

# Robust Zero-point Attraction Least Mean Square Algorithm on Near Sparse System Identification

Jian Jin, Qing Qu, and Yuantao Gu\*

Received Feb. 2012; accepted Feb. 2013.

This article appears in *IET Signal Processing*, 7(3):210-218, 2013.

## Abstract

The newly proposed  $l_1$  norm constraint zero-point attraction Least Mean Square algorithm (ZA-LMS) demonstrates excellent performance on exact sparse system identification. However, ZA-LMS has less advantage against standard LMS when the system is near sparse. Thus, in this paper, firstly the near sparse system modeling by Generalized Gaussian Distribution is recommended, where the sparsity is defined accordingly. Secondly, two modifications to the ZA-LMS algorithm have been made. The  $l_1$  norm penalty is replaced by a partial  $l_1$  norm in the cost function, enhancing robustness without increasing the computational complexity. Moreover, the zero-point attraction item is weighted by the magnitude of estimation error which adjusts the zero-point attraction force dynamically. By combining the two improvements, Dynamic Windowing ZA-LMS (DWZA-LMS) algorithm is further proposed, which shows better performance on near sparse system identification. In addition, the mean square performance of DWZA-LMS algorithm is analyzed. Finally, computer simulations demonstrate the effectiveness of the proposed algorithm and verify the result of theoretical analysis.

**Keywords:** LMS, Sparse system identification, Zero-point attraction, ZA-LMS, Generalized Gaussian distribution

## 1 Introduction

A sparse system is defined when impulse response contains only a small fraction of large coefficients compared to its ambient dimension. Sparse systems widely exist in many applications, such as Digital TV transmission channel [1] and the echo path [2]. Generally, they can be further classified into two categories: exact sparse system (ESS) and near sparse system (NSS). If most coefficients of the impulse response are exactly zero, it is defined as

---

\*This work was partially supported by National Natural Science Foundation of China (NSFC 60872087 and NSFC U0835003). The authors are with Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. The corresponding author of this paper is Yuantao Gu (Email: gyt@tsinghua.edu.cn).

an exact sparse system (Fig. 1. a); Instead, if most of the coefficients are close (not equal) to zero, it is a near sparse system (Fig. 1. b and c). Otherwise, a system is non-sparse if its most taps have large values (Fig. 1. d). For the simplicity of theoretical analysis, sparse systems are usually simplified into exact sparse. However, in real applications most systems are near sparse due to the ineradicable white noise. Therefore, it is necessary to investigate on near sparse system modeling and identification.

Among many adaptive filtering algorithms for system identification, Least Mean Square (LMS) algorithm [3], which was proposed by Widrow and Hoff in 60s of the past century, is the most attractive one for its simplicity, robustness and low computation cost. However, without utilizing the sparse characteristic, it shows no advantage on sparse system identification. In the past few decades, some modified LMS algorithms for sparse systems are proposed. M-Max Normalized LMS (MMax-NLMS) [4] and Sequential Partial Update LMS (S-LMS) [5] reduces the computational complexity and steady-state misalignment by partially updating the filter coefficients. Proportionate LMS (PLMS) [2] and its improved ones such as IPNLMS [6] and IIPNLMS [7] accelerate the convergence rate by updating each coefficient iteratively with different step size proportional to the magnitude of filter coefficient. Stochastic Tap-Normalized LMS (ST-NLMS) [8, 9] improves the performance on specific sparse system identification where large coefficients appear in clusters. It locates and tracks the non-zero coefficients by adjusting the filter length dynamically. However, its convergence performance largely depends on the span of clusters. If the span is too long or the system has multiple clusters, it shows no advantage compared with standard LMS algorithm.

More recently, inspired by the research of CS reconstruction problem [10, 11], a class of novel adaptive algorithms for sparse system identification have emerged based on the  $l_p$  ( $0 \leq p \leq 1$ ) norm constraint [12–14]. Especially, Zero-point Attraction LMS (ZA-LMS) algorithm [12] significantly improves the performance on exact sparse system identification by introducing a  $l_1$  norm constraint on the cost function of standard LMS, which exerts the same zero-point attraction force on all coefficients. However, for near sparse systems identification, the zero-point attractor can be a double-edged sword. Though it increases the convergence rate because by the  $l_1$  norm constraint, it also produces larger steady-state misalignment as it forces all coefficients to exact zero. Thus, it possesses less advantage against standard LMS algorithm when the system is near sparse. In this paper, firstly Generalized Gaussian Distribution (GGD) [15] is introduced to model the near sparse system. Then two improvements on the ZA-LMS algorithm is proposed. Above all, by adding a window on the  $l_1$  norm constraint, the steady-state misalignment is reduced without increasing the computational complexity. Furthermore, the zero-point attractor is weighted to adjust the zero-point attraction by utilizing the estimation error. By combining the two improvements, the Dynamic Windowing ZA-LMS (DWZA-LMS) algorithm is proposed which shows improved performance on the near sparse system identification.

The rest of the paper is organized as follows: In Section II, ZA-LMS algorithm based on

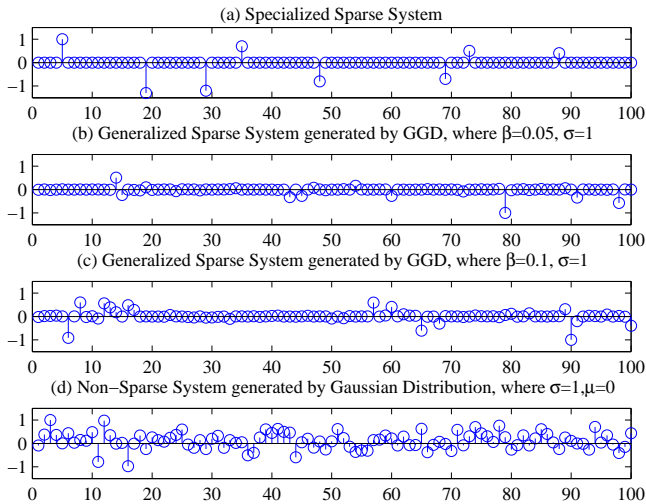


Figure 1: Sparse and non-sparse systems. (a) is a exact sparse system, (b) and (c) are near sparse system generated by Generalized Gaussian Distribution with  $\beta = 0.05$  and  $0.1$  respectively. (d) is a non-sparse system generated by Gaussian distribution.

$l_1$  norm constraint is reviewed, and the near sparse system is modeled. The new algorithm is proposed in Section III. In Section IV, the mean square convergence performance of DWZA-LMS is analyzed. The performances of the new algorithm and other improved LMS algorithms for sparse system identification are compared by simulation in Section V, where the effectiveness of our analysis is verified as well. Finally, Section VI concludes the paper.

