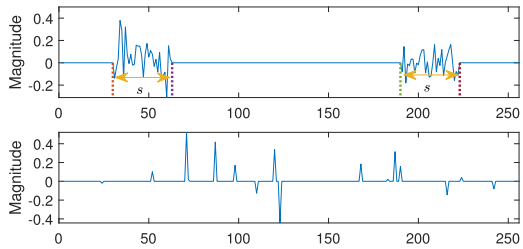


Fig. 2. Two types of sparse channels, where s represents the number of non-zeros in the block. (*Top*) A two-cluster block-sparse channel; (*Bottom*) A general sparse channel.

we obtain the proportional DNLMS algorithm

$$\mathbf{w}(i+1)=\mathbf{w}(i)+\mu\frac{\mathbf{P}(i)\overline{\mathbf{u}}(i)}{\overline{\mathbf{u}}^{\mathsf{T}}(i)\mathbf{P}(i)\overline{\mathbf{u}}(i)+\epsilon}\overline{e}(i), \tag{37}$$

where μ is introduced to enhance the robustness of the algorithm and $\overline{e}(i)=\overline{d}(i)-\overline{\mathbf{u}}^{\mathsf{T}}(i)\mathbf{w}(i)$. As a result of nesting different proportional matrices as described in Sections III-A and III-B, we can obtain the ABx-DPNLMS and ABy-DPNLMS methods, respectively.

*Remark 6 (**Discussion on Complexity**):* Table III summarizes the computational complexity of the NLMS, IPNLMS, IAF-PNLMS, enhanced IAF-PNLMS (EIAF-PNLMS) [16], B-IPNLMS, TG-NLMS, MPNLMS, ABx-PNLMS, ABy-PNLMS, ABx-DPNLMS and ABy-DPNLMS algorithms, in terms of the total number of additions, multiplications, square-roots and comparisons per input sample. It can be seen that, aside from the sorting operation, both ABx-PNLMS and B-IPNLMS require the same number of multiplications and additions. In comparison with the IPNLMS, IAF-PNLMS, EIAF-PNLMS, B-IPNLMS and ABx-PNLMS, the ABy-PNLMS algorithm is less complex.

*Remark 7 (**Extensions of DPNLMS Algorithm**):* As an alternative to (31), following the same procedures as in [37], the periodic recursive-least-squares can be used to devise a low-complexity implementation of the ABx-DPNLMS and ABy-DPNLMS.

VI. MONTE-CARLO SIMULATION STUDY

In all simulations, the initial weight vectors of all adaptive algorithms are set to zero. The measurement noise is a zero-mean

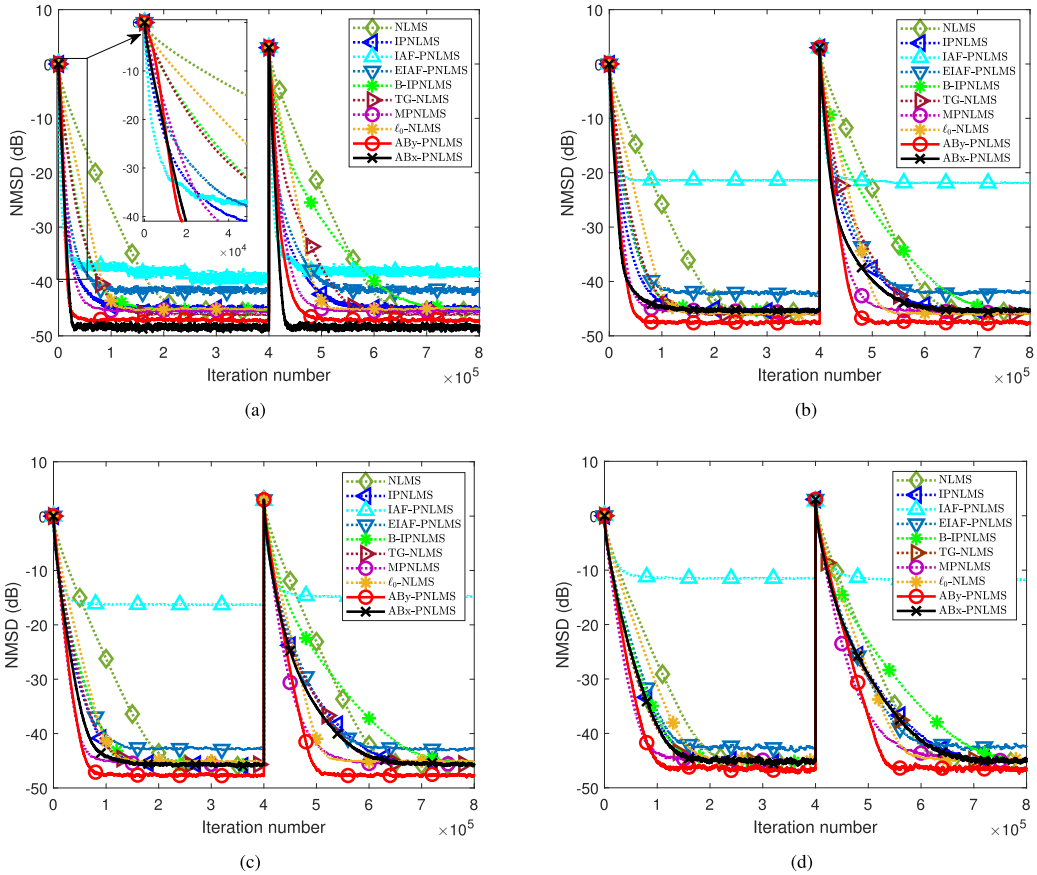


Fig. 3. NMSD learning curves for the NLMS, IPNLMS, IAF-PNLMS, EIAF-PNLMS, B-IPNLMS, TG-NLMS, MPNLMS, ℓ_0-NLMS, ABy-PNLMS, and ABx-PNLMS algorithms using colored input signal with two-cluster block-sparse channels. (a) $s = 3$ and $\xi = 0.93$; (b) $s = 8$ and $\xi = 0.85$; (c) $s = 16$ and $\xi = 0.76$; (d) $s = 32$ and $\xi = 0.64$.

