Developments of Adaptive Filter Algorithms for Sparse Channel Estimation

Yingsong Li

A dissertation submitted to
Kochi University of Technology
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Graduate School of Engineering

Kochi University of Technology

Kochi, Japan

March 2014

I would like to dedicate this thesis to my loving parents.

Acknowledgements

First of all, I would like to give my deep and sincere gratitude to my supervisor Prof. Masanori Hamamura for his great guidance and encouragement on my research and study. Prof. Masanori Hamamura devotes his time into my research and leads me into the mystical and magical field of signal processing for wireless communication. Furthermore, his research attitude, his strong enthusiasm, his constructive enlightenment and his broad foresight, as well as his broad and profound knowledge in wireless communication and signal processing, have given rise to a deep impact on me, and undoubtedly will have a far-reaching influence on my future research and work. The research results proposed in this thesis would not be carried out without his continuous guidance, his instructive suggestion and his heuristic opinion. I am grateful for his persevered effort to track the novel and interesting topics in my research.

Next, I am appreciated Kochi University of Technology (KUT) for giving me a chance to study in Japan and offering finance support for me under special scholarship program (SSP). Without SSP support for the international students, I cannot pursuit my Ph.D degree abroad. The financial support and the SSP support is appreciatively acknowledged.

My sincere thanks also turn to Prof. Kazunori Shimamura, Prof. Xiangshi Ren, Prof. Katsushi Iwashita and Prof. Masahiro Fukumoto for their enlightening discussion and constructive suggestion in my research and my dissertation. I also would like to thank Prof. Chaoyang Li for her kind help for my life in Japan. Moreover, I appreciate the secretaries of the School of Information and the staffs of the International Relations Division for their kind-hearted help.

I owe my loving thanks to my parents and my girlfriend who give me spiritual encouragement and support. I also would like to thank Dr. Yafei Hou gives me much encouragement and useful suggestion.

Finally, I am grateful for all the members in Hamamura Lab. Especially, Mihiro Nojima, Yuya Hayashi, Naoya Kitamura, Kenta Fukunaga, Ryo Utsunomiya, Shougo Kuwamura, Yoshimasa Hashida, Kento Hachikubo, Kittipong Warasup and Saifur Rahman Sabuj for their helpful discussion and useful advice. I would never forget their kind and warm help in my life and their encouragement in my research.

Abstract

Developments of Adaptive Filter Algorithms for Sparse Channel Estimation

Yingsong Li

Broadband signal transmission becomes a commonly used high-data-rate technique for next-generation wireless communication systems, and the coherent detection for such broadband communication systems strongly depends on the quality of the channel estimation which can be well estimated by using adaptive filters. Furthermore, channel measurements have shown that the broadband wireless multipath channels can often be described as sparse channels. Thus, developing sparse adaptive filter algorithms for broadband multipath estimation is becoming an increasing important research topic. In this dissertation, we aim to develop efficient sparse adaptive filter algorithms for sparse channel estimation applications in terms of the convergence speed and steady-state performance.

Firstly, we studies the two classes of important adaptive filter algorithms: classical adaptive filter algorithms including least-mean-square (LMS) and affine projection algorithm (APA), and sparsity-aware adaptive filter algorithms which include the zero attracting (ZA) algorithms and the proportionate-type adaptive filter algorithms. We found that the classical adaptive filters have good performance for non-sparse signal estimation while these algorithms perform poorly for the sparse channel estimation. Furthermore, ZA and

proportionate-type algorithms have been proposed for sparse system identifications and echo cancellation applications. However, most of the ZA algorithms were proposed based on the l_1-norm penalty and LMS algorithms while most of the improved proportionate-type algorithms were concentrated on variable step size technique and gain matrix modification.

Secondly, we proposed an adaptive reweighted zero-attracting sigmoid functioned variable step size LMS (ARZA-SVSS-LMS) algorithm based on variable step size technique and adaptive parameter adjustment method. In order to implement the ARZA-SVSS-LMS algorithm, it was described step-by-step. To begin with, a sigmoid functioned variable step size LMS (SVSS-LMS) algorithm was proposed, which was an improved variable step size (VSS) LMS algorithm. Next, the ZA techniques used in zero-attracting LMS (ZA-LMS) and reweighted ZA-LMS (RZA-LMS) algorithms were incorporated into the proposed SVSS-LMS algorithm in order to form the zero-attracting SVSS-LMS (ZA-SVSS-LMS) and reweighted ZA-SVSS-LMS (RZA-SVSS-LMS) algorithms, respectively. At last, an adaptive parameter adjustment method was adopted to form the ARZA-SVSS-LMS algorithm by adjusting the zero-attracting strength in the RZA-SVSS-LMS algorithm dynamically. The simulation results demonstrated that the proposed ARZA-SVSS-LMS algorithm can achieve faster convergence speed and smaller steady-state error in comparison with these of the standard LMS and previously proposed sparsity-aware LMS algorithms.

Thirdly, we proposed a smooth approximation l_0-norm-constrained affine projection algorithm (SL0-APA) to obtain the benefits of both the APA and ZA algorithms. The proposed SL0-APA algorithm was realized via incorporating a smooth approximation l_0-norm (SL0) into the cost function of the standard APA in order to construct a zero attractor, by which the convergence speed and the steady-state performance of the standard APA were significantly improved. Moreover, the theoretical analysis of the convergence speed and mean square error (MSE) were given to further understand the proposed SL0-APA. The simulation results showed that the proposed SL0-APA can achieve faster convergence speed and better steady-state performance than the standard APA, zero-attracting affine projection algorithm (ZA-APA) and reweighted ZA-APA (RZA-APA). In addition, we also proposed a discrete weighted zero-attracting affine projection algorithm (DWZA-APA) in order to reduce the computation complexity of the RZA-APA by the introduction of a piece-wise linear function instead of the sum-logarithm function in the RZA-APA. The simulation results demonstrated that the DWZA-APA reduced the multiplication complexity of the RZA-APA and had no channel estimation performance reduction in comparison with the RZA-APA in terms of the convergence speed and steady-state performance.

Finally, we proposed an l_p-norm-constrained PNLMS (LP-PNLMS) algorithm on the basis of the proportionate normalized least-mean-square (PNLMS) algorithm to avail both the benefits of the PNLMS algorithm and ZA techniques. The proposed LP-PNLMS algorithm was realized by incorporating a gain-matrix-weighted l_p-norm penalty into the cost function of the PNLMS algorithm in order to design a zero attractor. The simulation results showed that the proposed LP-PNLMS algorithm can achieve the same convergence speed as that of the PNLMS algorithm at the early iterations, and converged faster than that of the PNLMS algorithm after the convergence of the active taps. Furthermore, the LP-PNLMS algorithm has smaller steady-state error than the PNLMS and its related commonly used algorithms, namely improved PNLMS (IPNLMS) and μ-law PNLMS (MPNLMS) algorithms.

Keywords: Least-mean-square, adaptive filter, variable step size, affine projection algorithm, l_1-norm, smooth approximation l_0-norm, l_p-norm, proportionate normalized least-mean-square, sparse channel estimation, compressed sensing, zero-attracting

Contents

Contents **vii**

List of Figures **xi**

1 Introduction **1**

1.1 Background overview 2

1.1.1 Sparse signal definition 2

1.1.2 Broadband multipath channel 2

1.2 Sparse signal measurement 4

1.3 Sparse channel estimation techniques 5

1.3.1 Compressed sensing 6

1.3.2 Adaptive filters 7

1.4 Challenges for sparse channel estimation 8

1.5 Motivation 9

1.6 Contributions of this thesis 10

1.7 Outline 11

2 Adaptive filter algorithms for sparse channel estimation **13**

2.1 Purpose of adaptive channel estimation 13

2.2 LMS algorithm and its variants 14

2.2.1 LMS algorithm 14

2.2.2 VSS-LMS algorithm 16
2.2.3 NLMS algorithm 17
2.3 Zero-attracting LMS algoritms 17
2.3.1 ZA-LMS algorithm 17
2.3.2 RZA-LMS algorithm 19
2.3.3 ZA-NLMS algorithm 20
2.3.4 RZA-NLMS algorithm 21
2.4 APA and its zero-attracting algorithms 21
2.4.1 Affine projection algorithm (APA) 22
2.4.2 Zero-attracting affine projection algorithm (ZA-APA) . . . 26
2.4.3 Reweighted zero-attracting affine projection algorithm (RZA-APA) 28
2.5 Proportionate-type adaptive filters 30
2.5.1 Proportionate normalized least mean square algorithm . . 31
2.5.2 Improved IPNLMS algorithm 32
2.5.3 μ-law PNLMS algorithm 33
2.6 Conclusion 33

3 Zero-Attracting Variable Step Size Least-Mean-Square Algorithms 35

3.1 Introduction 35
3.2 Proposed ARZA-SVSS-LMS algorithms 36
3.2.1 Proposed SVSS-LMS algorithm 36
3.2.2 Proposed ZA-SVSS-LMS algorithm 39
3.2.3 Proposed RZA-SVSS-LMS algorithm 39
3.2.4 Proposed ARZA-SVSS-LMS algorithm 40
3.3 Analysis of the proposed ARZA-SVSS-LMS algorithm 42
3.4 Performance of the proposed sparse channel estimation algorithms 45
3.4.1 Effects of parameters on the ARZA-SVSS-LMS algorithm . 46
3.4.2 Effects of sparsity level on the ARZA-SVSS-LMS algorithm 49
3.5 Conclusion 54

4 Smooth Approximation l_0-Norm Constrained Affine Projection Algorithm **55**

4.1 Introduction 55

4.2 Proposed SL0-APA algorithm 56

4.3 Analysis of the proposed SL0-APA 59

4.4 Results and discussions 69

4.4.1 Performance of the proposed SL0-APA 69

4.4.2 Computational complexity 74

4.5 Conclusion 75

5 Discrete Weighted Zero-Attracting Affine Projection Algorithm **77**

5.1 Introduction 77

5.2 Proposed DWZA-APA 78

5.3 Results and discussions 79

5.4 Conclusion 84

6 L_p-Norm Constrained Proportionate Normalized Least-Mean-Square Algorithm **87**

6.1 Introduction 87

6.2 Proposed LP-PNLMS algorithm 88

6.3 Results and discussion 92

6.3.1 Estimation performance of the proposed LP-PNLMS algorithm 92

6.3.1.1 Effects of parameters on the proposed LP-PNLMS algorithm 92

6.3.1.2 Effects of sparsity level on the proposed LP-PNLMS algorithm 94

6.3.2 Computational complexity 97

6.4 Conclusion 97

7 Conclusions and Future Research Directions **99**
7.1 Conclusions of this dissertation 99
7.2 Future Research Directions 101
7.2.1 For the ARZA-SVSS-LMS algorithm 101
7.2.2 For the sparse APA algorithms 102
7.2.3 For the sparse proportionate-type algorithms 102
7.2.4 Sparse adaptive filter applications 103

A Proof for Theorem 1 **105**

B Publications **109**
B.1 Refereed journal papers 109
B.2 Conference proceedings 110

References **111**

List of Figures

1.1 Typical multipath communication environment. 3
1.2 Typical sparse multipath channel. 3

2.1 General adaptive filter configuration. 14
2.2 Typical sparse multipath communication system. 15
2.3 Typical sparse multipath communication system based on APA channel estimation algorithms. 22

3.1 Parameter effects on $\mu_{\text{SVSS}}(n)$. 37
3.2 Typical sparse multipath channel. 45
3.3 Effects of parameters on the proposed ARZA-SVSS-LMS algorithm. 47
3.4 Performance of algorithms with variable step size based on sigmoid function. 48
3.5 Performance of the proposed sparse channel estimation algorithms for SNR=20 dB. 50
3.6 Performance of the proposed sparse channel estimation algorithms for SNR=30 dB. 51
3.7 Performance of theARZA-SVSS-LMS algorithm for SNR=20 dB. . 52
3.8 Performance of the ARZA-SVSS-LMS algorithm for SNR=30 dB. 53

4.1 Convergence of the proposed SL0-APA. 70
4.2 Affine projection order effects on the SL0-APA. 71
4.3 Performance of the SL0-APA with different sparsity levels for $Q = 2$. 72
4.4 Performance of the SL0-APA with different sparsity levels for $Q = 3$. 72

4.5 Steady-state MSE performance of the SL0-APA with different step-size μ_{SL0} for $K = 4$. 73

5.1 Effects on the parameter δ_w on the proposed DWZA-APA with $K = 1$. 80
5.2 Convergence of the proposed DWZA-APA with $K = 1$ and $\delta_w = 2$. 81
5.3 Steady-state performance of the proposed DWZA-APA with $K = 1$. 82
5.4 Cluster-sparse multipath fading channel model. 83
5.5 Performance of the DWZA-APA over a cluster-sparse channel. . . 84

6.1 Effects of ρ_{LP} on the proposed LP-PNLMS algorithm. 93
6.2 Effects of p on the proposed LP-PNLMS algorithm. 94
6.3 Effects of sparsity on the proposed LP-PNLMS algorithm for $N = 64$. 95
6.4 Effects of sparsity on the proposed LP-PNLMS algorithm for $N =$ 128. 96

Chapter 1

Introduction

With the rapid development of wireless communication, there has been increasing demand for high-data transmission rates and a wide bandwidth in modern communication systems, which has led to the development of new standards for various wireless devices such as smartphones, laptops and iPads [1]. Given these requirements, broadband signal transmission is a strong candidate and an essential technique for next-generation wireless communication systems. The coherent detection of such broadband communication systems relies heavily on the quality of channel estimation. Fortunately, channel measurement results for these broadband communication systems show that such broadband channels can be regarded as sparse. On the other hand, sparse signal estimation has become an increasingly important research area in signal processing owing to its wide range of applications such as sparse channel estimation, echo cancellation and image processing. Consequently, corresponding algorithms have been proposed for these applications based on sparse signal estimation. In particular, with the development of compressed sensing (CS), many sparse algorithms have been developed and investigated for sparse signal reconstruction applications. Inspired by CS theory, the combination of CS theory and adaptive filters has attracted considerable attention for sparse signal recovery, particularly for sparse system identification. Thus, the development of adaptive filter algorithms for applications based on broadband sparse multipath channel estimation has been a hot topic in recent years.

1.1 Background overview

1.1.1 Sparse signal definition

A sparse signal is defined as a signal with most of its elements equal to zero or close to zero and only a few active components whose magnitudes are nonzero [2–4]. In fact, many practical real-world channels exhibit sparse characteristics [3, 4] such as multipath wireless channels dominated by a relatively small number of significant paths [5–9], frequency-selective channels with a large delay spread and most of the energy localized in small regions with a delay [6, 8, 10, 11], and multicarrier underwater acoustic channels with significant Doppler effects [12].

1.1.2 Broadband multipath channel

Broadband signal transmission is a commonly used high-data-rate technique for modern wireless communication systems [2, 5, 6] such as 3GPP long-term evolution (LTE) and worldwide interoperability for microwave access (WiMAX). Furthermore, the coherent detection for broadband communication systems strongly depends on quality of the channel estimation [8, 13, 14]. On the other hand, channel measurements have shown that broadband wireless multipath channels can often be described by only a small number of propagation paths with long delays [2, 8, 15, 16]. A case of the such broadband wireless communication which is used in hilly terrain environments, is shown in Fig. 1.1. Thus, a broadband multipath channel can be regarded as a sparse channel, having only a few active impulse responses that are dominant while the other inactive taps are zero or close to zero. Such channels, an example of which is given in Fig. 1.2, are described as sparse channels with a few large impulse responses and are encountered in a number of different applications. For instance, in high-definition television

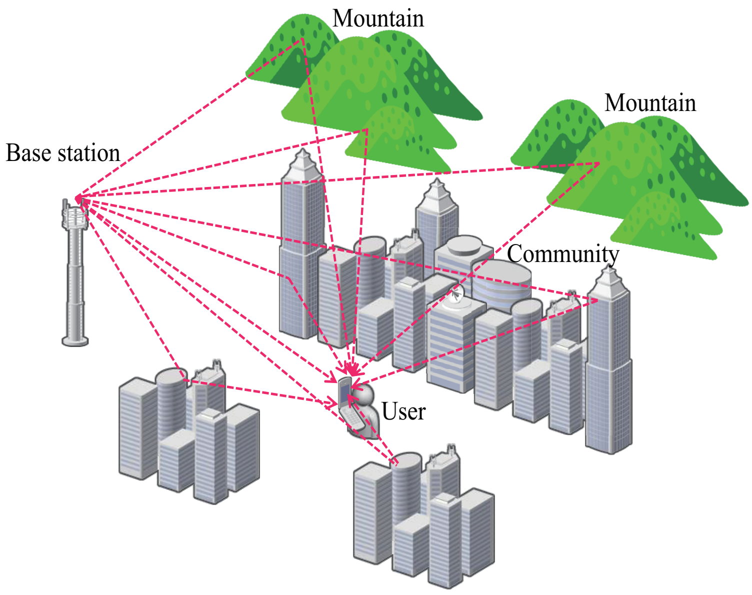


Figure 1.1: Typical multipath communication environment.

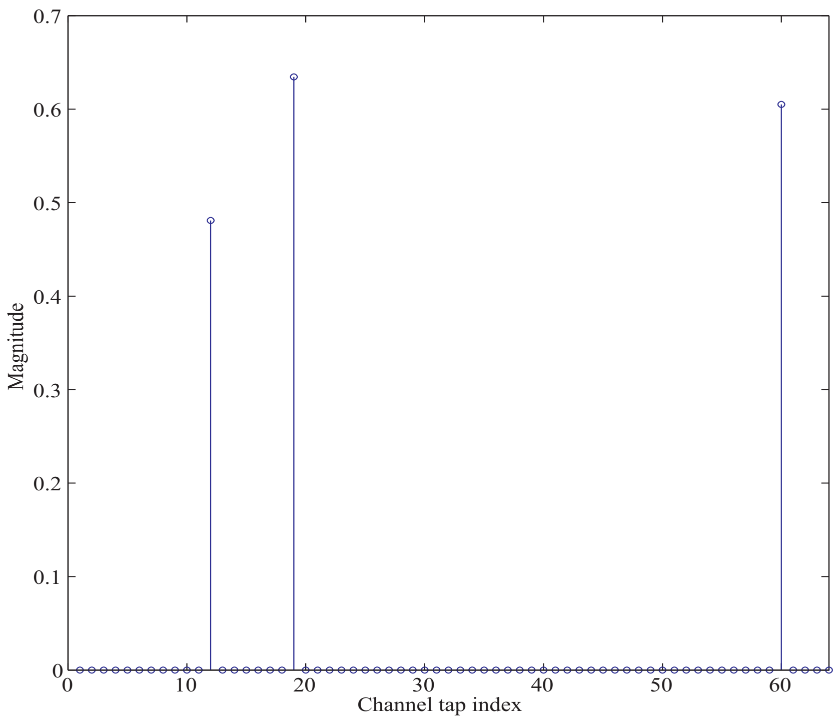


Figure 1.2: Typical sparse multipath channel.

(HDTV), there are few echoes but the channel response spans many hundreds of data symbols [8]. For broadband wireless communications, a “hilly terrain” (HT) delay profile consists of a sparsely distributed multipath, which is shown in Figs. 1.1 and 1.2. In addition, underwater acoustic channels also exhibit a similar response in shallow-sea communication systems [12].

1.2 Sparse signal measurement

Recently, a large number of methods have been proposed and used for sparse signal measurement [17, 18]. One class of effective sparse signal measurement methods involves utilizing vector norms, which are defined as follows. We consider the sparse channel vector

$$\mathbf{h} = [\ h_0 \quad h_1 \quad \cdots \quad h_{N-1}\ ]^T \neq \mathbf{0}, \tag{1.1}$$

where $N > 1$ is the length of the channel memory and $(\cdot)^T$ denotes the transposition operation. The following function is adopted to discuss these norms:

$$f(h_i) = \begin{cases} 1, & h_i \neq 0 \\ 0, & h_i = 0 \end{cases}, \tag{1.2}$$

The number of active taps whose magnitudes are nonzero in $\mathbf{h}$ can be obtained using the l_0-norm and is given by

$$\|\mathbf{h}\|_0 = \sum_{i=0}^{N-1} f(h_i). \tag{1.3}$$

For $\mathbf{h} \neq \mathbf{0}$, we have

$$1 \leq \|\mathbf{h}\|_0 \leq N. \tag{1.4}$$

When $\|\mathbf{h}\|_0 = N$, channel $\mathbf{h}$ is a nonsparse channel. Furthermore, the function $f(h_i)$ is not a continuous function and solving the l_0-norm is a non-polynomial

(NP)-hard problem [19]. In addition, as a sparse channel, many samples of $\mathbf{h}$ can be very small but not exactly zero because of the additive noise in the channel such as additive white Gaussian noise. This makes it difficult to use the l_0-norm defined in (1.3) to accurately measure the sparsity of the channel in practical engineering applications.

The l_1, l_2 and l_p norms of channel $\mathbf{h}$ are adopted to measure the channel and are respectively defined as

$$\|\mathbf{h}\|_1 = \sum_{i=0}^{N-1} |h_i|, \tag{1.5}$$

$$\|\mathbf{h}\|_2 = \sqrt{\sum_{i=0}^{N-1} h_i^2} = \sqrt{\mathbf{h}^T \mathbf{h}}, \tag{1.6}$$

$$\|\mathbf{h}\|_p = (\sum_{i=0}^{N-1} h_i^p)^{1/p}. \tag{1.7}$$

Consequently, these norms have been widely used in compressed sensing for sparse signal recovery applications [19–22]. In this dissertation, we use these norms to develop sparse adaptive channel estimation algorithms.

1.3 Sparse channel estimation techniques

To improve the performance of broadband communication systems, channel estimation schemes have been adopted to reduce the effects of propagation errors and noise in the channel [13, 23–26]. For instance, pilot schemes have been proposed and used in orthogonal frequency-division multiplexing (OFDM) channel estimation [13, 23, 24] to improve the performance of OFDM systems. However, most of the existing pilot-assisted channel estimation schemes rely on the use of a large number of pilots to improve the accuracy of the estimation, which reduces the spectral efficiency. In addition, it is difficult to design pilots and efficient estimation algorithms. In this section, we introduce two classes of popular channel estimation algorithms, namely, compressed sensing and adaptive filter algorithms, described in sections 1.3.1 and 1.3.2, respectively.

1.3.1 Compressed sensing

Recently, compressed sensing (CS), which is also known as compressive sensing or compressed sampling, has emerged as an important topic in signal processing when the signal is highly sparse [3, 4, 19–22, 27–30]. CS is a novel technique that combines signal compression and sparse signal recovery, and deals with the acquisition of sparse signals. The basic mathematical model can be expressed as

$$\mathbf{u} = \mathbf{\Phi s} + \mathbf{n}, \tag{1.8}$$

where $\mathbf{u}$ is the measurement vector, $\mathbf{s}$ is an $N \times 1$ sparse signal with K nonzero elements and $K \ll N$, $\mathbf{\Phi}$ is a measurement matrix with size $M \times N$, $\mathbf{n}$ is an $M \times 1$ noise vectors, and $M \ll N$. Then, the sparse signal can be estimated by using the CS reconstruction algorithms through incomplete measurement $\mathbf{u}$ if the measurement matrix satisfies the restricted isometry property (RIP) [31]. The reconstructed $\hat{\mathbf{s}}$ can be uniquely obtained by solving the following minimization problem:

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \left\{ \frac{1}{2} \|\mathbf{u} - \mathbf{\Phi s}\|_2^2 + \lambda_0 \|\mathbf{s}\|_0 \right\}, \tag{1.9}$$

where λ_0 is the regularized parameter, which is used for balancing the mean square error (MSE) and sparsity of $\mathbf{s}$. However, the l_0-norm is an NP-hard problem, making it difficult to use in practical applications. Consequently, the l_1-norm is adopted as a good approximation for measuring a sparse signal and has been widely investigated [19, 32]. A number of CS reconstruction algorithms have already been proposed for sparse signal estimation [22, 28, 33–37] such as orthogonal matching pursuit (OMP) [28, 33, 34], the least absolute shrinkage and selection operator (LASSO) [35] and iterative support detection (ISD) [36]. Recently, the CS technique has been widely used in image processing and wireless communication [38, 39], where highly sparse signals contain sufficient information to achieve an approximate or exact recovery. In wireless communication, an important application of CS is to estimate the sparse multipath channel $\mathbf{h}$ [39–42]. However, we found that these CS channel estimation algorithms were sensitive to the channel inferences. In addition, the measurement matrices are difficult to design because of the RIP condition and the CS-based channel estimation algorithms have high computational complexity.

1.3.2 Adaptive filters

Adaptive filters, such as least mean squares (LMS), recursive least squares (RLS) and Kalman filter algorithms, have been widely studied owing to their effectiveness for signal estimation and have been applied in channel estimation in wireless communication systems [43–61]. The ability of adaptive filters to satisfactorily operate in an unknown channel and track the time variation of channel statistics makes their elployment a powerful and useful method for channel estimation. The purpose of adaptive channel estimation is to compensate for signal distortion in a wireless multipath propagation channel. In wireless communication systems, a modulated signal is transmitted from one point to another across a communication channel such as a fiber-optic cable or a wireless radio link. During the transmission process, the transmitted signal, which contains important information, may become distorted because of the interference of noise in the communication channel, particularly in a wireless multipath communication channel. Adaptive channel estimation applies an adaptive filter algorithm to the multipath communication channel to compensate for this distortion. Adaptive filters acts as adaptive channel equalizers and have been investigated in code division multiple access (CDMA), orthogonal frequency-division multiplexing (OFDM) and multiple-input and multiple-output (MIMO) communication systems [59, 62–65]. Moreover, these adaptive filters are easy to implement in practical engineering applications. However, classical adaptive filters cannot utilize the sparsity of the broadband multipath communication channel and hence they perform poorly when dealing with sparse signals. Thus, it is necessary to develop effective sparse adaptive filter algorithms for sparse channel estimation to exploit the sparsity of wireless multipath channels.

1.4 Challenges for sparse channel estimation

With the increasing transmission rate and bandwidth in wireless communication, the wireless channel length has increased from tens to hundreds or even thousands, and thus conventional adaptive filters are facing new challenges. On the other hand, the wireless multipath channel is a sparse channel in broadband wireless communication systems for hilly and indoor environments. For these reasons, sparse channel estimation is facing the following challenges:

1. The channel estimation performance of conventional adaptive filter algorithms is reduced for these sparse channel applications in terms of the convergence speed and steady-state performance. Therefore, the development of adaptive filter algorithms for sparse channel estimation, which can utilize the sparsity of the channel and improve the channel estimation performance, is necessary and desirable.

2. The convergence speed of classical adaptive filter algorithms decreases with increasing filter length because of their inversely proportional relationship. The convergence speed can be predicted from the experimental formula $R = 10\mu(2-\mu)/N \ln 10$ [66], where μ is the step size of the adaptive filter. In particular, for the sparse channel estimation, there are many inactive taps that are zero or close to zero, which reduces the convergence speed of classical adaptive filters.

3. The estimation accuracy of conventional adaptive channel estimation algorithms deteriorates in the presence of noise. Furthermore, most of the previously proposed sparse channel estimation algorithms have a constant step size that cannot be adjusted to improve the estimation accuracy.

4. The computational complexity of conventional adaptive channel estimation algorithms increases with the channel length. This increases the cost of consumer devices such as personal wireless communication applications.

Recently, a number of adaptive filter algorithms and their variants have been proposed and investigated to overcome one or more of the above drawbacks. In particular, a class of new adaptive filter algorithms, zero-attracting (ZA) algorithms [67], have been proposed for sparse system identification applications and applied to sparse channel estimation [2]. However, ZA algorithms were mainly realized by the incorporation of the l_1-norm given by (1.5) into the cost function of the standard affine projection algorithm (APA) and LMS algorithms. In addition, they have a fixed step size, which limits the estimation accuracy for sparse channel estimation applications.

1.5 Motivation

In the hilly and indoor wireless communication environments, the wireless multipath channel can be assumed to be sparse, containing only a few active impulse responses with large coefficients interspersed among many inactive ones. On the basis of this prior knowledge, channel estimation can be improved by exploiting the sparsity of the channel. In the past few decades, a number of sparse channel estimation algorithms have been proposed that use a subset selection scheme during the filtering process [68–71], which is implemented via sequential partial updating. Another type of sparse adaptive channel estimation algorithm involves assigning independent step sizes to different taps according to their magnitudes such as the proportionate-type algorithms [66].

Driven by the recently developed CS algorithms [19, 30], some efforts has been made to incorporate CS techniques into adaptive filtering methods to design more accurate or less complex channel estimation algorithms. For instance, by combining a CS technique and the Kalman filter algorithm, a new algorithm named Kalman filtered compressed sensing (KF-CS) has been proposed and applied to magnetic resonance imaging (MRI) [72]. In this method, a Kalman filter estimates the support set, which has an important effect on the estimation error. Moreover, another algorithm combining the CS and least-squares techniques has been developed to improve the performance of the CS and least-square algorithms [73]. Recently, other effective sparse signal estimation algorithms, referred as as ZA algorithms, were proposed by incorporating the l_1-norm into the cost function of the standard LMS algorithm [67].

On the basis of the concept of ZA algorithms and the recent development of CS theory, we have developed several improved ZA algorithms based on the variable-step-size technique, affine projection algorithm (APA) [44] and proportionate normalized least-mean-square (PNLMS) algorithm [66, 74] to further exploit the sparsity of broadband multipath channels and to improve the channel estimation performance of classical adaptive filter algorithms.

1.6 Contributions of this thesis

In this dissertation, we mainly propose several sparse adaptive filters based on the variable-step-size technique, APA and PNLMS algorithm for broadband multipath channel estimation applications that enhance the convergence speed and steady-state performance. The main contributions of this dissertation are summarized as follows:

I. An improved variable-step-size LMS algorithm is proposed that is based on the modification of a sigmoid function used for step size adjustment. This algorithm is referred as the sigmoid functioned variable step size LMS (SVSS-LMS) algorithm.

II. A ZA sigmoid functioned variable step size LMS (ZA-SVSS-LMS) algorithm is proposed by incorporating the ZA technique into the proposed SVSS-LMS algorithm to improve its convergence speed for sparse channel estimation. Similarly to the reweighted ZA-LMS (RZA-LMS) algorithm, we also propose a reweighted ZA-SVSS-LMS (RZA-SVSS-LMS) algorithm, which further accelerates the convergence speed of the ZA-SVSS-LMS algorithm.

III. To enhance the robustness of the RZA-SVSS-LMS algorithm, an adaptive parameter adjustment method is adopted to provide an adaptive ZA strength, by which both the convergence speed and steady-state performance of the RZA-SVSS-LMS algorithm are significantly improved.

IV. A smooth approximation l_0-norm-constrained APA (SL0-APA) is proposed by the integration of a smooth approximation l_0-norm (SL0) into the cost function of the standard APA, which is equivalent to adding a zero attractor in its iterations.

V. We propose a low-complexity discrete weighted ZA affine projection algorithm (DWZA-APA), which is realized by using a piece wise linear function approximation instead of the sum-logarithm function used in the reweighted ZA-

APA (RZA-APA) [75].

VI. We propose an l_p-norm-constrained PNLMS (LP-PNLMS) algorithm, which is realized by the integration of the gain-matrix-weighted l_p-norm into the cost function of the PNLMS algorithm, to improve the convergence speed of the inactive taps in the PNLMS algorithm and hence reduce the steady-state error.

1.7 Outline

This dissertation is organized as follows:

Chapter 2 briefly reviews the conventional channel estimation algorithms, including the standard LMS algorithm and its variants, standard APA and its related ZA-APAs and the previously proposed PNLMS and improved PNLMS algorithms. After reviewing these channel estimation algorithms, we discuss their advantages and disadvantages. Although some of these algorithms have been developed for system identification and echo cancellation applications, their disadvantages can be mitigated by using the proposed techniques to exploit their advantages.

Chapter 3 describes our proposed ARZA-SVSS-LMS algorithm, which further enhances the robustness of the RZA-LMS algorithm in terms of the convergence speed and steady-state performance. The proposed algorithm is based on the VSS technique and the adaptive parameter adjustment method. The ARZA-SVSS-LMS algorithm is described in detail. To begin with, an SVSS-LMS algorithm is proposed, which is an improved VSS-LMS algorithm. Next, the zero-attracting (ZA) techniques used in the ZA-LMS and RZA-LMS algorithms are incorporated into the proposed SVSS-LMS algorithm to obtain ZA-SVSS-LMS and RZA-SVSS-LMS algorithms. Finally, an adaptive parameter adjustment method is adopted to form the ARZA-SVSS-LMS algorithm by providing an adaptive ZA strength in the RZA-SVSS-LMS algorithm. The proposed ARZA-SVSS-LMS algorithm and its related channel estimation algorithms are verified via a sparse channel to evaluate their channel estimation performance.

Chapter 4 introduces a novel algorithm referred as SL0-APA, which is realized via incorporating the SL0 into the cost function of the standard APA to form a zero attractor. The SL0-APA is mathematically derived and experimentally investigated on the basis of a multipath communication channel. Furthermore, its estimation performance is verified over a sparse channel and compared with the standard APA as well as the previously proposed ZA-APA and reweighted ZA-APA (RZA-APA). In addition, a convergence analysis of the SL0-APA is performed to predict the mean square error.

Chapter 5 presents a low-complexity DWZA-APA, which aims to reduce the computational complexity of the RZA-APA by utilizing a piece wise linear function approximation instead of the sum-logarithm. The DWZA-APA is investigated over a sparse channel and a sparse-cluster channel to evaluate its channel estimation performance in comparison with those of other popular sparse channel estimation algorithms.

