Research Article

Smooth Approximation l_0 -Norm Constrained Affine Projection Algorithm and Its Applications in Sparse Channel Estimation

Yingsong Li and Masanori Hamamura

Graduate School of Engineering, Kochi University of Technology, Kami-shi 782-8502, Japan

Correspondence should be addressed to Yingsong Li; liyingsong@ieee.org

Received 10 December 2013; Accepted 30 January 2014; Published 26 March 2014

Academic Editors: G. Jovanovic Dolecek, C. Saravanan, and D. Tay

Copyright © 2014 Y. Li and M. Hamamura. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a smooth approximation l_0 -norm constrained affine projection algorithm (SL0-APA) to improve the convergence speed and the steady-state error of affine projection algorithm (APA) for sparse channel estimation. The proposed algorithm ensures improved performance in terms of the convergence speed and the steady-state error via the combination of a smooth approximation l_0 -norm (SL0) penalty on the coefficients into the standard APA cost function, which gives rise to a zero attractor that promotes the sparsity of the channel taps in the channel estimation and hence accelerates the convergence speed and reduces the steady-state error when the channel is sparse. The simulation results demonstrate that our proposed SL0-APA is superior to the standard APA and its sparsity-aware algorithms in terms of both the convergence speed and the steady-state behavior in a designated sparse channel. Furthermore, SL0-APA is shown to have smaller steady-state error than the previously proposed sparsity-aware algorithms when the number of nonzero taps in the sparse channel increases.

1. Introduction

With the development of wireless communication, there have been increasing demands for higher transmission rates in modern communication systems. This has led to the development of new standards for various wireless devices, such as smartphones, laptops, and iPads [1–5]. Given these requirements, broadband signal transmission is an essential technique for next-generation wireless communication systems [6]. In broadband wireless communications, a “hilly terrain” (HT) delay profile consists of a sparsely distributed multipath channel in which most of taps are zero or close to zero, while only a few taps are dominant [4]. In this paper, we consider the communication problems which involve the estimation and equalization of channels with a large delay spread but with a small nonzero support, which is also known as sparse channel estimation.

Recently, a rising method for sparse channel estimation has been proposed and extensively investigated by the use of compressed sensing (CS) to improve the performance of such sparse wireless communication channels [7–9]. We

found that these CS channel estimation algorithms were sensitive to the channel interferences. Another effective class of methods that have been widely studied in channel estimation is adaptive filtering algorithms [10–13], such as least mean square (LMS), recursive least squares (RLS), and Kalman filter algorithms. However, these standard adaptive filtering algorithms cannot utilize the sparse property of the wireless communication channel and hence they perform poorly in dealing with the sparse signals. To utilize the sparse characteristic of such channels, some improved adaptive filtering algorithms by the use of partial updating techniques have been proposed and investigated in wireless communications [14–16]. However, this partial updating degraded the estimation performance in contrast to the standard LMS and RLS algorithms.

Motivated by the widely developed CS techniques [17, 18], some efforts have been put into combining the CS technique into the adaptive filtering algorithms in order to improve the performance of standard adaptive filtering performance for sparse signal recovery. For example, a Kalman filter compressed sensing (KF-CS) algorithm has

been proposed and applied in magnetic resonance imaging (MRI) by the combination of CS and standard Kalman filter [19]. In this algorithm, Kalman filter estimates the support set which has significant effect on the estimator errors. Furthermore, another algorithm denoted as least square compressed sensing (LS-CS) has been developed and well investigated by using the CS and RLS techniques [20, 21]. Unfortunately, these algorithms are highly complex because of the computational complexity of Kalman filter and RLS algorithms. LMS algorithm has attracted much more attention in recent years due to its low computational complexity and reliable recovery capability. Inspired by the CS theory [17, 18] and the KF-CS and LS-CS algorithms, several sparsity-aware LMS algorithms have been proposed with additional norm constrained terms in the cost function of standard LMS algorithms [6, 22–27]. It was found in these studies that these linear constrained sparsity-aware LMS algorithms can achieve faster convergence speed and better steady-state performance compared to the standard LMS algorithm. However, these sparsity-aware LMS algorithms are sensitive to the noise and the sparsity characteristics of the channel, which results in high steady-state misadjustment due to the estimation error that occurs in the adaptation. The affine projection algorithm (APA) is another popular method in adaptive filtering applications [28–31], with its complexity and estimation performance intermediary between the LMS and RLS algorithms. The APA reuses old data resulting in fast convergence, and is also an improved normalized LMS (NLMS) algorithm that converges faster than the standard LMS algorithm. Subsequently, l_1 -norm penalized APA has been proposed to render the standard APA suitable for sparse signal estimation applications [32]. However, these l_1 -norm penalized APAs impose the condition that the proportion of nonzero taps must be very small as compared to the proportion of dominant taps in the associated parameter vector in channel estimation.

In this paper, we propose a smooth approximation l_0 -norm constrained affine projection (SL0-APA) algorithm for sparse channel estimation. The proposed SL0-APA is similar to the algorithms proposed in [32], which are known as zero-attracting affine projection algorithm (ZA-APA) and reweighted zero-attracting affine projection algorithm (RZA-APA). It differs by the regularization term which is a smooth approximation l_0 -norm obtained from a continuous function that is an accurate approximation of l_0 -norm. By exploiting the information of the sparsity channel and using the concepts of the smooth approximation of l_0 -norm, we can improve the performance of the previous sparsity-aware APAs with respect to both the convergence speed and the steady-state performance. We also provide a convergence analysis and the mean-square-error analysis of our proposed SL0-APA. Furthermore, we experimentally investigate the effect of adding a smooth approximation l_0 -norm penalty term to the cost function on learning the convergence behavior and the steady-state error performance of the SL0-APA. Accordingly, we experimentally illustrate that the SL0-APA is superior to ZA-APA and RZA-APA in terms of steady-state error and the convergence speed. Besides, the theoretical analysis is also presented and compared to the computer

FIGURE 1: A typical sparse multipath communication system.

simulation results. Finally, the computational complexity of the proposed SL0-APA is mathematically given and is experimentally evaluated.

The remainder of the paper is structured as follows. Section 2 briefly reviews the standard APA, ZA-APA, and RZA-APA based on a sparse multipath communication system. In Section 3, we first propose a SL0-APA by the use of a smooth approximation l_0 -norm penalty on the cost function of the standard APA. Next, we provide a theoretical expression of the convergence analysis and the mean-square-error (MSE) analysis of our proposed SL0-APA based on the energy-conservation approach. In Section 4, the proposed SL0-APA is experimentally investigated over a sparse channel to demonstrate the estimation performance of the SL0-APA, including the convergence speed, steady-state error, and the computational complexity. Finally, Section 5 is the conclusion.

2. Conventional Channel Estimation Algorithms

In this section, we consider a sparse multipath communication system shown in Figure 1 to discuss traditional channel estimation algorithms. The input signal $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T$ containing the N most recent samples is transmitted over a finite impulse response (FIR) channel with channel impulse response (CIR) $\mathbf{h} = [h_0, h_1, \dots, h_{N-1}]^T$, where $(\cdot)^T$ denotes the transposition. The input signal $\mathbf{x}(n)$ is also used as an input for an adaptive filter $\hat{\mathbf{h}}(n)$ with N coefficients to produce an estimation output $\hat{y}(n)$, and the received signal $\mathbf{r}(n) = \mathbf{y}(n) + \mathbf{v}(n)$ is obtained at the receiver.

2.1. Affine Projection Algorithm (APA). The channel estimation technique called the standard APA estimates the unknown sparse channel $\mathbf{h}$ using the input signal $\mathbf{x}(n)$ and the output signal $\mathbf{y}(n)$. In the standard APA, let us assume that we

keep the last Q input signal $\mathbf{x}(n)$ to form the matrix $\mathbf{U}(n)$ as follows [28]:

$$\mathbf{U}(n) = \begin{bmatrix} \mathbf{x}^T(n) \\ \mathbf{x}^T(n-1) \\ \vdots \\ \mathbf{x}^T(n-Q+1) \end{bmatrix} = \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-N+1) \\ x(n-1) & x(n-2) & \cdots & x(n-N) \\ \vdots & \vdots & \ddots & \vdots \\ x(n-Q+1) & x(n-Q) & \cdots & x(n-N-Q+2) \end{bmatrix}, \quad (1)$$

where Q denotes the projection order of the APA. Furthermore, we also define some vectors representing reusing results at a given instant n , such as the output $\mathbf{y}(n)$ of the channel, the output $\hat{\mathbf{y}}(n)$ of the filter, the received signal $\mathbf{r}(n)$, and the additive white Gaussian noise vector $\mathbf{v}(n)$ and these vectors are expressed as

$$\mathbf{y}(n) = \mathbf{U}(n) \mathbf{h} = \begin{bmatrix} y(n) \\ y(n-1) \\ \vdots \\ y(n-Q+1) \end{bmatrix}, \quad (2)$$

$$\hat{\mathbf{y}}(n) = \mathbf{U}(n) \hat{\mathbf{h}}(n) = \begin{bmatrix} \hat{y}(n) \\ \hat{y}(n-1) \\ \vdots \\ \hat{y}(n-Q+1) \end{bmatrix}, \quad (3)$$

$$\mathbf{v}(n) = \begin{bmatrix} v(n) \\ v(n-1) \\ \vdots \\ v(n-Q+1) \end{bmatrix}, \quad (4)$$

$$\mathbf{r}(n) = \begin{bmatrix} r(n) \\ r(n-1) \\ \vdots \\ r(n-Q+1) \end{bmatrix}. \quad (5)$$

