Mixed Norm Constrained Sparse APA Algorithm for Satellite and Network Echo Channel Estimation

YINGSONG LI1,2,3, (Member, IEEE), ZHENGXIONG JIYANG1, OMER MOHAMMED OMER OSMAN1, XIAO HAN3, and JINGWEI YIN3, (Member, IEEE)

1 College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China
2 Key Laboratory of Microwave Remote Sensing, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China
3 Acoustic Science and Technology Laboratory, Harbin Engineering University, Harbin 150001, China

Corresponding author: Yingsong Li (e-mail: liyingsong@ieee.org).

This paper is supported by the National Key Research and Development Program of China (2016YFE0111100), Key Research and Development Program of Heilongjiang (GX17A016), the Science and Technology innovative Talents Foundation of Harbin (2016RAXXJ084), China Postdoctoral Science Foundation (2017M620918), the Natural Science Foundation of China (61701366,51779061) and the Fundamental Research Funds for the Central Universities (HEUCFM180806), in part by the Opening Fund of Acoustics Science and Technology Laboratory (SSKF2016001).

ABSTRACT We propose an improved proportionate affine projection algorithm (PAPA), which is realized by integrating a hybrid-norm constraint into the affine projection algorithm (APA), to estimate cluster sparse signals that are happened in satellite and network echo channels. The proposed algorithm optimizes the hybrid $l_{2,0}$-norm of the filter coefficients, and the PAPA is a special case of the proposed method. Moreover, an enhanced PAPA is modified for cluster sparse channel estimations. Various simulation experiments are performed to verify that the proposed algorithms are superior to the APA, PAPA and related sparse algorithms with different input signals and various parameters.

INDEX TERMS Cluster sparse signal; $l_{2,0}$-norm; proportionate affine projection algorithm; sparse system identification

I. INTRODUCTION

Adaptive filtering has been widely studied and used for various engineering applications, such as system identification, channel estimation, active noise control and echo cancellation [1], [2]. The normalized least mean square (NLMS) algorithm has been extensively exploited owing to its low complexity and easy implementation [3]–[5]. However, the NLMS suffers from slow convergence for colored inputs [6]. Therefore, the data reusing method has been presented to create the affine projection algorithm (APA) to provide faster convergence when the inputs are colored [6]. Furthermore, natural signals may have sparse structure, and in this case, most of the signal model coefficients are zero or near-zero while only a few coefficients are active. Thus, sparse system identification has attracted substantial attention in the area of adaptive filtering [7]–[11].

In recent years, proportionate-type adaptive filtering has been widely developed and used for different sparse signal estimations, including channel estimations and echo cancellations [7]–[9]. The first proportionate adaptive filtering is the proportionate NLMS (PNLMS), which updates each filter coefficient by proportionately assigning a different adaptation step size to each estimated filter coefficient [8]. The PNLMS converges rapidly in the initial stage and provides good tracking ability in comparison with the NLMS [8], [9]. However, the PNLMS might achieve a worse estimation mean square error (MSE) when it converges. Several improved PNLMSs have been reported and analyzed based on various weighting rules, such as $\mu$-law and $\varepsilon$-law rules, which stimulate iterations for complete convergence [12]–[15]. Moreover, partition block (PB) [16] and block sparse (BS) [17] PNLMSs have been proposed to identify the active and inactive regions in the echo channel response. PB-PNLMS uses different algorithms to track the channels with two partitioned blocks, while BS-PNLMS identifies the block channel impulse response (CIR) using the block scheme of the $l_2$-norm and utilizes the sparseness of the $l_1$-norm penalty [16], [17]. In fact, the presented PB-IPNLMS uses two different adaptive algorithms to implement the time-domain partitioned blocks.

Similar to the NLMS, the PNLMS algorithm suffers from slow convergence when the inputs are colored. The proportionate APA (PAPA) and the improved PAPA (IPAPA) have
been presented to exploit the sparseness of the echo channel [18]. Furthermore, the memory-improved PAPA method, which is realized by making use of the memory of proportionate taps [19], [20], has been reported to accelerate the convergence rate and to reduce the computational burden. Several improved PAPA methods have been reported and analyzed [19]–[23]. In fact, the PAPA can be obtained from the basis pursuit and convex optimization [24], [25], which renders the PAPA more powerful for practical applications with respect to the modern hardware platform. The network echo path and the satellite channels are usually characterized by a bulk delay caused by the network encoding, loading and jitter buffer delays [25], [26]. Thus, the echoes of sparse systems are divided into two types, as shown in FIGURE 1. FIGURE 1(a) is a typical cluster sparse network channel with an active region in the channel ranging from 8 ms to 12 ms, and the other channel taps are close to zero and are denoted as inactive regions. Thus, the CIR of the network echo is usually a single-cluster sparse system, as presented in FIGURE 1(a). Satellite communication develops so fast that it has attracted increased attention for satellite networks. The CIR from the echoes of satellite-linked communication systems always contains active regions with long flat delays and dispersive active regions [17], [27], [28]. In fact, single-side modulated satellite communication links usually include large far-end echo that has a short delay and a small near-end echo that has a longer delay. Thus, the satellite echo path can be regarded as the two-cluster sparse channel shown in FIGURE 1(b).