Chapter 6 proposes an LP-PNLMS algorithm to improve the convergence speed of the inactive taps in the basic PNLMS algorithm. The LP-PNLMS algorithm is realized by incorporating the gain-matrix-weighted l_p-norm into the cost function of the PNLMS algorithm, which is a type of ZA algorithm. The proposed LP-PNLMS algorithm can accelerate the convergence of the inactive taps and hence increase the convergence speed of the PNLMS algorithm. The simulation results obtained from a sparse channel estimation demonstrated that the LP-PNLMS algorithm can achieve a higher convergence speed and a smaller steady-state error than the PNLMS algorithm.

Chapter 7 gives a conclusion of the dissertations and suggests future tasks in developing sparse adaptive filter algorithms for channel estimation and other applications.

Chapter 2

Adaptive filter algorithms for sparse channel estimation

In this chapter, we review the previously proposed channel estimation algorithms such as LMS , variable step size LMS (VSS-LMS), normalized least-mean-square (NLMS) algorithms, and ZA algorithms including ZA-LMS and RZA-LMS algorithms and the proportionate NLMS (PNLMS) algorithms. These algorithms are discussed on the basis of a sparse multipath communication system.

2.1 Purpose of adaptive channel estimation

The basic task of adaptive channel estimation is to minimize a meaningful error function by proper setting the parameters of the adaptive filter algorithms. In the channel estimation, the error function is the difference between the channel output and the adaptive filter output signals. Thus, the optimal filter parameters are found via minimization of the cost function of the error signal. One of the useful methods is to minimize the mean square error of the error signal, which has been widely studied and investigated in various adaptive filters. The basic setup of an adaptive filter system is illustrated in Fig. 2.1.

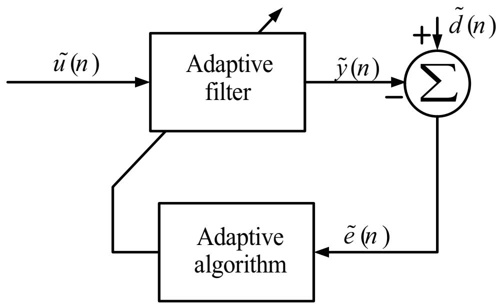


Figure 2.1: General adaptive filter configuration.

Here, $\tilde{u}(n)$ is the input signal at instant n, $\tilde{y}(n)$ is the output of the adaptive filter, and $\tilde{d}(n)$ denotes the desired signal. The error signal $\tilde{e}(n)$ is the difference of the desired signal $\tilde{d}(n)$ and the output of the adaptive filter $\tilde{y}(n)$, which is denoted as $\tilde{e}(n) = \tilde{d}(n) - \tilde{y}(n)$. Then the adaptive filters are to minimize the mean square error of the error signal $\tilde{e}(n)$. These adaptive filters have been widely studied and applied to channel estimation applications, adaptive control and echo cancellation. Next, we review several popular adaptive filters based on a typical sparse multipath communication system.

2.2 LMS algorithm and its variants

Based on the fundament of the principle of adaptive filter and the channel estimation task, a number of LMS and its variants are proposed and widely studied [43, 44, 76–79] and have been used for channel estimation applications. In this section, we review three classes LMS algorithms based on a multipath communication system, namely, standard LMS, VSS-LMS and NLMS algorithms.

2.2.1 LMS algorithm

We consider the sparse multipath communication system shown in Fig. 2.2. The input signal $\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^T$ which contains the N

most recent samples is transmitted over a finite impulse response (FIR) channel with channel impulse response (CIR) $\mathbf{h} = [h_0, h_1, \cdots, h_{N-1}]^T$, where $(\cdot)^T$ denotes transposition operation. Then the output signal of the channel is written as follows:

$$y(n) = \mathbf{h}^T \mathbf{x}(n), \tag{2.1}$$

where $\mathbf{h}$ is a sparse channel vector with K dominant active taps and $K \ll N$. To estimate the unknown sparse channel $\mathbf{h}$, an LMS adaptive filter uses the input signal $\mathbf{x}(n)$, the output signal $y(n)$, and the instantaneous estimation error $e(n)$, which is given by

$$e(n) = r(n) - \hat{\mathbf{h}}^T(n)\mathbf{x}(n), \tag{2.2}$$

where $\hat{\mathbf{h}}(n)$ is the LMS adaptive channel estimator, $r(n) = y(n) + v(n)$, and $v(n)$ is an additive noise at the receiver. On the basis of the LMS algorithm, the cost function $J_{\text{LMS}}(n)$ is given by

$$J_{\text{LMS}}(n) = \frac{1}{2}e^2(n). \tag{2.3}$$

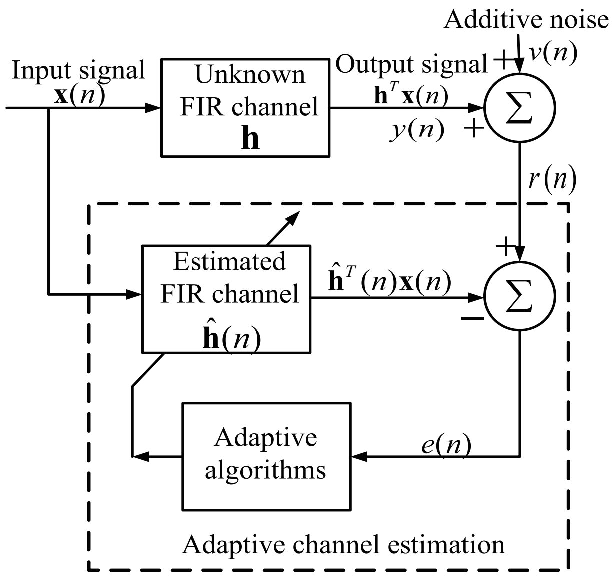


Figure 2.2: Typical sparse multipath communication system.

Therefore, the LMS adaptive channel estimation is achieved by minimizing $J_{\mathrm{LMS}}(n)$, and update function of the estimated channel can be written as

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) - \mu_{\mathrm{LMS}} \frac{\partial J_{\mathrm{LMS}}(n)}{\partial \hat{\mathbf{h}}(n)} = \hat{\mathbf{h}}(n) + \mu_{\mathrm{LMS}} e(n) \mathbf{x}(n), \tag{2.4}$$

where μ_{LMS} is the step-size such that $0 < \mu_{\mathrm{LMS}} < \frac{1}{\lambda_{\max}}$, with $\lambda_{\max}$ being the maximum eigenvalue of the covariance matrix $\mathbf{R} = \mathrm{E}\{\mathbf{x}(n)\mathbf{x}^T(n)\}$ of $\mathbf{x}(n)$, where $\mathrm{E}\{\cdot\}$ is the expectation operand. The gradient descent algorithm is adopted to guarantee the convergence of the LMS algorithm to converge to the optimum point under an appropriate value of μ_{LMS}. The detailed derivation of μ_{LMS} can be found in [2, 43, 44].

2.2.2 VSS-LMS algorithm

The following is a description of variable step size LMS (VSS-LMS) [76]-based adaptive channel estimation, and whose update equation is described as follows:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) - \mu_{\mathrm{VSS}}(n) \frac{\partial J_{\mathrm{LMS}}(n)}{\partial \hat{\mathbf{h}}(n)} = \hat{\mathbf{h}}(n) + \mu_{\mathrm{VSS}}(n) e(n) \mathbf{x}(n), \tag{2.5}$$

where

$$\mu_{\mathrm{VSS}}(n) = \begin{cases} \mu_{\max}, & \mu'(n) > \mu_{\max} \\ \mu_{\min}, & \mu'(n) < \mu_{\min} \\ \mu'(n), & \text{otherwise} \end{cases}. \tag{2.6}$$

$$\mu'(n) = \kappa \mu_{\mathrm{VSS}}(n-1) + \chi e^2(n-1), 0 < \kappa < 1, \chi > 0. \tag{2.7}$$

A constant $\mu_{\max}$ is normally selected near the point of instability of the standard LMS algorithm to provide required convergence speed, while the value of $\mu_{\min}$ is chosen as a compromise between the desired level of steady-state misadjustment and the required tracking capability of the VSS-LMS algorithm [76]. The parameter χ balances the convergence speed as well as the level of the misadjustment of the VSS-LMS algorithm. From the update equations (2.5)-(2.7), we can see that the step size can be adjusted using the mean square misalignment $e(n)$. At the early stage of the adaptive algorithm, the instantaneous error is large, causing the step size to increase, thus leading to rapid convergence. With the decrease in the instantaneous error, the step size decreases, yielding a smaller misadjustment that is close to the optimum value.

2.2.3 NLMS algorithm

On the basis of the LMS algorithm in Section 2.2.1, the update function of the NLMS[1] algorithm can be written as [77, 80, 81]

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\mathrm{NLMS}} \frac{e(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{x}(n) + \delta_{\mathrm{NLMS}}}, \tag{2.8}$$

where μ_{NLMS} is the step-size, and δ_{NLMS} is a small positive constant for preventing division by zero.

2.3 Zero-attracting LMS algoritms

2.3.1 ZA-LMS algorithm

The ZA-LMS algorithm is a type of sparsity-aware LMS algorithm with an l_1-penalty in its cost function. In the ZA-LMS algorithm, the cost function is defined by combining the instantaneous square error $e(n)$ with an l_1-penalty of the adaptive channel estimator and is given by

$$J_{\mathrm{ZA-LMS}}(n) = \frac{1}{2}e^2(n) + \gamma_{\mathrm{ZA-LMS}} \left\| \hat{\mathbf{h}}(n) \right\|_1, \tag{2.9}$$

[1] Here, NLMS algorithm can be regarded as a special case of standard LMS algorithm, and hence it is discussed as a part of the LMS algorithms.

where $\gamma_{\mathrm{ZA-LMS}} > 0$ is a regularization parameter used to balance the estimation error and the sparse penalty of $\hat{\mathbf{h}}(n)$. Using the gradient descent algorithm, the update equation of the ZA-LMS algorithm is obtained as follows:

$$\begin{aligned}\hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) - \mu_{\mathrm{ZA-LMS}} \frac{\partial J_{\mathrm{ZA-LMS}}(n)}{\partial \hat{\mathbf{h}}(n)} \\ &= \hat{\mathbf{h}}(n) + \mu_{\mathrm{ZA-LMS}} e(n)\mathbf{x}(n) - \rho_{\mathrm{ZA-LMS}} \mathrm{sgn}(\hat{\mathbf{h}}(n)),\end{aligned} \tag{2.10}$$

where $\rho_{\mathrm{ZA-LMS}} = \mu_{\mathrm{ZA-LMS}}\gamma_{\mathrm{ZA-LMS}}$, $0 < \rho_{\mathrm{ZA-LMS}} < 1.2 \times 10^{-3}$ for achieving good estimation performance [82], and $\mathrm{sgn}(\cdot)$ is a component wise sign function defined as

$$\mathrm{sgn}[\hat{h}_i(n)] = \begin{cases} 1, & \hat{h}_i(n) > 0 \\ -1, & \hat{h}_i(n) < 0 \\ 0, & \hat{h}_i(n) = 0 \end{cases}, \qquad 0 \le i \le N \tag{2.11}$$

Comparing the update equation (2.10) with the standard LMS update equation (2.4), we can see that the ZA-LMS algorithm has the additional term $\rho_{\mathrm{ZA-LMS}}\mathrm{sgn}(\hat{\mathbf{h}}(n))$, denoted as zero attractor, which attracts the small channel coefficients to zero with high probability. In other words, the zero attractor speeds up the convergence speed of the ZA-LMS algorithm when most of the channel taps are zero. Additionally, the attractor strength is controlled using the parameter $\rho_{\mathrm{ZA-LMS}}$.

For the ZA-LMS algorithm, it exploits the sparsity of the wireless multipath channel and can speed up the convergence speed of standard LMS algorithm. Furthermore, in the sparse channel, the number of dominant active taps are important, which has direct effect on the Carmer-Rao lower bound (CRLB) the computational complexity and has been verified by Theorem 1 and the detailed proof is given in **Appendix A** [2, 43].

Theorem 1 *Assume a channel vector* $\mathbf{h}$ *with length of* N*, and the step-size* μ_{LMS} *satisfies the* $0 < \mu_{\mathrm{LMS}} < \frac{1}{\lambda_{\max}}$*, then the mean square error (MSE) lower bound of the standard LMS channel estimator is* $B = \mu_{\mathrm{LMS}} PN/(2 - \mu_{\mathrm{LMS}}\lambda_{\min}) \sim \mathcal{O}(N)$*,* *where P denotes the unit power of gradient noise and $\lambda_{\min}$ denotes the minimum eigenvalue of $\mathbf{R}$. If the channel $\mathbf{h}$ is a sparse channel which contains K active taps and $K \ll N$, then the MSE lower bound channel estimator of the sparse channel can be described as $B_s = \mu_{\text{LMS}} PK/(2 - \mu_{\text{LMS}}\lambda_{\min}) \sim \mathcal{O}(K)$.*

According to the Theorem 1, it is very important to design adaptive sparse channel estimation algorithms, which can not only improve the channel estimation performance but also can reduce the complexity of these algorithms.

2.3.2 RZA-LMS algorithm

Since the ZA-LMS algorithm cannot distinguish the difference between zero taps and nonzero taps [67], the same penalty is applied to all the taps, which forces all the taps to become zero uniformly. Therefore, the performance is degraded for less sparse systems. Motivated by CS theory [19, 30, 35] and the reweighted l_1-norm minimization recovery algorithm [83], a heuristic approach to heighten the zero attractor was proposed, named the reweighted zero-attracting LMS (RZA-LMS) algorithm [67]. In the RZA-LMS algorithm, the cost function is defined as

$$J_{\text{RZA-LMS}}(n) = \frac{1}{2}e^2(n) + \gamma_{\text{RZA-LMS}} \sum_{i=1}^{N} \log(1 + \varepsilon_{\text{RZA-LMS}} \left|\hat{h}_i(n)\right|), \tag{2.12}$$

where $\gamma_{\text{RZA-LMS}} > 0$ is the regularization parameter and $\varepsilon_{\text{RZA-LMS}} > 0$ is the positive threshold. In the RZA-LMS algorithm, the $\sum_{i=1}^{N} \log(1 + \varepsilon_{\text{RZA-LMS}} \left|\hat{h}_i(n)\right|)$ is adopted instead of $\left\|\hat{\mathbf{h}}(n)\right\|_1$ in the ZA-LMS algorithm, because $\sum_{i=1}^{N} \log(1 + \varepsilon_{\text{RZA-LMS}} \left|\hat{h}_i(n)\right|)$ is more similar to the l_0-norm [67]. Then the ith channel coefficient $\hat{h}_i(n)$ can be updated using the following equation:

$$\begin{aligned}\hat{h}_i(n+1) &= \hat{h}_i(n) - \mu_{\mathrm{RZA-LMS}} \frac{\partial J_{\mathrm{RZA-LMS}}(n)}{\partial \hat{h}_i(n)} \\ &= \hat{h}_i(n) + \mu_{\mathrm{RZA-LMS}} e(n) x_i(n) - \rho_{\mathrm{RZA-LMS}} \frac{\mathrm{sgn}(\hat{h}_i(n))}{1 + \varepsilon_{\mathrm{RZA-LMS}} \left|\hat{h}_i(n)\right|}.\end{aligned} \tag{2.13}$$

Equation (2.13) can be expressed in vector form as

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\mathrm{RZA-LMS}} e(n)\mathbf{x}(n) - \rho_{\mathrm{RZA-LMS}} \frac{\mathrm{sgn}(\hat{\mathbf{h}}(n))}{1 + \varepsilon_{\mathrm{RZA-LMS}} \left|\hat{\mathbf{h}}(n)\right|}, \tag{2.14}$$

where $\rho_{\mathrm{RZA-LMS}} = \mu_{\mathrm{RZA-LMS}} \gamma_{\mathrm{RZA-LMS}} \varepsilon_{\mathrm{RZA-LMS}}$. Note that the reweighted zero attractor only effects on the taps whose magnitudes are comparable to the parameter $1/\varepsilon_{\mathrm{RZA-LMS}}$, while little shrinkage is exerted on taps whose magnitudes are much greater than $1/\varepsilon_{\mathrm{RZA-LMS}}$. In addition, in the RZA-LMS algorithm, $\varepsilon_{\mathrm{RZA-LMS}}$ is constant.

2.3.3 ZA-NLMS algorithm

From the discussions of the NLMS algorithm in Section 2.2.3 and the zero attractor in the ZA-LMS algorithms (Section 2.3.1), the update function of the ZA-NLMS algorithm can be written as [2, 84]

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\mathrm{ZA-NLMS}} \frac{e(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{x}(n) + \delta_{\mathrm{NLMS}}} - \rho_{\mathrm{ZA-NLMS}} \mathrm{sgn}(\hat{\mathbf{h}}(n)), \tag{2.15}$$

where $\mu_{\mathrm{ZA-NLMS}}$ is the step-size of the ZA-NLMS algorithm, and $\rho_{\mathrm{ZA-NLMS}}$ is a regularization parameter.

2.3.4 RZA-NLMS algorithm

On the basis of the NLMS and the zero-attracting algorithms [67], the concepts of reweighted zero-attracting is expanded to the NLMS algorithm in order to form the RZA-NLMS algorithm. Thus, the update function of the RZA-NLMS is [2, 84]

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\mathrm{RZA-NLMS}} \frac{e(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{x}(n) + \delta_{\mathrm{NLMS}}} - \rho_{\mathrm{RZA-NLMS}} \frac{\mathrm{sgn}(\hat{\mathbf{h}}(n))}{1 + \varepsilon_{\mathrm{RZA-NLMS}} \left|\hat{\mathbf{h}}(n)\right|}, \tag{2.16}$$

where $\mu_{\mathrm{RZA-NLMS}}$ is the step-size of the RZA-NLMS algorithm, and $\rho_{\mathrm{RZA-NLMS}}$ is a regularization parameter.

2.4 APA and its zero-attracting algorithms

The affine projection algorithm (APA) is another popular method in adaptive filtering applications [44, 61, 77, 85], 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 [75]. In this section, we discuss the APA and its zero-attracting algorithms based on a sparse multipath communication system shown in Fig. 2.3, which has a little difference from the Fig. 2.2 because of the reusing data scheme in APAs. The input signal $\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^T$ containing the N most recent samples is transmitted over a finite impulse response (FIR) channel with channel impulse response (CIR) $\mathbf{h} = [h_0, h_1, \cdots, h_{N-1}]^T$, 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 $\hat{\mathbf{y}}(n)$, and the received signal $\mathbf{r}(n) = \mathbf{y}(n) + \mathbf{v}(n)$ is obtained at the receiver.

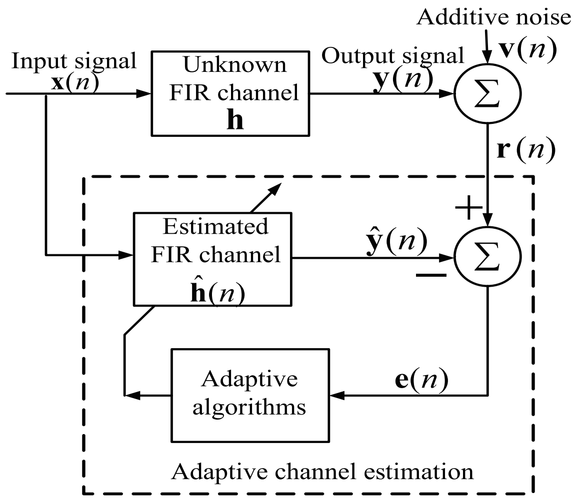


Figure 2.3: Typical sparse multipath communication system based on APA channel estimation algorithms.

2.4.1 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 [44]:

$$
\begin{aligned}
\mathbf{U}(n) &= \begin{bmatrix} \mathbf{x}^T(n) \\ \mathbf{x}^T(n-1) \\ \vdots \\ \mathbf{x}^T(n-Q+1) \end{bmatrix} \\
&= \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-N+1) \\ x(n-1) & x(n-2) & \cdots & x(n-N) \\ \vdots & \vdots & \ddots & \vdots \\ x(n-Q+1) & x(n-Q) & \cdots & x(n-N-Q+2) \end{bmatrix},
\end{aligned}
\tag{2.17}
$$

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

$$\mathbf{y}(n) = \mathbf{U}(n)\mathbf{h} = \begin{bmatrix} y(n) \\ y(n-1) \\ \vdots \\ y(n-Q+1) \end{bmatrix}, \tag{2.18}$$

$$\hat{\mathbf{y}}(n) = \mathbf{U}(n)\hat{\mathbf{h}}(n) = \begin{bmatrix} \hat{y}(n) \\ \hat{y}(n-1) \\ \vdots \\ \hat{y}(n-Q+1) \end{bmatrix}, \tag{2.19}$$

$$\mathbf{v}(n) = \begin{bmatrix} v(n) \\ v(n-1) \\ \vdots \\ v(n-Q+1) \end{bmatrix}, \tag{2.20}$$

$$\mathbf{r}(n) = \begin{bmatrix} r(n) \\ r(n-1) \\ \vdots \\ r(n-Q+1) \end{bmatrix}. \tag{2.21}$$

From the equations (2.17)-(2.21), the instantaneous error $\mathbf{e}(n)$ can be written as

$$\mathbf{e}(n) = \begin{bmatrix} e(n) \\ e(n-1) \\ \vdots \\ e(n-Q+1) \end{bmatrix} = \begin{bmatrix} r(n) - \hat{y}(n) \\ r(n-1) - \hat{y}(n-1) \\ \vdots \\ r(n-Q+1) - \hat{y}(n-Q+1) \end{bmatrix} = \mathbf{r}(n) - \hat{\mathbf{y}}(n). \tag{2.22}$$

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

$$\begin{aligned} & \left\| \hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n) \right\|^2 \\ & \text{subject to}: \\ & \mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n+1) = \mathbf{0} \end{aligned} . \tag{2.23}$$

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 posteriori error to zero at the same time. Here, the Lagrange multiplier method is used to find out the solution that minimizes the cost function $J_{\text{APA}}(n)$ of the APA

$$J_{\text{APA}}(n) = \left\| \hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n) \right\|^2 + [\mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n+1)]^T \boldsymbol{\lambda}_{\text{APA}}, \tag{2.24}$$

where $\boldsymbol{\lambda}_{\text{APA}}$ is a $Q \times 1$ vector of Lagrange multiplier and $\boldsymbol{\lambda}_{\text{APA}} = [\ \lambda_0 \quad \lambda_1 \quad \cdots \quad \lambda_{Q-1}\ ]^T$. The equation (2.24) can be rewritten as

$$J_{\text{APA}}(n) = [\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)]^T [\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)] + [\mathbf{r}^T(n) - \hat{\mathbf{h}}^T(n+1)\mathbf{U}^T(n)]\boldsymbol{\lambda}_{\text{APA}}. \tag{2.25}$$

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

$$\frac{\partial J_{\text{APA}}(n)}{\partial \hat{\mathbf{h}}(n+1)} = 2\hat{\mathbf{h}}(n+1) - 2\hat{\mathbf{h}}(n) - \mathbf{U}^T(n)\boldsymbol{\lambda}_{\text{APA}}. \tag{2.26}$$

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

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}^T(n)\boldsymbol{\lambda}_{\text{APA}}. \tag{2.27}$$

Multiplying $\mathbf{U}(n)$ on both sides of equation (2.27), we have

$$\mathbf{U}(n)\hat{\mathbf{h}}(n+1) = \mathbf{U}(n)\hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}(n)\mathbf{U}^T(n)\boldsymbol{\lambda}_{\text{APA}}. \tag{2.28}$$

By taking the constraint condition of equation (2.23) into consideration, we have

$$\mathbf{r}(n) = \mathbf{U}(n)\hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}(n)\mathbf{U}^T(n)\boldsymbol{\lambda}_{\text{APA}}. \tag{2.29}$$

Taking equations (2.19), (2.22) and (2.28) into account, we can get

$$\mathbf{e}(n) = \frac{1}{2}\mathbf{U}(n)\mathbf{U}^T(n)\boldsymbol{\lambda}_{\text{APA}}. \tag{2.30}$$

Then

$$\boldsymbol{\lambda}_{\text{APA}} = 2[\mathbf{U}(n)\mathbf{U}^T(n)]^{-1}\mathbf{e}(n). \tag{2.31}$$

The update equation is now given by (2.27) with $\boldsymbol{\lambda}_{\text{APA}}$ being the solution of (2.30) and is expressed as

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mathbf{U}^T(n)[\mathbf{U}(n)\mathbf{U}^T(n)]^{-1}\mathbf{e}(n) = \hat{\mathbf{h}}(n) + \mathbf{U}^{+}(n)\mathbf{e}(n), \tag{2.32}$$

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 [44]. 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 (2.32) can be rewritten as

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\text{APA}}\mathbf{U}^T(n)[\mathbf{U}(n)\mathbf{U}^T(n)]^{-1}\mathbf{e}(n) = \hat{\mathbf{h}}(n) + \mu_{\text{APA}}\mathbf{U}^{+}(n)\mathbf{e}(n). \tag{2.33}$$

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 $Q = 1$.

2.4.2 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 the equation (2.24), which is known as zero-attracting affine projection algorithm (ZA-APA) [75]. In the ZA-APA, the cost function is defined by combining the cost function $J_{\mathrm{APA}}(n)$ of standard APA with l_1-penalty of the channel estimator and is given by

$$\begin{aligned} J_{\mathrm{ZA-APA}}(n) = & \left\|\hat{\mathbf{h}}(n+1)-\hat{\mathbf{h}}(n)\right\|^2+[\mathbf{r}(n)-\mathbf{U}(n)\hat{\mathbf{h}}(n+1)]^T\boldsymbol{\lambda}_{\mathrm{ZA-APA}} \\ & +\gamma_{\mathrm{ZA-APA}}\left\|\hat{\mathbf{h}}(n+1)\right\|_1, \end{aligned} \tag{2.34}$$

where $\boldsymbol{\lambda}_{\mathrm{ZA-APA}}$ is the vector of Lagrange multiplier with $Q \times 1$. $\gamma_{\mathrm{ZA-APA}} > 0$ 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 $J_{\mathrm{ZA-APA}}(n)$, we use the Lagrange multiplier to calculate its gradient, which is expressed as

$$\frac{\partial J_{\mathrm{ZA-APA}}(n)}{\partial \hat{\mathbf{h}}(n+1)} = 2\hat{\mathbf{h}}(n+1)-2\hat{\mathbf{h}}(n)-\mathbf{U}^T(n)\boldsymbol{\lambda}_{\mathrm{ZA-APA}}+\gamma_{\mathrm{ZA-APA}}\mathrm{sgn}[\hat{\mathbf{h}}(n+1)], \tag{2.35}$$

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

$$\mathrm{sgn}[x] = \begin{cases} \dfrac{x}{|x|}, x \neq 0 \\ 0, x = 0 \end{cases}. \tag{2.36}$$

As is known to us all, the minimum is obtained by letting

$\partial J_{\text{ZA}-\text{APA}}(n)/\partial\hat{\mathbf{h}}(n+1) = \mathbf{0}$. Thus, we can get

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}^T(n)\boldsymbol{\lambda}_{\text{ZA}-\text{APA}} - \frac{1}{2}\gamma_{\text{ZA}-\text{APA}}\text{sgn}[\hat{\mathbf{h}}(n+1)]. \tag{2.37}$$

Multiplying both sides by $\mathbf{U}(n)$ of (2.37), we can obtain

$$\mathbf{U}(n)\hat{\mathbf{h}}(n+1) = \mathbf{U}(n)\hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}(n)\mathbf{U}^T(n)\boldsymbol{\lambda}_{\text{ZA}-\text{APA}} - \frac{1}{2}\gamma_{\text{ZA}-\text{APA}}\mathbf{U}(n)\text{sgn}[\hat{\mathbf{h}}(n+1)]. \tag{2.38}$$

Considering the constraint condition of equations (2.23), we can get the following expression

$$\mathbf{r}(n) = \mathbf{U}(n)\hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}(n)\mathbf{U}^T(n)\boldsymbol{\lambda}_{\text{ZA}-\text{APA}} - \frac{1}{2}\gamma_{\text{ZA}-\text{APA}}\mathbf{U}(n)\text{sgn}[\hat{\mathbf{h}}(n+1)]. \tag{2.39}$$

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

$$\boldsymbol{\lambda}_{\text{ZA}-\text{APA}} = [\mathbf{U}(n)\mathbf{U}^T(n)]^{-1}\{2\mathbf{e}(n) + \gamma_{\text{ZA}-\text{APA}}\mathbf{U}(n)\text{sgn}[\hat{\mathbf{h}}(n+1)]\}. \tag{2.40}$$

Substituting (2.40) into (2.37) 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

$$\begin{array}{l}\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mathbf{U}^+(n)\mathbf{e}(n) + \frac{1}{2}\gamma_{\text{ZA}-\text{APA}}\mathbf{U}^+(n)\mathbf{U}(n)\text{sgn}[\hat{\mathbf{h}}(n)] \\ \qquad\qquad -\frac{1}{2}\gamma_{\text{ZA}-\text{APA}}\text{sgn}[\hat{\mathbf{h}}(n)]\end{array}. \tag{2.41}$$

To balance the convergence speed and steady-state error, a step-size $\mu_{\text{ZA}-\text{APA}}$ is introduced and integrated into (2.41). Then, equation (2.41) can be rewritten as

$$\begin{aligned}\hat{\mathbf{h}}(n+1)=&\hat{\mathbf{h}}(n)+\mu_{\mathrm{ZA-APA}}\mathbf{U}^{+}(n)\mathbf{e}(n)+\frac{1}{2}\gamma_{\mathrm{ZA-APA}}\mathbf{U}^{+}(n)\mathbf{U}(n)\mathrm{sgn}[\hat{\mathbf{h}}(n)]\\&-\frac{1}{2}\gamma_{\mathrm{ZA-APA}}\mathrm{sgn}[\hat{\mathbf{h}}(n)].\end{aligned} \quad (2.42)$$

Comparing the update function (2.42) of the ZA-APA with the update function (2.33) of the standard APA, we find that there are two additional terms in (2.42) 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 $\gamma_{\mathrm{ZA-APA}}$. 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.

2.4.3 Reweighted zero-attracting affine projection algorithm (RZA-APA)

Unfortunately, the ZA-APA cannot distinguish the zero taps and the non-zero taps of the sparse channel, and it exerts the same penalty on all the channel taps, which forces all the taps to zero uniformly [67, 75]. 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 [83] and employed in [67, 75] to reinforce the zero attractor was proposed and was denoted as reweighted zero-attracting affine projection algorithm (RZA-APA). In the RZA-APA, $\sum_{i=1}^{N}\log(1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{h}_i(n)\right|)$ is adopted instead of $\left\|\hat{\mathbf{h}}(n)\right\|_1$. Thus, the cost function of the RZA-APA can be written as

$$\begin{aligned}J_{\mathrm{RZA-APA}}(n)=&\left\|\hat{\mathbf{h}}(n+1)-\hat{\mathbf{h}}(n)\right\|^2+[\mathbf{r}(n)-\mathbf{U}(n)\hat{\mathbf{h}}(n+1)]^T\boldsymbol{\lambda}_{\mathrm{RZA-APA}}\\&+\gamma_{\mathrm{RZA-APA}}\sum_{i=1}^{N}\log(1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{h}_i(n+1)\right|)\end{aligned}, \quad (2.43)$$

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

$$\begin{aligned}\frac{\partial J_{\mathrm{RZA-APA}}(n)}{\partial \hat{\mathbf{h}}(n+1)} &= 2\hat{\mathbf{h}}(n+1) - 2\hat{\mathbf{h}}(n) - \mathbf{U}^T(n)\boldsymbol{\lambda}_{\mathrm{RZA-APA}} \\ &+\gamma_{\mathrm{RZA-APA}} \frac{\mathrm{sgn}[\hat{\mathbf{h}}(n+1)]}{1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n+1)\right|}.\end{aligned} \tag{2.44}$$

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

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}^T(n)\boldsymbol{\lambda}_{\mathrm{RZA-APA}} - \frac{1}{2}\gamma_{\mathrm{RZA-APA}}\frac{\mathrm{sgn}[\hat{\mathbf{h}}(n)]}{1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|}. \tag{2.45}$$

By multiplying $\mathbf{U}(n)$ on both sides of (2.45), the following equation can be obtained

$$\begin{aligned}\mathbf{U}(n)\hat{\mathbf{h}}(n+1) &= \mathbf{U}(n)\hat{\mathbf{h}}(n) + \tfrac{1}{2}\mathbf{U}(n)\mathbf{U}^T(n)\boldsymbol{\lambda}_{\mathrm{RZA-APA}} \\ &-\frac{1}{2}\gamma_{\mathrm{RZA-APA}}\mathbf{U}(n)\frac{\mathrm{sgn}[\hat{\mathbf{h}}(n)]}{1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|}.\end{aligned} \tag{2.46}$$

Taking (2.23) and (2.46) into consideration, we can get

$$\begin{aligned}\mathbf{r}(n) &= \mathbf{U}(n)\hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}(n)\mathbf{U}^T(n)\boldsymbol{\lambda}_{\mathrm{RZA-APA}} \\ &-\frac{1}{2}\gamma_{\mathrm{RZA-APA}}\mathbf{U}(n)\frac{\mathrm{sgn}[\hat{\mathbf{h}}(n)]}{1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|}.\end{aligned} \tag{2.47}$$

Thus, the Lagrange multiplier vector $\boldsymbol{\lambda}_{\mathrm{RZA-APA}}$ is obtained

$$\boldsymbol{\lambda}_{\mathrm{RZA-APA}} = [\mathbf{U}(n)\mathbf{U}^T(n)]^{-1}\{2\mathbf{e}(n) + \gamma_{\mathrm{RZA-APA}}\mathbf{U}(n)\frac{\mathrm{sgn}[\hat{\mathbf{h}}(n)]}{1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|}\}, \tag{2.48}$$

where $\mathbf{e}(n) = \mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n)$. Substituting (2.48) into (2.45), we can get the update equation of the RZA-APA

$$\begin{aligned}\hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mathbf{U}^{+}(n)\mathbf{e}(n) + \frac{1}{2}\gamma_{\mathrm{RZA-APA}}\mathbf{U}^{+}(n)\mathbf{U}(n)\frac{\mathrm{sgn}[\hat{\mathbf{h}}(n)]}{1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|} \\ &\quad -\frac{1}{2}\gamma_{\mathrm{RZA-APA}}\frac{\mathrm{sgn}[\hat{\mathbf{h}}(n)]}{1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|}\end{aligned}. \tag{2.49}$$

Similarly, a step size $\mu_{\mathrm{RZA-APA}}$ is introduced and cooperated into (2.49) to balance the convergence speed and the steady-state error of the RZA-APA. Then, equation (2.49) can be rewritten as

$$\begin{aligned}&\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\mathrm{RZA-APA}}\mathbf{U}^{+}(n)\mathbf{e}(n) \\ &+\frac{1}{2}\gamma_{\mathrm{RZA-APA}}\mathbf{U}^{+}(n)\mathbf{U}(n)\frac{\mathrm{sgn}[\hat{\mathbf{h}}(n)]}{1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|} - \frac{1}{2}\gamma_{\mathrm{RZA-APA}}\frac{\mathrm{sgn}[\hat{\mathbf{h}}(n)]}{1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|}\end{aligned}. \tag{2.50}$$

From the analysis and the 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 effects taps whose magnitudes are comparable to $1/\varepsilon_{\mathrm{RZA-APA}}$ while has less shrinkage exerted on $\left|\hat{\mathbf{h}}(n)\right| \gg 1/\varepsilon_{\mathrm{RZA-APA}}$. Thus, the RZA-APA can improve steady-state performance compared to the ZA-APA.