## 2 Preliminary

### 2.1 Review of ZA-LMS algorithm

Let  $d(n)$  be a sample of the desired output signal

$$d(n) = \mathbf{h}^T \mathbf{x}(n) + v(n), \quad (1)$$

where  $\mathbf{h} = [h_0, h_1, \dots, h_{L-1}]^T$  is the unknown system with memory length  $L$ ,  $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L+1)]^T$  denotes the input vector, and  $v(n)$  is the observation noise assumed to be independent of  $x(n)$ . The estimation error  $e(n)$  between desired and output signal is defined as

$$e(n) = d(n) - \mathbf{w}^T(n) \mathbf{x}(n). \quad (2)$$

where  $\mathbf{w}(n)$  are the filter coefficients and  $\mathbf{w}(n) = [w_0(n), w_1(n), \dots, w_{L-1}(n)]^T$ . Thus the cost function of ZA-LMS is

$$\xi_{\text{ZA}}(n) = \frac{1}{2} |e(n)|^2 + \eta \|\mathbf{w}(n)\|_1, \quad (3)$$

where  $\|\mathbf{w}(n)\|_1$  denotes the  $l_1$  norm of the filter coefficients. Parameter  $\eta$  is the factor balancing the new penalty and the estimation error. By minimizing (3), the ZA-LMS algorithm updates its coefficients by

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n) - \rho \text{sgn}[\mathbf{w}(n)], \quad (4)$$

where  $\mu$  is the step size,  $\rho = \eta\mu$  is the zero-point attraction controller, and  $\text{sgn}[\cdot]$  is the component-wise sign function.

By observing (4), the recursion of filter coefficients for sparse system can be summarized as

$$\mathbf{w}_{\text{new}} = \mathbf{w}_{\text{prev}} + \mu f_{\text{GC}}(\mathbf{w}) + \rho f_{\text{ZA}}(\mathbf{w}), \quad (5)$$

where  $f_{\text{GC}}(\mathbf{w}) = e(n)\mathbf{x}$  denotes the gradient correction function, and  $f_{\text{ZA}}(\mathbf{w}) = -\text{sgn}(\mathbf{w})$  stands for zero-point attractor which is caused by the  $l_1$  norm penalty. For each iteration, the zero-point attractor forces the filter taps to decrease a little when it is positive, or otherwise to increase a little when it is negative.

## 2.2 Near Sparse System Modeling

Exact sparse system is appropriate for theoretical analysis, however, most physical systems are near sparse with widely existing white noise in real life, thus their modeling is of more significant importance. Generalized Gaussian Distribution (GGD) is one of the most prominent and widely used sparse distributions. For example, in multimedia communications such as image and speech coding, GGD is usually found to best fit the coefficients of the discrete sine and cosine transforms, the Walsh-Hadamard transform, and the wavelet transform [16]. In ultra wide bandwidth (UWB) systems, it has recently been found to fit the multiuser interference better [17]. These findings lead to applications of GGD in video coding, speech recognition, blind signal separation and UWB receiver design [18–20]. Therefore GGD is utilized to model the near sparse system in this study. It is a class of symmetry distribution with the Gaussian and Laplacian distribution as the special cases, with delta and uniformity distribution as limit. The probability density function of GGD is

$$f(x) = \frac{\beta}{2\lambda\Gamma(1/\beta)} \exp\left[-(|x - \mu|/\lambda)^\beta\right], \quad (6)$$

where  $\lambda = \sigma_g \sqrt{\frac{\Gamma(1/\beta)}{\Gamma(3/\beta)}}$ ,  $\Gamma(x)$  denotes the Gamma function,  $\mu$  and  $\sigma_g^2$  are called the mean and variance of GGD, respectively. Besides,  $\beta$  determines the decay rate of the density function and can be used to denote the sparsity of the system (see as Fig. 1. b, c; please notice that the system sparsity decreases as  $\beta$  increases). Especially, GGD is Gaussian Distribution when  $\beta = 2$ , and it turns into Laplacian Distribution when  $\beta = 1$ . By integrating  $f(x)$ , the

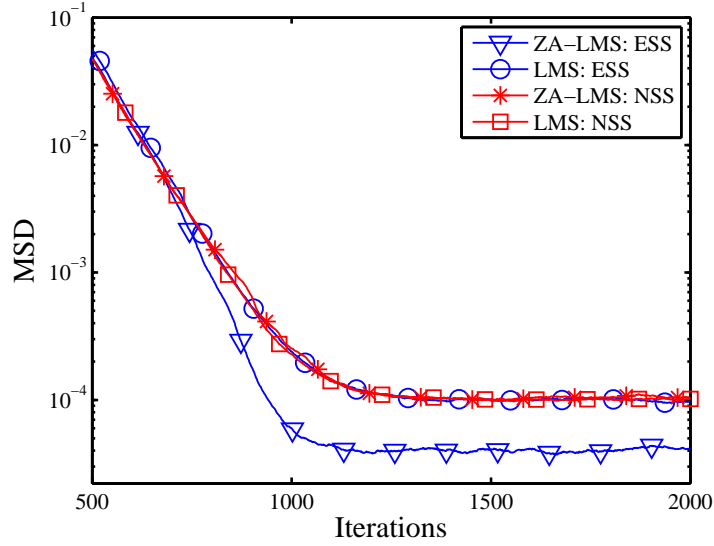


Figure 2: Comparison between ZA-LMS and LMS on Exact Sparse System and Near Sparse System.

distribution function is

$$F(x) = \frac{1}{2} + \text{sgn}(x - \mu) \frac{\Theta \left[ 1/\beta, \left( \frac{|x-\mu|}{\lambda} \right)^\beta \right]}{2\Gamma(1/\beta)}. \quad (7)$$

where  $\Theta(s, x) = \int_0^x t^{s-1} e^{-t} dt$  is called the lower incomplete function.

### 3 Improved ZA-LMS Algorithm

The ZA-LMS algorithm improves the performance by exerting an  $l_1$  norm penalty forcing small coefficients to zero iteratively. It is very effective for the exact sparse system identification. However, for the near sparse system with small white noise on all coefficients, most of the small coefficients are forced to zero. Thus ZA-LMS usually degrades with large steady-state misalignment, showing no improvement compared with standard LMS. The situation is more clearly illustrated in Fig. 2, where the exact sparse system is generated with 8 nonzero large coefficients with tap length of 100, and the near sparse system is the same with the former except that a power of  $1 \times 10^{-4}$  white noise is added on all coefficients. The step sizes for all algorithms are set as  $\mu = 0.01$ . The optimal parameter  $\rho$  is derived theoretically from (36) of [21] by minimizing the steady-state MSD.

From Fig. 2, it can be seen that by choosing the theoretically optimal parameter  $\rho$  the performance of ZA-LMS is much better than standard LMS for exact sparse system. However, the performance of ZA-LMS severely degrades on near sparse system identification. As stated above, the main reason is that the strong zero-point attraction forces near zero coefficients to zero that results in the large steady-state misalignment. Though empirically

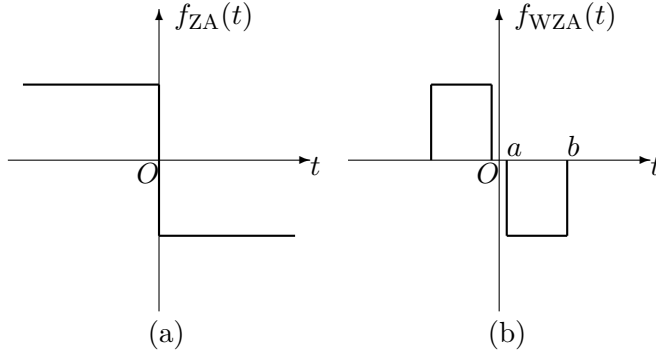


Figure 3: The zero-point attractor for sparse system, (a)ZA-LMS, (b)WZA-LMS.

decreasing  $\rho$  may alleviate this problem, in that case, ZA-LMS will degrade to standard LMS showing no improvement, as shown in Fig. 2. Thus to improve the performance of near sparse system identification, a new ZA-LMS based algorithm is put forward, the main modifications are as follows.