white Gaussian process. The signal-to-noise ratio (SNR) in dB and the degree of sparseness for an impulse response are defined as [15]

$$\text{SNR} = 10\log_{10}\left(\frac{\mathbb{E}\left\{(\mathbf{x}^{\mathsf{T}}(n)\mathbf{w}_{\text{opt}})^2\right\}}{\sigma_v^2}\right), \tag{38}$$

$$\xi = \frac{L}{L-\sqrt{L}}\left(1 - \frac{\|\mathbf{w}_{\text{opt}}\|_1}{\sqrt{L}\|\mathbf{w}_{\text{opt}}\|_2}\right). \tag{39}$$

The performance of the algorithms is measured by normalized MSD (NMSD) curves, calculated as

$$\text{NMSD}(i) = \frac{\|\mathbf{w}(i) - \mathbf{w}_{\text{opt}}\|_2^2}{\|\mathbf{w}_{\text{opt}}\|_2^2}. \tag{40}$$

A. Evaluation of ABx-PNLMS and ABy-PNLMS Algorithms

In the first experiment, we compare the ABx-PNLMS and ABy-PNLMS methods with the NLMS, IPNLMS, IAF-PNLMS, EIAF-PNLMS, B-IPNLMS, TG-NLMS, MPNLMS and ℓ_0-NLMS[5] [17], [46] algorithms for block-sparse channels with different sparsity, general sparse channels as well as non-sparse channels. The block-sparse channels satisfy the structure illustrated in Fig. 2. The colored input is obtained by

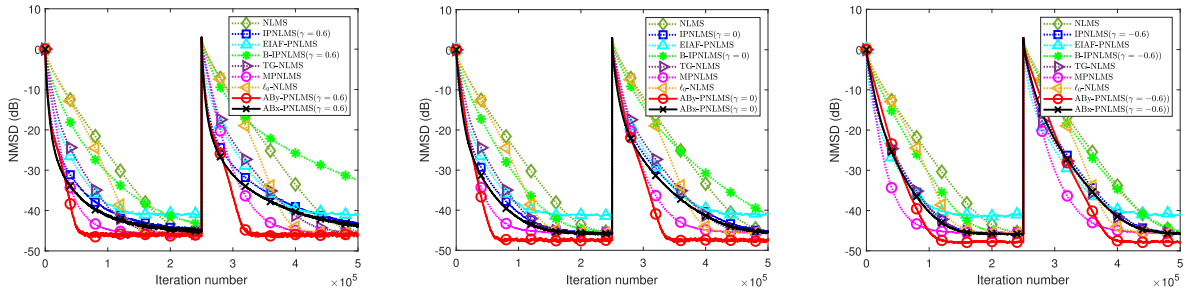
$$u(n) = au(n-1) + x(n), \tag{41}$$

where $x(n)$ is a zero-mean white Gaussian signal with variance 1, the parameter a represents the correlation index. The length of the filter is $L = 256$. The coefficients of $\mathbf{w}_{\text{opt}}$ are shifted to the right by 200 samples and multiplied by -1 at the middle of the iterations to test the tracking ability of the algorithms. All learning curves are the average of 50 independent Monte-Carlo runs.

Fig. 3 displays the NMSD learning curves for block-sparse channels. The unknown non-zero taps are generated randomly. The input is colored with $a = 0.8$. Observation noise is a white Gaussian signal with an SNR of 30 dB. The parameters of the ABx-PNLMS and ABy-PNLMS algorithms are $\mu = 0.05$, $\beta = 0.9995$, $P = 16$, $\epsilon = 0.001$, $\gamma = 0$ and $\kappa = 0.2$. The algorithmic parameters are $\mu = 0.05$, $P = 16$, $\epsilon = 0.001$, $\gamma = 0$, $K = 3 \times 10^{-3}$, $\tau = 0.05$ and $R = 4$ for the NLMS, IPNLMS, IAF-PNLMS, EIAF-PNLMS, B-IPNLMS, TG-NLMS and MPNLMS algorithms; $\mu = 0.09$ (Fig. 3(a), (b) and (c)), $\mu = 0.06$ (Fig. 3(d)), $\beta_z = 5$, $Q = 4$, $\kappa_z = 1 \times 10^{-7}$ (Fig. 3(a)), 2×10^{-8} (Fig. 3(b)), 1×10^{-8} (Fig. 3(c)) and 1×10^{-9} (Fig. 3(d)) for the ℓ_0-NLMS.

Fig. 4 shows the NMSD learning curves for general sparse channels with different γ. The number of non-zero elements

[5]Here, the partial updating version is used to reduce computational complexity.



(a) (b) (c)

Fig. 4. NMSD learning curves for the NLMS, IPNLMS, EIAF-PNLMS, B-IPNLMS, TG-NLMS, MPNLMS, ℓ_0-NLMS, ABy-PNLMS, and ABx-PNLMS algorithms with a general sparse channel. (a) $\gamma = 0.6$; (b) $\gamma = 0$; (c) $\gamma = -0.6$.