2.5 Proportionate-type adaptive filters

On the basis of the analysis and the prior knowledge, we know that PNLMS algorithm is another type of sparse adaptive filter algorithm, which has been proposed to exploit the sparsity in nature, and has been applied for echo cancellation in telephone networks. Recently, many improved PNLMS algorithms [74, 86–93] has been proposed by using variable step size technique and l_1-norm technique. Furthermore, improved PNLMS (IPNLMS) [86] and the μ-law PNLMS (MPNLMS) [87] are two commonly used algorithms. Thus, in this section, we review the PNLSM, IPNLMS and MPNLMS algorithms.

2.5.1 Proportionate normalized least mean square algorithm

The PNLMS algorithm, which is an NLMS algorithm improved by the use of a proportionate technique, has been proposed for sparse system identification and echo cancellation [66]. In this algorithm, each tap is assigned an individual step size, which is obtained from the previous estimation of the filter coefficient. According to the gain allocation rule in this algorithm, the greater the magnitude of the tap, the larger the step size assigned to it, and hence the active taps converge quickly. The update function of the PNLMS algorithm [66] is described by the following equation with reference to Fig. 2.2.

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\text{PNLMS}} \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) + \delta_{\text{PNLMS}}} \tag{2.51}$$

Here, $\mathbf{G}(n)$, which denotes as the gain matrix, is a diagonal matrix that modifies the step size of each tap, μ_{PNLMS} is the global step size of the PNLMS algorithm and $\delta_{\text{PNLMS}} = \delta_x^2/N$ is a regularization parameter to prevent division by zero at the initialization stage, where δ_x^2 is the power of the input signal $\mathbf{x}(n)$. In the PNLMS algorithm, the gain matrix $\mathbf{G}(n)$ is given by

$$\mathbf{G}(n) = \text{diag}(g_0(n), g_1(n), \cdots, g_{N-1}(n)), \tag{2.52}$$

where the individual gain $g_i(n)$ is defined as

$$g_i(n) = \frac{\gamma_i(n)}{\sum\limits_{i=0}^{N-1} \gamma_i(n)}, 0 \le i \le N-1 \tag{2.53}$$

with

$$\gamma_i(n) = \max[\rho_g \max[\delta_p, \left|\hat{h}_0(n)\right|, \left|\hat{h}_1(n)\right|, \cdots, \left|\hat{h}_{N-1}(n)\right|], \left|\hat{h}_i(n)\right|], \tag{2.54}$$

where the parameters δ_p and ρ_g are positive constants with typical values of $\delta_p = 0.01$ and $\rho_g = 5/N$. δ_p is used to regularize the updating at the initial stage when all the taps are initialized to zero, and ρ_g is used to prevent $\hat{h}_i(n)$ from stalling when it is much smaller than the largest coefficient.

2.5.2 Improved IPNLMS algorithm

The IPNLMS algorithm is a type of PNLMS algorithm used to improve the convergence speed of the PNLMS algorithm. It is a combination of the PNLMS and NLMS algorithms with the relative significance of each coefficient controlled by a factor α. The IPNLMS algorithm [86] adopts the l_1-norm to enable the smooth selection of (2.54), and the update equation of the IPNLMS algorithm is expressed as

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) - \mu_{\text{IPNLMS}} \frac{e(n)\mathbf{K}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{K}(n)\mathbf{x}(n) + \delta_{\text{IPNLMS}}}, \tag{2.55}$$

where $\mathbf{K}(n) = \text{diag}(k_0(n), k_1(n), \cdots, k_{N-1}(n))$ is a diagonal matrix used to adjust the step size of the IPNLMS algorithm, where

$$k_j(n) = \frac{1-\alpha}{2N} + (1+\alpha)\frac{\left|\hat{h}_j(n)\right|}{2\left\|\hat{\mathbf{h}}(n)\right\|_1 + \varepsilon}, 0 \le j \le N-1 \tag{2.56}$$

for a small positive constant ε and $-1 \leq \alpha \leq 1$. At the initial stage, the step size is multiplied by $(1-\alpha)/2N$, since all the filter coefficients are initialized to zero. Thus, in the IPNLMS algorithm, a regularization parameter δ_{IPNLMS} is introduced, which is given by

$$\delta_{\text{IPNLMS}} = \frac{1-\alpha}{2N}\delta_{\text{NLMS}}. \tag{2.57}$$

We can see that the IPNLMS is identical to the NLMS algorithm for $\alpha = -1$, while the IPNLMS behaves identically to the PNLMS algorithm when $\alpha = 1$. In practical engineering applications, a suitable value for α is 0 or -0.5.

2.5.3 μ-law PNLMS algorithm

The μ-law PNLMS algorithm (MPNLMS) [87] is another enhancement of the PNLMS algorithm that utilizes the logarithm of the magnitudes of the filter coefficients instead of using the magnitudes directly in the PNLMS algorithm. The update equation is the same as that in the PNLMS algorithm given by (2.51). In the MPNLMS algorithm,

$$\gamma_i(n) = \max[\rho_g \max[\delta_p F(\left|\hat{h}_0(n)\right|), F(\left|\hat{h}_1(n)\right|), \cdots, F(\left|\hat{h}_{N-1}(n)\right|)], F(\left|\hat{h}_i(n)\right|)], \tag{2.58}$$

where

$$F(\left|\hat{h}_i(n)\right|) = \log(1 + \vartheta\left|\hat{h}_i(n)\right|), \tag{2.59}$$

where ϑ is a large positive constant related to the estimation accuracy requirement, typically $\vartheta = 1000$.

2.6 Conclusion

In this chapter, a review of adaptive filter algorithms has been introduced on the basis of the sparse multipath communication system. We reviewed the standard LMS algorithm and its variants as well as its related zero-attracting algorithms, namely ZA-LMS and RZA-LMS algorithms. In addition, we also reviewed the APA and its related zero-attracting algorithms. Finally, we discussed the PNLMS, IPNLMS and MPNLMS algorithms, which are also developed for sparse signal estimation applications. Then, we discuss the proposed adaptive sparse channel estimation algorithms in the next four chapters.

Chapter 3

Zero-Attracting Variable Step Size Least-Mean-Square Algorithms

3.1 Introduction

Since the standard LMS and the VSS-LMS algorithms discussed in Sections 2.2.1 and 2.2.2 cannot utilize the property of sparse channels, we propose an adaptive reweighted zero-attracting sigmoid functioned variable step size LMS (ARZA-SVSS-LMS) algorithm [94] that is comprised of sigmoid functioned variable step size LMS (SVSS-LMS), zero-attracting SVSS-LMS (ZA-SVSS-LMS), reweighted ZA-SVSS-LMS (RZA-SVSS-LMS) algorithms and adaptive parameter adjustment method. In this algorithm, the SVSS-LMS is a type of VSS-LMS algorithms that utilizes a sigmoid function, while the ZA-SVSS-LMS and the RZA-SVSS-LMS algorithms are realized by incorporating ZA-LMS and RZA-

LMS techniques into our proposed SVSS-LMS algorithm. The ARZA-SVSS-LMS algorithm is implemented by using an adaptive parameter adjustment method in our proposed RZA-SVSS-LMS algorithm, by which the benefits of the ARZA-SVSS-LMS algorithm are twofold. First, the convergence speeds of the proposed algorithms are increased by the introducing sigmoid function step size control methods based on the information provided by the mean square estimation error. Second, the steady-state misalignment is reduced by the incorporating an adaptive parameter adjustment method [95] into our proposed RZA-SVSS-LMS algorithm, particularly in a high signal-to-noise (SNR) environment. In addition, the relevant parameters in the proposed algorithms, which control the convergence speed and steady-state misalignment, are discussed qualitatively. The proposed algorithms are used for sparse channel estimation, and the results of a simulation demonstrate that the proposed algorithms significantly outperform the ZA-LMS, RZA-LMS and standard LMS algorithms.

3.2 Proposed ARZA-SVSS-LMS algorithms

In this section, we propose a robust sparse channel estimation algorithm denoted as ARZA-SVSS-LMS step-by-step. First, we propose an SVSS-LMS algorithm which significantly improves the convergence speed and the steady-state performance of the VSS-LMS algorithm. Then, we propose the ZA-SVSS-LMS and RZA-SVSS-LMS algorithms on the basis of the SVSS-LMS, ZA-LMS and RZA-LMS algorithms. Finally, we present our ARZA-SVSS-LMS algorithm by the use of adaptive parameter adjustment method in our proposed RZA-SVSS-LMS algorithm [94].

3.2.1 Proposed SVSS-LMS algorithm

Inspired by the VSS-LMS algorithm [76], several VSS-LMS algorithms based on sigmoid functions have been proposed [96, 97]. One of the step size update functions in [96] is

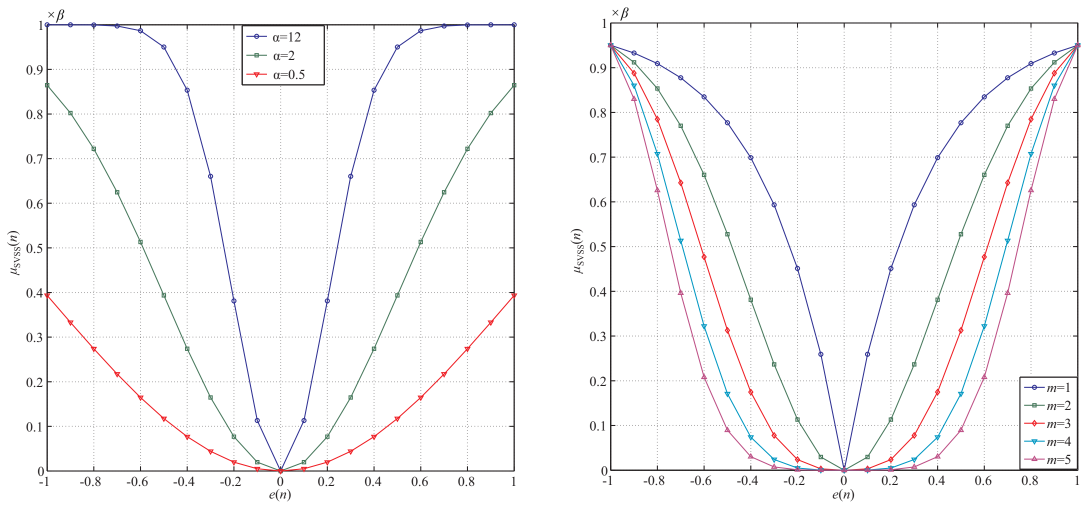
$$\mu(n) = \beta\left(\frac{1}{1 + e^{-\alpha|e(n)|}} - 0.5\right), \tag{3.1}$$

where $\alpha > 0$ and $\beta > 0$. Although this VSS-LMS algorithm speeds up the convergence and can achieve low steady-state misadjustment, its sigmoid function is complex and $\mu(n)$ changes so rapidly when $|e(n)|$ is close to zero. To reduce the complexity of the sigmoid function and improve the performance of the this VSS-LMS algorithm, we propose an SVSS-LMS algorithm with easy and flexible implementation by modifying the sigmoid function.

On the basis of the conventional VSS-LMS algorithm discussed in Section 2.2.2 and the concept of the sigmoid function mentioned above, an SVSS-LMS algorithm is proposed with a variable step size given by

$$\mu_{\mathrm{SVSS}}(n) = \beta(1 - e^{-\alpha|e(n)|^{m}}), \tag{3.2}$$

where $m > 0$, $\alpha > 0$ and $\beta > 0$, which are used to control $\mu_{\mathrm{SVSS}}(n)$. The relationship between the instantaneous error $e(n)$ and the step size $\mu_{\mathrm{SVSS}}(n)$ is



(a) α $(m = 2)$ (b) m $(\alpha = 3)$

Figure 3.1: Parameter effects on $\mu_{\mathrm{SVSS}}(n)$.

illustrated in Fig. 3.1 to give a better understanding of this sigmoid function.

It can be seen from Fig. 3.1a that in the early stages of adaptation, since the error $|e(n)|$ is large, a large $\mu_{\text{SVSS}}(n)$ is obtained. In this case, rapid convergence can be achieved for the proposed SVSS-LMS algorithm. On the other hand, with the increase in parameter α, the step size varies sharply, which may lead to a large mean squares misadjustment even when $|e(n)|$ is small. Moreover, we can also see that the step size $\mu_{\text{SVSS}}(n)$ is proportional to β. When β is small, a smaller $\mu_{\text{SVSS}}(n)$ is obtained, which results in a slower convergence of the system, lower capacity of tracking the time-varying channel and less robustness. In contrast, a larger β may lead to rapid changes $\mu_{\text{SVSS}}(n)$ which will cause the oscillation of the system. To improve the system's robustness and to make the system follow the time-varying channel rapidly with faster convergence in the initial stage and a smaller steady-state error in the steady stage, a parameter m is adopted to further control the step size $\mu_{\text{SVSS}}(n)$. We can see from Fig. 3.1b that a small $|e(n)|$ will result in a large $\mu_{\text{SVSS}}(n)$ when m is small. With increasing of m, the learning curve of $\mu_{\text{SVSS}}(n)$ is getting smoother. Thus, we conclude that the parameters α, β and m can be used to control the step size $\mu_{\text{SVSS}}(n)$. In practical applications, we should select these parameters carefully to balance the convergence speed and the steady-state error and to meet the time-varying channel-tracking requirements.

On the basis of the discussion of the proposed sigmoid function and the analysis of the VSS-LMS algorithm in Section 2.2.2, an SVSS-LMS algorithm is proposed for channel estimation. The update equation of the proposed SVSS-LMS is

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) - \mu_{\text{SVSS}}(n)\frac{\partial J_{\text{LMS}}(n)}{\partial \hat{\mathbf{h}}(n)} = \hat{\mathbf{h}}(n) + \mu_{\text{SVSS}}(n)e(n)\mathbf{x}(n), \tag{3.3}$$

where $\mu_{\text{SVSS}}(n)$ is given by (3.2). In the initial convergence phase, as $|e(n)|$ is large, a large $\mu_{\text{SVSS}}(n)$ is obtained to ensure that the SVSS-LMS algorithm converges more rapidly. When the algorithm reaches a steady state, $|e(n)|$ becomes smaller and $\mu_{\text{SVSS}}(n)$ reaches a minimum, meaning that we can obtain the best Wiener value. Thus, the performance is superior to those of the standard LMS and VSS-LMS algorithms because SVSS-LMS can effectively adjust the step-size as VSS-LMS by using a sigmoid function while maintaining the immunity against independent noise disturbance.

3.2.2 Proposed ZA-SVSS-LMS algorithm

According to the above discussions, the steady-state error is proportional to the step size while the convergence is inversely proportional to the step size. To improve the performance of the above LMS algorithms, many step-size control methods have been proposed and investigated. In this subsection, the SVSS-LMS method is adopted to improve the previously proposed sparse LMS algorithms. By combining the SVSS-LMS technique and the ZA-LMS method, we propose a ZA-SVSS-LMS algorithm which is a ZA-LMS algorithm with a sigmoid functioned variable step size discussed in Section 3.2.1. The cost function of the ZA-SVSS-LMS algorithm is given by (3.4) and its update equation is given by (3.5).

$$J_{\mathrm{ZA-LMS}}(n) = \frac{1}{2}e^2(n) + \gamma_{\mathrm{ZA-LMS}}\left\|\hat{\mathbf{h}}(n)\right\|_1, \tag{3.4}$$

$$\begin{aligned}\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) - \mu_{\mathrm{SVSS}}(n)\frac{\partial J_{\mathrm{ZA-LMS}}(n)}{\partial \hat{\mathbf{h}}(n)} = \hat{\mathbf{h}}(n) + \mu_{\mathrm{SVSS}}(n)e(n)\mathbf{x}(n) \\ -\rho_{\mathrm{ZA-LMS}}\mathrm{sgn}(\hat{\mathbf{h}}(n)).\end{aligned} \tag{3.5}$$

Here the step-size $\mu_{\mathrm{SVSS}}(n)$ is obtained from (3.2), $\gamma_{\mathrm{ZA-LMS}} > 0$ is a regularization parameter used to balance the estimation error and the sparse penalty of $\hat{\mathbf{h}}(n)$, $\rho_{\mathrm{ZA-LMS}} > 0$ is zero-attracting factor that is used to control the zero-attracting strength, and $\mathrm{sgn}(\cdot)$ is the component wise sign function given by (2.11). Since variable step-size (3.2) is used in the proposed ZA-SVSS-LMS algorithm, the performance of the ZA-LMS algorithm is effectively improved.

3.2.3 Proposed RZA-SVSS-LMS algorithm

Motivated by the concept and analysis of the RZA-LMS algorithm, we propose an RZA-SVSS-LMS algorithm by incorporating the reweighted zero-attracting

(RZA) technique [67] into the proposed SVSS-LMS algorithm. The RZA-SVSS-LMS algorithm is described as follows:

$$J_{\mathrm{RZA-LMS}}(n) = \frac{1}{2}e^2(n) + \gamma_{\mathrm{RZA-LMS}} \sum_{i=1}^{N} \log(1 + \varepsilon_{\mathrm{RZA-LMS}} \left|\hat{h}_i(n)\right|), \tag{3.6}$$

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\mathrm{SVSS}}(n)e(n)\mathbf{x}(n) - \rho_{\mathrm{RZA-LMS}} \frac{\mathrm{sgn}(\hat{\mathbf{h}}(n))}{1 + \varepsilon_{\mathrm{RZA-LMS}} \left|\hat{\mathbf{h}}(n)\right|}, \tag{3.7}$$

where $\gamma_{\mathrm{RZA-LMS}} > 0$ is the regularization parameter. Also the positive value of $\varepsilon_{\mathrm{RZA-LMS}}$ is the threshold and $\rho_{\mathrm{RZA-LMS}} > 0$ is used to control the zero-attracting strength of the RZA-SVSS-LMS algorithm. Equations (3.6) and (3.7) are the cost function and the update function of the proposed RZA-SVSS-LMS algorithm, respectively. Similarly, by the use of the sigmoid functioned variable step size $\mu_{\mathrm{SVSS}}(n)$ in the RZA-SVSS-LMS algorithm instead of the fixed $\mu_{\mathrm{RZA-LMS}}$ in the RZA-LMS algorithm, the behaviors of the RZA-LMS algorithm is significantly improved.

3.2.4 Proposed ARZA-SVSS-LMS algorithm

According to the results of previous papers [2, 67, 98], the performances for most of the proposed algorithms are degraded at a high SNR. To improve the performance and enhance the robustness of the proposed RZA-SVSS-LMS algorithm, we propose an adaptive RZA-SVSS-LMS (ARZA-SVSS-LMS) algorithm based on the concept in [95]. In the RZA-SVSS-LMS algorithm, the reweighted zero attractor only effects taps whose magnitudes are comparable to the parameter $1/\varepsilon_{\mathrm{RZA-LMS}}$, which is constant. To reach the optimum as rapidly as possible at each iteration, we introduce a variable $\varepsilon_{\mathrm{ARZA}}(n)$ into the term $\sum_{i=1}^{N} \log(1 + \varepsilon_{\mathrm{RZA-LMS}} \left|\hat{h}_i(n)\right|)$ to enhance the robustness of the RZA-SVSS-LMS algorithm. Initially, a large $\varepsilon_{\mathrm{ARZA}}(n)$ is adopted, which is then decreased until it meets the convergence requirement. The proposed ARZA-SVSS-LMS algorithm is expressed below.

To exploit the channel sparsity in the time domain, the cost function of the proposed ARZA-SVSS-LMS algorithm is given by

$$J_{\mathrm{ARZA}}(n) = \frac{1}{2}e^2(n) + \gamma_{\mathrm{RZA-LMS}} \sum_{i=1}^{N} \log(1 + \varepsilon_{\mathrm{ARZA}}(n) \left|\hat{h}_i(n)\right|), \tag{3.8}$$

where $\gamma_{\mathrm{RZA-LMS}} > 0$ is a regularization parameter used to balance the estimation error and the sparse penalty of $\hat{\mathbf{h}}(\mathbf{n})$, and $\varepsilon_{\mathrm{ARZA}}(n)$ is the threshold. The corresponding update equation of the proposed ARZA-SVSS-LMS algorithm is

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\mathrm{SVSS}}(n)e(n)\mathbf{x}(n) - \rho_{\mathrm{ARZA}} \frac{\mathrm{sgn}(\hat{\mathbf{h}}(n))}{1 + \varepsilon_{\mathrm{ARZA}}(n)\left|\hat{\mathbf{h}}(n)\right|}, \tag{3.9}$$

where $\rho_{\mathrm{ARZA}} > 0$ is used to control the zero-attracting strength. Then, an adaptive parameter adjustment method is introduced by relating $\varepsilon_{\mathrm{ARZA}}(n)$ to the current mean square error $e^2(n)$ to enhance the robustness of the ARZA-SVSS-LMS algorithm. We can form an estimate of $e^2(n)$ and use the following functions to update $\varepsilon_{\mathrm{ARZA}}(n)$.

$$U(n+1) = \xi U(n) + (1-\xi)e^2(n), \tag{3.10}$$

$$\bar{\varepsilon}_{\mathrm{ARZA}}(n+1) = \frac{U(n+1)}{\zeta}, \tag{3.11}$$

$$\varepsilon_{\mathrm{ARZA}}(n+1) = \left(\frac{\bar{\varepsilon}_{\mathrm{ARZA}}(n+1)}{N\delta_x^2}\right)^{-\frac{1}{2}} = \frac{1}{\theta}\left(\frac{U(n+1)}{N\delta_x^2}\right)^{-\frac{1}{2}} \tag{3.12}$$

where $U(n+1)$ is an estimation of $e^2(n)$, $0 \leq \xi \leq 1$, and δ_x^2 is the power of the input signal, and $\theta = (\zeta)^{-\frac{1}{2}}$ and ζ is a constant relating the current mean square error to $\bar{\varepsilon}_{\mathrm{ARZA}}(n+1)$, where $\bar{\varepsilon}_{\mathrm{ARZA}}(n+1)$ is the distance to the steady-state mean square error, also named the noise floor, and it is considered that convergence is achieved when noise floor reached to the required mean square error. In essence, $\varepsilon_{\mathrm{ARZA}}(n)$ is adopted by utilizing the estimation of the mean square error, which gives a large initial value of $\varepsilon_{\mathrm{ARZA}}$. With increasing time, the bound of the mean square error become tighter. Finally, $\varepsilon_{\mathrm{ARZA}}(n)$ with a smaller amplitude is obtained when the ARZA-SVSS-LMS algorithm performs effectively.

3.3 Analysis of the proposed ARZA-SVSS-LMS algorithm

In this subsection, we analyze the mean square convergence of the proposed ARZA-SVSS-LMS algorithm. First, the misalignment error vector is defined as

$$\boldsymbol{\Delta}(n) = \hat{\mathbf{h}}(n) - \mathbf{h}, \tag{3.13}$$

and the auto-covariance matrix of $\boldsymbol{\Delta}(n)$ is defined as

$$\mathbf{K}(n) = E\{\boldsymbol{\Delta}(n)\boldsymbol{\Delta}^T(n)\}. \tag{3.14}$$

Assuming that $\mathbf{h}$ is a real FIR channel vector, subtracting $\mathbf{h}$ from both sides of (3.9), we obtain

$$\hat{\mathbf{h}}(n+1) - \mathbf{h} = \hat{\mathbf{h}}(n) - \mathbf{h} + \mu_{\mathrm{SVSS}}(n)e(n)\mathbf{x}(n) - \rho_{\mathrm{ARZA}}\mathbf{m}(n), \tag{3.15}$$

where

$$\mathbf{m}(n) = \frac{\mathrm{sgn}(\hat{\mathbf{h}}(n))}{1 + \varepsilon_{\mathrm{ARZA}}(n)\left|\hat{\mathbf{h}}(n)\right|}. \tag{3.16}$$

By taking

$$r(n) = \mathbf{h}^T\mathbf{x}(n) + v(n) \tag{3.17}$$

and (2.2) and (3.13) into consideration [67], we obtain

$$\begin{aligned}&\boldsymbol{\Delta}(n+1) = \boldsymbol{\Delta}(n) + \mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)(r(n) - \hat{\mathbf{h}}^T(n)\mathbf{x}(n)) - \rho_{\mathrm{ARZA}}\mathbf{m}(n)\\ &=\boldsymbol{\Delta}(n) + \mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)v(n) - \mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)\mathbf{x}^T(n)\boldsymbol{\Delta}(n) - \rho_{\mathrm{ARZA}}\mathbf{m}(n)\\ &=(\mathbf{I} - \mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)\mathbf{x}^T(n))\boldsymbol{\Delta}(n) + \mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)v(n) - \rho_{\mathrm{ARZA}}\mathbf{m}(n)\end{aligned} . \tag{3.18}$$

Substituting (3.18) into (3.14) yields

$$\begin{aligned}\mathbf{K}(n+1) &= [\mathbf{A}(n)\boldsymbol{\Delta}(n) + \mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)v(n) - \rho_{\mathrm{ARZA}}\mathbf{m}(n)]\\ &\quad [\mathbf{A}(n)\boldsymbol{\Delta}(n) + \mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)v(n) - \rho_{\mathrm{ARZA}}\mathbf{m}(n)]^T\\ &= [\mathbf{A}(n)\boldsymbol{\Delta}(n) + \mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)v(n) - \rho_{\mathrm{ARZA}}\mathbf{m}(n)]\\ &\quad [\boldsymbol{\Delta}^T(n)\mathbf{A}^T(n) + \mu_{\mathrm{SVSS}}(n)v(n)\mathbf{x}^T(n) - \rho_{\mathrm{ARZA}}\mathbf{m}^T(n)]\\ &= \mathrm{E}\{\mathbf{A}(n)\boldsymbol{\Delta}(n)\boldsymbol{\Delta}^T(n)\mathbf{A}^T(n)\} + \mathrm{E}\{\mu_{\mathrm{SVSS}}(n)v(n)\mathbf{A}(n)\boldsymbol{\Delta}(n)\mathbf{x}^T(n)\}\\ &\quad +\mathrm{E}\{-\rho_{\mathrm{ARZA}}\mathbf{A}(n)\boldsymbol{\Delta}(n)\mathbf{m}^T(n)\}\\ &\quad +\mathrm{E}\{\mu_{\mathrm{SVSS}}(n)v(n)\mathbf{x}(n)\boldsymbol{\Delta}^T(n)\mathbf{A}^T(n)\}\\ &\quad +\mathrm{E}\{\mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)v(n)\mu_{\mathrm{SVSS}}(n)v(n)\mathbf{x}^T(n)\}\\ &\quad +\mathrm{E}\{-\rho_{\mathrm{ARZA}}\mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)v(n)\mathbf{m}^T(n)\}\\ &\quad +\mathrm{E}\{-\rho_{\mathrm{ARZA}}\mathbf{m}(n)\boldsymbol{\Delta}^T(n)\mathbf{A}^T(n)\}\\ &\quad +\mathrm{E}\{-\rho_{\mathrm{ARZA}}\mathbf{m}(n)\mu_{\mathrm{SVSS}}(n)v(n)\mathbf{x}^T(n)\}\\ &\quad +\mathrm{E}\{\rho^2_{\mathrm{ARZA}}\mathbf{m}(n)\mathbf{m}^T(n)\}\end{aligned} , \tag{3.19}$$

where $\mathbf{A}(n) = \mathbf{I}_N - \mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)\mathbf{x}^T(n)$ and $\mathbf{I}_N$ is the $N \times N$ identity matrix. By taking into consideration the statistical independence between $\mathbf{x}(n)$ and $v(n)$, (3.19) can be simplified to

$$\begin{aligned}&\mathbf{K}(n+1) = \mathrm{E}\{\mathbf{A}(n)\boldsymbol{\Delta}(n)\boldsymbol{\Delta}^T(n)\mathbf{A}^T(n)\} + \mathrm{E}\{-\rho_{\mathrm{ARZA}}\mathbf{A}(n)\boldsymbol{\Delta}(n)\mathbf{m}^T(n)\}\\ &+\mathrm{E}\{\mu_{\mathrm{SVSS}}(n)\mathbf{x}(n)v(n)\mu_{\mathrm{SVSS}}(n)v(n)\mathbf{x}^T(n)\} + \mathrm{E}\{-\rho_{\mathrm{ARZA}}\mathbf{m}(n)\boldsymbol{\Delta}^T(n)\mathbf{A}^T(n)\}\\ &+\mathrm{E}\{\rho^2_{\mathrm{ARZA}}\mathbf{m}(n)\mathbf{m}^T(n)\}\\ &= \mathrm{E}\{\mathbf{A}(n)\boldsymbol{\Delta}(n)\boldsymbol{\Delta}^T(n)\mathbf{A}^T(n)\} - \rho_{\mathrm{ARZA}}\mathrm{E}\{\mathbf{A}(n)\boldsymbol{\Delta}(n)\mathbf{m}^T(n)\} + \mu^2_{\mathrm{SVSS}}(n)\delta_v^2\delta_x^2\mathbf{I}_N\\ &- \rho_{\mathrm{ARZA}}\mathrm{E}\{\mathbf{m}(n)\boldsymbol{\Delta}^T(n)\mathbf{A}^T(n)\} + \rho^2_{\mathrm{ARZA}}\mathrm{E}\{\mathbf{m}(n)\mathbf{m}^T(n)\}\end{aligned} , \tag{3.20}$$

where δ_v^2 and δ_x^2 denote the powers of the input signal and additive noise, respectively. Using the property that the fourth-order moment of a Gaussian variable is three times the square of the variance [99], we obtain

$$\begin{aligned} \mathrm{E}\{\mathbf{A}(n)\boldsymbol{\Delta}(n)\boldsymbol{\Delta}^T(n)\mathbf{A}^T(n)\} = {} & (1-2\mu_{\mathrm{SVSS}}(n)\delta_x^2+2\mu_{\mathrm{SVSS}}^2(n)\delta_x^4)\mathbf{K}(n) \\ & +\mu_{\mathrm{SVSS}}^2(n)\delta_x^4 D(n)\mathbf{I}_N \end{aligned}, \tag{3.21}$$

where $D(n) = \mathrm{tr}[\mathbf{K}(n)]$. Also,

$$\begin{aligned} \mathrm{E}\{\mathbf{A}(n)\boldsymbol{\Delta}(n)\mathbf{m}^T(n)\} &= \mathrm{E}\{\mathbf{m}(n)\boldsymbol{\Delta}^T(n)\mathbf{A}^T(n)\}\}^T \\ &= (1-\mu_{\mathrm{SVSS}}(n)\delta_x^2)\mathrm{E}\{\boldsymbol{\Delta}(n)\mathbf{m}^T(n)\} \end{aligned}. \tag{3.22}$$

Combining (3.20)-(3.22), we obtain

$$\begin{aligned} &\mathbf{K}(n+1) = (1-2\mu_{\mathrm{SVSS}}(n)\delta_x^2+2\mu_{\mathrm{SVSS}}^2(n)\delta_x^4)\mathbf{K}(n)+\mu_{\mathrm{SVSS}}^2(n)\delta_x^4 D(n)\mathbf{I}_N \\ &+\mu_{\mathrm{SVSS}}(n)\delta_v^2\delta_x^2\mathbf{I}_N-\rho_{\mathrm{ARZA}}(1-\mu_{\mathrm{SVSS}}(n)\delta_x^2)\mathrm{E}\{\boldsymbol{\Delta}(n)\mathbf{m}^T(n)\} \\ &-\rho_{\mathrm{ARZA}}(1-\mu_{\mathrm{SVSS}}(n)\delta_x^2)\{\mathrm{E}\{\boldsymbol{\Delta}(n)\mathbf{m}^T(n)\}\}^T+\rho_{\mathrm{ARZA}}^2\mathrm{E}\{\mathbf{m}(n)\mathbf{m}^T(n)\} \end{aligned}. \tag{3.23}$$

By taking the trace on both sides of (3.23), it can be concluded that the adaptive filter is stable if and only if [99]

$$0 < (1-2\mu_{\mathrm{SVSS}}(n)\delta_x^2+(N+2)\mu_{\mathrm{SVSS}}^2(n)\delta_x^4) < 1, \tag{3.24}$$

which is simplified to

$$0 < \mu_{\mathrm{SVSS}}(n) < \frac{2}{(N+2)\delta_x^2}. \tag{3.25}$$

We can see that the proposed ARZA-SVSS-LMS algorithm has the same stability condition for the mean square convergence as the ZA-LMS and standard LMS [43]. Thus, we should choose a large initial step-size for $\mu_{\mathrm{SVSS}}(n)$ that is subjected to (3.25).