### 3.1 Windowing ZA-LMS

As the attracting range of ZA-LMS algorithm reaches infinity, all coefficients of the sparse system are attracted to zero point. However, the identical attraction on both large and small ones will lead to increase the computational complexity and steady-state misalignment. Thus, the first improvement lies in the constraint of zero-point attracting range. As shown in Fig. 3, the new zero-point attraction, which attracts the coefficients only in a certain range, is proposed by adding a window on the original zero-point attracting. The new recursion of coefficients of the proposed Windowing ZA-LMS (WZA-LMS) is

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n) - \rho \text{sgn}_w[\mathbf{w}(n)], \quad (8)$$

where  $\text{sgn}_w[\cdot]$  is the component-wise partial sign function, defined as

$$\text{sgn}_w[t] = -f_{\text{WZA}}(t) = \begin{cases} \text{sgn}(t) & a < |t| \leq b; \\ 0 & \text{elsewhere.} \end{cases} \quad (9)$$

where  $a$  and  $b$  are both positive constant which denotes the lower and upper threshold of the attraction range, respectively.

From Fig. 3, it can be concluded that ZA-LMS is the special case of WZA-LMS when  $a$  reaches 0 and  $b$  approaches infinity, respectively. Besides, by investigating (4), (8) and (9), it can be seen that the computational complexity of the two algorithms is approximately the same. And by adopting the new zero-point attractor and properly setting the threshold, the coefficients, whether too small or too large, will not be attracted any more. Thus, the steady-state misalignment is significant reduced especially for near sparse system.

### 3.2 Dynamic ZA-LMS

As mentioned above, the sparse constraint should be relaxed in order to reduce the steady-state misalignment when the updating procedure reaches the steady-state. Inspired by the idea of variable step size methods of standard LMS algorithm [23], the magnitude of estimation error, which denotes the depth of convergence, is introduced here to adjust the force of zero-point attraction dynamically. That is,

$$f_{\text{DZA}}(t) = f_{\text{WZA}}(t)|e(n)|. \quad (10)$$

At the beginning of iterations, large estimation error increases zero-point attraction force which also accelerates the convergence. When the algorithm is approaching the steady-state, the error decreases to a minor value accordingly. Thus the influence of zero-point attraction force on small coefficients is reduced that produce smaller steady-state misalignment. By implementation of this improvement on ZA-LMS algorithm, the algorithm is named as Dynamic ZA-LMS (DZA-LMS).

### 3.3 Dynamic Windowing ZA-LMS

Finally, by combining the two improvements, the final Dynamic Windowing ZA-LMS (DWZA-LMS) algorithm can be drew. The new recursion of filter coefficients is as follows,

$$w_i(n+1) = w_i(n) + \mu e(n)x(n-i) - \rho |e(n)| \text{sgn}_w[w_i(n)] \quad \forall 0 \leq i < L. \quad (11)$$

In addition, the new method can also improve the performance of ZA-NLMS, which is known for its robustness. The recursion of DWZA-NLMS is

$$w_i(n+1) = w_i(n) + \frac{1}{\epsilon + \mathbf{x}^T(n)\mathbf{x}(n)} \{ \mu e(n)x(n-i) - \rho |e(n)| \text{sgn}_w[w_i(n)] \} \quad \forall 0 \leq i < L. \quad (12)$$

where  $\epsilon > 0$  is the regularization parameter.

## 4 Analysis of the proposed Algorithm

The mean square convergence analysis of DWZA-LMS algorithm is carried out in this section. The analysis is based on the following assumptions.

1. The input signal  $x(n)$  is i.i.d zero-mean Gaussian. The observation noise  $v(n)$  is zero-mean white. The tap-input vectors  $\mathbf{x}(n)$  and the desired response  $d(n)$  follow the common independence assumption [22], which are generally used for performance analysis of LMS algorithm.
2. The unknown near sparse filter tap follows GGD. As stated in Section II, this assumption is made because GGD is the suitable sparse distribution for near sparse system modeling.

3. The steady state adaptive filter tap  $w_i(n)$  follows the same distribution with  $h_i$  ( $1 \leq i \leq L$ ). This is a reasonable assumption in that the error between the coefficients of the identified and the unknown real systems are very small when the algorithm converges.

Under these assumptions, the mean square convergence condition and the steady-state MSE of DWZA-LMS algorithm are derived. Also, the choice of parameters is discussed in the end of this section.

First of all, the misalignment vector is defined as

$$\mathbf{\Delta}(n) = \mathbf{w}(n) - \mathbf{h}, \quad (13)$$

and auto-covariance matrix of  $\mathbf{\Delta}(n)$  as

$$\mathbf{K}(n) = \mathbb{E} \{ \mathbf{\Delta}(n) \mathbf{\Delta}^T(n) \}. \quad (14)$$

#### 4.1 Mean Square Convergence of Misalignment

Combining (1), (2), (11) and (13), one derives

$$\mathbf{\Delta}(n+1) = \mathbf{A}(n) \mathbf{\Delta}(n) + \mu v(n) \mathbf{x}(n) - \rho \mathbf{m}(n), \quad (15)$$

where  $\mathbf{A}(n) = \mathbf{I} - \mu \mathbf{x}(n) \mathbf{x}^T(n)$ ,  $\mathbf{m}(n) = [m_0(n), m_1(n), \dots, m_{L-1}(n)]$ , and

$$m_i(n) = |e(n)| \text{sgn}_w[w_i(n)] \quad \forall 0 \leq i < L. \quad (16)$$

By utilizing the independence assumption [3], and substituting (15) into (14) yields

$$\begin{aligned} \mathbf{K}(n+1) &= \mathbb{E} \{ \mathbf{A}(n) \mathbf{\Delta}(n) \mathbf{\Delta}^T(n) \mathbf{A}^T(n) \} + \mu^2 \sigma_v^2 \sigma_x^2 \mathbf{I}_L \\ &\quad - \rho \mathbb{E} \{ \mathbf{m}(n) \mathbf{\Delta}^T(n) \mathbf{A}^T(n) \} + \rho^2 \mathbb{E} \{ \mathbf{m}(n) \mathbf{m}^T(n) \} \\ &\quad - \rho \mathbb{E} \{ \mathbf{A}(n) \mathbf{\Delta}(n) \mathbf{m}^T(n) \}, \end{aligned} \quad (17)$$

where  $\mathbf{I}_L$  is an  $L \times L$  unit matrix,  $\sigma_x^2$  and  $\sigma_v^2$  denote the power of input signal and observation noise, respectively. By utilizing the property that the fourth-order moment of a Gaussian variable is three times the variance square, one obtains

$$\mathbb{E} \{ \mathbf{A}(n) \mathbf{\Delta}(n) \mathbf{\Delta}^T(n) \mathbf{A}^T(n) \} = (1 - 2\mu\sigma_x^2 + 2\mu^2\sigma_x^4) \mathbf{K}(n) + \mu^2 \sigma_x^4 D(n) \mathbf{I}_L, \quad (18)$$

where  $D(n) = \text{tr} [\mathbf{K}(n)]$ . With (13), one has

$$\mathbb{E} \{ \mathbf{A}(n) \mathbf{\Delta}(n) \mathbf{m}^T(n) \} = \{ \mathbb{E} \{ \mathbf{m}(n) \mathbf{\Delta}^T(n) \mathbf{A}^T(n) \} \}^T = (1 - \mu\sigma_x^2) \mathbb{E} \{ \mathbf{\Delta}(n) \mathbf{m}^T(n) \}. \quad (19)$$