TABLE IV
COMPARISONS OF COMPUTATIONAL COMPLEXITY PER ITERATION

Algorithm (Fig. 4)	$\times/\div$	$+/-$	time (second)
NLMS	768	769	3.20×10^{-6}
IPNLMS	1280	1282	1.58×10^{-4}
EIAF-PNLMS	1287	1026	1.59×10^{-4}
B-IPNLMS	1302	1042	1.78×10^{-4}
TG-NLMS	1029	1025	1.59×10^{-4}
MPNLMS	1541	1279	1.77×10^{-4}
ℓ_0-NLMS	775	772	5.79×10^{-6}
ABx-PNLMS	1302	1042	1.92×10^{-4}
ABy-PNLMS	1041	1287	1.63×10^{-4}

is 16 and the degree of sparseness $\xi = 0.85$. The position of non-zero taps is randomly selected in every Monte-Carlo run. The input is also colored with $a = 0.8$. Compared with Fig. 3, the changed algorithm parameters are $\kappa = 1$ for the ABy-PNLMS; $\mu = 0.05$ and $\kappa_z = 6.2 \times 10^{-8}$ for the ℓ_0-NLMS.

As we can see from Figs. 3 and 4, the ABy-PNLMS algorithm has faster convergence rate than the NLMS, IPNLMS, EIAF-PNLMS, B-IPNLMS, TG-NLMS and ℓ_0-NLMS methods both for block-sparse and general sparse channels. When identifying strong sparse channels (see Fig. 3(a)), the performance of ABx-PNLMS outperforms ABy-PNLMS, in terms of convergence speed and steady-state error. In Fig. 3(b), (c) and (d), the ABy-PNLMS algorithm, as compared with the MPNLMS, has lower steady-state NMSDs while maintaining similar convergence and tracking rates. Tabel IV lists the detailed addition and multiplication numbers for each algorithm. The average running time was tested using matlab R2020a on a desktop computer with Intel(R) Core(TM) i5-8400 CPU @2.80 GHz.

Fig. 5 plots the NMSD learning curves for non-sparse channels, in which the unknown tap coefficients are produced randomly. The parameters are the same as those in Fig. 3. In this scenario, the IAF-PNLMS, IPNLMS, TG-NLMS, MPNLMS and ABx-PNLMS perform worse than NLMS, and other algorithms perform similarly to NLMS.

B. Echo Channel Identification Application

In the second experiment, the performance of the proposed methods are evaluated in an echo channel identification application for colored Gaussian and speech input signals. Three different echo channels are used; they are shown in Fig. 6. The

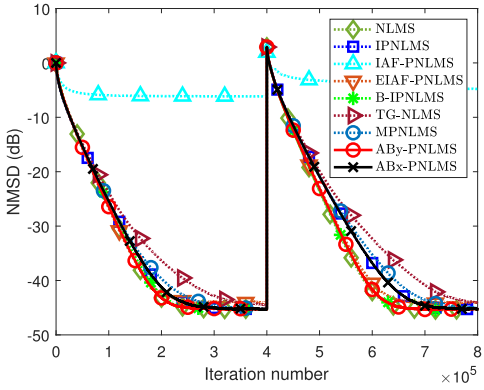


Fig. 5. NMSD learning curves for the NLMS, IPNLMS, IAF-PNLMS, EIAF-PNLMS, B-IPNLMS, TG-NLMS, MPNLMS, ABy-PNLMS, and ABx-PNLMS algorithms for non-sparse channels.

step-size of ℓ_0-NLMS is 0.12, and that of other algorithms is 0.1. Other parameter settings are $\beta = 0.9995$, $P = 64$, $\epsilon = 0.001$, $\gamma = 0$ and $\kappa = 1$ for the ABx-PNLMS and ABy-PNLMS algorithms; $P = 64, \epsilon = 0.001, \gamma = 0$ for the B-IPNLMS algorithm; $\tau = 0.05$ and $R = 4$ for the TG-NLMS; $\beta_z = 5$, $Q = 4$ and $\kappa_z = 1 \times 10^{-10}$ for the ℓ_0-NLMS; $\beta = 0.99$, $K = 3 \times 10^{-3}$ and initial activation factor 10^{-4} for the IAF-PNLMS and EIAF-PNLMS algorithms; $\mu_a = 0.1$, $K = 15$ for the ABx-DPNLMS and ABy-DPNLMS algorithms.

1) Using Colored Gaussian Input: The input signal is obtained from (41) with $a = 0.9$. The measurement noise is zero-mean independent and identically distributed (i.i.d.) with SNR = 40 dB. Fig. 7(a) illustrates the MSD curves.

2) Using Speech Input: As a comparison, Fig. 7(b) shows the NMSD results with a real speech signal. The speech sampled by 8kHz is also plotted in Fig. 7(b).

C. Evaluation of Theoretical Steady-State MSD

In the third experiment, the theoretical steady-state MSD is tested by simulations. The length of the filter is $L = 256$. The non-zero taps are $0.6022, -0.0661, -0.0902, 0.2643, -0.1369, 0.1002, 0.2640, 0.2120, 0.2573, 0.4557, -0.0917, -0.2756, -0.1756, 0.0296, -0.0519, -0.1377$ and their locations are randomly chosen. The input is a white Gaussian signal with mean 0 and variance 1. The variance of the noise is 0.001.

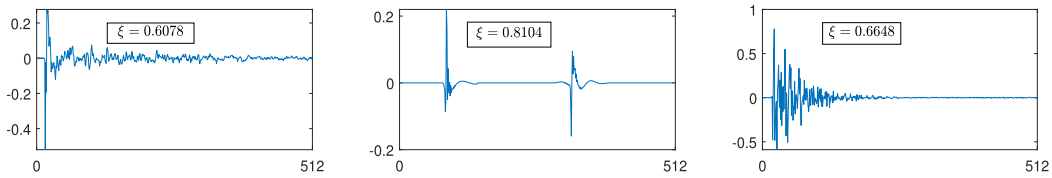


Fig. 6. Three echo channels with different sparsity. (*Left*) Echo channel 1 [48]; (*Middle*) Echo channel 2 [49]; (*Right*) Echo channel 3 [50].