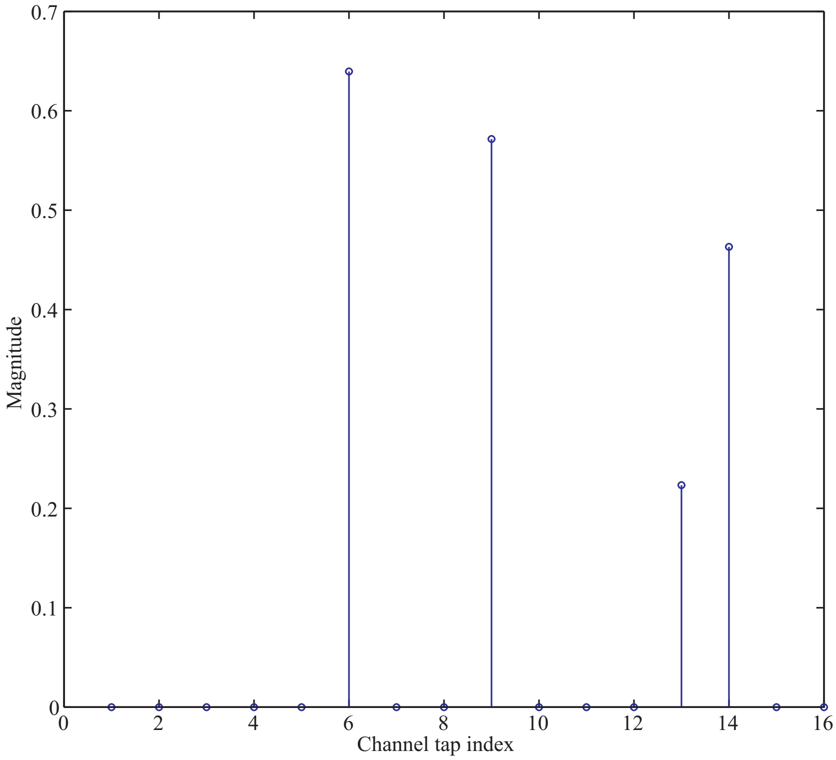


Figure 3.2: Typical sparse multipath channel.

3.4 Performance of the proposed sparse channel estimation algorithms

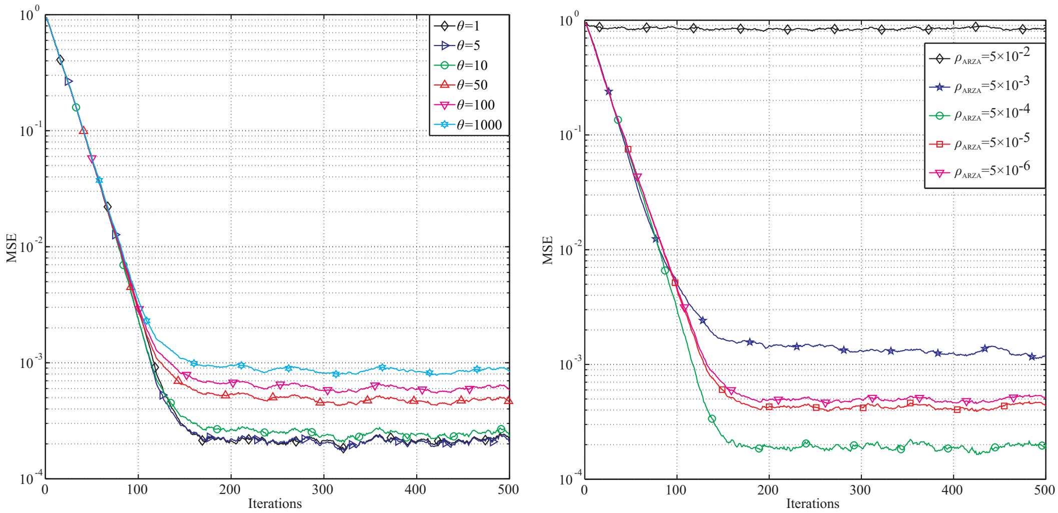
In this section, we report our investigation of the parameter effects and the estimation performances of the proposed ARZA-SVSS-LMS algorithm for sparse channel estimation by computer simulation. We consider a sparse multipath channel of length $N = 16$, whose number of dominant taps K is set to three different sparsity levels, namely $K = 1, K = 4$ and $K = 8$, similarly to [2, 67, 98]. The nonzero channel taps follow 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. An example of a typical sparse multipath channel with a channel length of 16 and 4 dominant taps is shown in Fig. 3.2. In the simulation, the input signal $\mathbf{x}(n)$ of the channel is Gaussian random signal while the output of the channel is corrupted by an independent white Gaussian noise $v(n)$. In the simulations, the power of the received signal is $E_b = 1$, while the noise power is given by $\delta_v^2 = 10^{-\frac{\text{SNR}}{10}}$ and the SNR is defined as SNR$= 10 \log \frac{E_b}{\delta_v^2}$. All the LMS algorithms are run 100 times. We investigate the relevant parameters for the proposed SVSS algorithms and compare the estimation behaviors of the proposed ARZA-SVSS-LMS algorithm with those of the previously proposed ZA-LMS, RZA-LMS, VSS-LMS and standard LMS algorithms at SNR=20 dB and 30 dB. In all the experiments, the difference between the actual and estimated channels based on the proposed sparsity-aware algorithms and the sparse channels mentioned above is evaluated by the MSE, defined as [2]

$$\text{MSE}\left\{\hat{\mathbf{h}}(n)\right\} = \text{E}\left\{\left\|\mathbf{h} - \hat{\mathbf{h}}(n)\right\|_2^2\right\}, \tag{3.26}$$

where $\mathbf{h}$ and $\hat{\mathbf{h}}(n)$ are the actual channel vector and its estimator, respectively.

3.4.1 Effects of parameters on the ARZA-SVSS-LMS algorithm

First, the parameters θ and ρ_{ARZA} are used to analyze the steady-state performance of the proposed ARZA-SVSS-LMS algorithm, and the obtained results are illustrated in Fig. 3.3. In the simulation, the parameters are $\xi = 0.999$, $\alpha = 100$, $\beta = 0.04$, $m = 0.1$, $K = 4$ and SNR $= 30$dB. It is observed from Fig. 3.3a that the steady-state performance of the proposed ARZA-SVSS-LMS algorithm is degraded with increasing θ when θ is greater than 10. Figure 3.3b shows that ρ_{ARZA} has an important effect on the steady-state performance of the proposed ARZA-SVSS-LMS algorithm. With decreasing ρ_{ARZA}, the steady-state performance is first improved and then degraded. This is because a small ρ_{ARZA} reduces the zero-attractor strength of the proposed ARZA-SVSS-LMS algorithm. On the other hand, a large ρ_{ARZA} results in a large deviation while a small ρ_{ARZA} means a

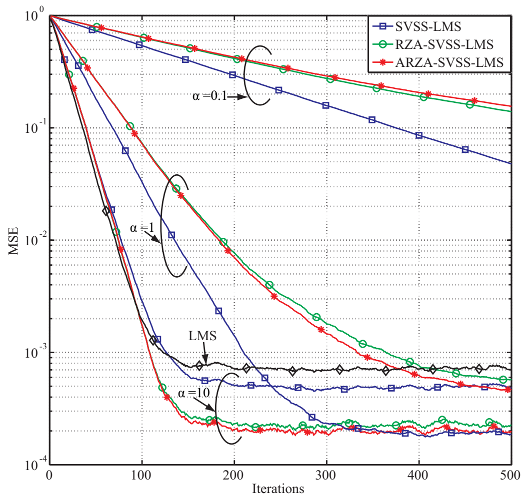


(a) Effects of θ ($\rho_{\text{ARZA}} = 5 \times 10^{-4}$) (b) Effects of ρ_{ARZA} ($\theta = 5$)

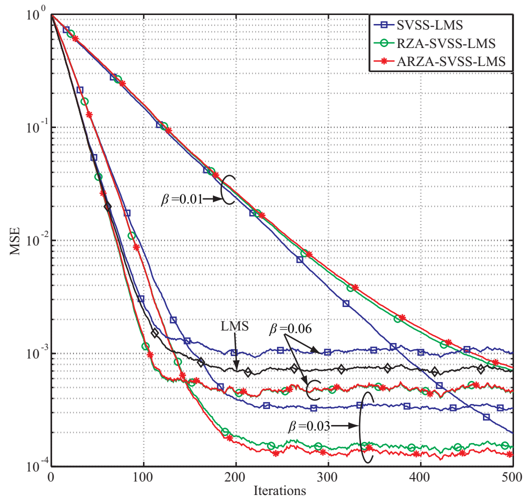
Figure 3.3: Effects of parameters on the proposed ARZA-SVSS-LMS algorithm.

weak zero attractor. Thus, in our proposed ARZA-SVSS-LMS algorithm, we can choose suitable values of ρ_{ARZA} and θ for the tradeoff between the convergence speed and steady-state error.

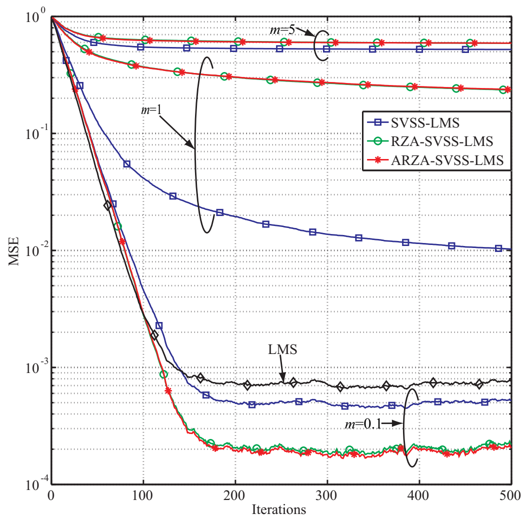
Next, we investigate the effects of varying parameters α, β, m and SNR on the proposed ARZA-SVSS-LMS algorithm where the obtained results are illustrated in Fig. 3.4. Since the proposed ARZA-SVSS-LMS algorithm is an improved RZA-SVSS-LMS, then RZA-SVSS-LMS, SVSS-LMS and standard LMS algorithms are also selected to study the parameter effects. The simulation parameters are $\mu_{\text{LMS}} = 0.05$, $\xi = 0.999$, $\rho_{\text{RZA-LMS}} = \rho_{\text{ARZA}} = 5 \times 10^{-4}$, $\alpha = 100$, $\beta = 0.04$, $m = 0.1$, $K = 4$, SNR $= 30$dB and $\theta = 5$. When we change one of these parameters, the other parameters are kept constant. The effects of α are shown in Fig. 3.4a. It can be seen from Fig. 3.4a that the convergence speeds of the proposed SVSS-LMS algorithms decrease rapidly with decreasing of α. When $\alpha = 10$, the proposed ARZA-SVSS-LMS and RZA-SVSS-LMS algorithms have nearly the same performances. Additionally, both algorithms outperform the proposed SVSS-LMS and standard LMS algorithms. However, with decreasing α, both the convergence speed and the steady-state performance of the ARZA-SVSS-LMS and RZA-SVSS-LMS algorithms are degraded, and the SVSS-LMS algorithm has a lower mean square error. The effects of β are shown in Fig. 3.4b.



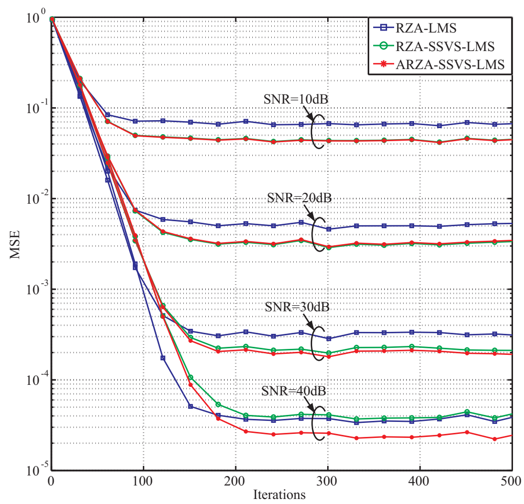
(a) Effects of α



(b) Effects of β



(c) Effects of m



(d) Effects of SNR

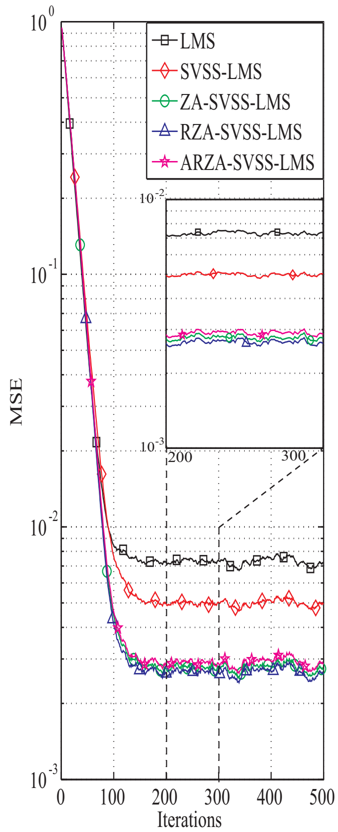
Figure 3.4: Performance of algorithms with variable step size based on sigmoid function.

It is found that with increasing β, both the convergence speed and the steady-state performance of all the proposed SVSS-LMS algorithms are initially improved. Also, the steady-state performance of the proposed ARZA-SVSS-LMS algorithm even exceeds those of the RZA-SVSS-LMS and SVSS-LMS algorithms when $\beta = 0.03$. When β is increased to 0.06, the steady-state performances of the proposed SVSS-LMS algorithms degrade. Also, the proposed ARZA-SVSS-LMS and RZA-SVSS-LMS algorithms are superior to the SVSS-LMS and standard LMS algorithms when $\beta = 0.06$ in terms of both the convergence speed and the steady-state behavior. The performances of the proposed SVSS-LMS algorithms with varying m are shown in Fig. 3.4c. Both the steady-state performance and the convergence speed decrease with increasing m. For the case of $m = 0.1$, both the convergence speed and the steady-state performances of the proposed ARZA-SVSS-LMS and RZA-SVSS-LMS algorithms are better than those of the SVSS-LMS and standard LMS algorithms, contrary to the case when $m = 1$ and $m = 5$. When m is less than 0.1, the estimation performance of these SVSS-LMS algorithms are nearly the same as the results of $m = 0.1$. The results of investigating the proposed ARZA-SVSS-LMS and RZA-SVSS-LMS algorithms by simulation with different noise level are shown in Fig. 3.4d. In this experiment, we evaluated the RZA-LMS, ARZA-SVSS-LMS and RZA-SVSS-LMS algorithms because the performance of the RZA-LMS algorithm is better than or the same as those of the standard LMS and ZA-LMS algorithms [67]. It is revealed that the performances of the proposed ARZA-SVSS-LMS and RZA-SVSS-LMS algorithms are superior to that of the RZA-LMS algorithm with respect to the steady-state error. When the SNR is greater than 30 dB, the proposed ARZA-SVSS-LMS algorithm can obtain a better steady-state performance than the RZA-SVSS-LMS and RZA-LMS algorithms because the adaptive adjustment of $\varepsilon_{\mathrm{ARZA}}(n)$ in the ARZA-SVSS-LMS algorithm can effectively adjust the zero-attracting strength. However, the convergence speeds of the proposed ARZA-SVSS-LMS and RZA-SVSS-LMS algorithms are slightly degraded. Thus, we can choose suitable parameters to balance the convergence speed and steady-state performance.

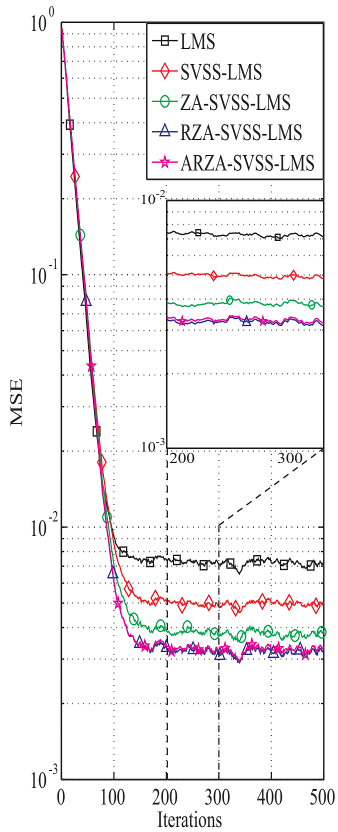
3.4.2 Effects of sparsity level on the ARZA-SVSS-LMS algorithm

Finally, we investigate the estimation performance of the proposed SVSS-LMS algorithms over a time-varying sparse multipath channel and the simulation results are shown in Figs. 3.5, 3.6, 3.7 and 3.8. The simulation parameters are

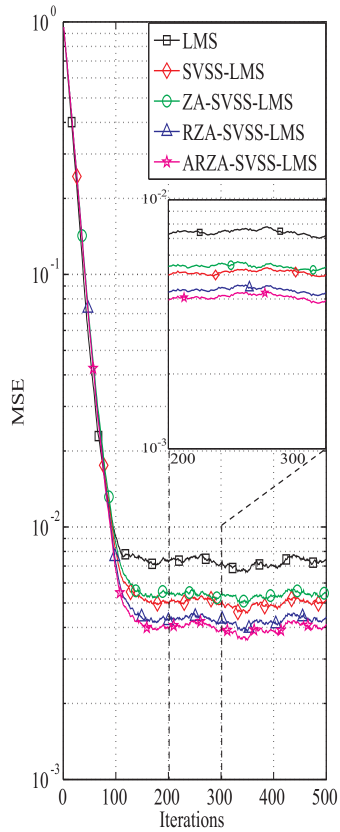
$\mu_{\text{LMS}} = \mu_{\text{ZA-LMS}} = \mu_{\text{RZA-LMS}} = 0.05$, $\rho_{\text{ZA-LMS}} = \rho_{\text{RZA-LMS}} = 5 \times 10^{-4}$, $\varepsilon_{\text{RZA-LMS}} = 10$ as suggested in [2, 67], $\rho_{\text{ARZA}} = 5 \times 10^{-4}$, $\xi = 0.999$, $\alpha = 100$, $\beta = 0.04$, $m = 0.1$, $\theta = 5$, $\mu_{max} = 0.5$, $\mu_{min} = 0.002$, $\kappa = 0.998$ and $\chi = 2 \times 10^{-3}$. We can see from Fig. 3.5 that our proposed SVSS-LMS algorithms are superior to the standard LMS algorithm in terms of both the steady-state error and the convergence speed for 20 dB. In particular, the proposed ARZA-SVSS-LMS, RZA-SVSS-LMS and ZA-SVSS-LMS algorithms have the lowest steady-state error when $K = 1$ shown in Fig. 3.5a. With the increasing of the number of non-zero taps K, the steady-state performance of the ARZA-SVSS-LMS, RZA-SVSS-LMS and ZA-SVSS-LMS are deteriorated. However, the proposed ARZA-SVSS-LMS and RZA-SVSS-LMS algorithms still have the best steady-state performance for



(a) K=1

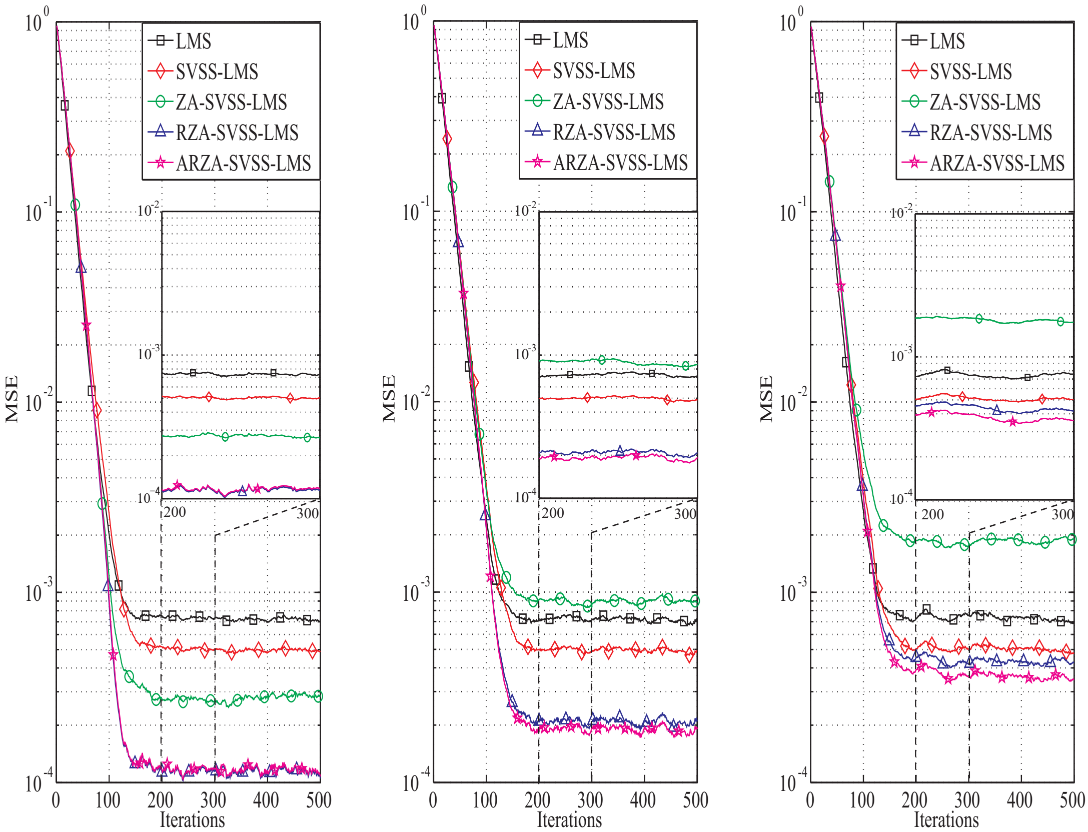


(b) K=4



(c) K=8

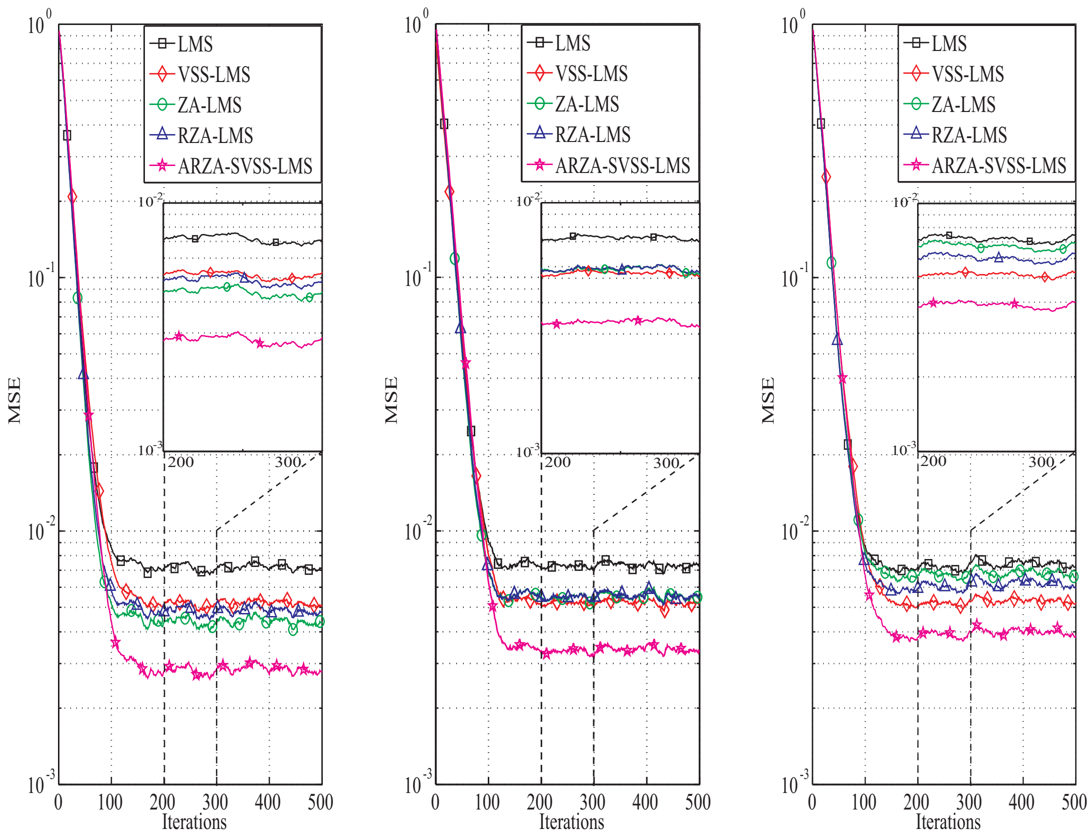
Figure 3.5: Performance of the proposed sparse channel estimation algorithms for SNR=20 dB.



(a) K=1 (b) K=4 (c) K=8

Figure 3.6: Performance of the proposed sparse channel estimation algorithms for SNR=30 dB.

$K = 4$. When K increases to 8, the ZA-SVSS-LMS is degraded and even worse than the SVSS-LMS algorithm because the ZA-SVSS-LMS cannot distinguish the non-zero taps and the zero taps, and uniformly exerts the l_1-penalty on all the channel taps. In this case, our proposed ARZA-SVSS-LMS algorithm still outperforms the other SVSS-LMS and standard LMS algorithms. When the SNR increases to 30 dB, we can see from Fig. 3.6 that the ARZA-SVSS-LMS and RZA-SVSS-LMS algorithms have the same steady-state performance for $K = 1$. When $K = 4$ and $K = 8$, the steady-state error of the RZA-SVSS-LMS algorithm is increased. However, our proposed ARZA-SVSS-LMS still has the best steady-state performance because of its adaptive parameter adjustment method which effectively changes $\varepsilon_{\text{ARZA}}(n)$ and further reduces the steady-state error. Next,

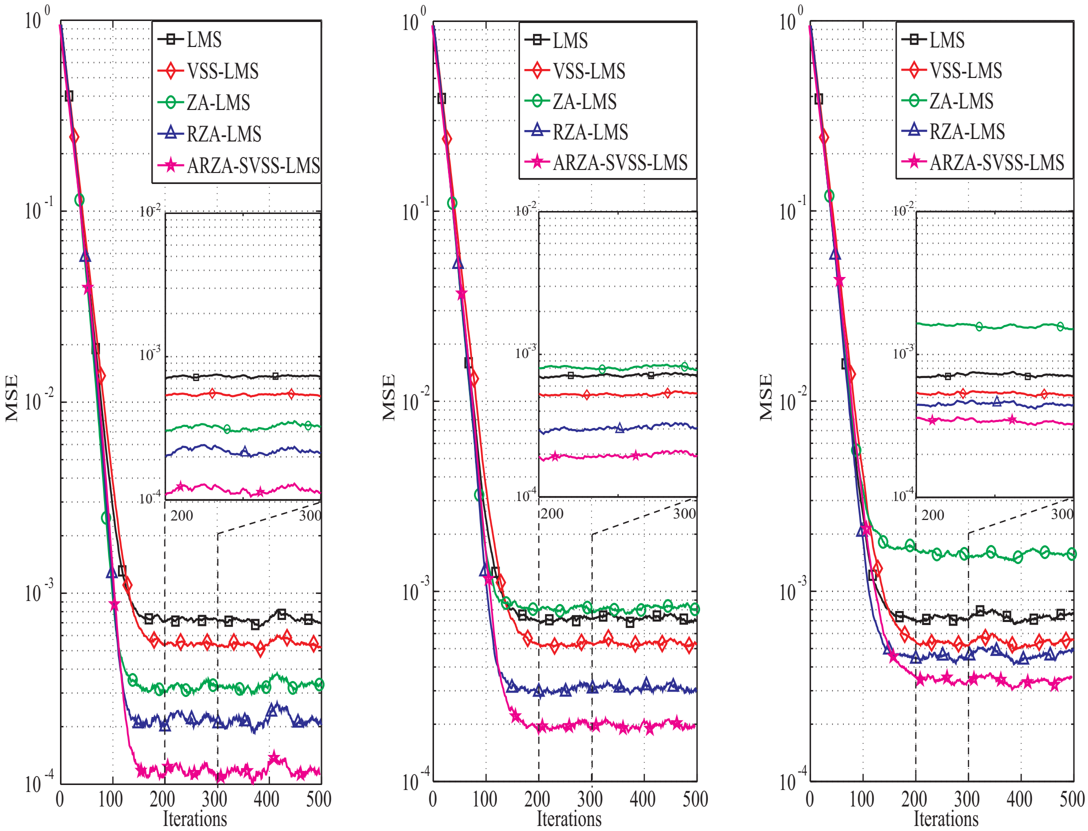


(a) K=1 (b) K=4 (c) K=8

Figure 3.7: Performance of theARZA-SVSS-LMS algorithm for SNR=20 dB.

we compare the estimation performance of our proposed ARZA-SVSS-LMS algorithm with the previously proposed VSS-LMS, ZA-LMS, RZA-LMS and standard LMS algorithms. The simulation results are shown in Figs. 3.7 and 3.8 for SNR=20 dB and SNR=30 dB, respectively. It is found that our proposed ARZA-SVSS-LMS algorithm has the best steady-state performance both at 20 dB and 30 dB. However, the convergence speed of the ARZA-SVSS-LMS algorithm is slightly deteriorated compared to the RZA-LMS algorithm only when $K = 8$ at 30 dB, which can be improved by the proper selection of the parameters in the ARZA-SVSS-LMS algorithm.

On the basis of the above discussions, it is expected that as the sparsity level K increases, the steady-state performances of the sparsity-aware channel estimation algorithms are degraded, which can indeed be observed in Figs. 3.5, 3.6, 3.7



(a) K=1 (b) K=4 (c) K=8

Figure 3.8: Performance of the ARZA-SVSS-LMS algorithm for SNR=30 dB.

and 3.8. It is observed that our proposed ARZA-SVSS-LMS algorithm not only improves the convergence speed of the standard LMS algorithm but also enhances steady-state performances of the sparse channel estimation algorithms compared with the previously proposed ZA-LMS and RZA-LMS algorithms. In these sparse channel estimation algorithms, the zero attractor accelerates the convergence speed when the majority of the channel taps are zero or nearly zero. The adaptive parameter adjustment method in the proposed ARZA-SVSS-LMS algorithm further improves the steady-state behavior of the RZA-SVSS-LMS algorithm at a high SNR. For the same parameters, the proposed ARZA-SVSS-LMS algorithm outperforms other sparse channel estimation algorithms. Thus, we can conclude that the ARZA-SVSS-LMS algorithm is not only stable in the case of unknown signals, but also robust to noise interference at both low and high SNRs when the signal is sparse.

3.5 Conclusion

In this chapter, we proposed a robust adaptive sparse channel estimation algorithm, ARZA-SVSS-LMS algorithm, using a sigmoid functioned variable-step-size and an adaptive parameter adjustment method, and the performances of the proposed ARZA-SVSS-LMS algorithm and its related sparse algorithms were investigated and compared with those of previous l_1-penalized LMS and standard LMS algorithms in a sparse multipath channel. The effects of relevant parameters in the ARZA-SVSS-LMS algorithm and the derived sparse channel estimation algorithms were investigated in detail. The simulation results demonstrated that the proposed ARZA-SVSS-LMS algorithm outperforms the previous l_1-penalized LMS algorithms. On the other hand, our proposed ARZA-SVSS-LMS algorithm is much more robust than the RZA-LMS algorithm. In addition, the proposed SVSS-LMS algorithms also have better performances than the VSS-LMS and standard LMS algorithms.

Chapter 4

Smooth Approximation l_0-Norm Constrained Affine Projection Algorithm

4.1 Introduction

In this chapter, 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 [100]. 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 non-zero taps in the sparse channel increases.

4.2 Proposed SL0-APA algorithm

On the basis of the discussion of the ZA-APA in Section 2.4.2 and RZA-APA in Section 2.4.3, we find that the RZA-APA can improve the performance of ZA-APA for sparse channel estimation because $\sum_{i=1}^{N} \log(1 + \varepsilon_{\mathrm{RZA-APA}} \left|\hat{h}_i(n+1)\right|)$ is more similar to l_0-norm [67, 82, 101]. On the other hand, solving l_0-norm $\left\|\hat{\mathbf{h}}(n+1)\right\|_0$ is a NP-hard problem [19]. Fortunately, smooth approximation l_0-norm (SL0) with low complexity have been proposed as a accurate approximation of $\left\|\hat{\mathbf{h}}(n+1)\right\|_0$ to reconstruct sparse signals in CS theory [22, 102, 103]. 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 [100].