Combining (17), (18) and (19), one derives

$$\begin{aligned} \mathbf{K}(n+1) &= (1 - 2\mu\sigma_x^2 + 2\mu^2\sigma_x^4) \mathbf{K}(n) + \mu^2 \sigma_x^4 D(n) \mathbf{I}_L \\ &\quad + \mu^2 \sigma_v^2 \sigma_x^2 \mathbf{I}_L - \rho(1 - \mu\sigma_x^2) \{ \mathbb{E} \{ \mathbf{\Delta}(n) \mathbf{m}^T(n) \} \}^T \\ &\quad - \rho(1 - \mu\sigma_x^2) \mathbb{E} \{ \mathbf{\Delta}(n) \mathbf{m}^T(n) \} + \rho^2 \mathbb{E} \{ \mathbf{m}(n) \mathbf{m}^T(n) \}. \end{aligned} \quad (20)$$



By taking trace on both sides of (20), it can be concluded that the adaptive filter is stable if and only if

$$0 < 1 - 2\mu\sigma_x^2 + (L + 2)\mu^2\sigma_x^4 < 1, \quad (21)$$

which is simplified to

$$0 < \mu < \frac{2}{(L + 2)\sigma_x^2}. \quad (22)$$

This implies that the proposed DWZA-LMS algorithm has the same stability condition for the mean square convergence as the ZA-LMS and standard LMS algorithm [3].

## 4.2 Steady-state Mean Square Error

In this subsection, the steady-state Mean Square Error (MSE) of DWZA-LMS algorithm is analyzed. By definition, MSE is

$$\text{MSE} = \text{E}\{e^2(\infty)\} = \sigma_v^2 + \text{tr}\{\mathbf{R}(\infty)\mathbf{K}(\infty)\}, \quad (23)$$

where  $\mathbf{R}(\infty) = \sigma_x^2\mathbf{I}_L$ , then (23) can be rewritten as

$$\text{MSE} = \sigma_v^2 + \sigma_x^2 D(\infty). \quad (24)$$

Thus, our work is to estimate  $D(\infty)$ . In (20), let  $n$  approach infinity, by observing the  $i$ th ( $0 \leq i < L$ ) element  $K_i(\infty)$  of the matrix  $\mathbf{K}(\infty)$ , one obtains

$$\begin{aligned} K_i(\infty) &= (1 - 2\mu\sigma_x^2 + 2\mu^2\sigma_x^4)K_i(\infty) + \mu^2\sigma_x^4 D(\infty) \\ &\quad + \mu^2\sigma_v^2\sigma_x^2 - 2\rho(1 - \mu\sigma_x^2)\text{E}\{\Delta_i m_i(\infty)\} \\ &\quad + \rho^2\text{E}\{m_i^2(\infty)\} \quad \forall 0 \leq i < L, \end{aligned} \quad (25)$$

With reference to (16), it is obvious that

$$\text{E}\{m_i^2(\infty)\} = \begin{cases} \text{MSE} & a < |w_i(\infty)| \leq b; \\ 0 & \text{elsewhere.} \end{cases} \quad (26)$$

To derive  $\text{E}\{\Delta_i(\infty)m_i(\infty)\}$ , by multiplying  $\mathbf{m}^T(n)$  on the right of each item of (15) and taking the expectation value on both sides as well as letting  $n$  approach infinity it yields

$$\text{E}\{\mathbf{\Delta}(\infty)\mathbf{m}^T(\infty)\} = -\frac{\rho}{\mu\sigma_x^2}\text{E}\{\mathbf{m}(\infty)\mathbf{m}^T(\infty)\}. \quad (27)$$

Thus, when  $|w_i(\infty)| \leq a$  or  $|w_i(\infty)| > b$  ( $0 \leq i < L$ ), it has

$$K_i(\infty) = \frac{\mu\sigma_x^2 D(\infty)}{2(1 - \mu\sigma_x^2)} + \frac{\mu^2\sigma_x^2\sigma_v^2}{2\mu\sigma_x^2(1 - \mu\sigma_x^2)}, \quad (28)$$

when  $a < |w_i(\infty)| \leq b$  ( $0 \leq i < L$ ), it has

$$K_i(\infty) = \frac{1}{2\mu\sigma_x^2(1 - \mu\sigma_x^2)} \left[ \mu^2\sigma_x^4 D(\infty) + \mu^2\sigma_v^2\sigma_x^2 + \left( \frac{2}{\mu\sigma_x^2} - 1 \right) \rho^2 \text{MSE} \right]. \quad (29)$$

According to Assumption (3), and combining (7), we have

$$\begin{aligned}
P_A &= \mathcal{P}\{a \leq |w_i(n)| \leq b\} \\
&= 2[F(b) - F(a)] \\
&= \left\{ \Theta \left[ 1/\beta, \left(\frac{|b|}{\lambda}\right)^\beta \right] - \Theta \left[ 1/\beta, \left(\frac{|a|}{\lambda}\right)^\beta \right] \right\} / \Gamma(1/\beta),
\end{aligned} \tag{30}$$

where  $P_A$  denotes the probability that the coefficients of adaptive filter will be attracted. On the other hand, the probability that they will not be attracted is  $P_{NA} = 1 - P_A$ . By combining (28) and (29) and summing up all the diagonal items of matrix  $\mathbf{K}(n)$ , it yields

$$D(\infty) = \frac{L \left[ \mu^2 \sigma_v^2 \sigma_x^2 + \rho^2 P_A \left( \frac{2}{\mu \sigma_x^2} - 1 \right) \text{MSE} \right]}{\mu \sigma_x^2 [2 - \mu \sigma_x^2 (L + 2)]}, \tag{31}$$

Combining (24) and (31), finally one has

$$\text{MSE} = \frac{\mu \sigma_v^2 [2 - \mu \sigma_x^2 (L + 2)] + L \mu^2 \sigma_x^2 \sigma_v^2}{2\mu - \mu^2 \sigma_x^2 (L + 2) - \rho^2 P_A L \left( \frac{2}{\mu \sigma_x^2} - 1 \right)}. \tag{32}$$

If  $\rho = 0$ , equation (32) is the same with MSE of standard LMS algorithm [3],

$$\text{MSE}_{\text{LMS}} = \frac{(2 - 2\mu \sigma_x^2) \sigma_v^2}{2 - \mu \sigma_x^2 (L + 2)}. \tag{33}$$

### 4.3 Parameter Analysis

The performance of the proposed algorithm is largely affected by the balancing parameter  $\rho$  and the thresholds  $a$  and  $b$ .