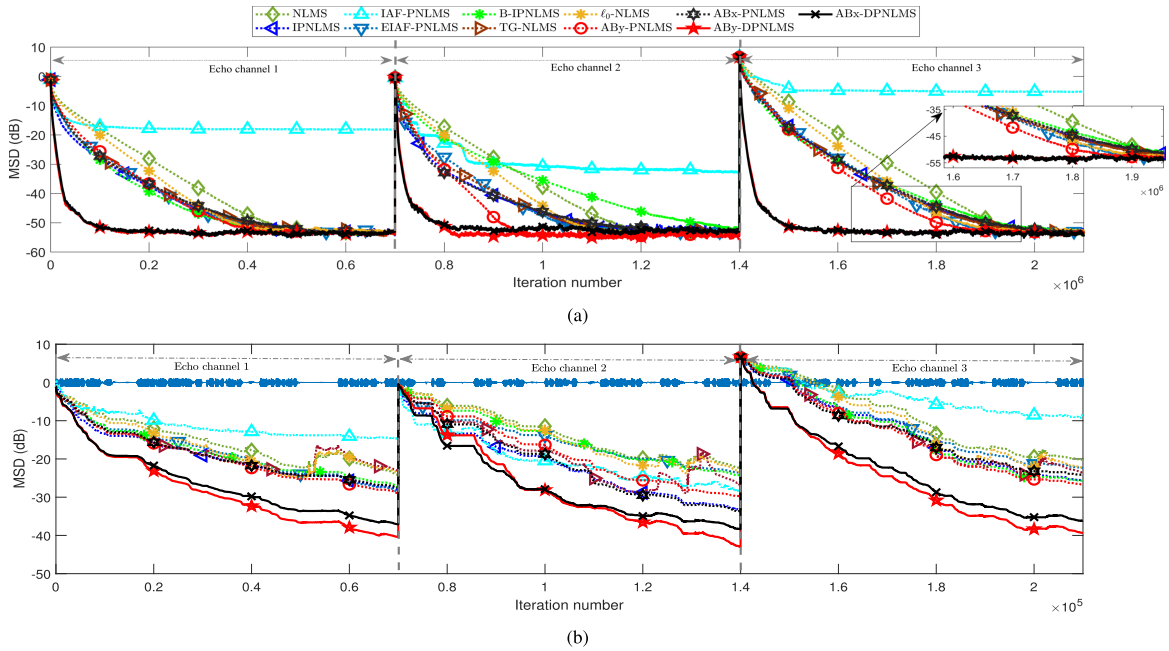


Fig. 7. MSD learning curves for the NLMS, IPNLMS, IAF-PNLMS, EIAF-PNLMS, B-IPNLMS, TG-NLMS, ℓ_0-NLMS, ABy-PNLMS, ABx-PNLMS, ABy-DPNLMS, and ABx-DPNLMS algorithms (single trial). (a) Gaussian input; (b) Speech input.

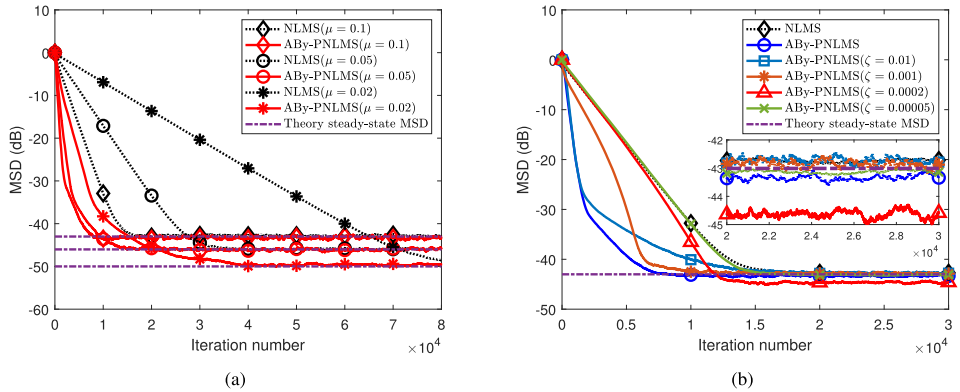


Fig. 8. MSD learning curves of the ABy-PNLMS algorithm with different step-sizes and ζ. (a) Using time-varying ζ; (b) Using fixed ζ and $\mu = 0.1$.

Fig. 8(a) compares simulated MSD and theoretical steady-state curves of the ABy-PNLMS with respect to three different step-sizes ($\mu = 0.1, 0.05, 0.02$). The threshold ζ is time-varying, i.e., using the update mechanism (18). The parameter settings are $\beta = 0.99$, $P = 16$, $\epsilon = 0.001$, $\gamma = 0$ and $\kappa = 1$. We can observe that in this case the theoretical steady-state results are in accordance with the simulated ones.

Fig. 8(b) draws the simulated MSD curves with fixed ζ but different values. It can be seen that our theoretical results predict the steady-state results of the algorithm well for $\zeta = 0.01, 0.001, 0.00005$. However, when using $\zeta = 0.0002$, in comparison to the NLMS algorithm, the ABy-PNLMS algorithm achieves lower steady-state MSD. A small gap exists between the simulated steady-state MSD and our theoretical value. This

may be due to the fact that the independence assumption between the small tap and the corresponding proportional gain is invalid. A brief discussion is provided in Appendix C. We will attempt to develop new analytical methods in the future to analyze algorithms in light of this fact.