To accurately estimate these sparse channels, the PAPA algorithm can be improved to further exploit the cluster sparse characteristics. In this paper, we propose a cluster sparse (CS) PAPA (CS-PAPA) and CS improved PAPA (CS-IPAPA) to make use of the cluster characteristics of the channels. The proposed CS-PAPA algorithm is realized by incorporating an $l_{2,0}$-norm into the cost function of the PAPA algorithm, where the $l_2$-norm divides the clusters and the $l_0$-norm aims to exploit the sparsity of the channels. The simulation results indicate that the proposed CS-PAPA and CS-IPAPA achieve faster convergence for handling cluster sparse signals under different inputs. Additionally, the basic APA and PAPA are special cases of the proposed cluster sparse PAPAs.

II. REVIEW OF THE PAPA AND IPAPA

In the classic adaptive filtering algorithm, the far-end signal is assumed to be $x(m) = [x(m), x(m - 1), x(m - 2), \ldots, x(m - K + 1)]^T$, where $m$ is the time index. The impulse response (IR) of the unknown system is $w(m) = [w_0(m), w_1(m), \ldots, w_{K-1}(m)]^T$, where the length of the coefficient vector is $K$. Then, the desired signal $d(m)$ containing the output of the echo channel and the near-end signal is given by [5]

$$d(m) = x^T(m)w(m) + u(m),$$

where $u(m)$ is the additive noise, which is assumed to be independent from the far-end signal $x(m)$, and $x^T(m)$ is the transpose operation of $x(m)$. The estimated IR vector is denoted to be $\hat{w}(m - 1) = [\hat{w}_0(m - 1), \hat{w}_1(m - 1), \ldots, \hat{w}_{K-1}(m - 1)]^T$. Therefore, the output of the filter is

$$y(m) = x^T(m)\hat{w}(m - 1).$$

Then, the estimation error is obtained

$$e(m) = d(m) - y(m).$$

A. PAPA

Inspired by the PNLM algorithm, PAPA introduces a data reusing scheme and proportionate scheme to reuse the input signals to improve the convergence speed and to exploit the sparsity. The input data matrix of the PAPA is defined as [18]

$$X(m) = [x(m), x(m - 1), \ldots, x(m - P + 1)],$$

where $P$ is the projection order of the PAPA. The output vector $y(m)$, the expected signal vector $d(m)$ and the estimation error vector $e(m)$ of the PAPA are given by

$$y(m) = X^T(m)\hat{w}(m - 1),$$

$$d(m) = [d(m), d(m - 1), \ldots, d(m - P + 1)]^T,$$

$$e(m) = d(m) - y(m).$$

The updated equation of the PAPA is described as [18]

$$\hat{w}(m) = \hat{w}(m - 1) + \mu G(m)X(m)e(m) \left(X^T(m)G(m)X(m) + \delta_{\text{PAPA}}I_P\right)^{-1},$$

where $\mu$ is the overall step size, $\delta_{\text{PAPA}}$ is a regularization parameter, $G(m)$ is the gain allocation matrix, which is a diagonal matrix, and $I_P$ is a $P \times P$ identity matrix. The proportionate matrix $G(m)$ is given by [7], [18]

$$G(m) = \text{diag}(g_0(m), g_1(m), \ldots, g_{K-1}(m)),\!$$

where the elements of $G(m)$ are defined as

$$g_k(m) = \frac{\varphi_k(m)}{\sum_{i=0}^{K-1} \varphi_i(m)}, 0 \leq k \leq K - 1,$$
\[ \varphi_k = \max \{ a \max \{ b, |\hat{w}_0|, |\hat{w}_k|, \ldots, |\hat{w}_{K-1}| \}, |\hat{w}_k| \}, \]  
(11)

where the time index is ignored. Parameter \( b > 0 \) is an initialization parameter that is used to prevent \( \hat{w}(m-1) \) from stalling at the initialization stage when \( \hat{w}(m-1) = 0_{K \times 1} \), while parameter \( a > 0 \) prevents \( \hat{w}_k(m-1) \) from stalling when they are smaller than the largest coefficient \[8\]. Usually, parameter \( a \) ranges from \( \frac{1}{K} \) to \( \frac{5}{K} \).