Similar to the ZA-APA and RZA-APA discussed in Section 2.4.2 and Section 2.4.3, respectively, the cost function of the SL0-APA is written as

$$J_{\mathrm{SL0}}(n) = \left\|\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)\right\|^2 + [\mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n+1)]^T \boldsymbol{\lambda}_{\mathrm{SL0}} + \gamma_{\mathrm{SL0}}\left\|\hat{\mathbf{h}}(n+1)\right\|_0, \tag{4.1}$$

where $\boldsymbol{\lambda}_{\mathrm{SL0}}$ is the vector of the Lagrange multiplier with size of $Q \times 1$ and $\gamma_{\mathrm{SL0}} > 0$ 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 $\left\|\hat{\mathbf{h}}(n+1)\right\|_0$ is a continuous function defined as follows [103]

$$\left\|\hat{\mathbf{h}}(n+1)\right\|_0 = \sum_{i=1}^{N-1} \frac{\left|\hat{h}_i(n+1)\right|}{\left|\hat{h}_i(n+1)\right| + \delta} = \frac{\left|\hat{\mathbf{h}}(n+1)\right|}{\left|\hat{\mathbf{h}}(n+1)\right| + \delta}, \tag{4.2}$$

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

$$\frac{\partial\left\|\hat{\mathbf{h}}(n+1)\right\|_0}{\partial\hat{\mathbf{h}}(n+1)} = \frac{\delta\mathrm{sgn}(\hat{\mathbf{h}}(n+1))}{(\left|\hat{\mathbf{h}}(n+1)\right| + \delta)^2}. \tag{4.3}$$

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

$$\frac{\partial J_{\mathrm{SL0}}(n)}{\partial\hat{\mathbf{h}}(n+1)} = 2\hat{\mathbf{h}}(n+1) - 2\hat{\mathbf{h}}(n) - \mathbf{U}^T(n)\boldsymbol{\lambda}_{\mathrm{SL0}} + \gamma_{\mathrm{SL0}}\frac{\delta\mathrm{sgn}(\hat{\mathbf{h}}(n+1))}{(\left|\hat{\mathbf{h}}(n+1)\right| + \delta)^2}. \tag{4.4}$$

Let the left hand side of the equation of (4.4) be equal to zero. We can get the following equation

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}^T(n)\boldsymbol{\lambda}_{\mathrm{SL0}} - \frac{1}{2}\gamma_{\mathrm{SL0}}\frac{\delta\mathrm{sgn}(\hat{\mathbf{h}}(n+1))}{(\left|\hat{\mathbf{h}}(n+1)\right| + \delta)^2}. \tag{4.5}$$

Multiplying $\mathbf{U}(n)$ on both sides of (4.5), we can get

$$\mathbf{U}(n)\hat{\mathbf{h}}(n+1) = \mathbf{U}(n)\hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}(n)\mathbf{U}^T(n)\boldsymbol{\lambda}_{\mathrm{SL0}} - \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{U}(n)\frac{\delta\mathrm{sgn}(\hat{\mathbf{h}}(n+1))}{(\left|\hat{\mathbf{h}}(n+1)\right| + \delta)^2}. \tag{4.6}$$

By taking the equation (2.23) into consideration, the equation (4.6) can be rewritten as

$$\mathbf{r}(n) = \mathbf{U}(n)\hat{\mathbf{h}}(n) + \frac{1}{2}\mathbf{U}(n)\mathbf{U}^T(n)\boldsymbol{\lambda}_{\mathrm{SL0}} - \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{U}(n)\frac{\delta \mathrm{sgn}(\hat{\mathbf{h}}(n+1))}{(\left|\hat{\mathbf{h}}(n+1)\right| + \delta)^2}. \tag{4.7}$$

From the discussion of the ZA-APA and RZA-APA, we can get the Lagrange multiplier vector $\boldsymbol{\lambda}_{\mathrm{SL0}}$ from (4.7) by taking $\mathbf{e}(n) = \mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n)$ into account

$$\boldsymbol{\lambda}_{\mathrm{SL0}} = [\mathbf{U}(n)\mathbf{U}^T(n)]^{-1}\{2\mathbf{e}(n) + \gamma_{\mathrm{SL0}}\mathbf{U}(n)\frac{\delta \mathrm{sgn}(\hat{\mathbf{h}}(n+1))}{(\left|\hat{\mathbf{h}}(n+1)\right| + \delta)^2}\}. \tag{4.8}$$

Substituting (4.8) into (4.5) and assuming that $\delta \mathrm{sgn}(\hat{\mathbf{h}}(n+1))/(\left|\hat{\mathbf{h}}(n+1)\right| + \delta)^2 \approx \delta \mathrm{sgn}(\hat{\mathbf{h}}(n))/(\left|\hat{\mathbf{h}}(n)\right| + \delta)^2$, the update function of the SL0-APA can be achieved

$$\begin{aligned}\hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mathbf{U}^+(n)\mathbf{e}(n) + \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{U}^+(n)\mathbf{U}(n)\frac{\delta \mathrm{sgn}(\hat{\mathbf{h}}(n))}{(\left|\hat{\mathbf{h}}(n)\right| + \delta)^2} \\ &\quad -\frac{1}{2}\gamma_{\mathrm{SL0}}\frac{\delta \mathrm{sgn}(\hat{\mathbf{h}}(n))}{(\left|\hat{\mathbf{h}}(n)\right| + \delta)^2} \\ &= \hat{\mathbf{h}}(n) + \mathbf{U}^+(n)\mathbf{e}(n) + \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{U}^+(n)\mathbf{U}(n)\mathbf{T}(n) - \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{T}(n)\end{aligned}, \tag{4.9}$$

where $\mathbf{T}(n) = \delta \mathrm{sgn}(\hat{\mathbf{h}}(n))/(\left|\hat{\mathbf{h}}(n)\right| + \delta)^2$. Similar to the ZA-APA and RZA-APA, a step size μ_{SL0} is introduced into (4.9) to create a balance between the convergence speed and steady-state error of the SL0-APA

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\mathrm{SL0}}\mathbf{U}^+(n)\mathbf{e}(n) + \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{U}^+(n)\mathbf{U}(n)\mathbf{T}(n) - \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{T}(n). \tag{4.10}$$

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 $\left\|\hat{\mathbf{h}}(n+1)\right\|_0$, which is proved to be an approximate and near-accurate approximation of l_0-norm in comparison with the sum-log function $\sum_{i=1}^{N}\log(1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{h}_i(n+1)\right|)$ 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[2].

4.3 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 [106–108] 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

$$\mathbf{r}(n) = \mathbf{U}(n)\mathbf{h} + \mathbf{v}(n), \tag{4.11}$$

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

$$\mathrm{MSE} \triangleq \lim_{n\to\infty} \mathrm{E}[|\mathbf{e}(n)|^2], \tag{4.12}$$

where $\mathrm{E}[\cdot]$ denotes the expectation and

$$\mathbf{e}(n) = \mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n) \tag{4.13}$$

[2]We derived the SL0-APA in this section by considering the l_0-norm-based LMS and NLMS [2, 84, 104] and the smooth approximation l_0-norm in [103]. In the paper [105], l_0-norm constrained APA was also derived around the same time as our paper was submitted for publication.

is the estimated error at time n. Taking the equation (4.11) and (4.13) into account, we obtain

$$\begin{aligned} \mathbf{e}(n) &= \mathbf{U}(n)\mathbf{h} + \mathbf{v}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n) \\ &= \mathbf{U}(n)[\mathbf{h} - \hat{\mathbf{h}}(n)] + \mathbf{v}(n) \end{aligned} . \tag{4.14}$$

Subtracting $\mathbf{h}$ from both sides of the SL0-APA update function (4.10), we get the misaligment vector

$$\begin{aligned} &\boldsymbol{\Delta}(n+1) = \mathbf{h} - \hat{\mathbf{h}}(n+1) \\ &= \mathbf{h} - \left\{ \hat{\mathbf{h}}(n) + \mu_{\mathrm{SL0}}\mathbf{U}^{+}(n)\mathbf{e}(n) + \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{U}^{+}(n)\mathbf{U}(n)\mathbf{T}(n) - \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{T}(n) \right\} \end{aligned} . \tag{4.15}$$

Substituting (4.14) into (4.15), we can get

$$\begin{aligned} \boldsymbol{\Delta}(n+1) &= \mathbf{h} - \hat{\mathbf{h}}(n) - \mu_{\mathrm{SL0}}\mathbf{U}^{+}(n)\left\{ \mathbf{U}(n)[\mathbf{h} - \hat{\mathbf{h}}(n)] + \mathbf{v}(n) \right\} \\ &\quad -\frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{U}^{+}(n)\mathbf{U}(n)\mathbf{T}(n) + \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{T}(n) \\ &= [\mathbf{I}_N - \mu_{\mathrm{SL0}}\mathbf{U}^{+}(n)\mathbf{U}(n)]\,\boldsymbol{\Delta}(n) - \mu_{\mathrm{SL0}}\mathbf{U}^{+}(n)\mathbf{v}(n) \\ &\quad -\frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{U}^{+}(n)\mathbf{U}(n)\mathbf{T}(n) + \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{T}(n) \end{aligned} . \tag{4.16}$$

Taking expectations on the both sides of (4.16), we get

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{\Delta}(n+1)\right] &= \mathrm{E}\left[\mathbf{I}_N - \mu_{\mathrm{SL0}}\mathbf{U}^{+}(n)\mathbf{U}(n)\right]\mathrm{E}\left[\boldsymbol{\Delta}(n)\right] - \mu_{\mathrm{SL0}}\mathrm{E}\left[\mathbf{U}^{+}(n)\mathbf{v}(n)\right] \\ &\quad -\frac{1}{2}\gamma_{\mathrm{SL0}}\mathrm{E}\left[\mathbf{U}^{+}(n)\mathbf{U}(n)\mathbf{T}(n)\right] + \frac{1}{2}\gamma_{\mathrm{SL0}}\mathrm{E}\left[\mathbf{T}(n)\right] \end{aligned} . \tag{4.17}$$

We assume the additive noise $\mathbf{v}(n)$ is statistically independent of the input signal $\mathbf{x}(n)$, and hence we have $\mathrm{E}\left[\mathbf{U}^{+}(n)\mathbf{v}(n)\right] = \mathbf{0}$. Therefore, the equation

(4.17) can be simplified as

$$
\begin{aligned}
\mathrm{E}\left[\boldsymbol{\Delta}(n+1)\right] = \mathrm{E}\left[\mathbf{I}_N - \mu_{\mathrm{SL0}}\mathbf{U}^{+}(n)\mathbf{U}(n)\right]\mathrm{E}\left[\boldsymbol{\Delta}(n)\right] \\
-\frac{1}{2}\gamma_{\mathrm{SL0}}\mathrm{E}\left[\mathbf{U}^{+}(n)\mathbf{U}(n)\mathbf{T}(n)\right] + \frac{1}{2}\gamma_{\mathrm{SL0}}\mathrm{E}\left[\mathbf{T}(n)\right]
\end{aligned}
. \tag{4.18}
$$

From the previous studies on sparse LMS algorithms [67, 109], in the steady-state we have

$$
\mathrm{E}\left\{\mathrm{sgn}\left[\hat{\mathbf{h}}(n)\right]\right\} \approx \mathrm{sgn}(\hat{\mathbf{h}}). \tag{4.19}
$$

Thus, the $\mathrm{E}\left[\mathbf{T}(n)\right]$ in (4.18) can be written as

$$
\mathrm{E}\left[\mathbf{T}(n)\right] = \mathrm{E}\left\{\frac{\delta\mathrm{sgn}(\hat{\mathbf{h}}(n))}{\left(\left|\hat{\mathbf{h}}(n)\right| + \delta\right)^2}\right\} = \frac{\delta\mathrm{sgn}(\hat{\mathbf{h}}(n))}{\left(\left|\hat{\mathbf{h}}(n)\right| + \delta\right)^2}. \tag{4.20}
$$

In addition, when the channel length is far larger than 1, $N \gg 1$, the $\mathrm{E}\left[\mathbf{U}^{+}(n)\mathbf{U}(n)\right]$ can be written as [107, 110, 111]

$$
\begin{aligned}
&\mathrm{E}\left[\mathbf{U}^{+}(n)\mathbf{U}(n)\right] = \mathrm{E}\left\{\mathbf{U}^{T}(n)\left[\mathbf{U}(n)\mathbf{U}^{T}(n)\right]^{-1}\mathbf{U}(n)\right\} \\
&\approx \mathrm{E}\left\{\mathbf{U}^{T}(n)\left\{\mathrm{E}\left[\mathbf{U}(n)\mathbf{U}^{T}(n)\right]\right\}^{-1}\mathbf{U}(n)\right\}
\end{aligned}
. \tag{4.21}
$$

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

$$\begin{aligned}
&\mathrm{E}[\mathbf{U}(n)\mathbf{U}^T(n)]\\
&= \mathrm{E}\left\{\begin{bmatrix} \mathbf{x}^T(n) \\ \mathbf{x}^T(n-1) \\ \vdots \\ \mathbf{x}^T(n-Q-1) \end{bmatrix} \begin{bmatrix} \mathbf{x}(n) & \mathbf{x}(n-1) & \cdots & \mathbf{x}(n-Q-1) \end{bmatrix}\right\}\\
&= \mathrm{E}\left\{\begin{bmatrix} \|\mathbf{x}(n)\|^2 & \mathbf{x}^T(n)\mathbf{x}(n-1) & \cdots & \mathbf{x}^T(n)\mathbf{x}(n-Q-1) \\ \mathbf{x}^T(n-1)\mathbf{x}(n) & \|\mathbf{x}(n-1)\|^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}^T(n-Q-1)\mathbf{x}(n) & \mathbf{x}^T(n-Q-1)\mathbf{x}(n-1) & \cdots & \|\mathbf{x}(n-Q-1)\|^2 \end{bmatrix}\right\}\\
&= \mathrm{E}\left\{\begin{bmatrix} \|\mathbf{x}(n)\|^2 & 0 & \cdots & 0 \\ 0 & \|\mathbf{x}(n-1)\|^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\mathbf{x}(n-Q-1)\|^2 \end{bmatrix}\right\}.
\end{aligned} \tag{4.22}$$

Here, we define

$$\mathbf{R} = \delta_x^2 \mathbf{I}_N, \tag{4.23}$$

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

$$\mathrm{E}\left[\mathbf{U}(n)\mathbf{U}^T(n)\right] \approx \mathrm{Tr}(\mathbf{R})\mathbf{I}_Q = N\delta_x^2\mathbf{I}_Q. \tag{4.24}$$

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

$$\mathrm{E}\left\{\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\right\} \approx \left\{\mathrm{E}\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]\right\}^{-1} = \frac{1}{N\delta_x^2}\mathbf{I}_Q. \tag{4.25}$$

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

$$\mathrm{E}\left\{\mathbf{U}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{U}(n)\right\} \approx \mathrm{E}\left\{\mathbf{U}^T(n)\frac{1}{N\delta_x^2}\mathbf{I}_Q\mathbf{U}(n)\right\} \approx \frac{Q\mathbf{R}}{N\delta_x^2}. \quad (4.26)$$

Therefore, the equation (4.18) can be rewritten as

$$\begin{aligned}\mathrm{E}\left[\boldsymbol{\Delta}(n+1)\right] = \mathrm{E}&\left[\mathbf{I}_N - \mu_{SL0}\frac{Q\mathbf{R}}{N\delta_x^2}\right]\mathrm{E}\left[\boldsymbol{\Delta}(n)\right] \\ &-\frac{1}{2}\gamma_{SL0}\frac{Q\mathbf{R}}{N\delta_x^2}\mathrm{E}\left[\mathbf{T}(n)\right] + \frac{1}{2}\gamma_{SL0}\mathrm{E}\left[\mathbf{T}(n)\right]\end{aligned}. \quad (4.27)$$

It is found that the matrix $\mathbf{T}(n)$ is approximatively bounded between $-\delta\mathbf{I}_N$ and $\delta\mathbf{I}_N$. Therefore, we see that such convergence is guaranteed only if $(\mathbf{I}_N - \mu_{\mathrm{SL0}}Q\mathbf{R}/N\delta_x^2)$ is less than 1 [44], which is given by

$$0 < \mu_{\mathrm{SL0}} < \frac{N\delta_x^2}{Q\lambda_{\max}}, \quad (4.28)$$

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 $n \to \infty$. Then, (4.27) can be rewritten as

$$\mathrm{E}\left[\boldsymbol{\Delta}(\infty)\right] = \left[\mathbf{I}_N - \mu_{\mathrm{SL0}}\frac{Q\mathbf{R}}{N\delta_x^2}\right]\mathrm{E}\left[\boldsymbol{\Delta}(\infty)\right] - \frac{1}{2}\gamma_{\mathrm{SL0}}\frac{Q\mathbf{R}}{N\delta_x^2}\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2} + \frac{1}{2}\gamma_{\mathrm{SL0}}\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2}. \quad (4.29)$$

From (4.29), we can obtain

$$\mathrm{E}\left[\boldsymbol{\Delta}(\infty)\right] = -\frac{\gamma_{\mathrm{SL0}}}{2\mu_{\mathrm{SL0}}}\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2} + \frac{\gamma_{\mathrm{SL0}}}{2\mu_{\mathrm{SL0}}}\frac{N\delta_x^2}{Q\mathbf{R}}\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2}, \quad (4.30)$$

which can be regarded as

$$\mathrm{E}\left[\hat{\mathbf{h}}(\infty)\right] = \mathbf{h} - \frac{\gamma_{\mathrm{SL0}}}{2\mu_{\mathrm{SL0}}} \frac{\delta \mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2} + \frac{\gamma_{\mathrm{SL0}}}{2\mu_{\mathrm{SL0}}} \frac{N\delta_x^2}{Q\mathbf{R}} \frac{\delta \mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2}. \tag{4.31}$$

Note that the equation (4.31) implies that the optimum solution of the SL0-APA is biased, as was also shown for zero-attracting least mean square (ZA-LMS) algorithms [67]. We then proceed to derive the steady-state MSE for our proposed SL0-APA. Firstly, multiplying both sides of (4.10) by $\mathbf{U}(n)$ from the left, we can get

$$\begin{array}{l} \mathbf{U}(n)\hat{\mathbf{h}}(n+1) = \mathbf{U}(n)\hat{\mathbf{h}}(n) + \mu_{\mathrm{SL0}}\mathbf{U}(n)\mathbf{U}^{+}(n)\mathbf{e}(n) \\ \qquad\qquad\qquad +\frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{U}(n)\mathbf{U}^{+}(n)\mathbf{U}(n)\mathbf{T}(n) - \frac{1}{2}\gamma_{\mathrm{SL0}}\mathbf{U}(n)\mathbf{T}(n) \end{array}. \tag{4.32}$$

Furthermore,

$$\mathbf{U}(n)\hat{\mathbf{h}}(n+1) = \mathbf{U}(n)\hat{\mathbf{h}}(n) + \mu_{\mathrm{SL0}}\mathbf{e}(n). \tag{4.33}$$

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

$$\begin{array}{l} \mathbf{e}_p(n) = \mathbf{U}(n)\mathbf{h} - \mathbf{U}(n)\hat{\mathbf{h}}(n+1) \\ \mathbf{e}_a(n) = \mathbf{U}(n)\mathbf{h} - \mathbf{U}(n)\hat{\mathbf{h}}(n) \end{array}. \tag{4.34}$$

Combining the equations (4.33) and (4.34), we have

$$\mathbf{e}_p(n) = \mathbf{e}_a(n) - \mu_{\mathrm{SL0}}\mathbf{e}(n). \tag{4.35}$$

In addition,

$$\begin{array}{l} \mathbf{e}(n) = \mathbf{r}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n) \\ = \mathbf{U}(n)\mathbf{h} + \mathbf{v}(n) - \mathbf{U}(n)\hat{\mathbf{h}}(n) \\ = \mathbf{e}_a(n) + \mathbf{v}(n) \end{array}. \tag{4.36}$$

By substituting (4.36) into (4.35), we have

$$\mathbf{e}_p(n) = (\mathbf{I} - \mu_{\text{SL0}})\mathbf{e}(n) - \mathbf{v}(n). \tag{4.37}$$

From the equation (4.35), we can also write the $\mathbf{e}(n)$ as follows:

$$\mathbf{e}(n) = \frac{1}{\mu_{\text{SL0}}}[\mathbf{e}_a(n) - \mathbf{e}_p(n)]. \tag{4.38}$$

Substituting (4.38) to (4.10), we have

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mathbf{U}^{+}(n)[\mathbf{e}_a(n) - \mathbf{e}_p(n)] + \frac{1}{2}\gamma_{\text{SL0}}\mathbf{U}^{+}(n)\mathbf{U}(n)\mathbf{T}(n) - \frac{1}{2}\gamma_{\text{SL0}}\mathbf{T}(n). \tag{4.39}$$

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 equation (4.39), using the steady-state condition $\mathrm{E}\left[\left\|\hat{\mathbf{h}}(n+1)\right\|^2\right] \approx \mathrm{E}\left[\left\|\hat{\mathbf{h}}(n)\right\|^2\right]$ when $n \to \infty$, 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

$$\begin{array}{l} \mathrm{E}\left[\mathbf{e}_p^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{e}_p(n)\right] = \mathrm{E}\left[\mathbf{e}_a^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{e}_a(n)\right] \\ -\frac{\gamma_{\text{SL0}}^2}{4}\mathrm{E}\left\{\mathbf{T}^T(n)\mathbf{U}^{+}(n)\mathbf{U}(n)\mathbf{T}(n)\right\} + \frac{\gamma_{\text{SL0}}^2}{4}\mathrm{E}[\mathbf{T}^T(n)\mathbf{T}(n)] \end{array}. \tag{4.40}$$

Substituting (4.37) into the left hand side (LHS) of (4.40), we have

$$\begin{array}{l} \text{LHS} = (1-\mu_{\text{SL0}})^2\mathrm{E}\left\{\mathbf{e}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{e}(n)\right\} \\ -(1-\mu_{\text{SL0}})\mathrm{E}\left\{\mathbf{e}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{v}(n)\right\} \\ -(1-\mu_{\text{SL0}})\mathrm{E}\left\{\mathbf{v}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{e}(n)\right\} \\ +\mathrm{E}\left\{\mathbf{v}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{v}(n)\right\} \end{array}. \tag{4.41}$$

Moreover, substituting (4.36) into the right hand side (RHS) of (4.40), we have

$$\begin{aligned}
&\mathrm{RHS} = \mathrm{E}\left\{\mathbf{e}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{e}(n)\right\} \\
&-\mathrm{E}\left\{\mathbf{e}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{v}(n)\right\} \\
&-\mathrm{E}\left\{\mathbf{v}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{e}(n)\right\} \\
&+\mathrm{E}\left\{\mathbf{v}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{v}(n)\right\} \\
&-\frac{\gamma_{\mathrm{SL0}}^2}{4}\mathrm{E}\left\{\mathbf{T}^T(n)\mathbf{U}^+(n)\mathbf{U}(n)\mathbf{T}(n)\right\} \\
&+\frac{\gamma_{\mathrm{SL0}}^2}{4}\mathrm{E}[\mathbf{T}^T(n)\mathbf{T}(n)]
\end{aligned} . \tag{4.42}$$

By combining the equation (4.41) and (4.42), we get

$$\begin{aligned}
&(2\mu_{\mathrm{SL0}} - \mu_{\mathrm{SL0}}^2)\mathrm{E}\left\{\mathbf{e}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{e}(n)\right\} \\
&= \mu_{\mathrm{SL0}}\mathrm{E}\left\{\mathbf{e}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{v}(n)\right\} \\
&+\mu_{\mathrm{SL0}}\mathrm{E}\left\{\mathbf{v}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{e}(n)\right\} \\
&+\frac{\gamma_{\mathrm{SL0}}^2}{4}\mathrm{E}\left\{\mathbf{T}^T(n)\mathbf{U}^+(n)\mathbf{U}(n)\mathbf{T}(n)\right\} \\
&-\frac{\gamma_{\mathrm{SL0}}^2}{4}\mathrm{E}[\mathbf{T}^T(n)\mathbf{T}(n)]
\end{aligned} . \tag{4.43}$$

We also assume that the additive Gaussian noise $\mathbf{v}(n)$ is statistically independent of the input signal $\mathbf{x}(n)$. Thus the equation (4.43) can be simplified as

$$\begin{aligned}
&\mathrm{E}\left\{\mathbf{e}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{e}(n)\right\} \\
&= \frac{1}{2-\mu_{\mathrm{SL0}}}\mathrm{E}\left\{\mathbf{v}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{v}(n)\right\} \\
&+\frac{\gamma_{\mathrm{SL0}}^2}{4(2\mu_{\mathrm{SL0}} - \mu_{\mathrm{SL0}}^2)}\mathrm{E}\left\{\mathbf{T}^T(n)\mathbf{U}^+(n)\mathbf{U}(n)\mathbf{T}(n)\right\} \\
&-\frac{\gamma_{\mathrm{SL0}}^2}{4(2\mu_{\mathrm{SL0}} - \mu_{\mathrm{SL0}}^2)}\mathrm{E}[\mathbf{T}^T(n)\mathbf{T}(n)]
\end{aligned} . \tag{4.44}$$

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 $\mathrm{E}\left[\mathbf{e}^T(n)\mathbf{e}(n)\right] = \mathbf{E}\left|e_t(n)\right|^2\mathbf{S}$ [106], where

$$\mathbf{S} \approx \begin{cases} \mathbf{I}, \mu_{\mathrm{SL0}} \ \text{is small} \\ \mathbf{1} \cdot \mathbf{1}^T, \mu_{\mathrm{SL0}} \ \text{is large} \end{cases}, \tag{4.45}$$

where $\mathbf{1}^T = [\ 1 \quad 0 \quad \cdots \quad 0\ ]$ and $e_t(n)$ is the top entry of $\mathbf{e}(n)$ [106]. Then, the LHS of the equation of (4.44) can be rewritten as

$$\begin{aligned} &\mathrm{E}\left\{\mathbf{e}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{e}(n)\right\} \\ &\approx \mathrm{Tr}\left\{\mathrm{E}\left[\mathbf{e}^T(n)\mathbf{e}(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\right]\right\} \\ &\approx \mathrm{E}|e_t(n)|^2\mathrm{Tr}\left\{\mathbf{S}\cdot\mathrm{E}\left[\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\right]\right\} \end{aligned}. \tag{4.46}$$

Similar to the calculation of the equation (4.46), the first term in the RHS of the equation (4.44) can be written as

$$\begin{aligned} &\mathrm{E}\left\{\mathbf{v}^T(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\mathbf{v}(n)\right\} \\ &= \mathrm{Tr}\left\{\mathrm{E}\left[\mathbf{v}^T(n)\mathbf{v}(n)\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\right]\right\} \\ &= Q\delta_v^2\mathrm{Tr}\left\{\mathrm{E}\left[\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\right]\right\} \end{aligned}. \tag{4.47}$$

In addition, the second term of RHS of the equation (4.44) can be rewritten as

$$\begin{aligned} &\frac{\gamma_{\mathrm{SL0}}^2}{4(2\mu_{\mathrm{SL0}}-\mu_{\mathrm{SL0}}^2)}\mathrm{E}\left\{\mathbf{T}^T(n)\mathbf{U}^+(n)\mathbf{U}(n)\mathbf{T}(n)\right\} \\ &\approx \frac{\gamma_{\mathrm{SL0}}^2}{4(2\mu_{\mathrm{SL0}}-\mu_{\mathrm{SL0}}^2)}\frac{\delta\mathrm{sgn}^T(\mathbf{h})}{(\mathbf{h}^T+\delta)^2}\mathrm{E}\left\{\mathbf{U}^+(n)\mathbf{U}(n)\right\}\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2} \end{aligned}. \tag{4.48}$$

Then the last term of the right hand side of the equation (4.44) can be expressed as

$$\frac{\gamma_{\mathrm{SL0}}^2}{4(2\mu_{\mathrm{SL0}}-\mu_{\mathrm{SL0}}^2)}\mathrm{E}[\mathbf{T}(n)^T\mathbf{T}(n)] = \frac{\gamma_{\mathrm{SL0}}^2}{4(2\mu_{\mathrm{SL0}}-\mu_{\mathrm{SL0}}^2)}\frac{\delta\mathrm{sgn}^T(\mathbf{h})}{(\mathbf{h}^T+\delta)^2}\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2}. \tag{4.49}$$

When the μ_{SL0} is small, we can get

$$\begin{aligned}&\mathrm{Tr}\left\{\mathbf{S}\cdot\mathrm{E}\left[\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\right]\right\}\\&=\mathrm{Tr}\left\{I\cdot\mathrm{E}\left[\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\right]\right\}\\&=\frac{Q}{N\delta_x^2}\end{aligned}. \tag{4.50}$$

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

$$\begin{aligned}&\mathrm{MSE}_{small}=\frac{1}{2-\mu_{\text{SL0}}}\delta_v^2\\&+\frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}}-1\mu_{\text{SL0}}^2)}\frac{\delta\mathrm{sgn}^T(\mathbf{h})}{(\mathbf{h}^T+\delta)^2}\frac{N\delta_x^2}{Q}\frac{Q\mathbf{R}}{N\delta_x^2}\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2}\\&-\frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}}-1\mu_{\text{SL0}}^2)}\frac{\delta\mathrm{sgn}^T(\mathbf{h})}{(\mathbf{h}^T+\delta)^2}\frac{N\delta_x^2}{Q}\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2}\\&=\frac{1}{2-\mu_{\text{SL0}}}\delta_v^2+\frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}}-\mu_{\text{SL0}}^2)}\frac{\delta\mathrm{sgn}^T(\mathbf{h})}{(\mathbf{h}^T+\delta)^2}(\mathbf{R}-\frac{N\delta_x^2}{Q}\mathbf{I})\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2}\end{aligned}. \tag{4.51}$$

When the step-size μ_{SL0} is large, $\mathbf{S}\approx\mathbf{1}\cdot\mathbf{1}^T$ [106]. In this case,

$$\mathrm{Tr}\left\{\mathbf{S}\cdot\mathrm{E}\left[\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]^{-1}\right]\right\}=\frac{1}{N\delta_x^2}. \tag{4.52}$$

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

$$\begin{aligned}&\mathrm{MSE}_{large}=\frac{1}{2-\mu_{\text{SL0}}}\delta_v^2 Q\\&+\frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}}-\mu_{\text{SL0}}^2)}\frac{\delta\mathrm{sgn}^T(\mathbf{h})}{(\mathbf{h}^T+\delta)^2}N\delta_x^2\frac{Q\mathbf{R}}{N\delta_x^2}\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2}\\&-\frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}}-\mu_{\text{SL0}}^2)}\frac{\delta\mathrm{sgn}^T(\mathbf{h})}{(\mathbf{h}^T(n)+\delta)^2}N\delta_x^2\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2}\\&=\frac{1}{2-\mu_{\text{SL0}}}\delta_v^2 Q+\frac{\gamma_{\text{SL0}}^2}{4(2\mu_{\text{SL0}}-\mu_{\text{SL0}}^2)}\frac{\delta\mathrm{sgn}^T(\mathbf{h})}{(\mathbf{h}^T+\delta)^2}(Q\mathbf{R}-N\delta_x^2\mathbf{I})\frac{\delta\mathrm{sgn}(\mathbf{h})}{(\mathbf{h}+\delta)^2}\end{aligned}. \tag{4.53}$$

4.4 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 4.3. 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.

4.4.1 Performance of the proposed SL0-APA

Firstly, we set up a simulation example to discussion the convergence speed of 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 [2, 67, 98, 104]. 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 $K = 4$ is shown in Fig. 3.2. 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:

$$\mathrm{MSE}(n) = 10 \log_{10} \mathrm{E}\left\{ \left\| \mathbf{h} - \hat{\mathbf{h}}(n) \right\|_2^2 \right\} (\mathrm{dB}). \tag{4.54}$$

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: $\mu_{\text{NLMS}} = 0.25$, $\mu_{\text{APA}} = 0.125$, $\mu_{\text{ZA-APA}} = 0.165$, $\mu_{\text{RZA-APA}} = 0.18$, $\mu_{\text{SL0}} = 0.21$, $\gamma_{\text{ZA-APA}} = 5 \times 10^{-5}$, $\gamma_{\text{RZA-APA}} = 8 \times 10^{-5}$, $\gamma_{\text{SL0}} = 3 \times 10^{-6}$, $\varepsilon_{\text{RZA-APA}} = 10$, $\delta_{\text{SL0}} = 0.001$, $Q = 2$, $\delta_v^2 = 10^{-3}$, where μ_{NLMS} is the step-size parameter for NLMS algorithm. It can be seen from Fig. 4.1 that our proposed SL0-APA possesses the fastest convergence speed compared to the previously proposed channel estimation algorithms used in this experiment 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 Fig. 4.2. It reveals that the convergence speed is improved

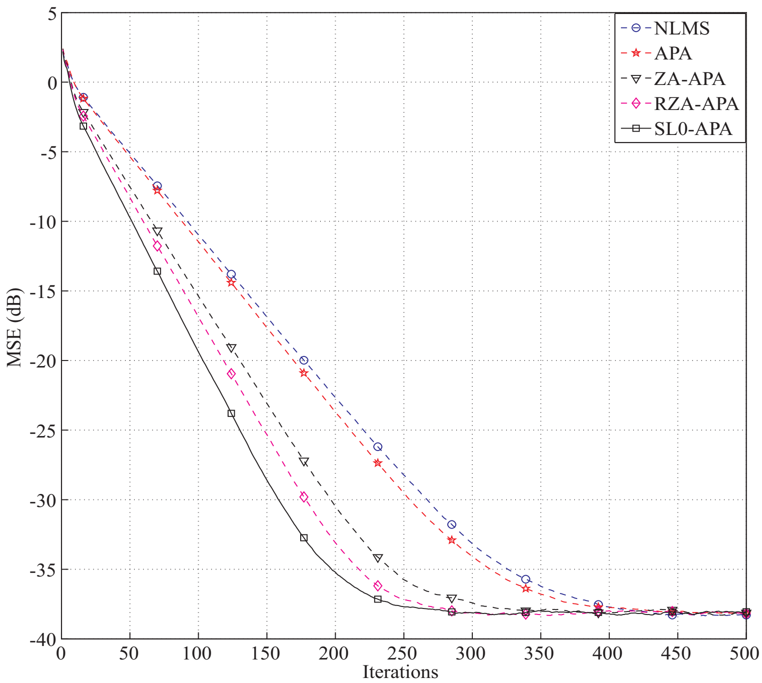


Figure 4.1: Convergence of the proposed SL0-APA.