VII. Conclusion

In this article, we designed the ABx-PNLMS and ABy-PNLMS algorithms. In these two algorithms, sorting and adaptive dynamic partitioning are used. The partitioning prevents that both large and small taps are assigned to the same block. The mean and mean-square behavior of ABy-PNLMS is performed. Additionally, using an adaptive separated-decorrelation method, the ABx-DPNLMS and ABy-DPNLMS have been developed to accelerate convergence for strongly correlated input signals. The simulation results validate the effectiveness of our algorithms. Developing new analytical methods for analyzing proportional algorithms more effectively will be part of future efforts.

Appendix A

The solution process follows the iterative reweighting framework in [9]. Define $\mathbf{Q}\mathbf{h}_s = \mathbf{w}_s$. Based on (10), we minimize the following objective function w.r.t. $\mathbf{h}_s$,

$$\min_{\mathbf{h}_s} \mathbb{E}\{(d(i) - \mathbf{u}_s^{\mathsf{T}}(i)\mathbf{Q}\mathbf{h}_s)^2\} + \rho G(\mathbf{h}_s), \tag{42}$$

where $G(\mathbf{h}_s) = \sum_{\ell=1}^{M} \|\mathbf{h}_{([\ell])}\|_2$ is a block sparsity-promoting regularization and $\mathbf{h}_s = [\mathbf{h}_{([1])}^{\mathsf{T}}, \cdots, \mathbf{h}_{([M])}^{\mathsf{T}}]^{\mathsf{T}}$.

Using the least-mean-square scheme, we solve (42) iteratively, resulting in

$$\mathbf{h}_s(i+1|i) = \mathbf{h}_s(i) + \mu'\mathbf{Q}\mathbf{u}_s(i)e(i) - \mu'\rho\Phi(i)\mathbf{h}_s(i), \tag{43}$$

where

$$\Phi(i) = \mathrm{diag}\left(\left[\frac{1}{\|\mathbf{h}_{([1])}(i)\|_2}\mathbb{1}_P, \ldots, \frac{1}{\|\mathbf{h}_{([M])}(i)\|_2}\mathbb{1}_P\right]\right) \tag{44}$$

and $\mathbf{h}_s(i) = [\mathbf{h}_{([1])}^{\mathsf{T}}(i), \ldots, \mathbf{h}_{([M])}^{\mathsf{T}}(i)]^{\mathsf{T}}$.

By premultiplying $\mathbf{Q}$ on both sides of (43) and using

$$\mu' = \mu\frac{1}{\mathbf{u}_s^{\mathsf{T}}(i)\mathbf{Q}^2\mathbf{u}_s(i)}, \tag{45}$$

we get

$$\mathbf{w}_s(i+1|i) = \mathbf{w}_s(i) + \mu\frac{\mathbf{Q}^2\mathbf{u}_s(i)e(i) - \rho\Phi(i)\mathbf{w}_s(i)}{\mathbf{u}_s^{\mathsf{T}}(i)\mathbf{Q}^2\mathbf{u}_s(i)}. \tag{46}$$

To avoid the computation of $\mathbf{h}_{([m])}(i)$, $m \in 1, \ldots, M$ in (44), proceeding as in [9], we set $\rho = 0$ in (46). While maintaining the stability of the proportional matrix, further replacing $\mathbf{Q}^2$ with its normalized version $\mathbf{P}_s = \mathbf{Q}^2/\mathrm{Tr}(\mathbf{Q}^2)$ in (46) leads to

$$\mathbf{w}_s(i+1|i) = \mathbf{w}_s(i) + \mu\frac{\mathbf{P}_s\mathbf{u}_s(i)}{\mathbf{u}_s^{\mathsf{T}}(i)\mathbf{P}_s\mathbf{u}_s(i)}e(i). \tag{47}$$

Applying the reweighting ℓ_2 method [9], [51], we obtain the m-th block diagonal element update of $\mathbf{Q}^2$:

$$q_m^2 = \|\mathbf{w}_{([m])}(i)\|_2,\ m = 1, 2, \ldots, M \tag{48}$$

To further embed the improved proportional scheme [11], [24], we have the proportional gain (12).

Appendix B

When the weight $w_\ell(i)$ obeys a Gaussian distribution, we calculate $h_\ell(i)$ as

$$h_\ell(i) = \int_{-\zeta}^{\zeta} f(w_\ell(i))dw_\ell(i), \tag{49}$$

where $f(\cdot)$ is the probability density function. Other expected terms in (26) can be approximated as

$$\mathbb{E}\{g_1(i)\big||w_\ell(i)| \le \zeta\} \approx \frac{1-\gamma}{2L} + \frac{(1+\gamma)\mathbb{E}\{a_1(i)\}}{2\|\mathbb{E}\{\mathbf{w}(i)\}\|_1 + \epsilon}, \tag{50}$$

$$\mathbb{E}\{g_2(i)\big||w_\ell(i)| > \zeta\} \approx \frac{1 - \mathbb{E}\{g_1(i)\}\mathbb{E}\{N(i)\}}{L - \mathbb{E}\{N(i)\}}, \tag{51}$$

where

$$\begin{aligned}\mathbb{E}\{a_1(i)\} &= \mathbb{E}\left\{\frac{1}{N(i)}\sum_{\ell\in\mathcal{N}_1(i)}|w_\ell(i)|\right\} \\ &\approx \frac{1}{\mathbb{E}\{N(i)\}}\mathbb{E}\left\{\sum_{\ell\in\mathcal{N}_1(i)}|w_\ell(i)|\right\}.\end{aligned} \tag{52}$$