From (1)–(5), the instantaneous error $\mathbf{e}(n)$ can be written as

$$\begin{aligned} \mathbf{e}(n) &= \begin{bmatrix} e(n) \\ e(n-1) \\ \vdots \\ e(n-Q+1) \end{bmatrix} \\ &= \begin{bmatrix} r(n) - \hat{y}(n) \\ r(n-1) - \hat{y}(n-1) \\ \vdots \\ r(n-Q+1) - \hat{y}(n-Q+1) \end{bmatrix} \\ &= \mathbf{r}(n) - \hat{\mathbf{y}}(n). \end{aligned} \quad (6)$$

As for the channel estimation, the purpose of the APA is to minimize

$$\begin{aligned} &\|\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)\|^2 \\ &\text{subject to:} \\ &\mathbf{r}(n) - \mathbf{U}(n) \hat{\mathbf{h}}(n+1) = \mathbf{0}. \end{aligned} \quad (7)$$

The APA maintains the next coefficient $\hat{\mathbf{h}}(n+1)$ as close as possible to the current coefficient $\hat{\mathbf{h}}(n)$ and minimizes the a posteriori error to zero at the same time. Here, the Lagrange multiplier is used to find out the solution that minimizes the cost function $J_{\text{APA}}(n)$ of the APA:

$$\begin{aligned} J_{\text{APA}}(n) &= \|\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)\|^2 \\ &+ [\mathbf{r}(n) - \mathbf{U}(n) \hat{\mathbf{h}}(n+1)]^T \boldsymbol{\lambda}_{\text{APA}}, \end{aligned} \quad (8)$$

where $\boldsymbol{\lambda}_{\text{APA}}$ is a $Q \times 1$ vector of Lagrange multiplier and $\boldsymbol{\lambda}_{\text{APA}} = [\lambda_0 \ \lambda_1 \ \cdots \ \lambda_{Q-1}]^T$. Equation (8) can be rewritten as

$$\begin{aligned} J_{\text{APA}}(n) &= [\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)]^T [\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)] \\ &+ [\mathbf{r}^T(n) - \hat{\mathbf{h}}^T(n+1) \mathbf{U}^T(n)] \boldsymbol{\lambda}_{\text{APA}}. \end{aligned} \quad (9)$$

Then, the gradient of $J_{\text{APA}}(n)$ with respect to $\hat{\mathbf{h}}(n+1)$ is given by

$$\frac{\partial J_{\text{APA}}(n)}{\partial \hat{\mathbf{h}}(n+1)} = 2\hat{\mathbf{h}}(n+1) - 2\hat{\mathbf{h}}(n) - \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{APA}}. \quad (10)$$

After setting the gradient of $J_{\text{APA}}(n)$ with respect to $\hat{\mathbf{h}}(n+1)$ equal to zero, we get

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{APA}}. \quad (11)$$

Multiplying $\mathbf{U}(n)$ on both sides of (11), we have

$$\mathbf{U}(n) \hat{\mathbf{h}}(n+1) = \mathbf{U}(n) \hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}(n) \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{APA}}. \quad (12)$$

By taking the constraint condition of (7) into consideration, we have

$$\mathbf{r}(n) = \mathbf{U}(n) \hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}(n) \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{APA}}. \quad (13)$$

Taking (3), (6), and (12) into account, we can get

$$\mathbf{e}(n) = \frac{1}{2} \mathbf{U}(n) \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{APA}}. \quad (14)$$

Then

$$\boldsymbol{\lambda}_{\text{APA}} = 2[\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n). \quad (15)$$

The update equation is now given by (11) with $\boldsymbol{\lambda}_{\text{APA}}$ being the solution of (14) and is expressed as

$$\begin{aligned} \hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mathbf{U}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n) \\ &= \hat{\mathbf{h}}(n) + \mathbf{U}^+(n) \mathbf{e}(n), \end{aligned} \quad (16)$$

where $\mathbf{U}^+(n) = \mathbf{U}^T(n)[\mathbf{U}(n)\mathbf{U}^T(n)]^{-1}$. The above update equation corresponds to the conventional APA with unity convergence factor [28]. In the practical engineering applications, a convergence factor μ_{APA} , also known as step-size, is adopted to tradeoff the mean square misadjustment and convergence speed, and thus, the update equation (16) can be rewritten as

$$\begin{aligned}\hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mu_{\text{APA}} \mathbf{U}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n) \\ &= \hat{\mathbf{h}}(n) + \mu_{\text{APA}} \mathbf{U}^+(n) \mathbf{e}(n).\end{aligned}\quad (17)$$

In general, the step-size μ_{APA} should be chosen in the range $0 < \mu_{\text{APA}} < 2$ to control the convergence speed and the steady-state behavior of the APA. It is worth noting that the APA becomes familiar normalized least mean square (NLMS) when the $Q = 1$.

2.2. Zero-Attracting Affine Projection Algorithm (ZA-APA). To improve the performance of the standard APA and to utilize the sparsity property of the sparse multipath communication channel, an l_1 -penalty term is cooperated into the cost function of (8), which is known as zero-attracting affine projection algorithm (ZA-APA) [32]. In the ZA-APA, the cost function is defined by combining the cost function $J_{\text{APA}}(n)$ of standard APA with l_1 -penalty of the channel estimator and is given by

$$\begin{aligned}J_{\text{ZA}}(n) &= \|\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)\|^2 + [\mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n+1)]^T \boldsymbol{\lambda}_{\text{ZA}} \\ &\quad + \gamma_{\text{ZA}} \|\hat{\mathbf{h}}(n+1)\|_1,\end{aligned}\quad (18)$$

where $\boldsymbol{\lambda}_{\text{ZA}}$ is the vector of Lagrange multiplier with $Q \times 1$. $\gamma_{\text{ZA}} > 0$ is a regularization parameter to balance the estimation error and the sparse l_1 -penalty of $\hat{\mathbf{h}}(n+1)$. In order to minimize the cost function $J_{\text{ZA}}(n)$, we use the Lagrange multiplier to calculate its gradient, which is expressed as

$$\begin{aligned}\frac{\partial J_{\text{ZA}}(n)}{\partial \hat{\mathbf{h}}(n+1)} &= 2\hat{\mathbf{h}}(n+1) - 2\hat{\mathbf{h}}(n) - \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{ZA}} \\ &\quad + \gamma_{\text{ZA}} \text{sgn}[\hat{\mathbf{h}}(n+1)],\end{aligned}\quad (19)$$

where $\text{sgn}[\cdot]$ is a component-wise sign function defined as

$$\text{sgn}[x] = \begin{cases} \frac{x}{|x|}, & x \neq 0 \\ 0, & x = 0. \end{cases}\quad (20)$$

As is known to us, the minimum is obtained by letting $\partial J_{\text{ZA}}(n)/\partial \hat{\mathbf{h}}(n+1) = \mathbf{0}$. Thus, we can get

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{ZA}} - \frac{1}{2} \gamma_{\text{ZA}} \text{sgn}[\hat{\mathbf{h}}(n+1)].\quad (21)$$

Multiplying both sides by $\mathbf{U}(n)$, we can obtain

$$\begin{aligned}\mathbf{U}(n) \hat{\mathbf{h}}(n+1) &= \mathbf{U}(n) \hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}(n) \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{ZA}} \\ &\quad - \frac{1}{2} \gamma_{\text{ZA}} \mathbf{U}(n) \text{sgn}[\hat{\mathbf{h}}(n+1)].\end{aligned}\quad (22)$$

Considering the constraint condition of (7), we can get the following expression:

$$\begin{aligned}\mathbf{r}(n) &= \mathbf{U}(n) \hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}(n) \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{ZA}} \\ &\quad - \frac{1}{2} \gamma_{\text{ZA}} \mathbf{U}(n) \text{sgn}[\hat{\mathbf{h}}(n+1)].\end{aligned}\quad (23)$$

From the above discussion, we know that $\mathbf{e}(n) = \mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n)$. Thus, the Lagrange multipliers vector $\boldsymbol{\lambda}_{\text{ZA}}$ is obtained:

$$\boldsymbol{\lambda}_{\text{ZA}} = [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \{2\mathbf{e}(n) + \gamma_{\text{ZA}} \mathbf{U}(n) \text{sgn}[\hat{\mathbf{h}}(n+1)]\}.\quad (24)$$

Substituting (24) into (21) and assuming that $\text{sgn}[\hat{\mathbf{h}}(n+1)] \approx \text{sgn}[\hat{\mathbf{h}}(n)]$, we can obtain the update function of the ZA-APA:

$$\begin{aligned}\hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mathbf{U}^+(n) \mathbf{e}(n) \\ &\quad + \frac{1}{2} \gamma_{\text{ZA}} \mathbf{U}^+(n) \mathbf{U}(n) \text{sgn}[\hat{\mathbf{h}}(n)] \\ &\quad - \frac{1}{2} \gamma_{\text{ZA}} \text{sgn}[\hat{\mathbf{h}}(n)].\end{aligned}\quad (25)$$

To balance the convergence speed and steady-state error, a step-size μ_{ZA} is introduced and integrated into (25). Then, (25) can be rewritten as

$$\begin{aligned}\hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mu_{\text{ZA}} \mathbf{U}^+(n) \mathbf{e}(n) \\ &\quad + \frac{1}{2} \gamma_{\text{ZA}} \mathbf{U}^+(n) \mathbf{U}(n) \text{sgn}[\hat{\mathbf{h}}(n)] \\ &\quad - \frac{1}{2} \gamma_{\text{ZA}} \text{sgn}[\hat{\mathbf{h}}(n)].\end{aligned}\quad (26)$$

Comparing the update equation (26) of the ZA-APA with the update (17) of the standard APA, we find that there are two additional terms in (26) which attract the tap coefficients to zero when the tap magnitudes of the sparse channel are close to zero. These two additional terms are zero attractors whose attracting strengths are controlled by γ_{ZA} . Intuitively, the zero attractor can speed the convergence of ZA-APA when the majority taps of the channel of $\mathbf{h}$ are zero or close to zero, such as sparse channel.