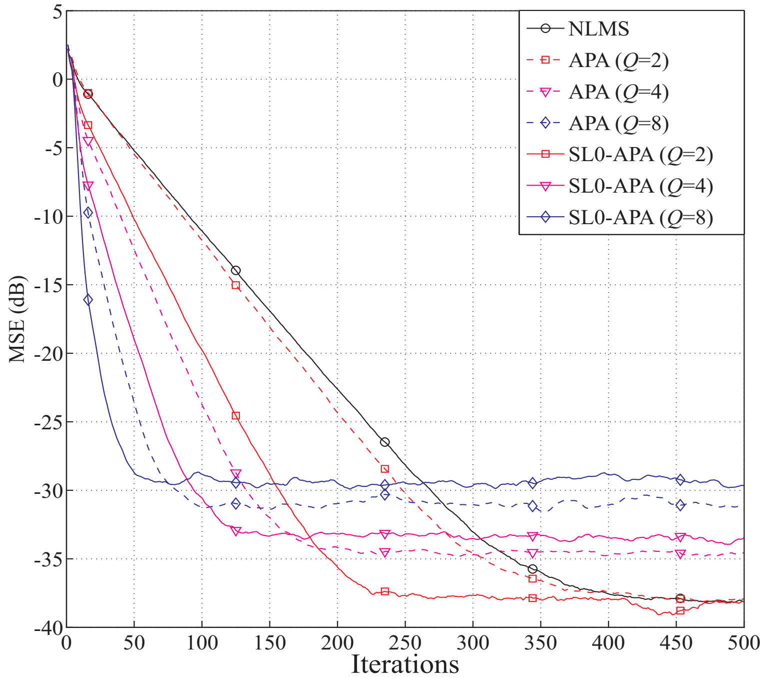
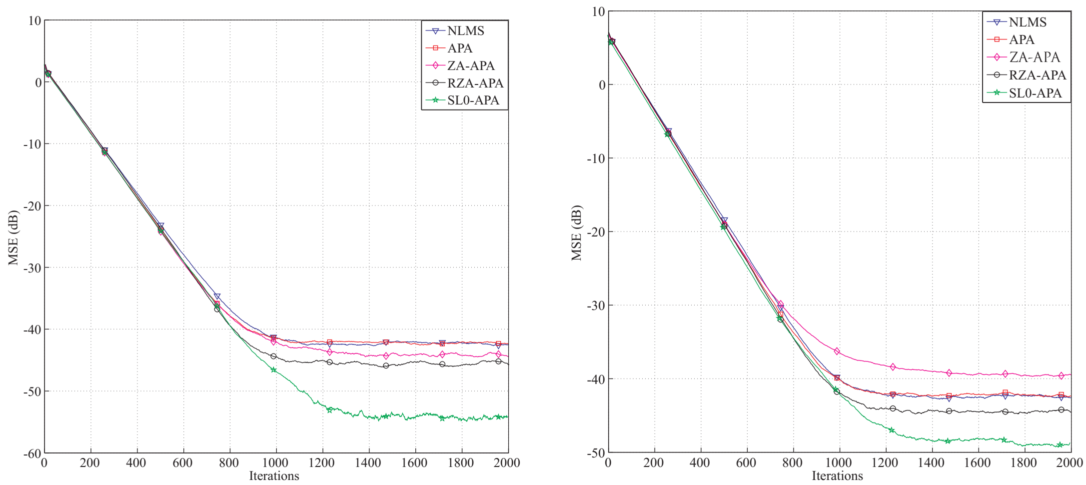


Figure 4.2: Affine projection order effects on the SL0-APA.

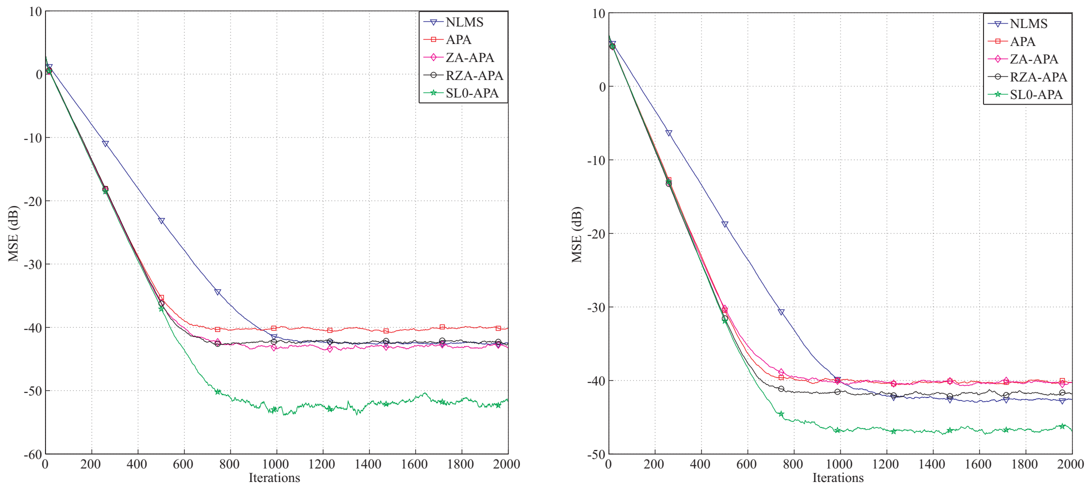
by the increment of the affine projection order Q. However, the steady-state performance has deteriorated from $Q = 2$ to $Q = 8$. Thus, in our proposed SL0-APA, the affine projection Q, the step-size μ_{SL0}, the regularization parameter γ_{SL0} and δ_{SL0} should be take into account to balance the 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: $\mu_{\mathrm{NLMS}} = 0.095$, $\mu_{\mathrm{APA}} = \mu_{\mathrm{ZA-APA}} = \mu_{\mathrm{RZA-APA}} = \mu_{\mathrm{SL0}} = 0.05$, $\gamma_{\mathrm{ZA-APA}} = 5\times10^{-5}$, $\gamma_{\mathrm{RZA-APA}} = 8 \times 10^{-5}$, $\gamma_{\mathrm{SL0}} = 4 \times 10^{-6}$, $\varepsilon_{\mathrm{RZA-APA}} = 10$, $\delta = 0.01$, $\delta_v^2 = 10^{-3}$. We can see from the Fig. 4.3 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 Fig. 4.3a. When the sparsity level K increases to 4, it is seen in Fig. 4.3b that our proposed SL0-APA still outperforms other algorithms, while its steady-state error increases in comparison with that of $K = 1$. When



(a) $K = 1$ (b) $K = 4$

Figure 4.3: Performance of the SL0-APA with different sparsity levels for $Q = 2$.

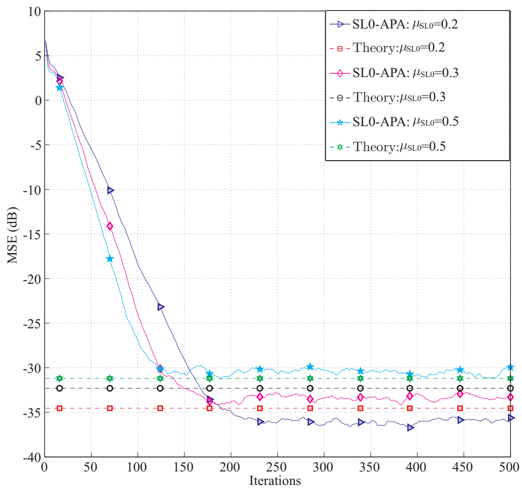


(a) $K = 1$ (b) $K = 4$

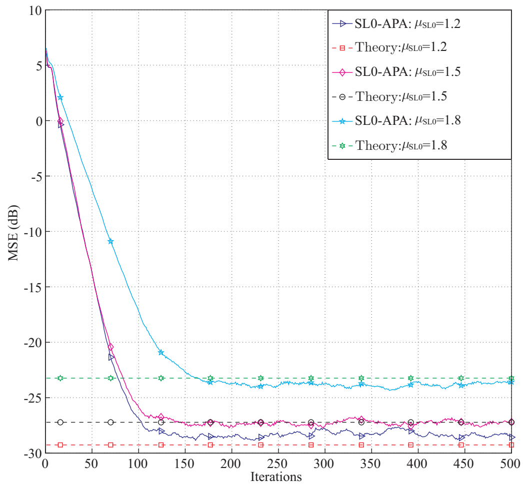
Figure 4.4: Performance of the SL0-APA with different sparsity levels for $Q = 3$.

the affine projection order increases to $Q = 3$, we can see from Fig. 4.4 that the convergence speed is significantly improved compared to that of $Q = 2$ shown in Fig. 4.3. 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 4.3 to predict



(a) small μ_{SL0}



(b) large μ_{SL0}

Figure 4.5: Steady-state MSE performance of the SL0-APA with different step-size μ_{SL0} for $K = 4$.

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 Fig. 4.5. The theoretical results are obtained from (4.51) and (4.53) 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 Fig. 4.5a, 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.

4.4.2 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 non-zero 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 4.1.

Table 4.1: 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 $+K-N+\mathrm{O}(Q^3)$	NQ^2+Q^2+NQ $+Q+\mathrm{O}(Q^3)$	Q
ZA-APA	NQ^2+Q^2+3NQ $-2Q-2N$ $+3K+\mathrm{O}(Q^3)$	NQ^2+2Q^2+3NQ $+2Q+K+\mathrm{O}(Q^3)$	Q
RZA-APA	NQ^2+Q^2+3NQ $-2Q-2N$ $+4K+\mathrm{O}(Q^3)$	NQ^2+2Q^2+3NQ $+Q+N+2K+\mathrm{O}(Q^3)$	$N+Q$
SL0-APA	NQ^2+Q^2+3NQ $-2Q-2N$ $+4K+\mathrm{O}(Q^3)$	NQ^2+2Q^2+3NQ $+Q+N^2-N$ $+2K+\mathrm{O}(Q^3)$	$N+Q$

According to Table 4.1, 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 are 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.

4.5 Conclusion

In this chapter, we proposed an SL0-APA to exploit the sparsity of sparse channel and to improve the performance on both the convergence speed and steady-state error of the APA, ZA-APA, 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 derived 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.

Chapter 5

Discrete Weighted Zero-Attracting Affine Projection Algorithm

5.1 Introduction

In this chapter, we propose a discrete weighted zero-attracting affine projection algorithm (DWZA-APA) with low complexity for broadband multipath channel estimation [112]. The proposed algorithm exploits the sparsity of the broadband multipath channel and utilizes the relationship between the inactive channel taps and discrete weighted coefficients to improve the convergence speed of the affine projection algorithm (APA). In the proposed DWZA-APA, the discrete weighted coefficients are mainly exerted on the inactive taps whose magnitudes are zero or close to zero. Compared to the previously proposed sparsity-aware algorithms, the proposed DWZA-APA can not only further improve the estimation performance but also reduce computational complexity with stable recovery compared with the RZA-APA. Simulation results demonstrate that the DWZA-APA outperforms the standard APA and its related sparsity-aware algorithm in terms of both the convergence speed and steady-state performance.

5.2 Proposed DWZA-APA

On the basis of the update function (2.50) of the RZA-APA, we propose a DWZA-APA which uses a piece-wise approximation instead of the $1/(1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|)$ to make use of the sparseness of the multipath channel [112]. The following piece-wise linear function, which is a segment function, is adopted to approximate the $1/(1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|)$ and hence it can be regarded as a fairly resemblance of the zero attractor term in (2.50)

$$f(\hat{h}_i) = \begin{cases} 350\left|\hat{h}_i\right|, \left|\hat{h}_i\right| < 0.005 \\ \delta_w, \text{elsewhere} \end{cases} \tag{5.1}$$

for $i = 1, 2, \cdots, N$. Where δ_w is a positive constant to adjust the weights effects on the non-zero taps. By taking the stochastic characteristic of the channel into account and substituting (5.1) into (2.50), we can get the update equation of the proposed DWZA-APA and it is given by

$$\begin{aligned} \hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) &+ \mu_{\mathrm{DWZA}}\mathbf{U}^{+}(n)\mathbf{e}(n) \\ &+ \frac{1}{2}\gamma_{\mathrm{DWZA}}\mathbf{U}^{+}(n)\mathbf{U}(n)f(\hat{\mathbf{h}}(n))\mathrm{sgn}[\hat{\mathbf{h}}(n)] \\ &- \frac{1}{2}\gamma_{\mathrm{DWZA}}f(\hat{\mathbf{h}}(n))\mathrm{sgn}[\hat{\mathbf{h}}(n)] \end{aligned}, \tag{5.2}$$

where μ_{DWZA} is the step size of the DWZA-APA and γ_{DWZA} is the regularization parameter. In contrast to the RZA-APA, the DWZA-APA eliminates the additional logic caused by the division from $1/(1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|)$ of the RZA-APA. In addition, $1/(1+\varepsilon_{\mathrm{RZA-APA}}\left|\hat{\mathbf{h}}(n)\right|)$ needs N multiplications and N divisions, the piece-wise linear function (5.1) of our proposed DWZA-APA only needs less than N multiplications and logic comparisons. Thus, our proposed discrete weighted method by the use of a piece-wise linear function reduces the computational complexity of the previous RZA-APA. On the other hand, our proposed DWZA-APA mainly exert penalty on the zero taps or close to zero taps, which can speeds up the convergence and reduces the steady-state error.

5.3 Results and discussions

In this section, the estimation performance of our proposed DWZA-APA are presented for channel estimations over a sparse channel with its length of N is equal to 16 and number of dominant taps is K. 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 Gaussian random signal while the output of the channel is corrupted by an independent white Gaussian noise $\mathbf{v}(n)$ with its power $\delta_v = 10^{-3}$. Moreover, the results are averaged over 100 independent trails for each algorithm. For the sake of comparison, the performance is evaluated by means of the mean-square-error (MSE) defined in (4.54).

Firstly, we set up an experiment to investigate the effects of the parameter δ_w and the simulation result is illustrated in Fig. 5.1. The specification parameters used in this experiment are $\mu_{\text{DWZA}} = 0.4$ and $\gamma_{\text{DWZA}} = 3 \times 10^{-4}$. It is found that the steady-state performance of the DWZA-APA is improved with the increment of the δ_w when $\delta_w < 1.5$. When δ_w increases to 2.5, the steady-state error of DWZA-APA is the same to these of $\delta_w = 2$ and $\delta_w = 1.5$. This is because the DWZA-APA has nearly the same weights to all the channel taps. Therefore, in this paper, $\delta_w = 2$ is adopted to verify the performance of the DWZA-APA.

Next, we discuss the convergence speed and the steady-state performance of the DWZA-APA and compere it with the previous sparse-aware algorithms, namely, ZA-LMS, RZA-LMS, ZA-NLMS, RZA-NLMS, ZA-APA, RZA-APA and the standard LMS, NLMS and APA. To compare the convergence speed of these algorithms, the specification parameters are optimized and shown in Table 5.1

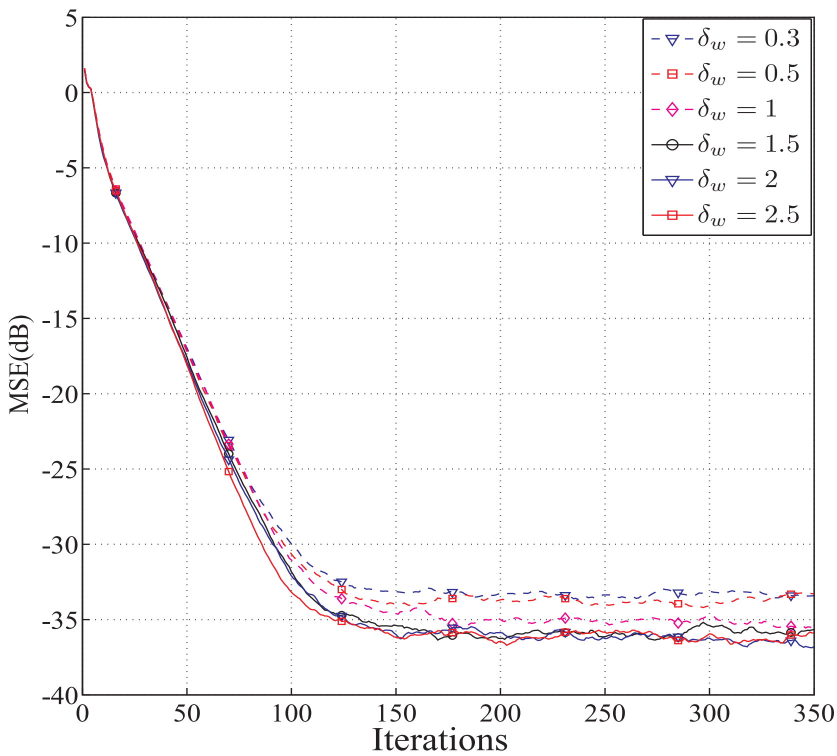


Figure 5.1: Effects on the parameter δ_w on the proposed DWZA-APA with $K = 1$.

for obtaining the same steady-state error floor. The obtained simulation result is demonstrated in Fig. 5.2. We can see from the Fig. 5.2 that the convergence

Table 5.1: Simulation parameters

Algorithms	Step-size	Regularization parameters
LMS	$\mu_{\text{LMS}} = 0.028$	-
ZA-LMS	$\mu_{\text{ZA-LMS}} = 0.031$	$\rho_{\text{ZA-LMS}} = 2 \times 10^{-6}$
RZA-LMS	$\mu_{\text{RZA-LMS}} = 0.034$	$\rho_{\text{RZA-LMS}} = 5 \times 10^{-5}$
NLMS	$\mu_{\text{NLMS}} = 0.42$	-
ZA-NLMS	$\mu_{\text{ZA-NLMS}} = 0.45$	$\rho_{\text{ZA-NLMS}} = 6 \times 10^{-6}$
RZA-NLMS	$\mu_{\text{RZA-NLMS}} = 0.5$	$\rho_{\text{RZA-NLMS}} = 5 \times 10^{-5}$
APA	$\mu_{\text{APA}} = 0.21$	-
ZA-APA	$\mu_{\text{ZA-APA}} = 0.3$	$\gamma_{\text{ZA-APA}} = 7 \times 10^{-5}$
RZA-APA	$\mu_{\text{RZA-APA}} = 0.4$	$\gamma_{\text{RZA-APA}} = 3 \times 10^{-4}$
DWZA-APA	$\mu_{\text{DWZA}} = 0.48$	$\gamma_{\text{DWZA}} = 3 \times 10^{-4}$

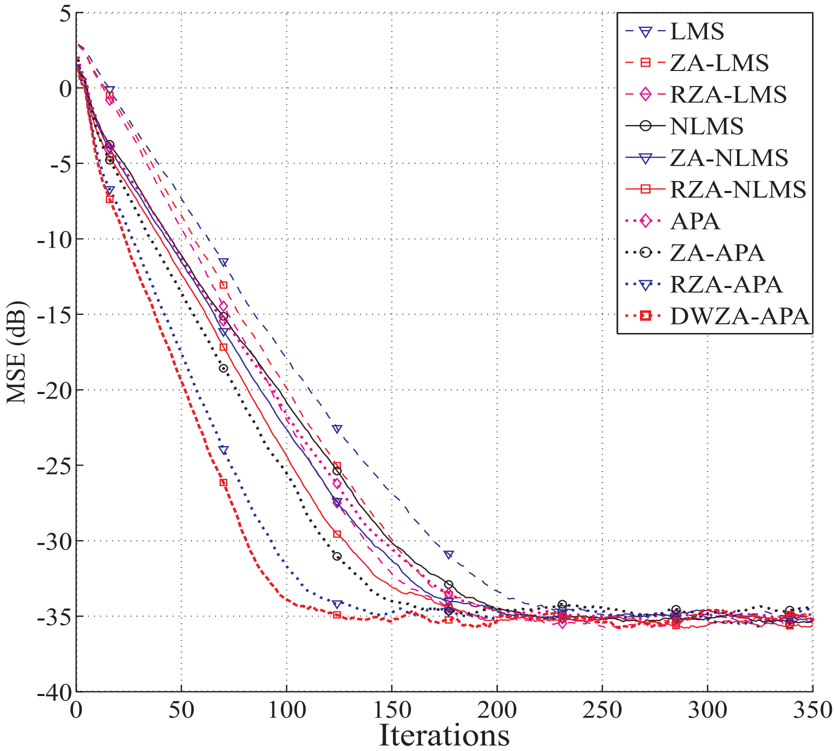


Figure 5.2: Convergence of the proposed DWZA-APA with $K = 1$ and $\delta_w = 2$.

speed of the DWZA-APA is much faster than the RZA-APA. This is because the RZA-APA mainly takes effects on the taps whose magnitudes are comparable to $1/\varepsilon_{\mathrm{RZA-APA}}$ while has less shrinkage exerted on $\left|\hat{\mathbf{h}}(n)\right| > 1/\varepsilon_{\mathrm{RZA-APA}}$. For the DWZA-APA, it exerts the penalty mainly on the taps whose magnitudes are less than 0.05 while it has weak zero attracting on the other taps whose magnitudes are greater than 0.05. Thus, the DWZA-APA has strong zero attracting on the inactive taps, which speeds up the convergence speed of these inactive taps. In addition, the convergence speed of the DWZA-APA and RZA-APA is superior to other sparsity-aware algorithms mentioned above.

Then, we exploit the steady-state performance of the DWZA-APA and compare it with the previously proposed channel estimation algorithms, including RZA-NLMS, ZA-APA, RZA-APA and the standard APA. In this experiment, the detailed parameters are listed in Table 5.2 to obtain the same convergence speed. The computer simulation results are shown in Fig. 5.3. Figure 5.3 shows that our proposed DWZA-APA is better than other algorithms in terms of the steady-state error due to that the DWZA-APA mainly penalizes the discrete weights on the zero taps. We can see that the MSE of DWZA-APA is about 1 dB better than that of the RZA-APA.

Finally, we show the performance of the DWZA-APA over a cluster-sparse multipath channel with the channel length of $N = 256$ [113, 114]. The sparsity

Table 5.2: Simulation parameters

Algorithms	Step-size	Regularization parameters
RZA-NLMS	$\mu_{\mathrm{RZA-NLMS}} = 0.4$	$\rho_{\mathrm{RZA-NLMS}} = 3 \times 10^{-5}$
APA	$\mu_{\mathrm{APA}} = 0.22$	-
ZA-APA	$\mu_{\mathrm{ZA-APA}} = 0.22$	$\gamma_{\mathrm{ZA-APA}} = 5 \times 10^{-5}$
RZA-APA	$\mu_{\mathrm{RZA-APA}} = 0.22$	$\gamma_{\mathrm{RZA-APA}} = 1 \times 10^{-4}$
DWZA-APA	$\mu_{\mathrm{DWZA}} = 0.22$	$\gamma_{\mathrm{DWZA}} 1 \times 10^{-4}$

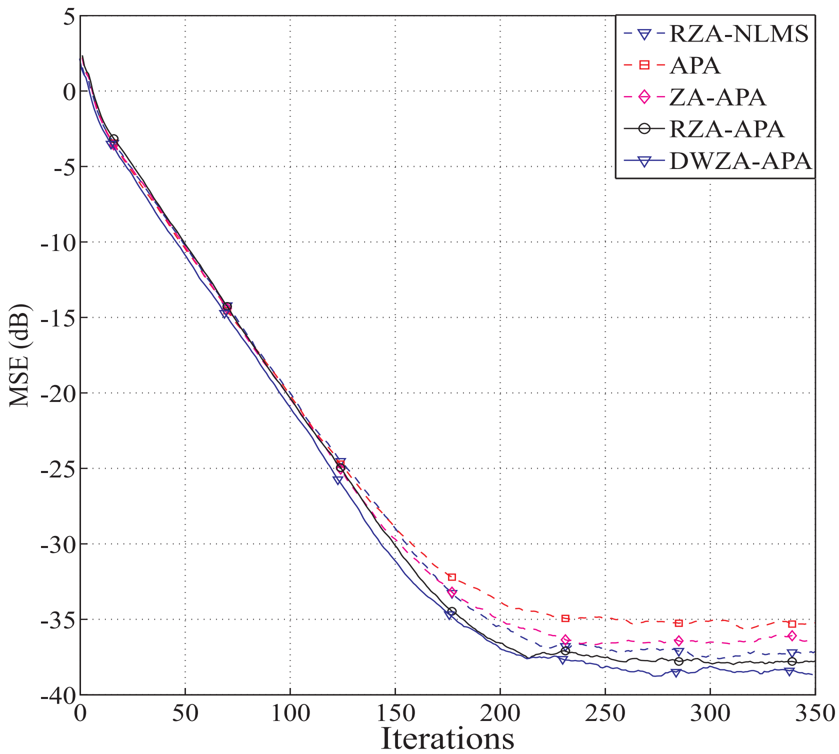


Figure 5.3: Steady-state performance of the proposed DWZA-APA with $K = 1$.

level is $K = 8$ and these 8 taps are always separated into two clusters with 4 taps for each cluster. An typical example of such cluster-sparse multipath channel [113] is shown in Fig. 5.4, where the length of the channel is 100 and the number of dominant taps is 8 which is divided into two clusters. Furthermore, the positions of the two clusters are randomly spaced along the length of the channel. In this experiment, the relevant parameters are listed as follows: $\mu_{\mathrm{RZA-NLMS}} = 0.4$, $\mu_{\mathrm{APA}} = \mu_{\mathrm{ZA-APA}} = \mu_{\mathrm{RZA-APA}} = \mu_{\mathrm{DWZA}} = 0.22$, $\rho_{\mathrm{RZA-NLMS}} = 6 \times 10^{-6}$, $\gamma_{\mathrm{APA}} = 1 \times 10^{-5}$, $\gamma_{\mathrm{ZA-APA}} = 1 \times 10^{-5}$, $\gamma_{\mathrm{RZA-APA}} = 7.5 \times 10^{-5}$, $\gamma_{\mathrm{DWZA}} = 3 \times 10^{-6}$. Figure 5.5 demonstrates the performance of the DWZA-APA over such cluster-sparse channel. We can see that the DWZA-APA is still better than other channel estimation algorithms with respect to the steady-state performance. Additionally, we found that all the algorithms have the same convergence at the initial stage. After several iterations, the RZA-APA converges faster than the DWZA-APA. After that, the convergence speed of the RZA-APA slows down, even it is slower

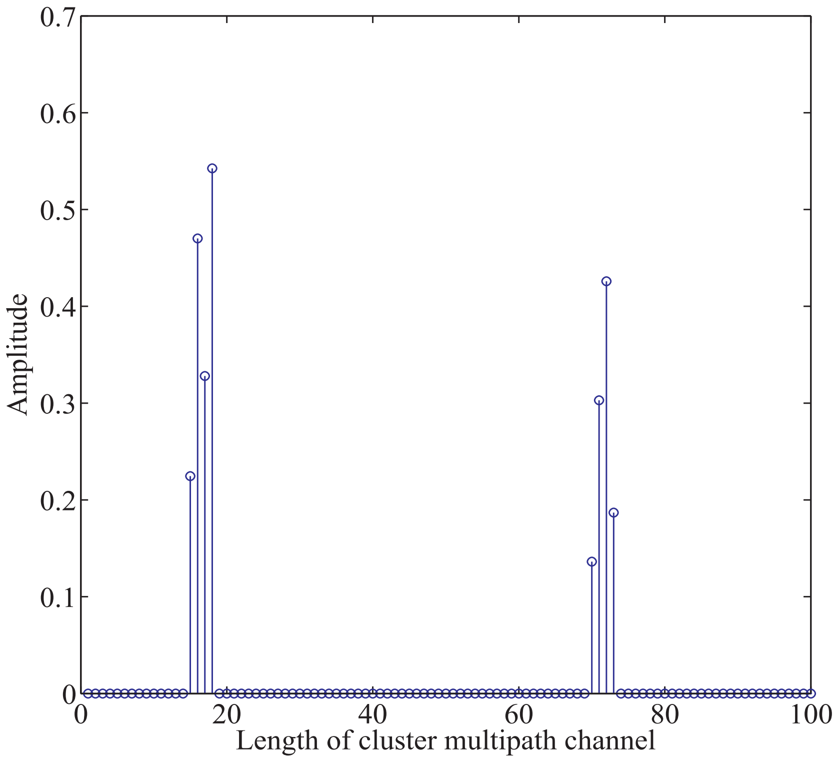


Figure 5.4: Cluster-sparse multipath fading channel model.

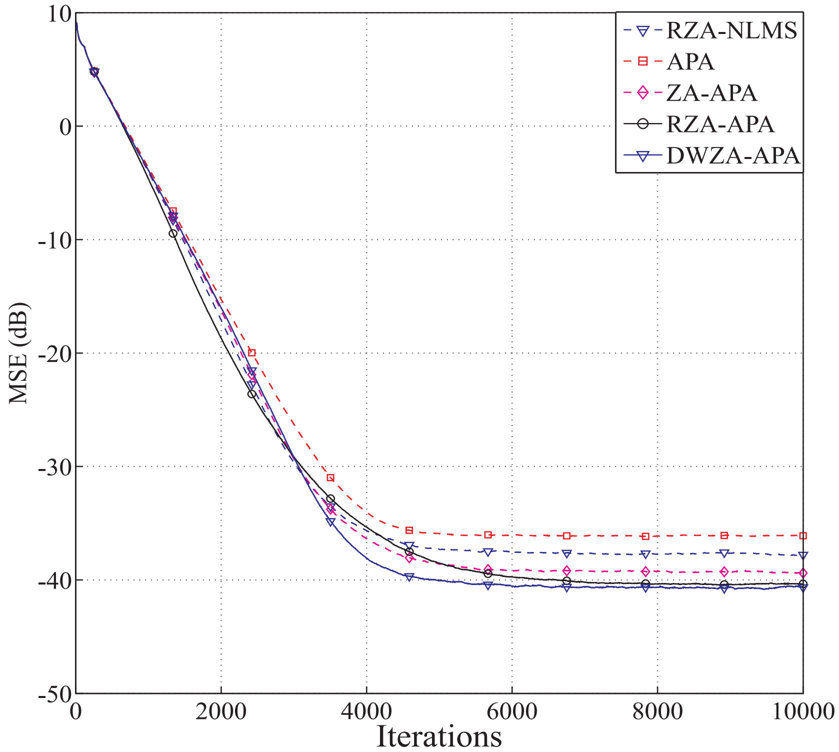


Figure 5.5: Performance of the DWZA-APA over a cluster-sparse channel.

than that of the ZA-APA. However, the DWZA-APA has better steady-state performance compared with RZA-APA at steady stage. This is caused by the introduction of piece-wise linear function in (5.1), which mainly exerts penalty on the inactive channel taps at the initialization, and hence the convergence speed of the zero taps are accelerated. Consequently, the DWZA-APA gives the same penalty on the active taps uniformly and thus its convergence speed is reduced.

5.4 Conclusion

In this chapter, we proposed a DWZA-APA for sparse channel estimation in broadband communication systems based on the approximate representation of the sum-log function in the RZA-APA. The behaviors of the DWZA-APA are evaluated over a sparse and a cluster-sparse channel. The simulation results show that the DWZA-APA has a faster convergence speed and a lower steady-state error than that of the RZA-APA. The introduction of discrete weighted scheme in the DWZA-APA, which is used to design the zero attractors, also reduces the computational complexity of the RZA-APA by eliminating the division and additional operations.

Chapter 6

L_p-Norm Constrained Proportionate Normalized Least-Mean-Square Algorithm

6.1 Introduction

To make use of the sparsity property of broadband multipath wireless communication channels, we have proposed an l_p-norm-constrained proportionate normalized least-mean-square (LP-PNLMS) sparse channel estimation algorithm [115]. A general l_p-norm is weighted by the gain matrix and is incorporated into the cost function of the proportionate normalized least-mean-square (PNLMS) algorithm. This integration is equivalent to adding a zero attractor to the iterations, by which the convergence speed and steady-state performance of the inactive taps are significantly improved. Our simulation results demonstrate that the proposed algorithm can effectively improve the estimation performance of the

PNLMS-based algorithm for sparse channel estimation applications.