According to the optimal taps, we divide all taps into two parts: $Z = \{\ell|w_{\mathrm{opt},\ell} = 0\}$ and $NZ = \{\ell|w_{\mathrm{opt},\ell} \neq 0\}$. Suppose ζ is small enough to satisfy $0 < \zeta \ll \min\{|w_{\mathrm{opt},\ell}|, \ell \in NZ\}$ at steady-state. Then, $\mathcal{N}_1(i)$ in (52) will belong to a subset of Z. Then, $\mathbb{E}\{a_1(i)\} \approx \int_{-\zeta}^{\zeta}|w_\ell(i)|f(w_\ell(i))dw_\ell(i)$ and $\mathbb{E}\{N(i)\} = |Z|h$, with h being the probability of occurrence, also calculated by (49).

Appendix C

Here, we compute $\mathbb{E}\{p_\ell(i)s_\ell(i)\}, \ell \in Z$ at steady-state. Based on (25) and (26), we get

$$\begin{aligned}\sum_{\ell\in Z}\mathbb{E}\{p_\ell(i)s_\ell(i)\} \approx &\sum_{\ell\in Z}h_\ell(i)\mathbb{E}\{g_1(i)s_\ell(i)\} \\ &+\sum_{\ell\in Z}(1-h_\ell(i))\mathbb{E}\{g_2(i)s_\ell(i)\},\end{aligned} \tag{53}$$

In steady-state, the probability $h_\ell(i)$ is assumed to be the same for all $\ell \in Z$, i.e., $h_\ell(i) = h$. Substituting

$$g_1(i) = \frac{1-\gamma}{2L} + \frac{(1+\gamma)\frac{1}{N(i)}\sum_{\ell\in\mathcal{N}_1(i)}|w_\ell(i)|}{2\|\mathbf{w}(i)\|_1 + \epsilon} \tag{54}$$

and $g_2(i) = (1 - g_1(i)N(i))/(L - N(i))$ into (53), we have

$$\begin{aligned}\sum_{\ell\in Z}h_\ell(i)\mathbb{E}\{g_1(i)s_\ell(i)\} &= h\mathbb{E}\{g_1(i)\sum_{\ell\in Z}s_\ell(i)\} \\ &= h\frac{(1-\gamma)\mathbb{E}\{\sum_{\ell\in Z}s_\ell(i)\}}{2L}\end{aligned}$$

$$+h(1+\gamma)\mathbb{E}\left\{\frac{\frac{1}{N(i)}\sum_{m\in\mathcal{N}_1(i)}|w_m(i)|}{2\|\mathbf{w}(i)\|_1+\epsilon}\sum_{\ell\in Z}s_\ell(i)\right\}, \tag{55}$$

and

$$\sum_{\ell\in Z}(1-h_\ell(i))\mathbb{E}\{g_2(i)s_\ell(i)\} = (1-h)\mathbb{E}\left\{\frac{1-g_1(i)N(i)}{L-N(i)}\sum_{\ell\in Z}s_\ell(i)\right\}, \tag{56}$$

Since $\mathcal{N}_1(i)$ is a subset of Z (when $0<\zeta\ll\min\{|w_{\text{opt},\ell}|,\ell\in NZ\}$) at steady-state, we assume that $w_m(i)$, $m\in Z$ is independent and identically distributed. Then, we have

$$\mathbb{E}\left\{\frac{1}{N(i)}\sum_{m\in\mathcal{N}_1(i)}|w_m(i)|\sum_{\ell\in Z}s_\ell(i)\right\} = \mathbb{E}\left\{\frac{1}{N(i)}\sum_{m\in\mathcal{N}_1(i)}|\widetilde{w}_m(i)|\sum_{\ell\in Z}s_\ell(i)\right\} \tag{57}$$

that is not equal to $\mathbb{E}\{\frac{1}{N(i)}\sum_{m\in\mathcal{N}_1(i)}|\widetilde{w}_m(i)|\}\mathbb{E}\{\sum_{\ell\in Z}s_\ell(i)\}$. Thus, the correlation between the tap and corresponding proportional matrix exists for zero taps.

REFERENCES

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Upper Saddle River, NJ, USA: Pearson, 1985.

[2] S. Haykin, *Adaptive Filter Theory*, 2nd ed.Hoboken, NJ, USA: Prentice Hall, 1991.

[3] A. H. Sayed, *Adaptive Filters*. Hoboken, NJ, USA: Wiley, 2008.

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

[5] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech, Audio Process*., vol. 8, pp. 508–518, Sep. 2000.

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

[7] P. S. Diniz, H. Yazdanpanah, and M. V. Lima, "Feature LMS algorithms," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process*., Calgary, Alberta, Canada, Apr. 2018, pp. 4144–4148.

[8] H. Yazdanpanah, P. S. R. Diniz, and M. V. S. Lima, "Feature adaptive filtering: Exploiting hidden sparsity," *IEEE Trans. Circuits Syst. I., Reg. Papers*, vol. 67, no. 7, pp. 2358–2371, Jul. 2020.

[9] C. H. Lee, B. D. Rao, and Harinath Garudadri, "Proportionate adaptive filtering algorithms derived using an iterative reweighting framework," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 171–186, 2020.

[10] S. L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," in *Proc. IEEE 32nd Asilomar Conf. Signals, Syst. Comput*., Monterey, CA, Nov. 1998, pp. 394–398.

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

[12] H. Deng and M. Doroslovacki, "Proportionate adaptive algorithms for network echo cancellation," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1794–1803, May 2006.