2.3. Reweighted Zero-Attracting Affine Projection Algorithm (RZA-APA). Unfortunately, the ZA-APA cannot distinguish the zero taps and the nonzero taps of the sparse channel, and it exerts the same penalty on all the channel taps, which forces all the taps to zero uniformly [22, 32]. Therefore, the performance of the ZA-APA is degraded when the channel is a less sparse one. In order to improve the performance of the ZA-APA and to solve this problem, a heuristic approach first reported in [33] and employed in [22, 32] to reinforce that the zero attractor was proposed and was denoted as reweighted zero-attracting affine projection algorithm (RZA-APA). In the RZA-APA, $\sum_{i=1}^N \log(1 + \epsilon_{\text{RZA}} |\hat{h}_i(n)|)$ is adopted instead

of $\|\hat{\mathbf{h}}(n)\|_1$. Thus, the cost function of the RZA-APA can be written as

$$J_{\text{RZA}}(n) = \|\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)\|^2 + [\mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n+1)]^T \boldsymbol{\lambda}_{\text{RZA}} + \gamma_{\text{RZA}} \sum_{i=1}^N \log(1 + \varepsilon_{\text{RZA}} |\hat{h}_i(n+1)|), \quad (27)$$

where $\gamma_{\text{RZA}} > 0$ is a regularization parameter, $\varepsilon_{\text{RZA}} > 0$ is a positive threshold, and $\boldsymbol{\lambda}_{\text{RZA}}$ is the vector of the Lagrange multiplier with size of $Q \times 1$. The Lagrange multiplier is used for calculating the minimization of $J_{\text{RZA}}(n)$ and the gradient of $J_{\text{RZA}}(n)$ can be expressed as

$$\frac{\partial J_{\text{RZA}}(n)}{\partial \hat{\mathbf{h}}(n+1)} = 2\hat{\mathbf{h}}(n+1) - 2\hat{\mathbf{h}}(n) - \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{RZA}} + \gamma_{\text{RZA}} \frac{\text{sgn}[\hat{\mathbf{h}}(n+1)]}{1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n+1)|}. \quad (28)$$

Let $\partial J_{\text{RZA}}(n)/\partial \hat{\mathbf{h}}(n+1) = \mathbf{0}$ and assume $\text{sgn}[\hat{\mathbf{h}}(n+1)]/(1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n+1)|) \approx \text{sgn}[\hat{\mathbf{h}}(n)]/(1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n)|)$, and then we can get

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{RZA}} - \frac{1}{2} \gamma_{\text{RZA}} \frac{\text{sgn}[\hat{\mathbf{h}}(n)]}{1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n)|}. \quad (29)$$

By multiplying $\mathbf{U}(n)$ on both sides of (29), the following equation can be obtained:

$$\mathbf{U}(n) \hat{\mathbf{h}}(n+1) = \mathbf{U}(n) \hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}(n) \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{RZA}} - \frac{1}{2} \gamma_{\text{RZA}} \mathbf{U}(n) \frac{\text{sgn}[\hat{\mathbf{h}}(n)]}{1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n)|}. \quad (30)$$

Taking (7) and (30) into consideration, we can get

$$\mathbf{r}(n) = \mathbf{U}(n) \hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}(n) \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{RZA}} - \frac{1}{2} \gamma_{\text{RZA}} \mathbf{U}(n) \frac{\text{sgn}[\hat{\mathbf{h}}(n)]}{1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n)|}. \quad (31)$$

Thus, the Lagrange multiplier vector $\boldsymbol{\lambda}_{\text{RZA}}$ is obtained:

$$\boldsymbol{\lambda}_{\text{RZA}} = [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \times \left\{ 2\mathbf{e}(n) + \gamma_{\text{RZA}} \mathbf{U}(n) \frac{\text{sgn}[\hat{\mathbf{h}}(n)]}{1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n)|} \right\}, \quad (32)$$

where $\mathbf{e}(n) = \mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n)$. Substituting (32) into (29), we can get the update equation of the RZA-APA:

$$\begin{aligned} \hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mathbf{U}^+(n) \mathbf{e}(n) \\ &+ \frac{1}{2} \gamma_{\text{RZA}} \mathbf{U}^+(n) \mathbf{U}(n) \frac{\text{sgn}[\hat{\mathbf{h}}(n)]}{1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n)|} \\ &- \frac{1}{2} \gamma_{\text{RZA}} \frac{\text{sgn}[\hat{\mathbf{h}}(n)]}{1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n)|}. \end{aligned} \quad (33)$$

Similarly, a step-size μ_{RZA} is introduced and cooperated into (33) to balance the convergence speed and the steady-state error of the RZA-APA. Then, (33) can be rewritten as

$$\begin{aligned} \hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mu_{\text{RZA}} \mathbf{U}^+(n) \mathbf{e}(n) \\ &+ \frac{1}{2} \gamma_{\text{RZA}} \mathbf{U}^+(n) \mathbf{U}(n) \frac{\text{sgn}[\hat{\mathbf{h}}(n)]}{1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n)|} \\ &- \frac{1}{2} \gamma_{\text{RZA}} \frac{\text{sgn}[\hat{\mathbf{h}}(n)]}{1 + \varepsilon_{\text{RZA}} |\hat{\mathbf{h}}(n)|}. \end{aligned} \quad (34)$$

From the analysis and the a priori knowledge of the sparse channel, we know that the RZA-APA is more sensitive to taps with small magnitudes. Note that the reweighted zero attractor mainly affects taps whose magnitudes are comparable to $1/\varepsilon_{\text{RZA}}$ while it has less shrinkage exerted on $|\hat{\mathbf{h}}(n)| \gg 1/\varepsilon_{\text{RZA}}$. Thus, the RZA-APA can improve steady-state performance compared to the ZA-APA.

3. Proposed Smooth Approximation l_0 -Norm Constrained Affine Projection Algorithm (SL0-APA)

On the basis of the discussion of the ZA-APA and RZA-APA, we find that the RZA-APA can improve the performance of ZA-APA for sparse channel estimation because $\sum_{i=1}^N \log(1 + \varepsilon_{\text{RZA}} |\hat{h}_i(n+1)|)$ is more similar to l_0 -norm [22, 32, 33]. On the other hand, solving l_0 -norm $\|\hat{\mathbf{h}}(n+1)\|_0$ is a NP-hard problem [18]. Fortunately, smooth approximation l_0 -norm (SL0) with low complexity has been proposed as an accurate approximation of $\|\hat{\mathbf{h}}(n+1)\|_0$ to reconstruct sparse signals in CS theory [34, 35]. Inspired by the SL0 algorithm and in order to exploit the sparse characteristic of the multipath channel in a more accurate way, a smooth approximation l_0 -norm constrained affine projection algorithm (SL0-APA) is proposed by exerting the SL0 on the cost function of standard APA to further improve the performance of the RZA-APA.

3.1. Proposed SL0-APA. Similar to the ZA-APA and RZA-APA discussed above, the cost function of the SL0-APA is written as

$$J_{\text{SL0}}(n) = \|\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)\|^2 + [\mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n+1)]^T \boldsymbol{\lambda}_{\text{SL0}} + \gamma_{\text{SL0}} \|\hat{\mathbf{h}}(n+1)\|_0, \quad (35)$$

where $\boldsymbol{\lambda}_{\text{SL0}}$ is the vector of the Lagrange multiplier with size of $Q \times 1$ and $\gamma_{\text{SL0}} > 0$ is a regularization parameter to tradeoff the estimation error and the sparse l_0 -penalty of $\hat{\mathbf{h}}(n+1)$. Here, the smooth approximation of l_0 -norm $\|\hat{\mathbf{h}}(n+1)\|_0$ is a continuous function defined as follows:

$$\|\hat{\mathbf{h}}(n+1)\|_0 = \sum_{i=1}^{N-1} \frac{|\hat{h}_i(n+1)|}{|\hat{h}_i(n+1)| + \delta} = \frac{|\hat{\mathbf{h}}(n+1)|}{|\hat{\mathbf{h}}(n+1)| + \delta}, \quad (36)$$

where δ is a small positive constant which is used for avoiding division by zero, and the gradient of this continuous functions for SL0 is obtained:

$$\frac{\partial \|\hat{\mathbf{h}}(n+1)\|_0}{\partial \hat{\mathbf{h}}(n+1)} = \frac{\delta \text{sgn}(\hat{\mathbf{h}}(n+1))}{(|\hat{\mathbf{h}}(n+1)| + \delta)^2}. \quad (37)$$