**B. IPAPA**

The IPAPA, which introduces the \( l_1 \)-norm to reassign the gains to each coefficient, is similar to the IPNLMS \[12\]. The updating equation of the IPAPA is the same as that of the PAPA and is presented in (8) \[9\], \[29\].

\[ \varphi_k(m) = (1 - \alpha) \left( \frac{1}{K} \right) \left| \hat{w}_k(m) \right| + (1 + \alpha) \left| \hat{w}_k(m) \right|, \]  
(12)

where \(-1 \leq \alpha < 1 \) and \( 0 \leq k \leq K - 1 \). Using Equation (12) to replace Equation (11), we obtain

\[ g_k(m) = \frac{1 - \alpha}{2L} + \frac{1}{2} \sum_{i=0}^{K-1} \left| \hat{w}_i(m) \right| + \frac{\varepsilon}{L}, \]  
(13)

Similar to the IPNLMS algorithm, the IPAPA turns into APA when \( \alpha = -1 \). When \( \alpha \) is close to 1, the IPAPA becomes the PAPA.

**III. PROPOSED CS-PAPA AND CS-IPAPA ALGORITHMS**

In this section, we propose the CS-PAPA and CS-IPAPA to exploit the cluster sparse characteristics in the satellite and network echo channels illustrated in **Figure 1**.

**A. THE PROPOSED CS-PAPA ALGORITHM**

The proposed CS-PAPA is obtained by minimizing the cost function obtained by adding a mixed \( l_2,0 \)-norm to the cost function of the PAPA to fully exploit the cluster sparse characteristics. The CS-PAPA solves the following problem

\[ \min \| \hat{w}(m) \|_{2,0} \text{ s.t. } d(m) - X^T(m) \hat{w}(m) = 0 \]  
(14)

where the \( l_{2,0} \)-norm is defined as

\[ \| \hat{w}(m) \|_{2,0} = \left\| \begin{array}{c} |\hat{w}_1|_2 \\ |\hat{w}_2|_2 \\ \vdots \\ |\hat{w}_N|_2 \end{array} \right\|_0, \]  
(15)

and \( N \) is the number of clusters, which is \( N = K/B \), where \( B \) is the number of channel coefficients per cluster. Since the solution of the \( l_0 \)-norm is an NP-hard problem, it can be approximated by \[30\]

\[ \| \hat{w}(m) \|_0 \approx \sum_{k=0}^{K-1} (1 - e^{-\beta |\hat{w}_k|}), \]  
(16)

where \( \beta > 0 \). From (15) and (16), we obtain

\[ \| \hat{w}(m) \|_{2,0} \approx \sum_{i=1}^{N} (1 - e^{-\beta |\hat{w}_i|}). \]  
(17)

The PAPA is gotten from the basis pursuit (BP) by solving \[24\], \[31\]

\[ \min \| \hat{w}(n) \|_1 \text{ s.t. } d(m) = X^T(m) \hat{w}(n) \]  
(18)

where \( \hat{w}(n) \) denotes the correction component, which is given by

\[ \hat{w}(n) = G(m-1)X(m) \left( X^T(m)G(m-1)X(m) \right)^{-1} d(m). \]  
(19)

If we consider \( \hat{G}(m) \approx G(m-1) \), a good approximation of (19) is written as

\[ \hat{w}(n) = G(m-1)X(m) \left( X^T(m)G(m-1)X(m) \right)^{-1} d(m). \]  
(20)

Then, a projection matrix \( R(m) \) is constructed for \( \hat{w}(n) \) to span in its nullspace, which is given by

\[ R(m) = I_K \]  
(21)

\[ -G(m-1)X(m) \left( X^T(m)G(m-1)X(m) \right)^{-1} X^T(m). \]  
(22)