[13] J. Yang and G. E. Sobelman, "A gradient-controlled proportionate technique for acoustic echo cancellation," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, 2013, pp. 1941–1945.

[14] P. Loganathan, A. W. H. Khong, and P. A. Naylor, "A class of sparseness-controlled algorithms for echo cancellation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 17, no. 8, pp. 1591–1601, Nov. 2009.

[15] F. C. de Souza, O. J. Tobias, R. Seara, and D. R. Morgan, "A PNLMS algorithm with individual activation factors," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2036–2047, Apr. 2010.

[16] F. das Chargas de Souza, R. Seara, and D. R. Morgan, "An enhanced IAF-PNLMS adaptive algorithm for sparse impulse response identification," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3301–3307, Jun. 2012.

[17] Y. Gu, J. Jin, and S. Mei, "ℓ_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Process. Lett.*, vol. 16, no. 9, Sep. 2009, pp. 774–777.

[18] L. Luo and A. Xie, "Steady-state mean-square deviation analysis of improved ℓ_0-norm-constraint LMS algorithm for sparse system identification," *Signal Process.*, vol. 175, Oct. 2020, Art. no. 107658.

[19] J. Meng et al., "A reweighted ℓ_0-norm-constraint LMS algorithm for sparse system identification," *Digit. Signal Process.*, vol. 123, Apr. 2022, Art. no. 103456.

[20] S. Chouvardas, K. Slavakis, Y. Kopsinis, and S. Theodoridis, "A sparsity-promoting adaptive algorithm for distributed learning," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5412–5425, Oct. 2012.

[21] S. Vlaski, L. Vandenberghe, and A. H. Sayed, "Regularized diffusion adaptation via conjugate smoothing," *IEEE Tran. Autom. Control*, vol. 67, no. 5, pp. 2343–2358, May 2022.

[22] P. D. Lorenzo and A. H. Sayed, "Sparse distributed learning based on diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1419–1433, Mar. 2013.

[23] A. G. Rusu, S. Ciochină, C. Paleologu, and J. Benesty, "An optimized differential step-size LMS algorithm," *Algorithms*, vol. 12, no. 8, Jul. 2019, Art. no. 147.

[24] J. Liu and S. L. Grant, "Proportionate adaptive filtering for block-sparse system identification," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 4, pp. 623–630, Apr. 2016.

[25] S. Jiang and Y. Gu, "Block-sparsity-induced adaptive filter for multi-clustering system identification," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5318–5330, Oct. 2015.

[26] B. K. Das, A. Mukherjee, and M. Chakraborty, "Block-sparsity-induced system identification using efficient adaptive filtering," in *Proc. IEEE Nat. Conf. Commun.*, Kharagpur, India, 2020, pp. 1–6.

[27] W. Wang and H. Zhao, "A novel block-sparse proportionate NLMS algorithm based on the $\ell_{2,0}$ norm," *Signal Process.*, vol. 176, Nov. 2020, Art. no. 107671.

[28] C. Paleologu, S. Ciochina, and J. Benesty, "An efficient proportionate affine projection algorithm for echo cancellation," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 165–168, Feb. 2010.

[29] Y. Zong and J. Ni, "Cluster-sparsity-induced affine projection algorithm and its variable step-size version," *Digit. Signal Process.*, vol. 195, Jun. 2022, Art. no. 108490.

[30] S. Zhang, F. Zhang, H. Chen, R. Merched, and A. H. Sayed, "Frequency-domain diffusion adaptation over networks," *IEEE Trans. Signal Process.*, vol. 69, pp. 5419–5430, Aug. 2021.

[31] R. Ahmad, A. W. Khong, and P. A. Naylor, "Proportionate frequency domain adaptive algorithms for blind channel identification," *Proc. IEEE ICASSP*, Toulouse, France, 2006, vol. 5, pp. V29–V32.

[32] B. N. M. Laska, R. A. Goubran, and M. Bolic, "Improved proportionate subband NLMS for acoustic echo cancellation in changing environments," *IEEE Signal Process. Lett.*, vol. 15, pp. 337–340, Mar. 2008.

[33] S. Zhang and W. X. Zheng, "Mean-square analysis of multi-sampled multiband-structured subband filtering algorithm," *IEEE Trans. Circuits Syst. I., Reg. Papers*, vol. 66, no. 3, pp. 1051–1062, Mar. 2019.

[34] M. S. E. Abadi, "Proportionate normalized subband adaptive filter algorithms for sparse system identification," *Signal Process.*, vol. 89, no. 7, pp. 1467–1474, 2009.

[35] J. Ni, X. Chen, and J. Yang, "Two variants of the sign subband adaptive filter with improved convergence rate," *Signal Process.*, vol. 96, no. 5, pp. 325–331, 2014.

[36] S. Zhang, H. C. So, W. Mi, and H. Han, "A family of adaptive decorrelation NLMS algorithms and its diffusion version over adaptive networks," *IEEE Trans. Circuits Syst. I., Reg. Papers*, vol. 65, no. 2, pp. 638–649, Feb. 2018.

[37] S. Zhang, J. Zhang, and H. C. So, "Low-complexity decorrelation NLMS algorithms: Performance analysis and AEC application," *IEEE Trans. Signal Process.*, vol. 68, pp. 6621–6632, Nov. 2020.

[38] S. Zhang and W. Zheng, "Distributed separated-decorrelation LMS algorithms over sensor networks with noisy inputs," *IEEE Trans. Signal Process.*, vol. 68, pp. 4163–4177, 2020.

[39] C. Paleologu, J. Benesty, and S. Ciochina, "Study of the general Kalman filter for echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 8, pp. 1539–1549, Aug. 2013.