6.2 Proposed LP-PNLMS algorithm

In this section, we propose an LP-PNLMS algorithm by incorporating the l_p-norm into the cost function of the PNLMS algorithm to create a zero attractor, making it a type of ZA algorithm. The difference between the LP-PNLMS algorithm and general ZA algorithms is that the gain-matrix-weighted l_p-norm is taken into account in designing the zero attractor [115]. On the other hand, the proposed LP-PNLMS algorithm is based on the commonly used PNLMS algorithm, which is also a sparse channel estimation algorithm and can improve the convergence for the active taps. Regarding channel estimation, the purpose of the LP-PNLMS algorithm is to minimize

$$\begin{array}{l} (\hat{\mathbf{h}}(n+1)-\hat{\mathbf{h}}(n))^T\mathbf{G}^{-1}(n)(\hat{\mathbf{h}}(n+1)-\hat{\mathbf{h}}(n))+\gamma_{\text{LP}}\left\|\mathbf{G}^{-1}(n)\hat{\mathbf{h}}(n+1)\right\|_p \\ \text{subject to} \\ r(n)-\hat{\mathbf{h}}^T(n+1)\mathbf{x}(n)=0 \end{array}, \quad (6.1)$$

where $\mathbf{G}^{-1}(n)$ is the inverse of the gain matrix $\mathbf{G}(n)$ in the PNLMS algorithm, $\gamma_{\text{LP}} > 0$ is a very small constant used to balance the estimation error and the sparse l_p-norm penalty of $\hat{\mathbf{h}}(n+1)$, $\|\cdot\|_p$ is the p-norm defined as $\left\|\hat{\mathbf{h}}\right\|_p = (\sum_i \hat{h}_i^p)^{1/p}$ in (1.7) and $0 \le p \le 1$. Note that in (6.1), we introduce an l_p-norm penalty to $\hat{\mathbf{h}}(n+1)$ after scaling the gain matrix by $\mathbf{G}^{-1}(n)$, which is different from the previously proposed ZA LMS algorithms.

To minimize (6.1), the Lagrange multiplier method is adopted, and the cost function $J_{\text{LP}}(n+1)$ of the proposed LP-PNLMS algorithm is expressed as

$$\begin{array}{r} J_{\text{LP}}(n+1)=(\hat{\mathbf{h}}(n+1)-\hat{\mathbf{h}}(n))^T\mathbf{G}^{-1}(n)(\hat{\mathbf{h}}(n+1)-\hat{\mathbf{h}}(n)) \\ +\gamma_{\text{LP}}\left\|\mathbf{G}^{-1}(n)\hat{\mathbf{h}}(n+1)\right\|_p+\lambda(r(n)-\hat{\mathbf{h}}^T(n+1)\mathbf{x}(n)) \end{array}, \quad (6.2)$$

where λ is the Lagrange multiplier.

By calculating the gradient of the cost function $J_{\mathrm{LP}}(n+1)$ of the LP-PNLMS algorithm and assuming $\hat{\mathbf{h}}(n+1) \approx \hat{\mathbf{h}}(n)$ in the steady stage, we have

$$\frac{\partial J_{\mathrm{LP}}(n+1)}{\partial \hat{\mathbf{h}}(n+1)} = \mathbf{0} \text{ and } \frac{\partial J_{\mathrm{LP}}(n+1)}{\partial \lambda} = 0 \tag{6.3}$$

and

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \lambda \mathbf{G}(n)\mathbf{x}(n) - \gamma_{\mathrm{LP}} \frac{\left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\left|\hat{\mathbf{h}}(n)\right|^{1-p}}. \tag{6.4}$$

In practice, a small positive constant is necessary for the final term in (6.4) to cope with the situation that an entry of $\hat{\mathbf{h}}(n)$ approaches zero, which is the case for a sparse CIR at the initialization. Then the update equation (6.4) of the LP-PNLMS algorithm is modified to

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \lambda \mathbf{G}(n)\mathbf{x}(n) - \gamma_{\mathrm{LP}} \frac{\left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\left|\hat{\mathbf{h}}(n)\right|^{1-p} + \varepsilon_p}, \tag{6.5}$$

where ε_p is a small value to prevent division by zero. By multiplying both sides of (6.5) by $\mathbf{x}^T(n)$, we obtain

$$\begin{aligned} \mathbf{x}^T(n)\hat{\mathbf{h}}(n+1) = {} & \mathbf{x}^T(n)\hat{\mathbf{h}}(n) + \lambda \mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) \\ & - \gamma_{\mathrm{LP}} \frac{\mathbf{x}^T(n)\left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\left|\hat{\mathbf{h}}(n)\right|^{1-p} + \varepsilon_p}. \end{aligned} \tag{6.6}$$

From (2.2), (6.3) and (6.5), we obtain

$$e(n) = -\gamma_{\mathrm{LP}} \frac{\mathbf{x}^T(n) \left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\left|\hat{\mathbf{h}}(n)\right|^{1-p} + \varepsilon_p} + \lambda \mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n). \tag{6.7}$$

Then, the Lagrange multiplier λ is given as follows by solving (6.7):

$$\lambda = \frac{e(n) + \gamma_{\mathrm{LP}} \dfrac{\mathbf{x}^T(n) \left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\left|\hat{\mathbf{h}}(n)\right|^{1-p} + \varepsilon_p}}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)}. \tag{6.8}$$

Substituting (6.8) into (6.5), we have

$$\begin{aligned}
\hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) - \gamma_{\mathrm{LP}} \frac{\left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\left|\hat{\mathbf{h}}(n)\right|^{1-p} + \varepsilon_p} \\
&\quad + \frac{e(n) + \gamma_{\mathrm{LP}} \dfrac{\mathbf{x}^T(n) \left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\left|\hat{\mathbf{h}}(n)\right|^{1-p} + \varepsilon_p}}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)} \mathbf{G}(n)\mathbf{x}(n) \\
&= \hat{\mathbf{h}}(n) + \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)} - \gamma_{\mathrm{LP}} \left\{ \mathbf{I} - \frac{\mathbf{G}(n)\mathbf{x}(n)\mathbf{x}^T(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)} \right\} \frac{\left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\left|\hat{\mathbf{h}}(n)\right|^{1-p} + \varepsilon_p}
\end{aligned} \tag{6.9}$$

It is found that the magnitudes of the elements in the matrix $\mathbf{G}(n)\mathbf{x}(n)\mathbf{x}^T(n)\{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)\}^{-1}$ are much smaller than 1 for broadband multipath channel estimation. Therefore, the update equation (6.9) of the proposed

LP-PNLMS algorithm is rewritten as

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)} - \gamma_{\mathrm{LP}} \frac{\left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \mathrm{sgn}(\hat{\mathbf{h}}(n))}{\left|\hat{\mathbf{h}}(n)\right|^{1-p} + \varepsilon_p}. \tag{6.10}$$

Here, we neglect the effects of the matrix $\mathbf{G}(n)\mathbf{x}(n)\mathbf{x}^T(n)\{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)\}^{-1}$ and assume that the filter order is large. Similarly to the PNLMS algorithm, a step size μ_{LP} is introduced to balance the convergence speed and the steady-state error of the proposed LP-PNLMS algorithm, and a small positive constant $\varepsilon_{\mathrm{LP}} = \delta_x^2/N$ is employed to prevent division by zero. Thus, the update function (6.10) can be modified to

$$\begin{aligned}\hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) + \mu_{\mathrm{LP}} \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) + \varepsilon_{\mathrm{LP}}} - \rho_{\mathrm{LP}} \frac{\left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \mathrm{sgn}(\hat{\mathbf{h}}(n))}{\left|\hat{\mathbf{h}}(n)\right|^{1-p} + \varepsilon_p}, \\ &= \hat{\mathbf{h}}(n) + \mu_{\mathrm{LP}} \frac{e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) + \varepsilon_{\mathrm{LP}}} - \rho_{\mathrm{LP}}\mathbf{T}(n)\end{aligned} \tag{6.11}$$

where $\rho_{\mathrm{LP}} = \mu_{\mathrm{LP}}\gamma_{\mathrm{LP}}$ and $\mathbf{T}(n) = \left\|\hat{\mathbf{h}}(n)\right\|_p^{1-p} \mathrm{sgn}(\hat{\mathbf{h}}(n))\{\left|\hat{\mathbf{h}}(n)\right|^{1-p} + \varepsilon_p\}^{-1}$. Comparing update function (6.11) of the proposed LP-PNLMS algorithm with the the update function (2.51) of the PNLMS algorithm, we see that our proposed LP-PNLMS algorithm has the additional term $\gamma_{\mathrm{LP}}\mathbf{T}(n)$, also defined as the zero attractor, which attracts the small channel taps to zero with high probability. In a words, in our proposed LP-PNLMS algorithm, the gain matrix $\mathbf{G}(n)$ assigns a large step size to the active channel taps of the sparse channel, while the zero attractor mainly exerts the l_p-penalty on the inactive taps whose taps are zero or close to zero. Thus, our proposed LP-PNLMS algorithm can further improve the convergence speed of the PNLMS algorithm after the convergence of the large taps which are active taps.

6.3 Results and discussion

In this section, we present the results of computer simulations carried out to illustrate the estimation performance of the proposed LP-PNLMS algorithm over a sparse multipath communication channel and compare it with those of the preciously proposed IPNLMS, MPNLMS, PNLMS and NLMS algorithms. We consider a sparse channel $\mathbf{h}$ whose length N is 64 or 128 and whose number of dominant active taps K is set to three different sparsity levels, namely $K = 2, 4$ and 8 similarly to previous studies [6,22,25-26]. The dominant active channel taps are obtained from a Gaussian distribution with $\|\mathbf{h}\|_2^2 = 1$, and the positions of the dominant channel taps are randomly spaced along 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 $v(n)$. An example of a typical sparse multipath channel with a channel length of $N = 64$ and a sparsity level of $K = 3$ is shown in Fig. 1.2. In the simulations, the power of the received signal is $E_b = 1$, while the noise power is given by δ_v^2 and the signal-to-noise ratio is defined as $\mathrm{SNR} = 10 \log \frac{E_b}{\delta_v^2}$. In all the simulations, 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 in (4.54).

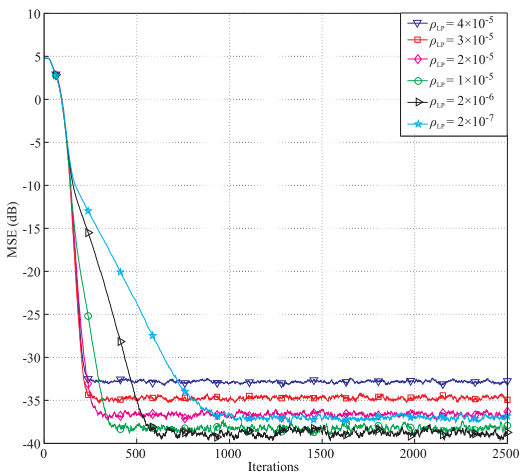
In these simulations, the simulation parameters are chosen to be $\mu_{\mathrm{NLMS}} = \mu_{\mathrm{PNLMS}} = \mu_{\mathrm{IPNLMS}} = \mu_{\mathrm{LP}} = 0.5$, $\delta_{\mathrm{NLMS}} = 0.01$, $\varepsilon = 0.001$, $\alpha = 0$, $\varepsilon_p = 0.05$, $\rho_{\mathrm{LP}} = 1 \times 10^{-5}$, $\delta_p = 0.01$, $\rho_g = 5/N$, $\vartheta = 1000$, $p = 0.5$, SNR $= 30$ dB. When we change one of these parameters, the other parameters remain constant.

6.3.1 Estimation performance of the proposed LP-PNLMS algorithm

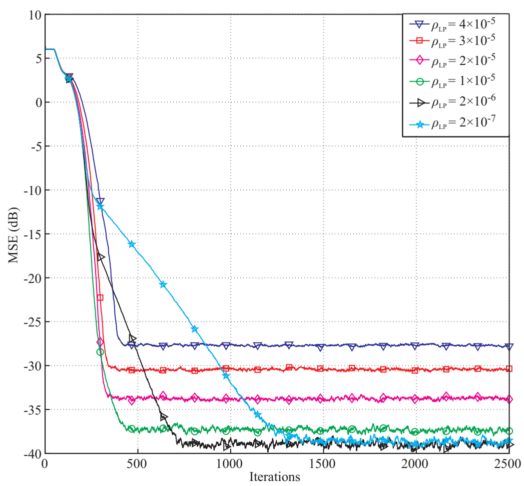
6.3.1.1 Effects of parameters on the proposed LP-PNLMS algorithm

In the proposed LP-PNLMS algorithm, there are two extra parameters, p and ρ_{LP}, compared with the PNLMS algorithm, which are introduced to design the zero attractor. Next, we show how these two parameters affect the proposed LP-PNLMS algorithm over a sparse channel with $N = 64$ or 128 and $K = 4$. The simulation results for different values of ρ_{LP} and p are shown in Figs. 6.1 and 6.2, respectively. According to the Fig. 6.1a, we can see that the steady-state error of the LP-PNLMS algorithm decreases with decreasing ρ_{LP} when $\rho_{\text{LP}} \geq 2 \times 10^{-6}$, while it increases again when ρ_{LP} is less than 2×10^{-6}. Furthermore, the convergence speed of the LP-PNLMS algorithm rapidly decreases when ρ_{LP} is less than 1×10^{-5}. This is because a small ρ_{LP} results in a low ZA strength, which consequently reduces the convergence speed. In the case of $N = 128$ shown in Fig. 6.1b, we observe that both the convergence speed and the steady-state performance are improved with decreasing ρ_{LP} for $\rho_{\text{LP}} \geq 1 \times 10^{-5}$. When $\rho_{\text{LP}} < 1 \times 10^{-5}$, the convergence speed of the LP-PNLMS algorithm decreases while the steady-state error remains constant.

Figure 6.2 demonstrates the effects of the parameter p. We can see from Fig. 6.2a that the convergence speed of the proposed LP-PNLMS algorithm rapidly decreases with increasing p for $N = 64$. Moreover, the steady-state error is reduced with p ranging from 0.45 to 0.5, while it remains constant for $p = 0.6$, 0.7 and 0.8. However, the steady-state performance for $p = 1$ is inferior to that for $p = 0.8$. This is because the proposed LP-PNLMS algorithm is an l_1-

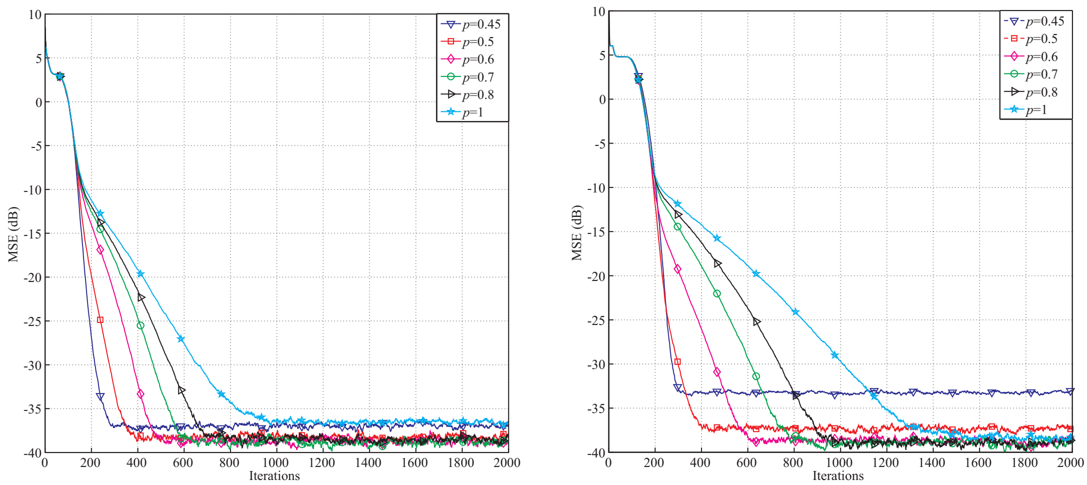


(a) $N = 64$



(b) $N = 128$

Figure 6.1: Effects of ρ_{LP} on the proposed LP-PNLMS algorithm.



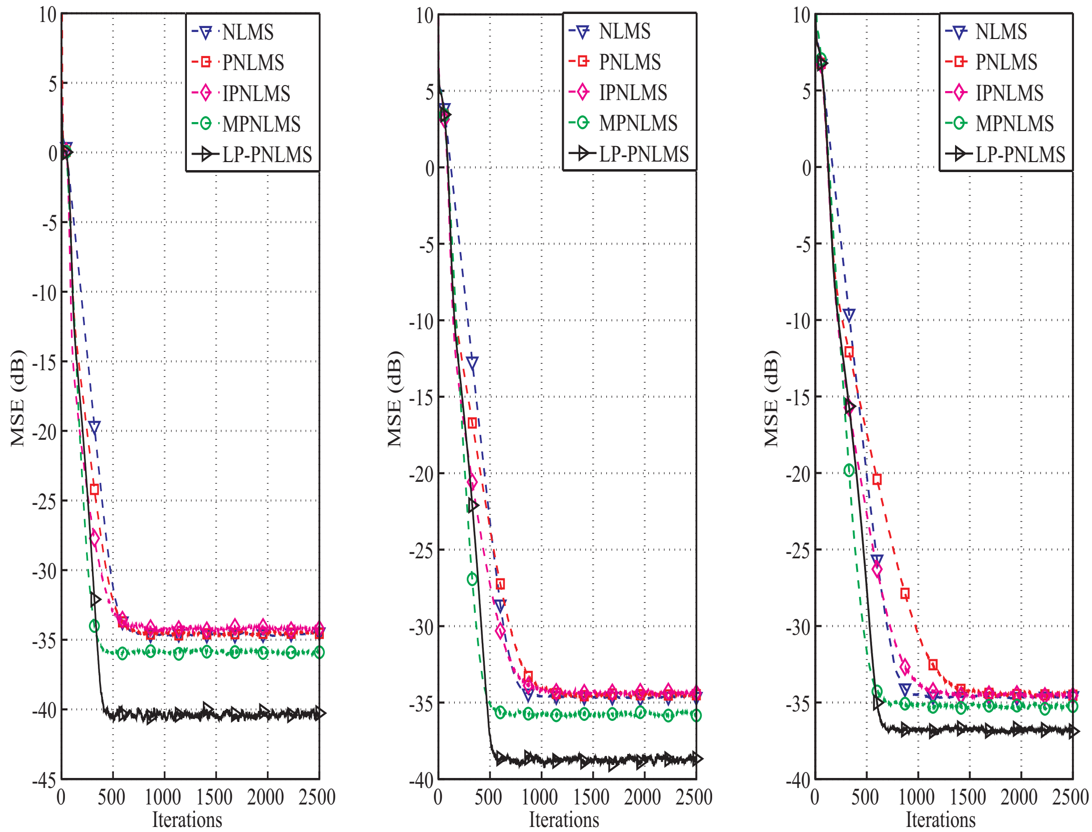
(a) $N = 64$ (b) $N = 128$

Figure 6.2: Effects of p on the proposed LP-PNLMS algorithm.

norm-penalized PNLMS algorithm, which cannot distinguish between active taps and inactive taps, reducing its convergence speed and steady-state performance. When $N = 128$, as shown in Fig. 6.2b, the steady-state performance is improved as p increases from 0.45 to 0.6. Thus, we should carefully select the parameters ρ_{LP} and p to balance the convergence speed and steady-state performance for the proposed LP-PNLMS algorithm.

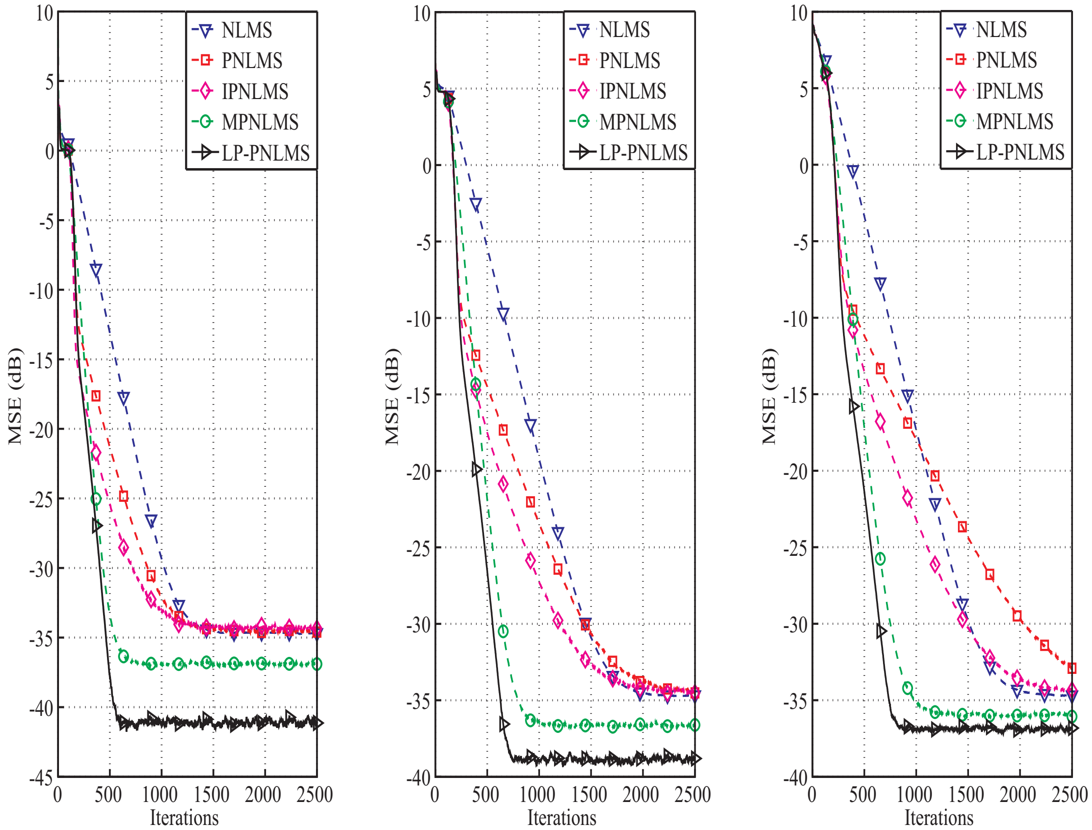
6.3.1.2 Effects of sparsity level on the proposed LP-PNLMS algorithm

On the basis of the results discussed discussed in Section 6.3.1.1 for our proposed LP-PNLMS algorithm, we choose $p = 0.5$ and $\rho_{\mathrm{LP}} = 1 \times 10^{-5}$ to evaluate the channel estimation performance of the LP-PNLMS algorithm over a sparse channel with different channel lengths of $N = 64$ and 128, for which the obtained simulation results are given in Figs. 6.3 and 6.4, respectively. From Fig. 6.3, we see that our proposed LP-PNLMS algorithm has the same convergence speed as the PNLMS algorithm at the initial stage. The proposed LP-PNLMS algorithm converges faster than the PNLMS algorithm as well as the IPNLMS and NLMS algorithms for all sparsity levels K, while its convergence is slightly slower than that of the MPNLMS algorithm before it reaches a steady stage. However, the proposed LP-PNLMS algorithm has the smallest steady-state error for $N = 64$. When $N = 128$, we see from Fig. 6.4 that our proposed LP-PNLMS algorithm not only has the highest convergence speed but also possesses the best steady-state performance. This is because with increasing sparsity, our proposed LP-PNLMS algorithm attracts the inactive taps to zero quickly and hence the convergence speed is significantly improved, while the previously proposed PNLMS algorithms mainly adjust the step size of the active taps and thus they only impact on the convergence speed at the early iteration stage. Additionally, we see from Figs. 6.3 and 6.4 that both the convergence speed and the steady-state performance of all the PNLMS algorithms deteriorate when the sparsity level K increases for both $N = 64$ and 128. In particular, when $K = 8$, the convergence speeds of the PNLMS and IPNLMS algorithms are greater than that of the NLMS algorithm



(a) K=2 (b) K=4 (c) K=8

Figure 6.3: Effects of sparsity on the proposed LP-PNLMS algorithm for $N = 64$.



(a) K=2 (b) K=4 (c) K=8

Figure 6.4: Effects of sparsity on the proposed LP-PNLMS algorithm for $N = 128$.

at the early iteration stage, while after this fast initial convergence, their convergence speeds decrease to less than that of the NLMS algorithm before reaching a steady stage. Furthermore, we observe that the MPNLMS algorithm is sensitive to the length N of the channel, and its convergence speed for $N = 128$ is less than that for $N = 64$ at the same sparsity level K and less than that of the proposed LP-PNLMS algorithm. Thus, we conclude that our proposed LP-PNLMS algorithm is superior to the previously proposed PNLMS algorithms in terms of both the convergence speed and the steady-state performance with the appropriate selection of the related parameters p and ρ_{LP}. From the above discussion, we believe that the gain-matrix-weighted l_p-norm method in the LP-PNLMS algorithm can be used to further improve the channel estimation performance of the IPNLMS and MPNLMS algorithms.

6.3.2 Computational complexity

Finally, we discuss the computational complexity of the proposed LP-PNLMS algorithm and compare it with those of the NLMS, PNLMS, IPNLMS and MPNLMS algorithms. The computational complexity is the arithmetic complexity, which includes additions, multiplications and divisions. The computational complexities of the proposed LP-PNLMS algorithm and the related PNLMS and NLMS algorithms are shown in Table 6.1.

Table 6.1: Computational complexity

Algorithms	Additions	Multiplications	Divisions
NLMS	$3N$	$3N+1$	1
PNLMS	$4N+3$	$6N+3$	$N+2$
IPNLMS	$4N+7$	$5N+5$	$N+2$
MPNLMS	$5N+3$	$7N+3$	$N+3$
LP-PNLMS	$4N+4$	$9N+4$	$2N+2$

According to Table 6.1, the computational complexity of our proposed LP-PNLMS algorithm is slightly higher than those of the MPNLMS and PNLMS algorithms, which is due to the calculation of the gradient of the l_p-norm. Furthermore, the MPNLMS algorithm has an additional logarithm operation, which increases its complexity but is not included in the Table 6.1. However, the LP-PNLMS algorithm noticeably increases the convergence speed and significantly improves the steady-state performance of the PNLMS algorithm. In addition, it also has a higher convergence speed and lower steady-state error than the IPNLMS and MPNLMS algorithms when the channel length is large.

6.4 Conclusion

In this chapter, we have proposed an LP-PNLMS algorithm to exploit the sparsity of broadband multipath channels and to improve both the convergence speed and steady-state performance of the PNLMS algorithm. This algorithm was mainly developed by incorporating the gain-matrix-weighted l_p-norm into the cost function of the PNLMS algorithm, which significantly improves its convergence speed and steady-state performance. The simulation results demonstrated that our proposed LP-PNLMS algorithm, which has an acceptable increase in computational complexity, increases the convergence speed and reduces the steady-state error compared with the previously proposed PNLMS algorithms.

Chapter 7

Conclusions and Future Research Directions

7.1 Conclusions of this dissertation

To fulfill the requirements of a high data rate and a high-quality service, broadband communication techniques have attracted much attention in both industry and academia for their use in future communication systems. The broadband wireless multipath channel, which can be regarded as a sparse channel, has important effects on coherent detection. Effective sparse channel estimation techniques which can exploit the sparsity of the broadband wireless multipath channel and provide sufficient gain for wireless communication systems, are attractive for applications. Furthermore, adaptive channel estimation using adaptive filtering algorithms is effective and easy to implement and has been widely studied. In this dissertation, we have developed four sparse adaptive channel estimation algorithms by utilizing the sparseness of the broadband multipath channel.

For sparse channel estimation, the conventional LMS algorithm cannot utilize the sparseness of the channel and performs poorly. On the basis of CS theory, variable step size techniques and ZA techniques [67], we have proposed an ARZA-SVSS-LMS algorithm using the sigmoid functioned variable step size and adaptive parameter adjustment methods. Simulation results showed that the ARZA-SVSS-LMS algorithm can achieve better channel estimation performance than the previously proposed ZA-LMS and RZA-LMS algorithms. In this algorithm, the SVSS technique reduces the steady-state error while the adaptive parameter adjustment method dynamically changes the ZA strength of the ARZA-SVSS-LMS algorithm, thus significantly increases the convergence speed.

Chapter 4 proposed the SL0-APA to improve both the convergence speed and steady-state performance compared with the ZA-APAs and NLMS algorithms. The proposed SL0-APA can provide a zero attractor by integrating the SL0 into the cost function of the APA, as a result of which both the convergence speed and steady-state performance are improved. Furthermore, we provided a convergence analysis of the SL0-APA. Simulation results were in good agreement with ones obtained from the theoretical analysis for different values of μ_{SL0}. The computational complexity of the SL0-APA was also discussed and compared with previously proposed channel estimation algorithms. The SL0-APA has acceptable computational complexity, a high convergence speed and a low steady-state error.

Chapter 5 focused on reducing the computational complexity of the RZA-APA. For this purpose, we presented the DWZA-APA algorithm to reduce the computational complexity, which was implemented by introducing a piece-wise linear segment function instead of the sum-logarithm function in the RZA-APA. The DWZA-APA achieves a higher convergence speed and nearly the same steady-state error compared with the RZA-APA. In addition, additions and divisions are removed from the RZA-APA, thus reducing its computational complexity.

Chapter 6 presented a novel perspective for proportionate adaptive algorithms. To the best of our knowledge, most of the existing proportionate adaptive algorithms exploit the sparsity of a sparse signal by assigning a proportionate step size to individual coefficients. Although these modified PNLMS algorithms can increase the convergence speed and improve the steady-state performance by employing a suitable variable step size or gain matrix, they are mainly focused on how to determine an appropriate step size and gain matrix $\mathbf{G}(n)$. Inspired by the ZA algorithms, we proposed the LP-PNLMS algorithm by incorporating a gain-matrix-weighted l_p-norm into the cost function of the PNLMS algorithm to obtain the benefits of both the PNLMS algorithm and the ZA algorithms. The update equation of the LP-PNLMS algorithm shows that the zero attractor is independent of $\mathbf{G}(n)$. Thus, the LP-PNLMS algorithm assigns a large step size to large coefficients and applies a zero attractor to the inactive taps. The simulation results demonstrated that the LP-PNLMS algorithm can achieve the same convergence speed as the PNLMS algorithm in the initial stage and that it converges faster than the PNLMS algorithm when the active taps have converged. The main advantage of this approach is that only the zero attractor must be designed, which can be easily achieved in practical engineering applications.

7.2 Future Research Directions

From the investigations and discussions of this dissertation, we can conclude that the proposed sparse adaptive filter algorithms demonstrate favorable performance for sparse channel estimation applications. In the future, the following issues can be further investigated.

7.2.1 For the ARZA-SVSS-LMS algorithm

In this dissertation, we proposed the ARZA-SVSS-LMS algorithm, which had greater computational complexity than the ZA-LMS and RZA-LMS algorithms. Although some low-complexity LMS algorithms have been proposed such as partial update LMS algorithms, they are not considered to be ZA algorithms. In [116], a segment ZA LMS algorithm was proposed and discussed. This algorithm reduces the complexity of the RZA-LMS algorithm for echo cancellation applications. Thus, the development of low-complexity ZA LMS algorithms with high channel estimation performance in terms of both the convergence speed and the steady-state performance is necessary and desirable. In addition, the complexity of the ARZA-SVSS-LMS algorithm can be reduced by using a suitable segment linear function. Furthermore, the proposed μ_{SVSS} can also be improved by the use of a nonparametric or optimal step size method. In addition, the proposed ZA LMS algorithm can be expanded to leaky-LMS and block-LMS algorithms.

7.2.2 For the sparse APA algorithms

We proposed the SL0-APA to improve the convergence speed and steady-state performance of the previously proposed ZA algorithms. This algorithm is realized by incorporating the SL0 into the cost function of the APA. In the future, the l_p-norm can also be used to design the zero attractor for sparse channel estimation on the basis of the investigation of the ZA algorithms. Also, the APA has higher complexity than the LMS algorithm, which is due to the use of the matrix inverse of $\mathbf{U}(n)$ and the calculation of the gradient for SL0. Although a fast APA (FAPA) [117] has been proposed and applied to related APAs, the complexity of the matrix inverse can also be reduced. The calculation of the gradient for SL0 can be reduced by selecting a suitable continuous function for smooth approximation of the l_0-norm. Furthermore, the calculation of the gradient for SL0 can be obtained from a Taylor series, such as the SL0 in [104], which can reduce the computational complexity of obtaining the gradient. Finally, the proposed method and the previously proposed ZA algorithms can be expanded to a set-membership affine projection sign algorithm to render them suitable for sparse channel estimation applications. In addition, we also proposed the DWZA-APA to reduce the complexity of the RZA-APA. However, in the future we hope to design a ZA block APA for sparse-cluster signal applications.

7.2.3 For the sparse proportionate-type algorithms

Proportionate adaptive algorithms have favorable performance and have been used for echo cancellation in telecommunication networks. We proposed the LP-PNLMS algorithm to improve the convergence speed of the inactive taps and the steady-state performance of the PNLMS algorithm. The computational complexity can be reduced by using partial-updating approaches [68–71], which update part of the coefficients. Moreover, the proposed gain-matrix-weighted l_p-norm can be expanded to μ-law PNLMS (MPNLMS) and IPNLMS algorithms to further improve the channel estimation performance of the PNLMS algorithms. In addition, the ZA technique can also be integrated into proportionate affine projection algorithms (PAPA) such as the improved PAPA (IPAPA) and μ-law PAPA

(MPAPA). In the LP-PNLMS algorithm, a zero attractor is added to the update function of the PNLMS algorithm, which is independent of the step size μ_{LP} and gain matrix $\mathbf{G}(n)$. Thus, the variable step size techniques and gain allocation methods can be employed to further improve the estimation performance of the LP-PNLMS algorithm. On the other hand, analysis of the convergence characteristics of the ZA sparse proportionate-type channel estimation algorithms should be considered in future research.