From the BS perspective, the updating equation of the CS-PAPA is

\[ \hat{w}(m) = R(m) \hat{w}(m-1) + \hat{w}(m-1), \]  
(23)

which can be rewritten as

\[ \hat{w}(m) = \hat{w}(m-1) + \mu G(m-1)X(m) \left( X^T(m)G(m-1)X(m) \right)^{-1} \]  
(24)

\[ \epsilon(m), \]  
(25)

where the diagonal matrix \( G(m) \) is given by

\[ G(m) = \text{diag} \{ g_1(m-1) \, 1_B, g_2(m-1) \, 1_B, \ldots, g_N(m-1) \, 1_B \}, \]  
(26)

where \( 1_B \) is a row vector with \( B \) elements, and all of its elements are ones. Moreover, \( g_s(m-1) \) is defined as

\[ g_s(m-1) = \frac{\varphi_s(m-1)}{\sum_{i=1}^{N} \varphi_i(m-1)}, 1 \leq s \leq N, \]  
(27)

where

\[ \varphi_s(m-1) = \max \{ \rho \max \{ \alpha_1, \alpha_2, \ldots, \alpha_N \} \}, \]  
(28)

and

\[ \alpha_s(\hat{w}) = 1 - e^{-\beta |\hat{w}_i|}, \]  
(29)

where the time index is ignored in (27). The proposed CS-PAPA is summarized as **Algorithm1**.
B. THE PROPOSED CS-IPAPA ALGORITHM

The CS-IPAPA is proposed by incorporating the $l_2,0$-norm into the cost function of the IPAPA. The updating equation of the CS-IPAPA is the same as that of CS-PAPA, and the difference is the gain assign matrix, which is

$$g_s(m) = \frac{1-\alpha}{2L} + (1+\alpha) \frac{1-\beta\|\hat{w}(m)\|_2}{\sum_{i=1}^{N} (1-e^{-\beta\|\hat{w}(m)\|_2})+\varepsilon}.$$  \hspace{1cm} (29)

In contrast to the PAPA and IPAPA, the proposed CS-PAPA and CS-IPAPA utilize only the priori information to define the cluster known for network echo and satellite communication channels. Furthermore, it is worth noting that the APA and PAPA are special cases of the proposed scheme.

![Normalized Misalignment (dB)](image)

**FIGURE 2:** Parameter $\beta$ effects on the CS-PAPA and CS-IPAPA with colored input.

<table>
<thead>
<tr>
<th>Algorithm 1 CS-PAPA Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> $K, P, B, \mu, \beta, \rho, \sigma$, in</td>
</tr>
<tr>
<td><strong>Output:</strong> $\hat{w}$ out</td>
</tr>
<tr>
<td><strong>Initialization:</strong></td>
</tr>
<tr>
<td>1: $\hat{w} = \text{zeros}(K, 1)$;</td>
</tr>
<tr>
<td><strong>Loop Process:</strong></td>
</tr>
<tr>
<td>2: for $m=1, 2, \ldots \text{do}$</td>
</tr>
<tr>
<td>3: $e(m) = d(m) - y(m)$;</td>
</tr>
<tr>
<td>4: for $i=1, 2, \ldots, N \text{ do}$</td>
</tr>
<tr>
<td>5: $q_i(\hat{w}) = 1 - e^{-\beta|\hat{w}(m)|_2}$;</td>
</tr>
<tr>
<td>6: end for</td>
</tr>
<tr>
<td>7: for $s=1, 2, \ldots, N \text{ do}$</td>
</tr>
<tr>
<td>8: $\varphi_s(m-1) = \max{\rho\max{\sigma, q_1(\hat{w}), \ldots, q_N(\hat{w})}, q_s(\hat{w})}$;</td>
</tr>
<tr>
<td>9: end for</td>
</tr>
<tr>
<td>10: for $j=1, 2, \ldots, N \text{ do}$</td>
</tr>
<tr>
<td>11: $g_j(m-1) = \frac{\varphi_j(m-1)}{\sum_{i=1}^{N} \varphi_i(m-1)}$;</td>
</tr>
<tr>
<td>12: end for</td>
</tr>
<tr>
<td>13: $G(m-1) = diag{g_1(m-1) 1_B, \ldots, g_N(m-1) 1_B}$</td>
</tr>
<tr>
<td>$\hat{w}(m) = \hat{w}(m-1) + \mu G(m-1) X(m) (X^T(m) G(m-1) X(m) + \delta_{CS} I)^{-1} e(m)$</td>
</tr>
<tr>
<td>14: end for</td>
</tr>
<tr>
<td>15: return $\hat{w}$</td>
</tr>
</tbody>
</table>

IV. SIMULATION RESULTS

To verify the performance of the CS-PAPA and CS-IPAPA, various experiments are conducted to assess their tracking behaviors, which are compared with the APA, PAPA, BS-PAPA and BS-IPAPA in the context of system identification. In all the experiments, the length of the adaptive filter is $K = 1024$, and all the mentioned algorithms are investigated in the presence of white Gaussian noise (WGN), colored noise and speech signals. The colored noise is obtained from WGN through a first-order system with a pole of 0.8, and the sampling frequency of the colored noise and speech signals is 8 kHz. Noise is added to the background of the unknown system at a signal-to-noise ratio (SNR) of 30 dB. We use normalized misalignment (NM) to measure the performance, and the NM is defined as $10 \log (\|\hat{w} - \hat{w}\|^2_2 / \|w\|^2_2)$. The two cluster-sparse IR systems shown in FIGURE 1(a) and FIGURE 1(b) are considered. The first IR in FIGURE 1(a) has 32 nonzero taps and a single cluster with active coefficients distributed in [267, 298]. The second IR in FIGURE 1(b) has two clusters with active taps distributed at [267, 298] and [779, 810], which