According to (3) and (4), it can be seen that the parameter  $\rho$  determines the importance of the  $l_1$  norm and the intensity of zero-point attraction. In a certain range, a larger  $\rho$ , which indicates stronger attraction intensity, will improve the convergence performance by forcing small coefficients toward zero with fewer iterations. However, according to (32), a larger  $\rho$  also results in a larger steady-state misalignment. So the parameter  $\rho$  can balance the tradeoff between adaptation speed and quality. Moreover, the optimal parameter  $\rho$  empirically satisfies  $\rho \ll \mu \ll 1$ . By analyzing steady-state MSE in (32) under such circumstance, it can be seen that

$$\left| \rho^2 P_A L \left( \frac{2}{\mu \sigma_x^2} - 1 \right) \right| \ll |2\mu - \mu^2 \sigma_x^2 (L + 2)|. \tag{34}$$

According to (34), the influence of the last term in the denominator of (32) can be ignored, which means that the steady-state MSE of the proposed algorithm is approximately the same with standard LMS for near sparse systems identification. The same conclusion can also be drawn from (11) intuitively: when the adaptation reaches steady-state, the small  $e(n)$  renders the value of  $\rho|e(n)|$  trivial compared to  $\mu$ , letting the relaxation of zero-point

attraction constraint. On the other hand, with large  $e(n)$  in the beginning and the process of adaptation which indicates larger zero-point attraction force, the zero-point attractor adjusts the small taps more effectively than ZA-LMS, forcing them to zero with fewer iterations, which accelerate the convergence rate significantly.

The thresholds  $a$  and  $b$  determine the zero-point attraction range together. The parameter  $a$  is set to avoid forcing all small coefficients to exact zero, it is suggested to be set as the mean amplitude of those near zero coefficients of the real system. Specifically, for exact sparse systems, as most coefficients of the unknown system are exactly zero except some large ones, accordingly  $a = 0$  is set to force most small coefficients to exact zero. For exact sparse systems contaminated by small Gaussian white noise,  $a$  should be set as the standard deviation of the noise. For near sparse systems generated by GGD, as the mean amplitude of the small coefficient is hard to derive, we empirically choose  $a$  for the proposed algorithm. As a small sparsity indicator  $\beta$  in GGD usually means smaller mean amplitude of the small coefficient, we choose smaller  $a$  when  $\beta$  is smaller. According to the simulations,  $a$  is chosen in the range  $1 \times 10^{-3}$  to  $1 \times 10^{-2}$  for GGD with  $\beta$  varying from 0.05 to 0.5. The parameter  $b$  is chosen to reduce the unnecessary attraction of large coefficients in ZA-LMS, therefore, empirically any constant  $b$ , which is much larger than the deviation of small coefficients and much smaller than infinity, should be appropriate. Various simulations demonstrate that the parameter  $b$  can be set as a constant around 1 for most near sparse systems. This choice of  $b$  is quite standard for most applications.

## 5 Simulations

In this section, first we demonstrate the convergence performance of our proposed algorithm on two near sparse systems and a exact sparse system in Experiment 1-4, respectively. Second, Experiment 5-7 are designed to verify the derivation and discussion in Section IV. Besides the proposed algorithm, Standard NLMS, ZA-LMS, IPNLMS [6] and IIPNLMS [7] are also simulated for comparison. To be noticed, the normalized variants of ZA-LMS and the proposed algorithm are adopted to guarantee a fair comparison in all experiments except the fourth, where DWZA-LMS is simulated to verify the theoretical analysis result.

The first experiment is to test the convergence and tracking performance of the proposed algorithm on near sparse system driven by Gaussian white signal and correlated input, respectively. The unknown system is generated by GGD which has been shown in Fig. 1. b with filter length  $L = 100$ , it is initialized randomly with  $\beta = 0.05$  and  $\sigma_g^2 = 1$ . For the white and correlated input, the system is regenerated following the same distribution after 1700 and 4500 iterations, respectively. For the white input, the signal  $x(n)$  is generated by white Gaussian noise with power  $\sigma_x^2 = 1$ . For the correlated input, the signal is generated by white Gaussian noise  $y(n)$  driving a first-order Auto-Regressive (AR) filter,  $x(n) = 0.8x(n-1) + y(n)$ , and  $x(n)$  is normalized. Besides, the power of observation noise is  $\sigma_v^2 = 1 \times 10^{-4}$  for both input. The five algorithms are simulated 100 times respectively with

Table 1: Comparison of computational complexity of IPNLMS, IIPNLMS and DWZA-NLMS

Algorithms	Multiplies			Comparisons
	Convolution	Tap update	Total	
IPNLMS	$L$	$3L$	$4L$	0
IIPNLMS	$L$	$3L$	$4L$	$4L$
ZA-NLMS	$L$	$2L$	$3L$	0
DWZA-NLMS	$L$	$(1 + P_1)L$	$(2 + P_1)L$	$2L$

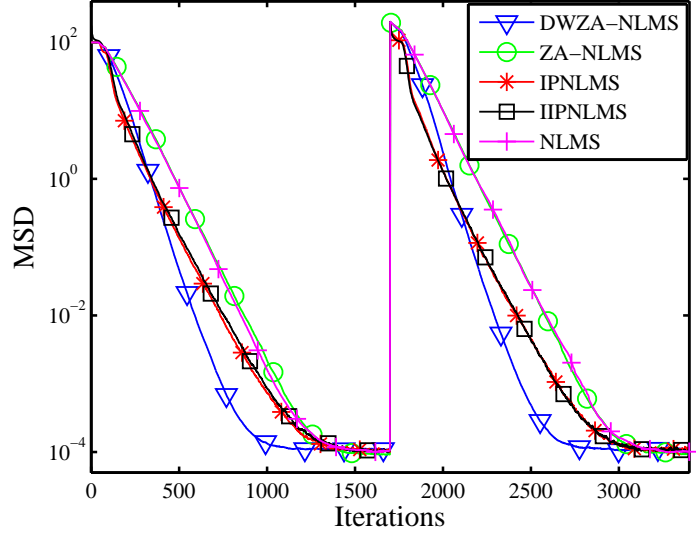
Where  $P_1 \in [0, 1]$  denotes the fraction of coefficients in the attracting range of DWZA-NLMS.

parameter  $\mu = 1$  in both cases. The other parameters are as follows

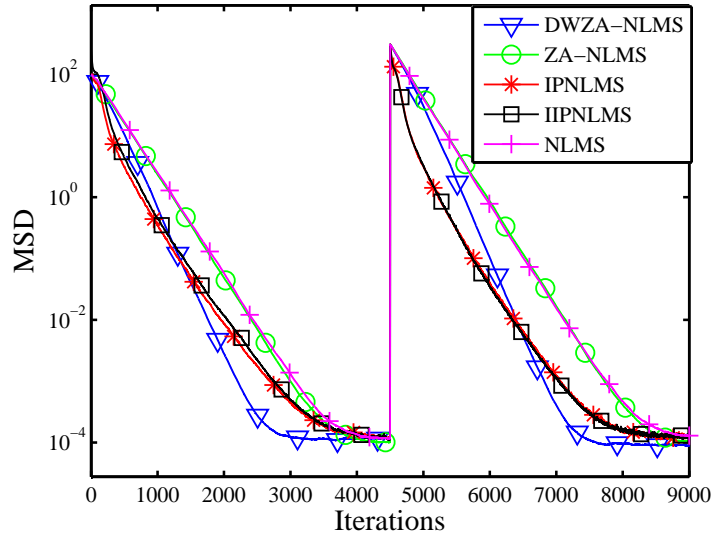
- IPNLMS and IIPNLMS with white input:  $\alpha_P = -0.5$ ,  $\rho = 0.4$ ,  $\alpha_{P1} = -0.5$ ,  $\alpha_{P2} = 0.5$ ,  $\Gamma = 0.1$ ;
- IPNLMS and IIPNLMS with correlated input:  $\alpha_P = -0.5$ ,  $\rho = 0.2$ ,  $\alpha_{P1} = -0.5$ ,  $\alpha_{P2} = 0.5$ ,  $\Gamma = 0.1$ ;
- ZA-NLMS and DWZA-NLMS with white input:  $\rho_{ZA} = 3 \times 10^{-4}$ ,  $\rho_{DWZA} = 6 \times 10^{-2}$ ,  $a = 1 \times 10^{-3}$ ,  $b = 0.8$ .
- ZA-NLMS and DWZA-NLMS with correlated input:  $\rho_{ZA} = 3 \times 10^{-4}$ ,  $\rho_{DWZA} = 3 \times 10^{-2}$ ,  $a = 1 \times 10^{-3}$ ,  $b = 0.8$ .