7.2.4 Sparse adaptive filter applications

The proposed sparse adaptive filters were developed while focusing on broadband multipath channel estimation applications. The proposed sparse channel estimation algorithms can be integrated into the practical wireless systems, such as the OFDM and MIMO systems, to investigate the bit-error-ratio (BER) performance. Moreover, another area for investigation is the application of the proposed algorithms in other scenarios different from channel estimation, such as adaptive beamforming, echo cancellation and adaptive networks. On the other hand, the development of 2D sparse adaptive filter algorithms based on the ZA techniques for 2D sparse signal applications such as image processing should be considered in the future. In addition, the sparse adaptive filter algorithms can be applied in compressed sensing for sparse signal processing with high recovery probability.

Appendix A

Proof for Theorem 1

Proof: We define the estimation error $\mathbf{\Delta}(n+1)$ at the $(n+1)$th iteration

$$\begin{aligned} &\mathbf{\Delta}(n+1) = \hat{\mathbf{h}}(n+1) - \mathbf{h} \\ &= \mathbf{\Delta}(n) + \mu_{\text{LMS}}\mathbf{x}(n)e(n) \\ &= \mathbf{\Delta}(n) - \frac{1}{2}\mu_{\text{LMS}}\mathbf{\Gamma}(n) \end{aligned} , \tag{A.1}$$

where

$$\mathbf{\Gamma}(n) = \frac{\partial J(n)}{\partial \hat{\mathbf{h}}(n)} = -2\mathbf{x}(n)e(n) , \tag{A.2}$$

where $J(n) = e^2(n)$ and $\mathbf{\Gamma}(n)$ is defined as the joint gradient error function that includes the channel estimator error and the additive white Gaussian noise error [2, 43–45]. Two gradient errors should be separated in order to obtain the lower bound of the channel estimator. Hence, we split $\mathbf{\Gamma}(n)$ into two terms:

$$\mathbf{\Gamma}(n) = \tilde{\mathbf{\Gamma}}(n) + 2\mathbf{w}(n), \tag{A.3}$$

where $\tilde{\mathbf{\Gamma}}(n)$ is the gradient error and can be expressed as

$$\tilde{\mathbf{\Gamma}}(n) = 2(\mathbf{R}\hat{\mathbf{h}}(n) - \mathbf{R}\mathbf{h}), \tag{A.4}$$

and $\mathbf{w}(n) = [w_0(n), w_1(n) \cdots, w_{N-1}(n)]^T$ is the gradient noise error with its expectation $\mathrm{E}\{\mathbf{w}(n)\} = 0$ [2, 43], and hence

$$\begin{aligned} \mathbf{x}(n)e(n) &= -\frac{1}{2}\boldsymbol{\Gamma}(n) \\ &= -(\mathbf{R}\tilde{\mathbf{h}}(n) - \mathbf{R}\mathbf{h}) - \mathbf{w}(n) \\ &= -\mathbf{R}(\tilde{\mathbf{h}}(n) - \mathbf{h}) - \mathbf{w}(n) \\ &= -\mathbf{R}\boldsymbol{\Delta}(n) - \mathbf{w}(n) \end{aligned} . \tag{A.5}$$

Substituting equation (A.5) into the equation (A.1), we can get

$$\begin{aligned} \boldsymbol{\Delta}(n+1) &= \boldsymbol{\Delta}(n) + \mu_{\mathrm{LMS}}\mathbf{x}(n)e(n) \\ &= \boldsymbol{\Delta}(n) - \mu_{\mathrm{LMS}}\mathbf{R}\boldsymbol{\Delta}(n) - \mu_{\mathrm{LMS}}\mathbf{w}(n) \\ &= (\mathbf{I}_N - \mu_{\mathrm{LMS}}\mathbf{R})\boldsymbol{\Delta}(n) - \mu_{\mathrm{LMS}}\mathbf{w}(n) \\ &= (\mathbf{I}_N - \mu_{\mathrm{LMS}}\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T)\boldsymbol{\Delta}(n) - \mu_{\mathrm{LMS}}\mathbf{w}(n) \\ &= \mathbf{Q}(\mathbf{I}_N - \mu_{\mathrm{LMS}}\boldsymbol{\Lambda})\mathbf{Q}^T\boldsymbol{\Delta}(n) - \mu_{\mathrm{LMS}}\mathbf{w}(n) \end{aligned} , \tag{A.6}$$

where the covariance matrix $\mathbf{R}$ of the input signal $\mathbf{x}(n)$ can be decomposed as $\mathbf{R} = \mathbf{Q}\Lambda\mathbf{Q}^T$. Here, $\mathbf{Q}$ is a $N \times N$ unitary matrix and $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_{N-1}, \lambda_N\}$ is an $N \times N$ diagonal matrix with the eigenvalue of covariance matrix $\mathbf{R}$ located at its diagonal. $\mathbf{I}_N$ is a $N \times N$ identity matrix. By multiplying the $\mathbf{Q}^T$ at the both side of the equation (A.6), we can rewrite it as

$$\mathbf{Q}^T\boldsymbol{\Delta}(n+1) = (\mathbf{I}_N - \mu_{\mathrm{LMS}}\boldsymbol{\Lambda})\mathbf{Q}^T\boldsymbol{\Delta}(n) - \mu_{\mathrm{LMS}}\mathbf{Q}^T\mathbf{w}(n). \tag{A.7}$$

Assuming $\bar{\boldsymbol{\Delta}}(n+1) = \mathbf{Q}^T\boldsymbol{\Delta}(n+1)$, $\bar{\boldsymbol{\Delta}}(n) = \mathbf{Q}^T\boldsymbol{\Delta}(n)$ and $\bar{\mathbf{w}}(n) = \mathbf{Q}^T\mathbf{w}(n)$, the equation (A.7) can be rewritten as

$$\bar{\boldsymbol{\Delta}}(n+1) = (\mathbf{I}_N - \mu_{\mathrm{LMS}}\boldsymbol{\Lambda})\bar{\boldsymbol{\Delta}}(n) - \mu_{\mathrm{LMS}}\bar{\mathbf{w}}(n). \tag{A.8}$$

Thus, the MSE lower bound of LMS can be obtained as

$$\begin{aligned}
B &= \lim_{n\to\infty} \mathrm{E}\left\{\left\|\bar{\mathbf{\Delta}}(n+1)\right\|_2^2\right\} \\
&= \lim_{n\to\infty} \mathrm{E}\left\{\left[(\mathbf{I}_N-\mu_{\text{LMS}}\mathbf{\Lambda})\bar{\mathbf{\Delta}}(n)-\mu_{\text{LMS}}\bar{\mathbf{w}}^T(n)\right]^T\left[(\mathbf{I}_N-\mu_{\text{LMS}}\mathbf{\Lambda})\bar{\mathbf{\Delta}}(n)-\mu_{\text{LMS}}\bar{\mathbf{w}}(n)\right]\right\} \\
&= \lim_{n\to\infty}\left\{(\mathbf{I}_N-\mu_{\text{LMS}}\mathbf{\Lambda})^2\mathrm{E}\left[\left\|\bar{\mathbf{\Delta}}(n)\right\|_2^2\right]+\mu_{\text{LMS}}^2\mathrm{E}\left[\bar{\mathbf{w}}^T(n)\bar{\mathbf{w}}(n)\right]\right\} \\
&= \lim_{n\to\infty}\left\{(\mathbf{I}_N-\mu_{\text{LMS}}\mathbf{\Lambda})^{2n}\mathrm{E}\left[\left\|\bar{\mathbf{\Delta}}(0)\right\|_2^2\right]+\mu_{\text{LMS}}^2\sum_{i=0}^{n}(\mathbf{I}_N-\mu_{\text{LMS}}\mathbf{\Lambda})^{2i}\mathrm{E}\left[\bar{\mathbf{w}}^T(n)\bar{\mathbf{w}}(n)\right]\right\} \\
&\geq \lim_{n\to\infty}\left\{\mu_{\text{LMS}}^2\sum_{i=0}^{N}(\mathbf{I}_N-\mu_{\text{LMS}}\mathbf{\Lambda})^{2i}\mathrm{E}\left[\bar{\mathbf{w}}^T(n)\bar{\mathbf{w}}(n)\right]\right\}
\end{aligned} \tag{A.9}$$

In the above equation, we use the recursion of $\left[\left\|\bar{\mathbf{\Delta}}(n)\right\|_2^2\right]$ and the $\lim_{n\to\infty}(\mathbf{I}_N-\mu_{\text{LMS}}\Lambda)^{2n}E\left[\left\|\bar{\Delta}(0)\right\|_2^2\right]\to 0$ when $|\mathbf{I}_N-\mu_{\text{LMS}}\lambda_i|<1$ to analyze the MSE lower bound [2]. By considering the ith channel taps b_i $(i=0,1,\cdots,N-2,N-1)$, we can get

$$b_i=\lim_{n\to\infty}\mu_{\text{LMS}}^2\sum_{i=0}^{N-1}(1-\mu_{\text{LMS}}\lambda_i)^{2i}\mathrm{E}\left\{|\bar{\mathbf{w}}_i(n)|^2\right\}=\frac{\mu_{\text{LMS}}P}{2-\mu_{\text{LMS}}\lambda_i}, \tag{A.10}$$

where P is the gradient noise power and $\mathrm{E}\left\{|\bar{\mathbf{w}}_i(n)|^2\right\}=\lambda_i P$. For the standard LMS channel estimation algorithm, it cannot utilize the sparsity of the multipath channel. Thus, the MSE lower bound should be obtained from all the channel taps and it can be expressed as

$$B=\sum_{i=0}^{N-1}b_i=\sum_{i=0}^{N-1}\frac{\mu_{\text{LMS}}P}{2-\mu_{\text{LMS}}\lambda_i}\geq\sum_{i=0}^{N-1}\frac{\mu_{\text{LMS}}P}{2-\mu_{\text{LMS}}\lambda_{\min}}=\frac{\mu_{\text{LMS}}NP}{2-\mu_{\text{LMS}}\lambda_{\min}}, \tag{A.11}$$

where $\lambda_{\min}$ is the minimum eigenvalue of covariance matrix $\mathbf{R}$ of the input signal $\mathbf{x}(n)$. From the equation (A.11), we can see that the CRLB of non-sparse channel estimator is $B\sim\mathcal{O}(N)$. Then we consider a sparse channel, which is comprised of K non-zero taps and $(N-K)$ non-zero taps. Let $\mathcal{A}=\{n\in\mathbb{N}\mid 1\leq n\leq N\}$ be the taps of the sparse channel and let $\mathcal{C}=\{c_1,c_2,\cdots,c_K\}\subset\mathcal{A}$ $(|\mathcal{C}|=K$, where $|\cdot|$ is the cardinality, that is, the number of elements in the set) be the non-zero taps of the sparse channel $\mathbf{h}$. Therefore, the sparse channel can be expressed as

$$h_i = \begin{cases} c_K, \text{ nonzero taps} \\ 0, \text{otherwise} \end{cases}. \tag{A.12}$$

Thus, the MSE lower bound of such sparse channel can be obtained from the equation (A.10)

$$B_s = \sum_{i=0, i\in\mathcal{C}}^{N-1} b_i = \sum_{i\in\mathcal{C}} \frac{\mu_{\text{LMS}}P}{2-\mu_{\text{LMS}}\lambda_i} \geq \sum_{i\in\mathcal{C}} \frac{\mu_{\text{LMS}}P}{2-\mu_{\text{LMS}}\lambda_{\min}} = \frac{\mu_{\text{LMS}}KP}{2-\mu_{\text{LMS}}\lambda_{\min}} \sim \mathcal{O}(K). \tag{A.13}$$

From the above discussions and the proof for **Theorem 1**, the K non-zero taps has important effects on the LMS-type sparse channel estimation. Thus, design sparse channel estimation algorithms is a attractive and desirable topic.

Appendix B

Publications

B.1 Refereed journal papers

[1]. R. Hayashi, Y. Li, and M. Hamamura, Two-stage decoding algorithm for unmodulated parallel-combinatory high-compaction multicarrier modulation signals, EURASIP Journal on Wireless Communications and Networking, vol.2013, no.3, Article ID:75, 2013.
[2]. Y. Li and M. Hamamura, Smooth approximation l_0-norm constrained affine projection algorithm and its applications in sparse channel estimation, Scientific World Journal, vol.2014, Article ID:937252, 2014.
[3]. Y. Li and M. Hamamura, An improved proportionate normalized least mean square algorithm for broadband multipath channel estimation, Scientific World Journal, vol.2014, Article ID:572969, 2014.
[4]. Y. Li and M. Hamamura, Zero-attracting variable-step-size least mean square algorithms for adaptive sparse channel estimation, Submitted to Mathematical Problems in Engineering.

B.2 Conference proceedings

[1]. Y. Li and M. Hamamura, Sparse signal detection for unmodulated parallel-combinatory high-compaction multicarrier modulation system using compressed sensing, Proc. IB2COM 2012, pp.188-189, Sydney, Australia, Nov. 2012.
[2]. Y. Li and M. Hamamura, A compressed sensing based signal recovery in parallel combinatory high compaction multicarrier modulation system," Proc. SJCIEE 2012, p.205, Kagawa, Japan.

References

[1] D. Raychaudhuri and N. B. Mandayam. Frontiers of wireless and mobile communications. *Proc. IEEE*, 100(4):824–840, 2012.

[2] G. Gui and F. Adachi. Improved least mean square algorithm with application to adaptive sparse channel estimation. *EURASIP J. Wireless Commun. Networks*, 2013:1–18, 2013.

[3] J. A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50(10):2231–2242, 2004.

[4] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. *IEEE Trans. Signal Process.*, 45(3):600–616, 1997.

[5] J. G. Proakis. *Digital Communications*. McGraw-Hill, fourth edition, 2001.

[6] L. Korowajczuk. *LTE, WiMAX and WLAN Network Design, Optimization and Performance Analysis*. John Wiley, 2011.

[7] F. Adachi, G. Grag, S. Takaoka, and K. Takeda. Broadband CDMA techniques. *IEEE Wirel. Commun.*, 12(2):8–18, 2005.

[8] S. F. Cotter and B. D. Rao. Sparse channel estimation via matching pursuit with application to equalization. *IEEE Trans. Commun.*, 50(3):374–377, 2002.

[9] F. Adachi, D. Grag, S. Takaoka, and K. Takeda. New direction of broadband CDMA techniques. *Wireless Commun. Mobile Comput.*, 50(3):374–377, 2002.

[10] F. Adachi and E. Kudoh. New direction of broadband wireless technology. *Wirel. Commun. Mob. Com.*, 7(8):969–983, 2007.

[11] J. Yang. *Adaptive filter design for sparse signal estimation.* University of Minnesota, 2011.

[12] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett. Sparse channel estimation for multicarrier underwater acoustic communication: from subspace methods to compressed sensing. *IEEE Trans. Signal Process.*, 58(3):1708–1721, 2010.

[13] M. A. R. Baissas and A. M. Sayeed. Pilot-based estimation of time-varying multipath channels for coherent CDMA receivers. *IEEE Trans. Signal Process.*, 50(8):2037–2049, 2002.

[14] K. Zhang, X. Lei, and S. Li. Iterative channel estimation and data detection for MIMO-OFDM systems operating in time-frequency dispersive channels under unknown background noise. *EURASIP J. Wireless Commun. Networks*, 2013:1–14, 2013.

[15] A. F. Molisch. Ultrawideband propagation channels-theory, measurement, and modeling. *IEEE Trans. Veh. Technol.*, 54(5):1528–1545, May 2005.

[16] P. Maechler, P. Dreisen, B. Sporrer, and S. Steiner. Implementation of greedy algorithms for LTE sparse channel estimation. *Proc. ASILOMAR'10*, 1:400–405, 2010.

[17] C. Paleologu, J. Benesty, and S. Ciochină. *Sparse adaptive filters for echo cancellation.* Morgan & Claypool Publishers, 2010.

[18] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing.* Springer, 2010.

[19] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, 2001.

[20] Y. Kabashima, T. Wadayama, and T. Tanaka. A typical reconstruction limit for compressed sensing based on l_p-norm minimization. *J. Stat. Mech*, 2009:1–11, 2009.

[21] D. Donoho, I. Johnstone, A. SMaleki, and A. SMontanari. Compressed sensing over l_p-balls: minimax mean square error. *Proc. ISIT'11*, pages 129–133, 2011.

[22] H. Mohimani, M. Babaie-Zadeh, and C. Jutten. A fast approach for over-complete sparse decomposition based on smoothed l^0 norm. *IEEE Trans. Signal Process.*, 57(1):289–301, 2009.

[23] Y. Li. Pilot-symbol-aided channel estimation for OFDM in wireless systems. *IEEE Trans. Veh. Technol.*, 49(4):1207–1215, 2010.

[24] I. Prodan, T. Obara, F. Adachi, and H. Gacanin. Performance of pilot-assisted channel estimation without feedback for broadband ANC systems using OFDM access. *EURASIP J. Wireless Commun. Networks*, 2012:1–9, 2012.

[25] F. Wan, W. P. Zhu, and M. N. S. Swamy. Semiblind sparse channel estimation for MIMO-OFDM systems. *IEEE Trans. Veh. Technol.*, 60(6):2569–2582, 2011.

[26] W. Nie, J. Zhang, Y. Liu, and F. Sun. A robust channel estimation for broadband OFDM systems with virtual tones. *IEEE Proc. VTC 2010-fall*, pages 1–5, 2010.

[27] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(3):21–30, 2008.

[28] D. Needell and R. Vershynin. Signal recovery from inaccurate and incomplete measurements via regularized orthogonal matching pursuitt. *IEEE J. Sel. Top. Signal Process.*, 4(2):310–316, 2010.

[29] K. Hayashi, M. Nagahara, and T. Tanaka. A user's guide to compressed sensing for communications systems. *IEICE Trans. Commun.*, (3):685–712, 2013.

[30] D. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

[31] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math.*, 346(9).

[32] E. J. Candès and J. Romberg. l_1-magic: recovery of sparse signals via convex programming. *http://users.ece.gatech.edu/ justin/l1magic/downloads /l1magic.pdf*, pages 1–19, 2005.

[33] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory*, 53(12):4655–4666, 2007.

[34] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Found. Comput. Math.*, 9(3):317–334, 2009.

[35] R. Tibshirani. Regression shringage and selection via the lasso. *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, 58(1):267–288, 1996.

[36] Y. Wang and W. Yin. Sparse signal reconstruction via iterative support detection. *SIAM J. Imaging Sci.*, 3(3):462–491, 2010.

[37] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inf. Theory*, 55(5):2230–2249, May 2009.

[38] M. Lustig, D. Donodo, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58(6):1182–1195, 2007.

[39] J. Meng, W. Yin, Y. Li, N. T. Nguyen, and H. Zhu. Compressive sensing based high-resolution channel estimation for OFDM system. *IEEE J. Sel. Top. Signal Process.*, 6(1):15–25, 2012.

[40] C. R. Berger, Z. Wang, and S. Zhou. Application of compressive sensing to sparse channel estimation. *IEEE Commun. Mag.*, 48(11):164–174, 2010.

[41] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak. Compressed channel sensing: a new approach to estimating sparse multipath channels. *Proc. IEEE*, 98(6):1058–1076, 2010.

[42] G. Tauböck and F. Hlawatach. A compressed sensing technique for OFDM channel estimation in mobile environment: exploting channel sparsity for reducing pilots. *Proc. IEEE ICASSP'08*, 1:2885–2888, 2008.

[43] A. H. Sayed. *Adaptive filters*. Wiley-IEEE Press, 2008.

[44] P. S. R. Diniz. *Adaptive filtering algorithms and practical implementation*. Spring, fourth edition, 2013.

[45] B. Wirow and S. D. Stearns. *Adaptive signal processing*. Prentice Hall, 1985.

[46] D. T. M. Slock. On the convergence behavior of the LMS and the normalized LMS algorithms. *IEEE Trans. Signal Process.*, 41(9):2811–2825, 1993.

[47] W. P. Ang and B. Farhang-Boroujeny. A new class of gradient adaptive step-size LMS algorithms. *IEEE Trans. Signal Process.*, 49(4):805–810, 2001.

[48] M. A. Mohammadi, M. Ardabilipour, B. Moussakhani, and Z. Mobini. Performance comparison of RLS and LMS channel estimation techniques with optimum training sequences for MIMO-OFDM systems. *Proc. WOCN'08*, 1:1–5, May 2008.

[49] Y. Yapc and A. O. Yilmaz. Joint channel estimation and decoding with low-complexity iterative structures in time-varying fading channels. *Proc. PIRMC'09*, 1:1943–1947, 2009.

[50] A. K. Kohli and D. K. Mehra. Adaptive multiuser channel estimation using reduced Kalman/LMS algorithm. *Wireless Pers. Commun.*, 46(4):507–521, 2008.

[51] K. S. Ahn, J. Cho, and H. K. Baik. Blind adaptive channel equalization using multichannel linear prediction-based cross-correlation vector estimation. *IEEE Trans. Consum. Electron.*, 50(4):1026–1032, 2004.

[52] A. M. A. Filho, E. L. Pinto, and J. F. Galdino. Variable step-size LMS algorithm for estimation of time-varying and frequency-selective channels. *Electronics Lett.*, 40(20):1312–1313, 2004.

[53] M. M. Rana and J. Kim. Lms based blind channel estimation of SC-FDMA systems using variable step size and phase information. *Electronics Lett.*, 47(5):346–348, 2011.

[54] J. Homer, I. Mareels, and C. Hoang. Enhanced detection-giuded NLMS estimation of sparse FIR-modeled signal channels. *IEEE Trans. Circuits Syst. I-Regul. Pap.*, 53(8):1783–1791, 2006.

[55] H. Zayyani, M. Babaie-zadeh, and C. Jutten. Compressed sensing block MAP-LMS adaptive filter for sparse channel estimation and a Bayesian Cramer-Rao bound. *IEEE Proc. MLSP'09*, pages 1–6, 2009.

[56] Md. M. Rana, J. Kim, and W. K. Cho. LMS based adaptive channel estimation for LTE uplink. *Radioengineering*, 19(4):678–688, 2010.

[57] B. Babadi, N. Kalouptsidis, and V. Tarokn. SPARLS: The sparse RLS algorithm. *IEEE Trans. Signal Process.*, 58(8):4013–4025, 2010.

[58] 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 Commun.*, 1(4):795–801, 2007.

[59] T. K. Akino. Optimum-weighted RLS channel estimation for rapid fading MIMO channels. *IEEE Trans. Wirel. Commun.*, 7(11):4248–4260, 2008.

[60] R. Niazadeh, S. H. Ghalehjegh, M. Babaie-Zadeh, and C. Jutten. ISI sparse channel estimation based on SL0 and its application in ML sequence-by-sequence equalization. *Signal Process.*, 92(8):1875–1885, 2012.

[61] R. Arablouei and K. Dogancay. Affine projection algorithm with selective projections. *Signal Process.*, 92(9):2253–2263, 2012.

[62] A. Mirbagheri and Y. C. Yoon. A blind adaptive receiver for interference suppression and multipath reception in long-code DS-CDMA. *Proc. IEEE ICC'02*, 1:242–246, 2002.

[63] S. Chompoo, C. Benjangkaprasert, N. Anatrasiarchai, and K. Janchitrapongvej. Adaptive feedback equalization receiver for DS-CDMA with turbo coded systems. *Proc. ICACT'02*, 1:1068–1073, 2005.

[64] X. G. Doukopoulos and G. V. Moustakides. Blind adaptive channel estimation in ofdm systems. *IEEE Trans. Commun.*, 5(7):1716–1725, 2006.

[65] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel. Multi-input multi-output fading channel tracking and equalization using Kalman estimation. *IEEE Trans. Signal Process.*, 50(5):1065–1076, May 2002.

[66] D. L. Duttweiler. Proportionate normalized least-mean-squares adaptive in echo cancelers. *IEEE Trans. Speech and Audio Process.*, 8(5):508–518, 2000.

[67] Y. Chen, Y. Gu, and A. O. Hero. Sparse LMS for system identification. *Proc. IEEE ICASSP'09*, 1:3125–3128, 2009.

[68] M. Godavarti and A. O. Hero. Partial update LMS algorithm. *IEEE Trans. Signal Process.*, 53(7):2382–2399, 2005.

[69] Y. F. Hung and J. H. Wen. An analysis on partial PIC multi-user detection with LMS algorithms for CDMA. *Proc. IEEE PIMRC'03*, 1:17–21, 2003.

[70] S. Werner, M. L. R. Campos, and P. S. R. Diniz. Partial-update NLMS algorithms with data-selective updating. *IEEE Trans. Signal Process.*, 52(4):938–949, 2004.

[71] P. A. Naylor and A. W. H. Khong. Affine projection and recursive least squares adaptive filters employing partial updates. *Proc. ACSSC*, 1:950–954, 2004.

[72] N. Vaswani. Kalman filtered compressed sensing. *Proc. ICIP'08*, 1:893–896, 2008.

[73] E. Eksioglu. Sparsity regularised recursive least squares adaptive filtering. *IET Signal Process.*, 5(2):480–487, 2011.

[74] A. W. H. Khong and P. A. Naylor. Efficient use of sparse adaptive filters. *Proc. ACSSC'06*, 1:1375–1379, 2006.

[75] R. Meng, R. C. Lamare, and V. H. Nascimento. Sparsity-aware affine projection adaptive algorithms for system identification. *Proc. SSPD'11*, 1:1–5, 2011.

[76] T. Aboulnasr and K. Mayyas. A robust variable step-size LMS-type algorithm: analysis and simulation. *IEEE Trans. Signal Process.*, 45(3):631–639, 1997.

[77] H. C. Shin, A. H. Sayed, and W. J. Song. Variable step size NLMS and affine projection algorithms. *IEEE Signal Process. Lett.*, 11(2):132–135, 2004.

[78] S. C. Douglas. Performance comparison of two implementations of the leaky LMS adaptive filter. *IEEE Trans. Signal Process.*, 45(8):2125–2129, 1997.

[79] M. S. Salman, M. N. S. Jahromi, A. Hocanin, and O. Kukrer. A weighted zero-attracting leaky-LMS algorithm. *Proc. SoftCOM'12*, 1:1–3, 2012.

[80] B. Widrow and S. D. Stearns. *Adaptive signal processing.* Prentice-Hall, 1985.

[81] D. P. Mandic. A generalized normalized gradient descent algorithm. *IEEE Signal Process. Lett.*, 11(2):115–118, 2004.

[82] K. Shi and P. Shi. Convergence analysis of sparse LMS algorithms with l_1-norm penalty based on white input signal. *Signal Process.*, 90(12):3289–3293, 2010.

[83] E. J. Candés, M. B. Wakin, and S. P. Noyd. Enhancing sparsity by reweighted l_1-minimization. *J. Fourier Anal. Appl.*, 15(5):877–905, 2008.

[84] Y. Chen, Y. Gu, and A. O. Hero. Regularized least-mean-square algorithms. *http://arxiv.org/pdf/1012.5066.pdf*, pages 1–9S, 2010.

[85] S. Werner, J. A. Apolinario Jr., M. L. R. Campos, and P. S. R. Diniz. Low complexity comstrained affine-projection algorithms. *IEEE Trans. Signal Process.*, 53(12):454–4555, 2005.

[86] J. Benesty and S. Gay. An improved PNLMS algorithm. *Proc. IEEE ICASSP'02*, 1:1881–1884, May 2002.

[87] H. Dong and M. Doroslovački. Improving convergence of the PNLMS algorithm for sparse impulse response identification. *IEEE Signal Process. Lett.*, 12(3):181–184, 2005.

[88] R. L. Das and M. Chakraborty. A zero attracting proportionate normalized least mean square algorithm. *Proc. APSIPA ASC'12*, 1:1–4, 2012.

[89] L. Liu, M. Fukumoto, S. Saiki, and S. Zhang. A variable step-size proportionate NLMS algorithm for identification of sparse impulse response. *IEICE Trans. Fundamentals*, (1):233–242, 2010.

[90] L. Liu, M. Fukumoto, and S. Saiki. An improved mu-law proportionate NLMS algorithm. *Proc. ICASSP'08*, 1:3797–3800, 2008.

[91] L. Liu, M. Fukumoto, and S. Saiki. Proportionate normalized least mean square algorithms based on coefficient difference. *IEICE Trans. Fundamentals*, E93-A(5):972–975, May 2010.

[92] M. Nekuii and M. Atarodi. A fast converging algorithm for network echo cancelation. *IEEE Signal Process. Lett.*, 11(4):427–430, 2004.

[93] K. Do gancay. *Partial-Update Adaptive Signal Processing: Design, Analysis and Implementation*. Academic Press, 2008.

[94] Y. Li and M. Hamamura. Zero-attracting variable-step-size least mean square algorithms for adaptive sparse channel estimation. *Submitted to Mathematical Problems in Engineering*.

[95] K. T. Wagner and M. I. Doroslovachi. Gain allocation in proportionate-type NLMS algorithms for fast decay of output error at all times. *Proc. IEEE ICASSP'09*, 1:2117–2120, 2009.

[96] J. F. Tan and J. Z. Ouyang. A novel variable step-size LMS adaptive filtering algorithm based on sigmoid function. *Journal of Data Acquisition & Processing*, 12(3):171–174, 1997.

[97] Y. Gao and S. Xie. A variable step-size LMS adaptive filtering algorithm and its analysis. *Acta Electronica Sinica*, 29(8):1094–1097, 2001.

[98] O. Taheri and S. A. Vorobyov. Sparse channel estimation with l_p-norm and reweighted l_1-norm penalized least mean squares. *Proc. IEEE ICASSP'11*, 1:2864–2867, May 2011.

[99] J. Jin, Y. Gu, and S. Mei. A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework. *IEEE J. Sel. Top. Signal Process.*, 4(2):409–420, 2010.

[100] Y. Li and M. Hamamura. Smooth approximation l_0-norm constrained affine projection and its applications in sparse channel estimation. *The Scientific World Journal*, 2014, 2014.

[101] F. Y. Wu and F. Tong. Gradient optimization p-norm-like constraint LMS algorithm for sparse system estimation. *Signal Process.*, 93(4):967–971, 2013.

[102] M. Hyder and K. Mahata. An approximate l_0-norm minimization algorithm for compressed sensing. *Proc. IEEE ICASSP'09*, 1:3365–3368, 2009.

[103] R. E. Carrillo and K. E. Barner. Iteratively re-weighted least squares for sparse signal reconstruction from noisy measurements. *Proc. CISS'09*, 1:448–453, 2009.

[104] Y. Gu, J. Jin, and S. Mei. L_0 norm constraint LMS algorithms for sparse system identification. *IEEE Signal Process. Lett.*, 16(9):774–777, 2009.

[105] M. V. S. Lima, W. A. Martins, and P. S. Z. Diniz. Affine projection algorithms for sparse system identification. *Proc. ICASSP'13*, 1:5666–5670, 2013.

[106] H. C. Shin and A. H. Sayed. Mean-square performance of a family of affine projection algorithms. *IEEE Signal Process. Lett.*, 52(1):90–102, 2004.

[107] R. Meng. *Sparsity-aware adaptive filtering algorithms and application to system identification*. University of York, 2011.

[108] H. C. Shin, W. J. Song, and A. H. Sayed. Mean-square performance of data-reusing adaptive algorithms. *IEEE Signal Process. Lett.*, 12(12):851–854, 2005.

[109] G. Su, J. Jin, Y. Gu, and J. Wang. Performance analysis of l_0 norm constraint least mean square algorithm. *IEEE Trans. Signal Process.*, 60(5):2223–2235, May 2013.

[110] G. Barrault, M. H. Costa, J. C. M. Bermudez, and A. Lenzi. A new analytical model for the NLMS algorithm. *Proc. IEEE ICASSP'05*, 4:41–44, 2005.

[111] M. H. Costa and J. C. M. Bermudez. An improved model for the normalized LMS algorithm with gaussian inputs and large number of coefficients. *Proc. IEEE ICASSP'02*, 2:1385–1388, 2002.

[112] Y. Li and M. Hamamura. Discrete weighted zero-attracing affine projection algorithm for sparse channel estimation. *Being Prepared*.

[113] G. Gui, N. Zheng, N. Wang, A. Mehbodniya, and F. Adachi. Compressive estimation of cluster-sparse channels. *Progress In Electromagnetics Research C*, 24:251–263, 2011.

[114] Y. C. Eldar, P. Kuppinger, and H. Bolcskei. Block-sparse signals: uncertainty relations and efficient recovery. *IEEE Trans. Signal Process.*, 58(6):3042–3054, 2010.

[115] Y. Li and M. Hamamura. An improved proportionate normalized least-mean-square algorithm for broadband multipath channel estimation. *Submitted to The Scientific World Journal*.

[116] J. Yang and G. E. Sobelman. Sparse LMS with segment zero attractors for adaptive estimation of sparse signals. *Proc. APCCAS'10*, 1:422–425, 2010.

[117] S. L. Gay and S. Tavathia. The fast affine projection algorithm. *Proc. ICASSP'95*, 1:3023–3026, May 1995.