<table>
<thead>
<tr>
<th>TABLE 1: Simulation parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
</tr>
<tr>
<td>PAPA</td>
</tr>
<tr>
<td>IPAPA</td>
</tr>
<tr>
<td>BS-PAPA</td>
</tr>
<tr>
<td>CS-PAPA</td>
</tr>
<tr>
<td>CS-IPAPA</td>
</tr>
</tbody>
</table>
contain 64 nonzero taps. In all the experiments, the step sizes of the discussed algorithms are set to $\mu = 0.01$ when the input signal is WGN or colored noise, and the step size is set to $\mu = 0.02$ for speech inputs. $\varepsilon = 0.01$ and $\alpha = 5/K$ are used in the related algorithms. The regularization parameters are $\delta_{APA} = 0.01$ and $\delta_{IPAPA} = 1 - \alpha^2 \delta_{PAPA} = 1 - \alpha^2 L \delta_{APA}$ [32]. Other simulation parameters are shown in Table 1. In all the experiments, the first 30,000 iterations represent the performance of the single-cluster system, and the second 30,000 iterations indicate the performance of the double-cluster system.

A. PERFORMANCE OF THE CS-PAPA AND CS-IPAPA WITH DIFFERENT $\beta$

The effect of $\beta$ on the performance of the CS-PAPA and CS-IPAPA is investigated with different inputs, namely, WGN, colored noise and speech signals. Herein, the cluster size is $B = 4$, and $\alpha$ is set to 0.5 for CS-IPAPA. The results with $\beta$ values of 2, 5, 10 and 20 are presented in FIGURES 3, 2 and 4 for different inputs.

The experimental results presented in FIGURE 3(a) indicate that the NM of the CS-PAPA is reduced with the increase of $\beta$ for both single-cluster and double-cluster systems when the input is a WGN signal. The CS-IPAPA algorithm shown in FIGURE 3(b) achieves the smallest NM for single-cluster
B. PERFORMANCE OF CS-PAPA AND CS-IPAPA WITH DIFFERENT $B$

The investigation of $\beta$ showed that $\beta$ has important effects on the NM of the CS-PAPA and CS-IPAPA algorithms. Another important parameter is $B$, which is used to determine the block size used in the proposed CS algorithms. Thus, four different block sizes with $B=4, 16, 32, 64$ are selected to evaluate the performance of the CS-PAPA and CS-IPAPA. In these experiments, $\beta$ is set to 20 when the input signal is WGN or colored noise, and $\beta$ is set to 2 when the input signal is a speech signal. The simulation results are given in FIGURES 5, 6 and 7. When the input signal is WGN and colored noise, the convergence and the NM of the CS-PAPA and CS-IPAPA deteriorate with increasing $B$ for both the single- and double-cluster systems, as shown in FIGURES 5 and 6. When $B = 4$, the CS-PAPA and CS-IPAPA achieve the best performance. When the speech signal is the inputs,
the CS-PAPA achieves the fastest convergence and lowest NM for $B = 16$, as shown in FIGURE 7(a). FIGURE 7(b) shows that the CS-IPAPA has smallest NM for $B = 4$. Thus, $\beta$ and $B$ should be carefully selected and optimized to achieve superior performance with respect to the convergence and NM for both the CS-PAPA and CS-IPAPA. However, the proposed CS-PAPA and CS-IPAPA outperform the APA, PAPA and BS-IPAPA [23] for estimating single- and double-cluster channels.

V. CONCLUSIONS

In this paper, the CS-PAPA and CS-IPAPA have been proposed for cluster-sparse satellite and network echo channel estimation applications. The devised algorithms are implemented by using the mixed $l_2, \alpha$-norm in the PAPA and IPAPA to make use of the prior cluster sparse information. Different $\beta$, $B$ and inputs are taken into account to investigate the performance of the CS-PAPA and CS-IPAPA for estimating single- and double-cluster systems. The simulation results confirm that the proposed CS-PAPA and CS-IPAPA are superior to the APA, PAPA and the BS-IPAPA in various experiments.

REFERENCES