All the parameters are particularly selected to keep their steady-state error in the same level. The MSD of these algorithms for both white and correlated input are shown in Fig. 4(a) and Fig. 4(b), respectively. The simulation results show that all algorithms converge more slowly in the color input driven scenario than in the white noise driven case. However, the ranks or their relative performances are similar and the proposed algorithm reaches the steady state first with both white and correlated input. On the other hand, the performance of ZA-NLMS degenerate to standard NLMS as the system is near sparse. Furthermore, the computational complexity of the proposed algorithm is also smaller compared with improved PNLMS algorithms (Table 1). Besides, when the system is changed abruptly, the proposed algorithm also reaches the steady-state first in both cases.

The second experiment is to demonstrate the proposed algorithm on near-sparse systems other than GGD. The near-sparse system with 100 taps is generated in the following manner. First, 8 large coefficients following Gaussian distribution  $\mathcal{N}(0, 1)$  are generated, where all



(a) White Input



(b) Correlated Input

Figure 4: Comparisons of convergence rate and tracking ability of five different algorithms for sparse system identification driven by (a) white input and (b) correlated input, respectively.

their tap positions follow Uniform distribution. Second, white Gaussian noise with variance  $\sigma_h^2 = 1 \times 10^{-4}$  is added to all taps, enforcing the system to be near-sparse. The signal  $x(n)$  is generated by white Gaussian noise with power  $\sigma_x^2 = 1$ . Five algorithms, the same as in Experiment 1, are simulated 100 times respectively with parameter  $\mu = 1$ . The parameters are set to the same values as in the white input case of Experiment 1 except  $a = 1 \times 10^{-2}$  for the proposed algorithm. From Fig. 5, we can conclude that the proposed algorithm

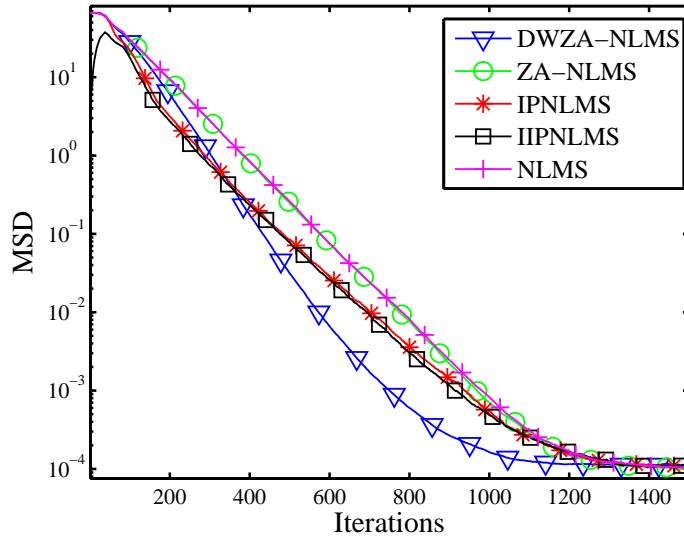


Figure 5: Comparisons of convergence rate on NLMS, ZA-NLMS, IPNLMS, IIPNLMS, and the proposed algorithm on near sparse system.

reaches the steady-state first in such near sparse system.

The third experiment demonstrates the effectiveness of the proposed modification on ZA-LMS algorithm on near sparse system. The system and the signal are generated in the same way as in Experiment 2. The proposed DZA-NLMS, DWZA-NLMS are compared with ZA-NLMS for the system with 100 simulations. The step length for all algorithms are set as  $\mu = 1$ . We particularly choose parameters  $\rho_{DZA} = 0.05$  and  $\rho_{ZA} = 6.5 \times 10^{-4}$  for DZA-NLMS and ZA-NLMS to ensure their steady-state mean square error in the same level. We set  $\rho_{DWZA} = 0.05$  for DWZA-NLMS algorithm for a fair comparison with DZA-NLMS. The parameter  $a$  and  $b$  in DWZA-NLMS are chosen as  $a = 0.01$  and  $b = 0.8$ , respectively.

From Fig. 6, we can see that with the dynamic zero-point attractor DZA-NLMS convergences faster than ZA-NLMS. By adding another window constraint on the zero-point attractor, The DWZA-NLMS not only preserves the property of fast convergence of DZA-NLMS, but also shows smaller steady-state mean square error than both ZA-NLMS and DZA-NLMS.

The fourth experiment shows that the proposed improvement is still effective on the exact sparse system identification. The unknown system is shown in Fig. 1. a, where filter length  $L = 100$ . Besides, 8 large coefficients is uniformly distributed and generated by Gaussian distribution  $\mathcal{N}(0, 1)$ , and all other tap coefficients are exactly zero. The input signal is generated by white Gaussian noise with power  $\sigma_x^2 = 1$ , and the power of observation noise is  $\sigma_v^2 = 1 \times 10^{-4}$ . The proposed algorithm is compared with ZA-NLMS and standard NLMS, where each algorithm is simulated 100 times with 2000 iterations. The step size  $\mu$  is set to 0.65 for NLMS, and  $\mu = 1$  for both the ZA-NLMS and the proposed algorithms.

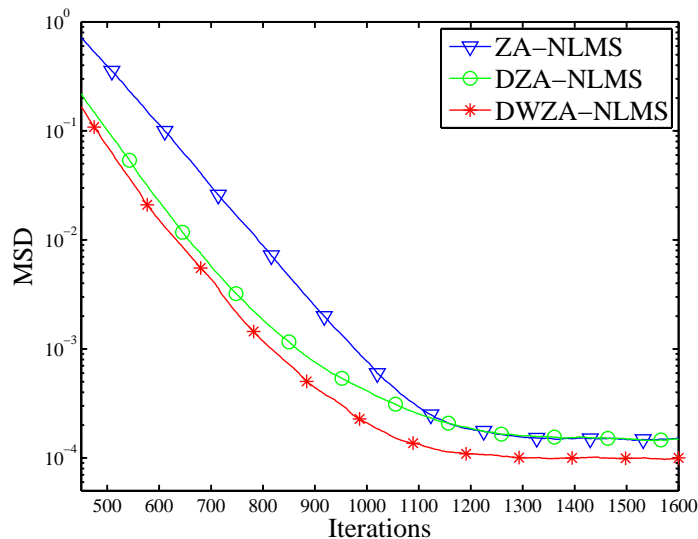


Figure 6: Comparisons of convergence speed on ZA-NLMS, DZA-NLMS and DWZA-NLMS.