To obtain the minimum of the $J_{\text{SL0}}(n)$, we use Lagrange multiplier to calculate the gradient of $J_{\text{SL0}}(n)$. Then the gradient of the cost function of the SL0-APA is written as

$$\begin{aligned} \frac{\partial J_{\text{SL0}}(n)}{\partial \hat{\mathbf{h}}(n+1)} &= 2\hat{\mathbf{h}}(n+1) - 2\hat{\mathbf{h}}(n) - \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{SL0}} \\ &+ \gamma_{\text{SL0}} \frac{\delta \text{sgn}(\hat{\mathbf{h}}(n+1))}{(|\hat{\mathbf{h}}(n+1)| + \delta)^2}. \end{aligned} \quad (38)$$

Let the left-hand side of (38) be equal to zero. We can get the following equation:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{SL0}} - \frac{1}{2} \gamma_{\text{SL0}} \frac{\delta \text{sgn}(\hat{\mathbf{h}}(n+1))}{(|\hat{\mathbf{h}}(n+1)| + \delta)^2}. \quad (39)$$

Multiplying $\mathbf{U}(n)$ on both sides of (39), we can get

$$\begin{aligned} \mathbf{U}(n)\hat{\mathbf{h}}(n+1) &= \mathbf{U}(n)\hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}(n) \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{SL0}} \\ &- \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}(n) \frac{\delta \text{sgn}(\hat{\mathbf{h}}(n+1))}{(|\hat{\mathbf{h}}(n+1)| + \delta)^2}. \end{aligned} \quad (40)$$

By taking (7) into consideration, (40) can be rewritten as

$$\begin{aligned} \mathbf{r}(n) &= \mathbf{U}(n)\hat{\mathbf{h}}(n) + \frac{1}{2} \mathbf{U}(n) \mathbf{U}^T(n) \boldsymbol{\lambda}_{\text{SL0}} \\ &- \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}(n) \frac{\delta \text{sgn}(\hat{\mathbf{h}}(n+1))}{(|\hat{\mathbf{h}}(n+1)| + \delta)^2}. \end{aligned} \quad (41)$$

From the discussion of the ZA-APA and RZA-APA, we can get the Lagrange multiplier vector $\boldsymbol{\lambda}_{\text{SL0}}$ from (41) by taking $\mathbf{e}(n) = \mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n)$ into account:

$$\begin{aligned} \boldsymbol{\lambda}_{\text{SL0}} &= [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \\ &\times \left\{ 2\mathbf{e}(n) + \gamma_{\text{SL0}} \mathbf{U}(n) \frac{\delta \text{sgn}(\hat{\mathbf{h}}(n+1))}{(|\hat{\mathbf{h}}(n+1)| + \delta)^2} \right\}. \end{aligned} \quad (42)$$

Substituting (42) into (39) and assuming that $\delta \text{sgn}(\hat{\mathbf{h}}(n+1))/(|\hat{\mathbf{h}}(n+1)| + \delta)^2 \approx \delta \text{sgn}(\hat{\mathbf{h}}(n))/(|\hat{\mathbf{h}}(n)| + \delta)^2$, the update function of the SL0-APA can be achieved:

$$\begin{aligned} \hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mathbf{U}^+(n) \mathbf{e}(n) \\ &+ \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}^+(n) \mathbf{U}(n) \frac{\delta \text{sgn}(\hat{\mathbf{h}}(n))}{(|\hat{\mathbf{h}}(n)| + \delta)^2} \\ &- \frac{1}{2} \gamma_{\text{SL0}} \frac{\delta \text{sgn}(\hat{\mathbf{h}}(n))}{(|\hat{\mathbf{h}}(n)| + \delta)^2} \\ &= \hat{\mathbf{h}}(n) + \mathbf{U}^+(n) \mathbf{e}(n) \\ &+ \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n) - \frac{1}{2} \gamma_{\text{SL0}} \mathbf{T}(n), \end{aligned} \quad (43)$$

where $\mathbf{T}(n) = \delta \text{sgn}(\hat{\mathbf{h}}(n))/(|\hat{\mathbf{h}}(n)| + \delta)^2$. Similar to the ZA-APA and RZA-APA, a step-size μ_{SL0} is introduced into (43) to create a balance between the convergence speed and steady-state error of the SL0-APA:

$$\begin{aligned} \hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mu_{\text{SL0}} \mathbf{U}^+(n) \mathbf{e}(n) \\ &+ \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n) - \frac{1}{2} \gamma_{\text{SL0}} \mathbf{T}(n). \end{aligned} \quad (44)$$

It is important to mention that our proposed SL0-APA is superior to APA, ZA-APA, and RZA-APA for sparse channel estimation because we utilize a smooth approximation of $\|\hat{\mathbf{h}}(n+1)\|_0$, which is proved to be an approximate and near-accurate approximation of l_0 -norm in comparison with the sum-log function $\sum_{i=1}^N \log(1 + \epsilon_{\text{RZA}} |\hat{h}_i(n+1)|)$ in the RZA-APA. Moreover, it is easy to calculate the gradient, as we can easily find a continuous gradient for this smoothed l_0 -norm function.

3.2. Analysis of the Proposed SL0-APA. In this section, we analyze the mean-square-error (MSE) behavior of the SL0-APA. Here, energy-conservation approach [36–38] is employed to obtain the theoretical expressions for the MSE of the SL0-APA. Let us consider the received signal $\mathbf{r}(n)$ that is derived from the following linear model:

$$\mathbf{r}(n) = \mathbf{U}(n) \mathbf{h} + \mathbf{v}(n), \quad (45)$$

where $\mathbf{h}$ is the sparse channel vector of the multipath communication system that we wish to estimate and $\mathbf{v}(n)$ is

the additive Gaussian noise at instant n . Our objective is to evaluate the steady-state MSE performance of the proposed SL0-APA. The steady-state MSE is defined as

$$\text{MSE} \triangleq \lim_{n \rightarrow \infty} E[|\mathbf{e}(n)|^2], \quad (46)$$

where $E[\cdot]$ denotes the expectation and

$$\mathbf{e}(n) = \mathbf{r}(n) - \mathbf{U}(n) \hat{\mathbf{h}}(n) \quad (47)$$

is the estimated error at time n . Taking (45) and (47) into account, we obtain

$$\begin{aligned} \mathbf{e}(n) &= \mathbf{U}(n) \mathbf{h} + \mathbf{v}(n) - \mathbf{U}(n) \hat{\mathbf{h}}(n) \\ &= \mathbf{U}(n) [\mathbf{h} - \hat{\mathbf{h}}(n)] + \mathbf{v}(n). \end{aligned} \quad (48)$$

Subtracting $\mathbf{h}$ from both sides of the SL0-APA update function (44), we get the misalignment vector:

$$\begin{aligned} \Delta(n+1) &= \mathbf{h} - \hat{\mathbf{h}}(n+1) \\ &= \mathbf{h} - \left\{ \hat{\mathbf{h}}(n) + \mu_{\text{SL0}} \mathbf{U}^+(n) \mathbf{e}(n) \right. \\ &\quad \left. + \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n) - \frac{1}{2} \gamma_{\text{SL0}} \mathbf{T}(n) \right\}. \end{aligned} \quad (49)$$

Substituting (48) into (49), we can get

$$\begin{aligned} \Delta(n+1) &= \mathbf{h} - \hat{\mathbf{h}}(n) - \mu_{\text{SL0}} \mathbf{U}^+(n) \\ &\quad \times \left\{ \mathbf{U}(n) [\mathbf{h} - \hat{\mathbf{h}}(n)] + \mathbf{v}(n) \right\} \\ &\quad - \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n) + \frac{1}{2} \gamma_{\text{SL0}} \mathbf{T}(n) \\ &= [\mathbf{I}_N - \mu_{\text{SL0}} \mathbf{U}^+(n) \mathbf{U}(n)] \Delta(n) - \mu_{\text{SL0}} \mathbf{U}^+(n) \mathbf{v}(n) \\ &\quad - \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n) + \frac{1}{2} \gamma_{\text{SL0}} \mathbf{T}(n). \end{aligned} \quad (50)$$

Taking expectations on both sides of (50), we get

$$\begin{aligned} E[\Delta(n+1)] &= E[\mathbf{I}_N - \mu_{\text{SL0}} \mathbf{U}^+(n) \mathbf{U}(n)] E[\Delta(n)] \\ &\quad - \mu_{\text{SL0}} E[\mathbf{U}^+(n) \mathbf{v}(n)] \\ &\quad - \frac{1}{2} \gamma_{\text{SL0}} E[\mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n)] \\ &\quad + \frac{1}{2} \gamma_{\text{SL0}} E[\mathbf{T}(n)]. \end{aligned} \quad (51)$$

We assume that the additive noise $\mathbf{v}(n)$ is statistically independent of the input signal $\mathbf{x}(n)$, and hence we have $E[\mathbf{U}^+(n) \mathbf{v}(n)] = \mathbf{0}$. Therefore, (51) can be simplified as

$$\begin{aligned} E[\Delta(n+1)] &= E[\mathbf{I}_N - \mu_{\text{SL0}} \mathbf{U}^+(n) \mathbf{U}(n)] E[\Delta(n)] \\ &\quad - \frac{1}{2} \gamma_{\text{SL0}} E[\mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n)] \\ &\quad + \frac{1}{2} \gamma_{\text{SL0}} E[\mathbf{T}(n)]. \end{aligned} \quad (52)$$