[40] F. Huang, J. Zhang, S. Zhang, H. Chen, and H. V. Poor, "Diffusion Bayesian subband adaptive filters for distributed estimation over sensor networks," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6909–6925, Oct. 2021.
[41] F. L. Perez, C. A. Pitz, and R. Seara, "A two-gain NLMS algorithm for sparse system identification," *Signal Process.*, vol. 200, Nov. 2022, Art. no. 108636.
[42] D. B. Haddad and M. R. Petraglia, "Transient and steady-state MSE analysis of the IMPNLMS algorithm," *Digit. Signal Process.*, vol. 33, pp. 50–59, Oct. 2014.
[43] E. V. Kuhn, F. d. C. de Souza, R. Seara, and D. R. Morgan, "On the steady-state analysis of PNLMS-type algorithms for correlated Gaussian input data," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1433–1437, Nov. 2014.
[44] Z. Qin, J. Tao, Y. Xia, and L. Yang, "Proportionate RLS with ℓ_1 norm regularization: Performance analysis and fast implementation," *Digit. Signal Process.*, vol. 122, Apr. 2022, Art. no. 103366.
[45] F. Huang, J. Zhang, and S. Zhang, "Complex-valued proportionate affine projection Versoria algorithms and their combined-step-size variants for sparse system identification under impulsive noises," *Digit. Signal Process.*, vol. 118, Nov. 2021, Art. no. 103209.
[46] S. Zhang and J. Zhang, "Transient analysis of zero attracting NLMS algorithm without Gaussian inputs assumption," *Signal Process.*, vol. 97, pp. 100–109, Apr. 2014.
[47] S. Zhang, J. Zhang, and H. C. So, "Mean square deviation analysis of LMS and NLMS algorithms with white reference inputs," *Signal Process.*, vol. 131, pp. 20–26, Feb. 2017.
[48] J.-H. Kim, J. Kim, J. H. Jeon, and S. W. Nam, "Delayless individual-weighting-factors sign subband adaptive filter with band-dependent variable step-sizes," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 7, pp. 1526–1534, Jul. 2017.
[49] C. Paleologu, J. Benesty, and S. Ciochina, "Linear system identification based on a kronecker product decomposition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 10, pp. 1793–1808, May 2018.
[50] F. Huang, J. Zhang, and S. Zhang, "Affine projection versoria algorithm for robust adaptive echo cancellation in hands-free voice communications," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11924–11935, Dec. 2018.
[51] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.

Sheng Zhang (Senior Member, IEEE) received the B.S. degree from the College of Mathematics and Information Sciences, North China University of Water Resources and Electric Power, Zhengzhou, China, in 2010, and the Ph.D. degree in signal and information processing from Southwest Jiaotong University, Chengdu, China, in 2016. From 2016 to 2017, he was a Research Fellow with the School of Computing, Engineering and Mathematics, Western Sydney University, Sydney, NSW, Australia. Since 2017, he has been with the School of Information Science and Technology, Southwest Jiaotong University, where he is currently an Associate Professor. His research interests include distributed adaptation and learning theories. He is currently an Associate Editor for *Circuits, Systems, and Signal Processing*. His homepage is at https://faculty.swjtu.edu.cn/ShengZHANG/en/.

Hongyang Chen (Senior Member, IEEE) received the B.S. and M.S. degrees from Southwest Jiaotong University, Chengdu, China, in 2003 and 2006, respectively, and the Ph.D. degree from The University of Tokyo, Tokyo, Japan, in 2011. From 2004 to 2006, he was a Research Assistant with the Institute of Computing Technology, Chinese Academy of Science, Beijing, China. In 2009, he was a Visiting Researcher with Adaptive Systems Laboratory, University of California, Los Angeles, CA, USA. From 2011 to 2020, he was a Researcher with Fujitsu Ltd., Tokyo. He is currently a Senior Research Expert with Zhejiang Lab, Hangzhou, China. He has authored and coauthored 100 refereed journal and conference papers in the *ACM Transactions on Sensor Networks*, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE MILCOM, IEEE GlOBECOM, and IEEE ICC, and has been granted or filed more than 50 PCT patents. His research interests include IoT, data-driven intelligent networking and systems, machine learning, localization, location-based Big Data, B5G, and statistical signal processing. He was the Symposium Chair or Special Session Organizer for some flagship conferences, including the IEEE PIMRC, IEEE MILCOM, IEEE GLOBECOM, and IEEE ICC. He was the recipient of the Best Paper Award from the IEEE PIMRC'09. He was the Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, Associate Editor for the IEEE COMMUNICATIONS LETTERS, and leading Guest Editor of the IEEE JOURNAL ON SELECTED TOPICS OF SIGNAL PROCESSING on tensor decomposition. He is currently an Associate Editor for the IEEE INTERNET OF THINGS JOURNAL. From 2021 to 2022, he was selected as the Distinguished Lecturer of the IEEE Communication Society. He is an Adjunct Professor with Hangzhou Institute for Advanced Study, The University of Chinese Academy of Sciences, China.

Ali H. Sayed (Fellow, IEEE) is the Dean of Engineering at EPFL, Switzerland, where he also leads the Adaptive Systems Laboratory (https://asl.epfl.ch/). He is a member of the US National Academy of Engineering and the World Academy of Sciences. He was the President of the IEEE Signal Processing Society in 2018 and 2019, respectively. His research interests include adaptation and learning theories, data and network sciences, statistical inference, and multi-agent systems. His work has been recognized with several major awards from IEEE and other professional societies, including more recently the 2022 IEEE Fourier Technical Field Award and the 2020 IEEE Norbert Wiener Society Award.