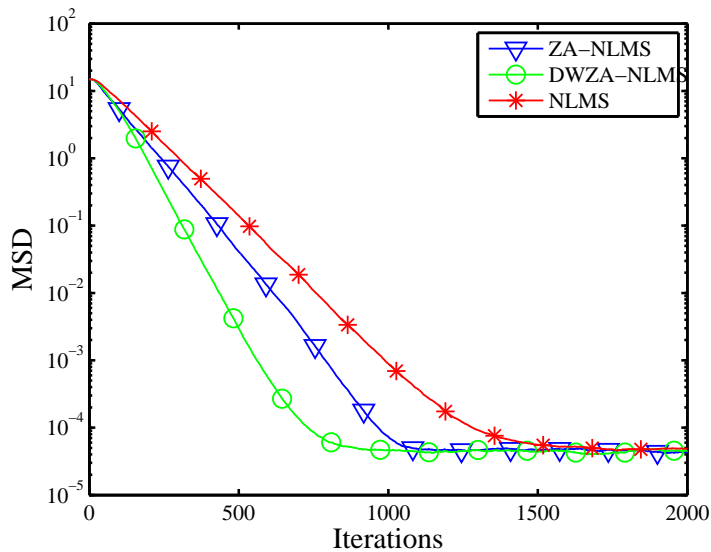


Figure 7: Comparisons of convergence rate on NLMS, ZA-NLMS, and the proposed algorithm on exact sparse system.

The parameter  $\rho_{ZA}$  is set to  $6 \times 10^{-4}$  and  $\rho_{DWZA} = 6 \times 10^{-2}$ . All the parameters are chosen to make sure that the steady-state error is the same for comparison. According to Fig. 7, it can be seen that the convergence performance is also improved compared with ZA-NLMS via the proposed method on exact sparse system, thus the proposed improvement on the algorithms are robust.

The fifth experiment is to test the sensitivity to sparsity of the proposed algorithm. All conditions are the same with the first experiment except the sparsity. The parameter

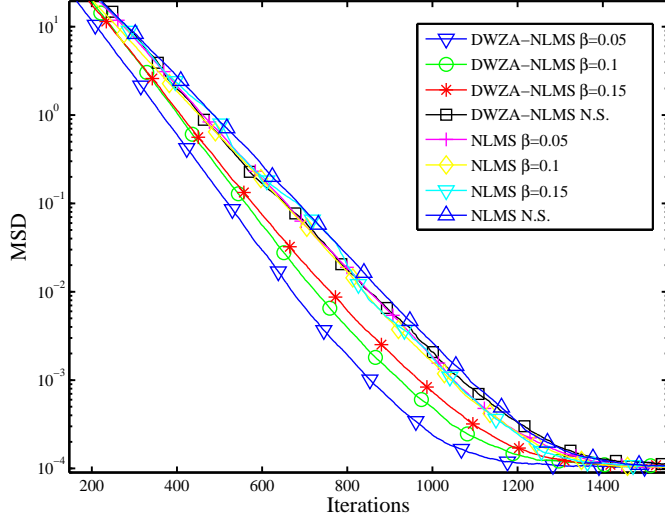


Figure 8: Comparisons of convergence rate of proposed algorithm for different sparsity.

$\beta$  is selected as 0.05, 0.1, and 0.15, respectively. Besides, we also compared our algorithm when the system is non-sparse which is generated by Gaussian distribution. For each  $\beta$ , both the proposed algorithm and NLMS are simulated 50 times with 1700 iterations. The step size  $\mu$  is 1 for both algorithms, and  $\rho = 4 \times 10^{-2}$ ,  $a = 1 \times 10^{-2}$ ,  $b = 0.8$  for the proposed algorithm. The simulated MSD curves are shown in Fig. 8. The steady-state MSD remains approximately the same for the proposed algorithm with varying parameter  $\beta$  which denotes the sparsity of systems, meanwhile the convergence rate decreases as the sparsity decreases. For the non-sparse case, our algorithm degenerates and shows similar behavior with standard NLMS. However, for each  $\beta$  the proposed algorithm is never slower than standard NLMS. It should be noticed that NLMS is independent on the system sparsity and behaves similar when  $\beta$  varies.

The sixth experiment is to test the steady-state MSE with different parameters. The coefficients of unknown system follows GGD with filter length  $L = 100$ , the sparsity and variance are chosen as  $\sigma_g^2 = 1$  and  $\beta = 0.1$ , respectively. The input is generated by white Gaussian noise with normalized power  $\sigma_x^2 = 1$ , and the power of observation noise is  $1 \times 10^{-4}$ . Under such circumstance, the steady-state MSD is tested. Here  $a = 1 \times 10^{-2}$  and  $b = 0.8$  are set for each simulation. The step size  $\mu$  is varied from 0 to  $1.1 \times 10^{-3}$  for given  $\rho = 2 \times 10^{-4}$ . And  $\rho$  is changed from 0 to  $1 \times 10^{-3}$  for given  $\mu = 1 \times 10^{-2}$ . Fig. 9 and Fig. 10 show that the analytical results accord with the simulated ones of different parameters for variable values. Specifically, in Fig. 9, the steady-state MSE goes up as the step size increases, whose trend is the same with standard LMS. Fig. 10 shows that the analytical steady-state MSE matches with simulated one as the parameter  $\rho$  gets larger, which verifies the result in Section V.



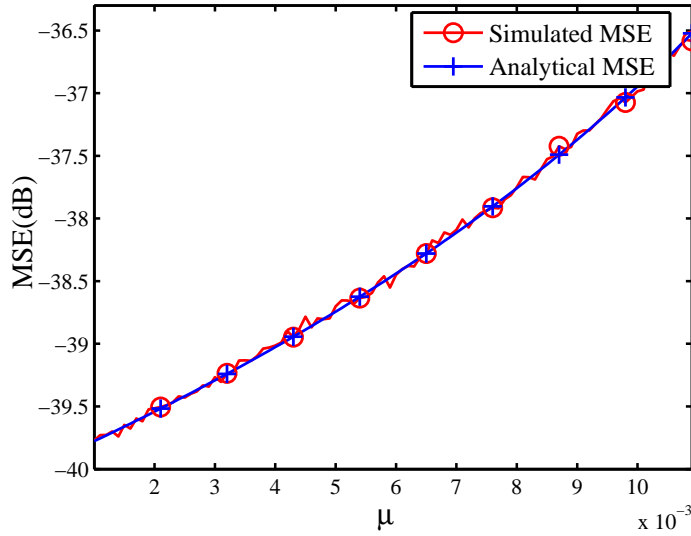


Figure 9: The steady-state MSD of the proposed algorithm with different step size