From previous studies on sparse LMS algorithms [22, 39], in the steady state, we have

$$E\{\text{sgn}[\hat{\mathbf{h}}(n)]\} \approx \text{sgn}(\hat{\mathbf{h}}). \quad (53)$$

Thus, the $E[\mathbf{T}(n)]$ in (52) can be written as

$$E[\mathbf{T}(n)] = E\left\{ \frac{\delta \text{sgn}(\hat{\mathbf{h}}(n))}{(|\hat{\mathbf{h}}(n)| + \delta)^2} \right\} = \frac{\delta \text{sgn}(\hat{\mathbf{h}}(n))}{(|\hat{\mathbf{h}}(n)| + \delta)^2}. \quad (54)$$

In addition, when the channel length is far larger than 1, $N \gg 1$, the $E[\mathbf{U}^+(n) \mathbf{U}(n)]$ can be written as [37, 40, 41]

$$\begin{aligned} E[\mathbf{U}^+(n) \mathbf{U}(n)] &= E\left\{ \mathbf{U}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{U}(n) \right\} \\ &\approx E\left\{ \mathbf{U}^T(n) \{E[\mathbf{U}(n) \mathbf{U}^T(n)]\}^{-1} \mathbf{U}(n) \right\}. \end{aligned} \quad (55)$$

Since $E[\mathbf{x}^T(n) \mathbf{x}(n-1)] = 0$ for sparse channel estimation, the inner expectation reduces to

$$\begin{aligned} E[\mathbf{U}(n) \mathbf{U}^T(n)] &= E\left\{ \begin{bmatrix} \mathbf{x}^T(n) \\ \mathbf{x}^T(n-1) \\ \vdots \\ \mathbf{x}^T(n-Q-1) \end{bmatrix} [\mathbf{x}(n) \ \mathbf{x}(n-1) \ \cdots \ \mathbf{x}(n-Q-1)] \right\} \\ &= E\left\{ \begin{bmatrix} \|\mathbf{x}(n)\|^2 & \mathbf{x}^T(n) \mathbf{x}(n-1) & \cdots & \mathbf{x}^T(n) \mathbf{x}(n-Q-1) \\ \mathbf{x}^T(n-1) \mathbf{x}(n) & \|\mathbf{x}(n-1)\|^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}^T(n-Q-1) \mathbf{x}(n) & \mathbf{x}^T(n-Q-1) \mathbf{x}(n-1) & \cdots & \|\mathbf{x}(n-Q-1)\|^2 \end{bmatrix} \right\} \end{aligned}$$

$$= E \left\{ \begin{bmatrix} \|\mathbf{x}(n)\|^2 & 0 & \cdots & 0 \\ 0 & \|\mathbf{x}(n-1)\|^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\mathbf{x}(n-Q+1)\|^2 \end{bmatrix} \right\}. \quad (56)$$

Here, we define

$$\mathbf{R} = \delta_x^2 \mathbf{I}_N, \quad (57)$$

where δ_x^2 is the power of the input signal. Thus,

$$E[\mathbf{U}(n) \mathbf{U}^T(n)] \approx \text{Tr}(\mathbf{R}) \mathbf{I}_Q = N\delta_x^2 \mathbf{I}_Q, \quad (58)$$

where $\text{Tr}(\cdot)$ is the trace of matrix and $\mathbf{I}_Q$ is the $Q \times Q$ identity matrix. Moreover, we can obtain

$$E\left\{[\mathbf{U}(n) \mathbf{U}^T(n)]^{-1}\right\} \approx \left\{E[\mathbf{U}(n) \mathbf{U}^T(n)]\right\}^{-1} = \frac{1}{N\delta_x^2} \mathbf{I}_Q. \quad (59)$$

Then we can approximate $E\{\mathbf{U}^T(n)[\mathbf{U}(n)\mathbf{U}^T(n)]^{-1}\mathbf{U}(n)\}$ by

$$\begin{aligned} E\left\{\mathbf{U}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{U}(n)\right\} &\approx E\left\{\mathbf{U}^T(n) \frac{1}{N\delta_x^2} \mathbf{I}_Q \mathbf{U}(n)\right\} \\ &\approx \frac{\mathbf{Q}\mathbf{R}}{N\delta_x^2}. \end{aligned} \quad (60)$$

Therefore, (52) can be rewritten as

$$\begin{aligned} E[\Delta(n+1)] &= E\left[\mathbf{I}_N - \mu_{\text{SL0}} \frac{\mathbf{Q}\mathbf{R}}{N\delta_x^2}\right] E[\Delta(n)] \\ &\quad - \frac{1}{2} \gamma_{\text{SL0}} \frac{\mathbf{Q}\mathbf{R}}{N\delta_x^2} E[\mathbf{T}(n)] + \frac{1}{2} \gamma_{\text{SL0}} E[\mathbf{T}(n)]. \end{aligned} \quad (61)$$

It is found that the matrix $\mathbf{T}(n)$ is approximately bounded between $-\delta \mathbf{I}_N$ and $\delta \mathbf{I}_N$. Therefore, we see that such convergence is guaranteed only if $(\mathbf{I}_N - \mu_{\text{SL0}} \mathbf{Q}\mathbf{R}/N\delta_x^2)$ is less than 1 [28], which is given by

$$0 < \mu_{\text{SL0}} < \frac{N\delta_x^2}{Q\lambda_{\max}}, \quad (62)$$

where $\lambda_{\max}$ is the maximum eigenvalue of the autocorrelation matrix $\mathbf{R}$ of $\mathbf{x}(n)$. We can observe that the stability condition of the SL0-APA is independent of the parameter γ_{SL0} . We assume that the estimated vector $\hat{\mathbf{h}}(n)$ converges when $n \rightarrow \infty$. Then, (61) can be rewritten as

$$\begin{aligned} E[\Delta(\infty)] &= \left[\mathbf{I}_N - \mu_{\text{SL0}} \frac{\mathbf{Q}\mathbf{R}}{N\delta_x^2}\right] E[\Delta(\infty)] \\ &\quad - \frac{1}{2} \gamma_{\text{SL0}} \frac{\mathbf{Q}\mathbf{R}}{N\delta_x^2} \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2} + \frac{1}{2} \gamma_{\text{SL0}} \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2}. \end{aligned} \quad (63)$$

From (63), we can obtain

$$E[\Delta(\infty)] = -\frac{\gamma_{\text{SL0}}}{2\mu_{\text{SL0}}} \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2} + \frac{\gamma_{\text{SL0}}}{2\mu_{\text{SL0}}} \frac{N\delta_x^2}{\mathbf{Q}\mathbf{R}} \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2}, \quad (64)$$

which can be regarded as

$$E[\hat{\mathbf{h}}(\infty)] = \mathbf{h} - \frac{\gamma_{\text{SL0}}}{2\mu_{\text{SL0}}} \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2} + \frac{\gamma_{\text{SL0}}}{2\mu_{\text{SL0}}} \frac{N\delta_x^2}{\mathbf{Q}\mathbf{R}} \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2}. \quad (65)$$

Note that (65) implies that the optimum solution of the SL0-APA is biased, as was also shown for zero-attracting least mean square (ZA-LMS) algorithms [22]. We then proceed to derive the steady-state MSE for our proposed SL0-APA. Firstly, multiplying both sides of (44) by $\mathbf{U}(n)$ from the left, we can get

$$\begin{aligned} \mathbf{U}(n) \hat{\mathbf{h}}(n+1) &= \mathbf{U}(n) \hat{\mathbf{h}}(n) + \mu_{\text{SL0}} \mathbf{U}(n) \mathbf{U}^+(n) \mathbf{e}(n) \\ &\quad + \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}(n) \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n) \\ &\quad - \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}(n) \mathbf{T}(n). \end{aligned} \quad (66)$$

Furthermore,

$$\mathbf{U}(n) \hat{\mathbf{h}}(n+1) = \mathbf{U}(n) \hat{\mathbf{h}}(n) + \mu_{\text{SL0}} \mathbf{e}(n). \quad (67)$$

Additionally, we define the a posteriori error vector $\mathbf{e}_p(n)$ and the a priori error vector $\mathbf{e}_a(n)$ as

$$\begin{aligned} \mathbf{e}_p(n) &= \mathbf{U}(n) \mathbf{h} - \mathbf{U}(n) \hat{\mathbf{h}}(n+1) \\ \mathbf{e}_a(n) &= \mathbf{U}(n) \mathbf{h} - \mathbf{U}(n) \hat{\mathbf{h}}(n). \end{aligned} \quad (68)$$

Combining (67) and (68), we have

$$\mathbf{e}_p(n) = \mathbf{e}_a(n) - \mu_{\text{SL0}} \mathbf{e}(n). \quad (69)$$

In addition,

$$\begin{aligned} \mathbf{e}(n) &= \mathbf{r}(n) - \mathbf{U}(n) \hat{\mathbf{h}}(n) \\ &= \mathbf{U}(n) \mathbf{h} + \mathbf{v}(n) - \mathbf{U}(n) \hat{\mathbf{h}}(n) \\ &= \mathbf{e}_a(n) + \mathbf{v}(n). \end{aligned} \quad (70)$$

By substituting (70) into (69), we have

$$\mathbf{e}_p(n) = (\mathbf{I} - \mu_{\text{SL0}}) \mathbf{e}(n) - \mathbf{v}(n). \quad (71)$$