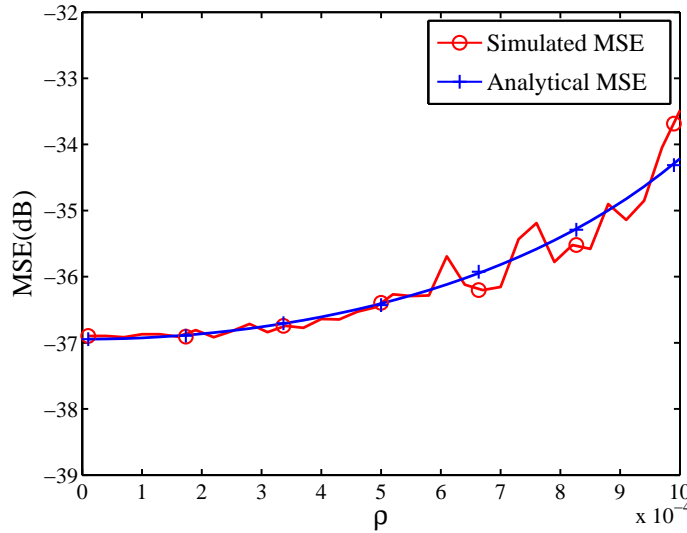
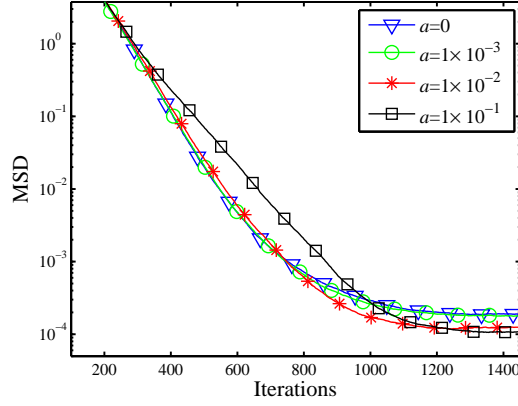
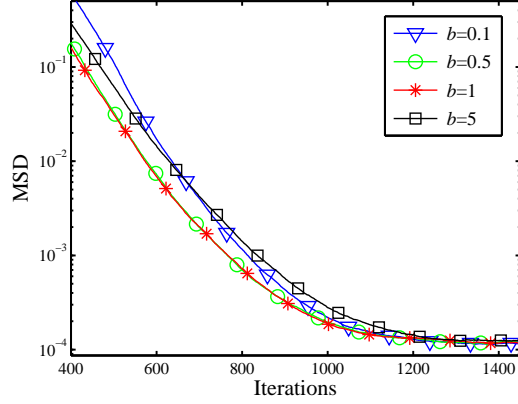


Figure 10: The steady-state MSD of the proposed algorithm algorithm with different parameters  $\rho$

The seventh experiment is designed to test the behavior of the proposed algorithm with respect to different parameters of  $a$  and  $b$ . All conditions of the proposed algorithm are the same with the second experiment except the parameters  $a$  and  $b$ . First, we set  $b = 0.8$  and vary  $a$  as  $a = 0, 1 \times 10^{-3}, 1 \times 10^{-2}$ , and  $1 \times 10^{-1}$ . Second, we set  $a = 1 \times 10^{-2}$  and vary  $b$  as  $b = 0.1, 0.5, 1$ , and  $5$ . From Fig. 11(a), we can see that the optimal  $a$  is chosen as the variance of the small coefficients. Smaller  $a$  will result in larger misalignment, and larger  $a$  will cause slow convergence. From Fig. 11(b), we conclude that  $b = 1$  shows the best



(a) Fix  $b = 0.8$  while varying  $a$  from 0 to 0.1.



(b) Fix  $a = 1 \times 10^{-2}$  while varying  $b$  from 0.1 to 5.

Figure 11: The effect of parameters  $a$  and  $b$  on the proposed algorithm for near sparse system.

performance. Either too large or too small  $b$  will result in slower convergence.

## 6 Conclusion

In order to improve the performance of ZA-LMS for near sparse system identification, an improved algorithm, DWZA-LMS algorithm, is proposed in this paper by adding a window to the zero-point attractor in ZA-LMS algorithm and utilizing the magnitude of estimation error to weight the zero-point attractor. Such improvement can adjust the zero-point attraction force dynamically to accelerate the convergence rate with no computational complexity increased. In addition, the mean square convergence condition, steady-state MSE and parameter selection of the proposed algorithm are theoretically analyzed. Finally, computer simulations demonstrate the improvement of the proposed algorithm and effectiveness of the analysis.

## References

- [1] W. F. Schreiber, Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions, *Proceedings of the IEEE*, vol. 83, no. 6, pp. 958-981, 1995.
- [2] D. L. Duttweiler, Proportionate normalized least-mean-squares adaptation in echo cancelers, *IEEE Trans. on Speech Audio Processing*, vol. 8, no. 5, pp. 508-518, 2000.
- [3] B. Widrow and S. D. Stearns, *Adaptive signal processing*, New Jersey. Prentice Hall, 1985.
- [4] M. Abadi, J. Husoy, Mean-square performance of the family of adaptive filters with selective partial updates, *Signal Processing*, vol. 88, no. 8, pp. 2008-2018, 2008.
- [5] M. Godavarti and A. O. Hero, Partial update LMS algorithms, *IEEE Trans. on Signal Processing*, vol. 53, no. 7, pp. 2382-2399, 2005.
- [6] J. Benesty and S. L. Gay, An improved PNLMS algorithm, *Proc. IEEE ICASSP*, vol. 2, pp. 1881-1884, Minneapolis, USA, 2002.
- [7] P. A. Naylor, J. Cui, M. Brookes, Adaptive algorithms for sparse echo cancellation, *Signal Processing*, vol. 86, no. 6, pp. 1182-1192, 2006.
- [8] Y. Gu, Y. Chen, K. Tang, Network echo canceller with active taps stochastic localization, *Proceedings of ISCIT*, pp. 556-559, 2005.
- [9] Y. Li, Y. Gu, and K. Tang, Parallel NLMS filters with stochastic active taps and step-sizes for sparse system identification, *Proc. IEEE ICASSP*, vol. 3, pp. 109-112, Toulouse, France, 2006.
- [10] D. Donoho, Compressed sensing, *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289-1306, 2006.
- [11] J. Jin, Y. Gu, S. Mei, A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework, *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, pp. 409-420, Apr. 2010.
- [12] Y. Chen, Y. Gu, A. O. Hero, Sparse LMS for system identification, *Proc. IEEE ICASSP*, pp. 3125-3128, Taipei, Taiwan, 2009.
- [13] Y. Gu, J. Jin, S. Mei,  $l_0$  norm constraint LMS algorithm for sparse system identification, *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 774-777, 2009.
- [14] G. Su, J. Jin, Y. Gu, and J. Wang, Performance Analysis of  $l_0$  Norm Constraint Least Mean Square Algorithm, *IEEE Transactions on Signal Processing*, 60(5): 2223-2235, 2012.

- [15] K. Song, A globally convergent and consistent method for estimating the shape parameter of a generalized Gaussian distribution, *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 510-527, 2006.
- [16] S. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. on Pattern Recognition Machine Intel*, vol.11, no.7, pp. 674-693, 1989.
- [17] Y. Chen and N. C. Beaulieu, Novel low complexity estimators for the shape parameter of the generalized Gaussian distribution, *IEEE Trans. on Veh. Technol.*, vol. 58, no. 4, pp. 2067-2071, 2009.
- [18] H.C. Wu and J. Principe, Minimum entropy algorithm for source separation, *Proc. IEEE Midwest Symp. Syst. Circuits*, pp. 242-245, 1998.
- [19] K. Sharifi and A. Leon-Garcia, Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video, *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 5, pp. 52-56, 1995.
- [20] S. Gazor, W. Zhang, Speech probability distribution, *IEEE Signal Processing Letters*, vol. 10, no.7, pp. 204-207, 2003.
- [21] K.Shi and P.Shi, Convergence analysis of sparse LMS algorithms with l1-norm penalty based on white input signal, *Signal Processing*, vol.90, no. 12, pp.3289-3293, 2010.
- [22] S. Haykin, *Adaptive Filter Theory*, Chapter 9, pp. 390-405. Upper Saddle River, NJ: Prentice-Hall, 1996.
- [23] R.H. Kwong and E.W. Johnston, A variable step size LMS algorithm, *IEEE Trans. on Signal Processing*, vol. 40, pp. 1633-1642, 1992.