From (69), we can also write the $\mathbf{e}(n)$ as follows:

$$\mathbf{e}(n) = \frac{1}{\mu_{\text{SL0}}} [\mathbf{e}_a(n) - \mathbf{e}_p(n)]. \quad (72)$$

Substituting (72) to (44), we have

$$\begin{aligned} \hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mathbf{U}^+(n) [\mathbf{e}_a(n) - \mathbf{e}_p(n)] \\ &+ \frac{1}{2} \gamma_{\text{SL0}} \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n) - \frac{1}{2} \gamma_{\text{SL0}} \mathbf{T}(n). \end{aligned} \quad (73)$$

On the basis of the discussion mentioned above, we notice that $\mathbf{U}(n)\mathbf{U}^+(n) = \mathbf{U}(n)\mathbf{U}^T(n)[\mathbf{U}(n)\mathbf{U}^T(n)]^{-1} = \mathbf{I}$. By considering the power of both sides of (73), using the steady-state condition $E[\|\hat{\mathbf{h}}(n+1)\|^2] \approx E[\|\hat{\mathbf{h}}(n)\|^2]$ when $n \rightarrow \infty$, and assuming that $\mathbf{e}_a(n)$, $\mathbf{e}_p(n)$, and $\hat{\mathbf{h}}(n)$ are independent of $\mathbf{x}(n)$ in the steady state, we get

$$\begin{aligned} E[\mathbf{e}_p^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}_p(n)] \\ = E[\mathbf{e}_a^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}_a(n)] \\ - \frac{\gamma_{\text{SL0}}^2}{4} E\{\mathbf{T}^T(n) \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n)\} \\ + \frac{\gamma_{\text{SL0}}^2}{4} E[\mathbf{T}^T(n) \mathbf{T}(n)]. \end{aligned} \quad (74)$$

Substituting (71) into the left-hand side (LHS) of (74), we have

$$\begin{aligned} \text{LHS} &= (1 - \mu_{\text{SL0}})^2 E\{\mathbf{e}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n)\} \\ &- (1 - \mu_{\text{SL0}}) E\{\mathbf{e}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{v}(n)\} \\ &- (1 - \mu_{\text{SL0}}) E\{\mathbf{v}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n)\} \\ &+ E\{\mathbf{v}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{v}(n)\}. \end{aligned} \quad (75)$$

Moreover, substituting (70) into the right-hand side (RHS) of (74), we have

$$\begin{aligned} \text{RHS} &= E\{\mathbf{e}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n)\} \\ &- E\{\mathbf{e}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{v}(n)\} \\ &- E\{\mathbf{v}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n)\} \\ &+ E\{\mathbf{v}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{v}(n)\} \\ &- \frac{\gamma_{\text{SL0}}^2}{4} E\{\mathbf{T}^T(n) \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n)\} \\ &+ \frac{\gamma_{\text{SL0}}^2}{4} E[\mathbf{T}^T(n) \mathbf{T}(n)]. \end{aligned} \quad (76)$$

By combining (75) and (76), we get

$$\begin{aligned} (2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2) E\{\mathbf{e}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n)\} \\ = \mu_{\text{SL0}} E\{\mathbf{e}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{v}(n)\} \\ + \mu_{\text{SL0}} E\{\mathbf{v}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n)\} \\ + \frac{\gamma_{\text{SL0}}^2}{4} E\{\mathbf{T}^T(n) \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n)\} \\ - \frac{\gamma_{\text{SL0}}^2}{4} E[\mathbf{T}^T(n) \mathbf{T}(n)]. \end{aligned} \quad (77)$$

We also assume that the additive Gaussian noise $\mathbf{v}(n)$ is statistically independent of the input signal $\mathbf{x}(n)$. Thus (77) can be simplified as

$$\begin{aligned} E\{\mathbf{e}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n)\} \\ = \frac{1}{2 - \mu_{\text{SL0}}} E\{\mathbf{v}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{v}(n)\} \\ + \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} E\{\mathbf{T}^T(n) \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n)\} \\ - \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} E[\mathbf{T}^T(n) \mathbf{T}(n)]. \end{aligned} \quad (78)$$

Here, we also assume that the $\mathbf{U}(n)$ is statistically independent of $\mathbf{e}(n)$ at the steady state. Moreover, we use the definition of $E[\mathbf{e}^T(n)\mathbf{e}(n)] = E[e_t(n)]^2 \mathbf{S}$ [36], where

$$\mathbf{S} \approx \begin{cases} \mathbf{I}, & \mu_{\text{SL0}} \text{ is small} \\ \mathbf{1} \cdot \mathbf{1}^T, & \mu_{\text{SL0}} \text{ is large,} \end{cases} \quad (79)$$

where $\mathbf{1}^T = [1 \ 0 \ \dots \ 0]$ and $e_t(n)$ is the top entry of $\mathbf{e}(n)$ [36]. Then, the LHS of (78) can be rewritten as

$$\begin{aligned} E\{\mathbf{e}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{e}(n)\} \\ \approx \text{Tr}\{E[\mathbf{e}^T(n) \mathbf{e}(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1}]\} \\ \approx E[e_t(n)]^2 \text{Tr}\{\mathbf{S} \cdot E[[\mathbf{U}(n) \mathbf{U}^T(n)]^{-1}]\}. \end{aligned} \quad (80)$$

Similar to the calculation of (80), the first term in the RHS of (78) can be written as

$$\begin{aligned} E\{\mathbf{v}^T(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1} \mathbf{v}(n)\} \\ = \text{Tr}\{E[\mathbf{v}^T(n) \mathbf{v}(n) [\mathbf{U}(n) \mathbf{U}^T(n)]^{-1}]\} \\ = Q\delta_v^2 \text{Tr}\{E[[\mathbf{U}(n) \mathbf{U}^T(n)]^{-1}]\}. \end{aligned} \quad (81)$$

In addition, the second term of RHS of (78) can be rewritten as

$$\begin{aligned} & \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} \mathbb{E} \left\{ \mathbf{T}^T(n) \mathbf{U}^+(n) \mathbf{U}(n) \mathbf{T}(n) \right\} \\ & \approx \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} \frac{\delta \text{sgn}^T(\mathbf{h})}{(\mathbf{h}^T + \delta)^2} \mathbb{E} \left\{ \mathbf{U}^+(n) \mathbf{U}(n) \right\} \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2}. \end{aligned} \quad (82)$$

Then the last term of the right-hand side of (78) can be expressed as

$$\begin{aligned} & \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} \mathbb{E} \left[\mathbf{T}(n)^T \mathbf{T}(n) \right] \\ & = \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} \frac{\delta \text{sgn}^T(\mathbf{h}) \delta \text{sgn}(\mathbf{h})}{(\mathbf{h}^T + \delta)^2 (\mathbf{h} + \delta)^2}. \end{aligned} \quad (83)$$

When the μ_{SL0} is small, we can get

$$\begin{aligned} & \text{Tr} \left\{ \mathbf{S} \cdot \mathbb{E} \left[\left[\mathbf{U}(n) \mathbf{U}^T(n) \right]^{-1} \right] \right\} \\ & = \text{Tr} \left\{ \mathbf{I} \cdot \mathbb{E} \left[\left[\mathbf{U}(n) \mathbf{U}^T(n) \right]^{-1} \right] \right\} \\ & = \frac{Q}{N\delta_x^2}. \end{aligned} \quad (84)$$

Therefore, the MSE of the proposed SL0-APA with small step-size μ_{SL0} can be written as

$$\begin{aligned} \text{MSE}_{\text{small}} & = \frac{1}{2 - \mu_{\text{SL0}}} \delta_v^2 \\ & + \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} \frac{\delta \text{sgn}^T(\mathbf{h})}{(\mathbf{h}^T + \delta)^2} \frac{N\delta_x^2}{Q} \frac{\mathbf{Q}\mathbf{R}}{N\delta_x^2} \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2} \\ & - \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} \frac{\delta \text{sgn}^T(\mathbf{h})}{(\mathbf{h}^T + \delta)^2} \frac{N\delta_x^2}{Q} \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2} \\ & = \frac{1}{2 - \mu_{\text{SL0}}} \delta_v^2 + \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} \frac{\delta \text{sgn}^T(\mathbf{h})}{(\mathbf{h}^T + \delta)^2} \\ & \quad \times \left(\mathbf{R} - \frac{N\delta_x^2}{Q} \mathbf{I} \right) \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2}. \end{aligned} \quad (85)$$

When the step-size μ_{SL0} is large, $\mathbf{S} \approx \mathbf{1} \cdot \mathbf{1}^T$ [36]. In this case,

$$\text{Tr} \left\{ \mathbf{S} \cdot \mathbb{E} \left[\left[\mathbf{U}(n) \mathbf{U}^T(n) \right]^{-1} \right] \right\} = \frac{1}{N\delta_x^2}. \quad (86)$$

FIGURE 2: A typical sparse multipath channel.

Thus, the MSE of the proposed SL0-APA with large step-size μ_{SL0} can be written as

$$\begin{aligned} \text{MSE}_{\text{large}} & = \frac{1}{2 - \mu_{\text{SL0}}} \delta_v^2 Q \\ & + \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} \frac{\delta \text{sgn}^T(\mathbf{h})}{(\mathbf{h}^T + \delta)^2} N\delta_x^2 \frac{\mathbf{Q}\mathbf{R}}{N\delta_x^2} \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2} \\ & - \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} \frac{\delta \text{sgn}^T(\mathbf{h})}{(\mathbf{h}^T + \delta)^2} N\delta_x^2 \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2} \\ & = \frac{1}{2 - \mu_{\text{SL0}}} \delta_v^2 Q + \frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}} - \mu_{\text{SL0}}^2)} \frac{\delta \text{sgn}^T(\mathbf{h})}{(\mathbf{h}^T + \delta)^2} \\ & \quad \times \left(\mathbf{Q}\mathbf{R} - N\delta_x^2 \mathbf{I} \right) \frac{\delta \text{sgn}(\mathbf{h})}{(\mathbf{h} + \delta)^2}. \end{aligned} \quad (87)$$

4. Results and Discussions

In this section, we present the computer simulation results to illustrate the performance of the proposed SL0-APA over a sparse multipath communication channel. Moreover, the simulation results for predicting the mean-square error of the proposed SL0-APA are also provided to verify the effectiveness of the theoretical expressions obtained in Section 3.2. In addition, the computational complexity of the SL0-APA is presented and compared with past sparsity-aware algorithms, namely, the ZA-APA, RZA-APA, and standard APA, NLMS algorithms.

4.1. Performance of the Proposed SL0-APA. Firstly, we set up a simulation example to discuss the convergence speed of

FIGURE 3: Convergence of the proposed SL0-APA.

FIGURE 4: Affine projection order effects on the SL0-APA.

the proposed SL0-APA in comparison with the previously proposed sparse channel estimation algorithms including the APA, ZA-APA, RZA-APA, and NLMS algorithms. In the setup of this experiment, we consider a sparse multipath communication channel $\mathbf{h}$ whose length N is equal to 16 and whose number of dominant taps K is set to two different

sparsity levels, namely, $K = 1, K = 4$, similarly to [6, 22, 25, 26]. The dominant channel taps are obtained from a Gaussian distribution subjected to $\|\mathbf{h}\|_2^2 = 1$, and the positions of the dominant channel taps are random within the length of the channel. The input signal $\mathbf{x}(n)$ of the channel is a Gaussian random signal, while the output of the channel is corrupted by an independent white Gaussian noise $\mathbf{v}(n)$. An example of a typical sparse multipath channel with a channel length of $N = 16$ and a sparsity level of $K = 3$ is shown in Figure 2. In the simulations, the power of the received signal is $E_b = 1$, while the noise power is given by δ_v^2 . In all the experiments, the difference between the actual and estimated channels based on the sparsity-aware algorithms and the sparse channel mentioned above is evaluated by the MSE defined as follows:

$$\text{MSE}(n) = 10 \log_{10} E \left\{ \|\mathbf{h} - \hat{\mathbf{h}}(n)\|_2^2 \right\} \text{ (dB)}. \quad (88)$$

In this subsection, we aim to investigate the convergence speed and the steady-state performance of the SL0-APA. The simulation parameters used to compare the convergence speed while maintaining the same MSE are listed as follows: $\mu_{\text{NLMS}} = 0.25, \mu_{\text{APA}} = 0.125, \mu_{\text{ZA}} = 0.165, \mu_{\text{RZA}} = 0.18, \mu_{\text{SL0}} = 0.21, \gamma_{\text{ZA}} = 5 \times 10^{-5}, \gamma_{\text{RZA}} = 8 \times 10^{-5}, \gamma_{\text{SL0}} = 3 \times 10^{-6}, \epsilon_{\text{RZA}} = 10, \delta_{\text{SL0}} = 0.001, Q = 2$, and $\delta_v^2 = 10^{-3}$, where μ_{NLMS} is the step-size parameter for NLMS algorithm. It can be seen from Figure 3 that our proposed SL0-APA possesses the fastest convergence speed compared to the previously proposed channel estimation algorithms at the same steady-state error floor. In addition, all the affine projection algorithms, namely, APA, ZA-APA, RZA-APA, and SL0-APA, converge much more quickly in comparison with NLMS algorithm, because the affine projection algorithms reuse the old data signal that is implemented by the use of parameter Q . Thus, we discuss the effects of the affine projection order Q for SL0-APA and compare it with the APA and NLMS algorithms. The computer simulation results with different values of Q are shown in Figure 4. It reveals that the convergence speed is improved by the increment of the affine projection order Q . However, the steady-state performance has deteriorated from $Q = 2$ to $Q = 8$. Thus, in our proposed SL0-APA, the affine projection Q , the step-size μ_{SL0} , the regularization parameter γ_{SL0} , and δ_{SL0} should be taken into account to balance the convergence speed and the steady-state behavior.

Next, we show the effects of the sparsity levels on the steady-state performance of the proposed SL0-APA at $K = 1$ and $K = 4$. To obtain the same convergence speed, the simulation parameters used in this experiment are listed as follows: $\mu_{\text{NLMS}} = 0.095, \mu_{\text{APA}} = \mu_{\text{ZA}} = \mu_{\text{RZA}} = \mu_{\text{SL0}} = 0.05, \gamma_{\text{ZA}} = 5 \times 10^{-5}, \gamma_{\text{RZA}} = 8 \times 10^{-5}, \gamma_{\text{SL0}} = 4 \times 10^{-6}, \epsilon_{\text{RZA}} = 10, \delta = 0.01$, and $\delta_v^2 = 10^{-3}$. We can see from Figure 5 that our proposed SL0-APA has the best steady-state performance compared to the ZA-APA, RZA-APA, APA, and NLMS algorithms. The SL0-APA can achieve 10 dB smaller MSE than the RZA-APA for $K = 1$ and $Q = 2$ shown in Figure 5(a). When the sparsity level K increases to 4, it is seen in Figure 5(b) that our proposed SL0-APA still outperforms other algorithms, while its steady-state error increases in comparison with that of $K = 1$. When the affine projection

FIGURE 5: Performances of the SL0-APA with different sparsity levels for $Q = 2$.

order increases to $Q = 3$, we can see from Figure 6 that the convergence speed is significantly improved compared to that of $Q = 2$ shown in Figure 5. However, the steady-state error is also slightly increased when the Q increases. Furthermore, our proposed SL0-APA still has the best convergence speed and lowest steady-state error.

Finally, we use the theoretical expressions obtained in Section 3.2 to predict the mean-square-error (MSE) of the proposed SL0-APA with different μ_{SL0} and compare the theoretical results with the simulation ones. The MSE comparisons of the SL0-APA as a function of the step-size μ_{SL0} for the designated sparse multipath communication channel with the simulation parameters of $\gamma_{\text{SL0}} = 4 \times 10^{-6}$, $\delta = 0.01$, $\delta_v^2 = 10^{-3}$, $Q = 3$, and $K = 1$ are shown in Figure 7. The theoretical results are obtained from (85) to (87) for small values of μ_{SL0} and large values of μ_{SL0} , respectively, while the simulation results are obtained by averaging 50 independent trials. We can see that the simulation results exhibit good agreement with the theoretical expressions with different step-size μ_{SL0} . In addition, we can see that the steady-state misadjustment between the computer simulation and the theory predicting is becoming larger with the decrement of the μ_{SL0} for small μ_{SL0} shown in Figure 7(a), but the steady-state error is becoming lower. For the large μ_{SL0} , both the steady-state error and the convergence speed are deteriorated by the increment of the step-size μ_{SL0} . Generally speaking, as μ_{SL0} increases, the MSE increases. Although a large zero attractor can help the SL0-APA to converge faster, it will lead to a higher misadjustment. Thus, in the most cases, we should choose the step-size μ_{SL0} carefully in order to balance convergence speed and steady-state performance.

4.2. Computational Complexity. In this subsection, we present the computational complexity of the proposed SL0-APA and compare it with the conventional sparsity-aware channel estimation algorithms, including the APA, ZA-APA, and RZA-APA. It is worth noting that when the affine projection order Q is equal to 1, these three affine projection algorithms converge to familiar NLMS, ZA-NLMS, and RZA-NLMS algorithms, respectively. Here, the computational complexity is the arithmetic complexity, which includes additions, multiplications, and divisions. We assume K nonzero taps in a sparse channel model as an FIR filter with N coefficients, and the order of these affine projection algorithms is Q . The computational complexity of the proposed SL0-APA and the relevant sparsity-aware algorithms are shown in Table 1.

From Table 1, we see that our proposed SL0-APA with the best steady-state performance and fastest convergence speed needs more calculations than the RZA-APA. The additional computational complexity comes from the continuous function for SL0 approximation, which can be reduced by proper selection of this continuous function. Furthermore, the complexity of all the APAs is higher than the NLMS algorithms. In addition, the sparsity property of the channel can also help to reduce the computational complexity of the proposed SL0-APA.

5. Conclusion

In this paper, we proposed an SL0-APA to exploit the sparsity of sparse channel and to improve the performance on both

FIGURE 6: Performances of the SL0-APA with different sparsity levels for $Q = 3$. FIGURE 7: Steady-state MSE performance of the SL0-APA with different step-size μ_{SL0} for $K = 4$.

TABLE 1: Computational complexity.

Algorithms	Additions	Multiplications	Divisions
NLMS	$3N$	$3N + 1$	1
ZA-NLMS	$N + 3K$	$N + 3K + 1$	1
RZA-NLMS	$N + 4K$	$N + 4K + 1$	$N + 1$
APA	$NQ^2 + NQ + K - N + O(Q^3)$	$NQ^2 + Q^2 + NQ + Q + O(Q^3)$	Q
ZA-APA	$NQ^2 + Q^2 + 3NQ - 2Q - 2N + 3K + O(Q^3)$	$NQ^2 + 2Q^2 + 3NQ + 2Q + K + O(Q^3)$	Q
RZA-APA	$NQ^2 + Q^2 + 3NQ - 2Q - 2N + 4K + O(Q^3)$	$NQ^2 + 2Q^2 + 3NQ + Q + N + 2K + O(Q^3)$	$N + Q$
SL0-APA	$NQ^2 + Q^2 + 3NQ - 2Q - 2N + 4K + O(Q^3)$	$NQ^2 + 2Q^2 + 3NQ + Q + N^2 - N + 2K + O(Q^3)$	$N + Q$

the convergence speed and steady-state error of the APA, ZA-APA, and RZA-APA. This algorithm is mainly developed by introducing a smooth approximation l_0 -norm, which has a significant impact on the sparsity due to the incorporation of SL0 into the cost function of the standard APA as an additional constraint. The improvement can evidently accelerate the convergence speed by exerting such additional regularization term on the zero taps of the sparse channel. Then, we provided a mathematical analysis for predicting the mean square error of our proposed SL0-APA. We also showed the convergence behavior and the steady-state performance in comparison with the standard APA and relevant sparsity-aware channel estimation algorithms. In summary, the simulation results demonstrated that the proposed SL0-APA with moderate computational complexity accelerates convergence speed and improves steady-state performance in a designated sparse channel.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors derived the SL0-APA in Section 3.1 by considering the l_0 -norm-based LMS and NLMS [6, 26] and the smooth approximation l_0 -norm in [21]. In the paper [42], l_0 -norm constrained APA was also derived around the same time as their paper was submitted for publication. The authors thank the editors for the constructive comments to clarify this paper.

References

- [1] L. Korowajczuk, *LTE, WiMAX and WLAN Network Design, Optimization and Performance Analysis*, John Wiley & Sons, New York, NY, USA, 2011.
- [2] J. G. Proakis, *Digital Communications*, McGraw-Hill, 4th edition, 2001.
- [3] F. Adachi, D. Grag, S. Takaoka, and K. Takeda, "New direction of broadband wireless technology," *Wireless Communications and Mobile Computing*, vol. 7, no. 8, pp. 969–983, 2007.
- [4] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 374–377, 2002.
- [5] P. Maechler, P. Greisen, B. Sporrer, S. Steiner, N. Felber, and A. Burg, "Implementation of greedy algorithms for LTE sparse channel estimation," in *Proceedings of the Conference Record of the 44th Asilomar Conference on Signals, Systems and Computers (ASILOMAR '10)*, pp. 400–405, Pacific Grove, Calif, USA, November 2010.
- [6] G. Gui and F. Adachi, "Improved least mean square algorithm with application to adaptive sparse channel estimation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, article 204, 2013.
- [7] J. Meng, W. Yin, Y. Li, N. T. Nguyen, and H. Zhu, "Compressive sensing based high-resolution channel estimation for OFDM system," *IEEE Journal on Selected Topics in Signal Processing*, vol. 6, no. 1, pp. 15–25, 2012.
- [8] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: a new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.
- [9] G. Tauböck and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: exploiting channel sparsity for reducing pilots," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 2885–2888, Las Vegas, Nev, USA, April 2008.
- [10] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 4th edition, 2001.
- [11] Y. Huang and J. Wen, "An analysis on partial PIC multi-user detection with LMS algorithms for CDMA," in *Proceedings of the 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '03)*, vol. 1, pp. 17–21, Beijing, China, September 2003.
- [12] M. Huang, X. Chen, L. Xiao, S. Zhou, and J. Wang, "Kalman-filter-based channel estimation for orthogonal frequency-division multiplexing systems in time-varying channels," *IET Communications*, vol. 1, no. 4, pp. 795–801, 2007.
- [13] T. K. Akino, "Optimum-weighted RLS channel estimation for rapid fading MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4248–4260, 2008.
- [14] M. Godavarti and A. O. Hero, "Partial update LMS algorithms," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2382–2399, 2005.
- [15] S. Werner, M. L. R. de Campos, and P. S. R. Diniz, "Partial-update NLMS algorithms with data-selective updating," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 938–949, 2004.
- [16] P. A. Naylor and A. W. H. Khong, "Affine projection and recursive least squares adaptive filters employing partial updates," in *Proceedings of the 38th Asilomar Conference on Signals, Systems*

- and Computers, vol. 1, pp. 950–954, Pacific Grove, Calif, USA, November 2004.
- [17] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
 - [18] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
 - [19] N. Vaswani, “Kalman filtered compressed sensing,” in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP '08)*, pp. 893–896, San Diego, Calif, USA, October 2008.
 - [20] E. M. Eksioğlu, “Sparsity regularised recursive least squares adaptive filtering,” *IET Signal Processing*, vol. 5, no. 5, pp. 480–487, 2011.
 - [21] R. E. Carrillo and K. E. Barner, “Iteratively re-weighted least squares for sparse signal reconstruction from noisy measurements,” in *Proceedings of the 43rd Annual Conference on Information Sciences and Systems (CISS '09)*, pp. 448–453, Baltimore, Md, USA, March 2009.
 - [22] Y. Chen, Y. Gu, and A. O. Hero, “Sparse LMS for system identification,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 3125–3128, Taipei, Taiwan, April 2009.
 - [23] K. Shi and P. Shi, “Convergence analysis of sparse LMS algorithms with l_1 -norm penalty based on white input signal,” *Signal Processing*, vol. 90, no. 12, pp. 3289–3293, 2010.
 - [24] F. Y. Wu and F. Tong, “Gradient optimization p -norm-like constraint LMS algorithm for sparse system estimation,” *Signal Processing*, vol. 93, no. 4, pp. 967–971, 2013.
 - [25] O. Taheri and S. A. Vorobyov, “Sparse channel estimation with L_p -norm and reweighted L_1 -norm penalized least mean squares,” in *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 2864–2867, Prague, Czech Republic, May 2011.
 - [26] Y. Gu, J. Jin, and S. Mei, “ L_0 norm constraint LMS algorithm for sparse system identification,” *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 774–777, 2009.
 - [27] R. Niazaideh, S. H. Ghalehjagh, M. Babaie-Zadeh, and C. Jutten, “ISI sparse channel estimation based on SL0 and its application in ML sequence-by-sequence equalization,” *Signal Processing*, vol. 92, no. 8, pp. 1875–1885, 2012.
 - [28] P. S. R. Diniz, *Adaptive Filtering Algorithms and Practical Implementation*, Springer, 4th edition, 2013.
 - [29] H. Shin, A. H. Sayed, and W. Song, “Variable step-size NLMS and affine projection algorithms,” *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 132–135, 2004.
 - [30] R. Arablouei and K. Dogançay, “Affine projection algorithm with selective projections,” *Signal Processing*, vol. 92, no. 9, pp. 2253–2263, 2012.
 - [31] S. Werner, J. A. Apolinário Jr., M. L. R. de Campos, and P. S. R. Diniz, “Low-complexity constrained affine-projection algorithms,” *IEEE Transactions on Signal Processing*, vol. 53, no. 12, pp. 4545–4555, 2005.
 - [32] R. Meng, R. C. Lamare, and V. H. Nascimento, “Sparsity-aware affine projection adaptive algorithms for system identification,” in *Proceedings of the Sensor Signal Processing for Defence Conference (SSPD '11)*, pp. 1–5, London, UK, September 2011.
 - [33] E. J. Candès, M. B. Wakin, and S. P. Noyd, “Enhancing sparsity by reweighted l_1 -minimization,” *Journal of Fourier Analysis and Applications*, vol. 15, no. 5-6, pp. 877–905, 2008.
 - [34] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, “A fast approach for overcomplete sparse decomposition based on smoothed l^0 norm,” *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, 2009.
 - [35] M. Hyder and K. Mahata, “An approximate l_0 norm minimization algorithm for compressed sensing,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 3365–3368, Taipei, Taiwan, April 2009.
 - [36] H. Shin and A. H. Sayed, “Mean-square performance of a family of affine projection algorithms,” *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 90–102, 2004.
 - [37] R. Meng, *Sparsity-Aware Adaptive Filtering Algorithms and Application to System Identification*, University of York, 2011.
 - [38] H. Shin, W. Song, and A. H. Sayed, “Mean-square performance of data-reusing adaptive algorithms,” *IEEE Signal Processing Letters*, vol. 12, no. 12, pp. 851–854, 2005.
 - [39] G. Su, J. Jin, Y. Gu, and J. Wang, “Performance analysis of l_0 norm constraint least mean square algorithm,” *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2223–2235, 2012.
 - [40] G. Barrault, M. H. Costa, J. C. M. Bermudez, and A. Lenzi, “A new analytical model for the NLMS algorithm,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 4, pp. IV41–IV44, Philadelphia, Pa, USA, March 2005.
 - [41] M. H. Costa and J. C. M. Bermudez, “An improved model for the normalized LMS algorithm with Gaussian inputs and large number of coefficients,” in *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP '02)*, vol. 2, pp. 1385–1388, Orlando, Fla, USA, May 2002.
 - [42] M. V. S. Lima, W. A. Martins, and P. S. Z. Diniz, “Affine projection algorithms for sparse system identification,” in *Proceedings of the IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP '13)*, pp. 5666–5670, Vancouver, Canada, May 2013.