



A Robust M-Shaped Error Weighted Algorithms for Censored Regression

Feng Zhao^{1,2} · Haiquan Zhao^{1,2} · Wenyuan Wang^{1,2}

Received: 24 December 2018 / Revised: 10 June 2019 / Accepted: 10 June 2019 / Published online: 25 June 2019
© The Author(s) 2019

Abstract

In reality, the range of sensor response is limited in many sensor systems due to the saturation characteristics of the sensor. That is, the value exceeding the sensor response range is not observed. Using traditional adaptive algorithms to identify the system of this type may lead to the performance degradation. To address this problem, the censored regression algorithms have been proposed. However, when the mixed sub-Gaussian and super-Gaussian/impulsive noises occur, these algorithms may fail to work. To overcome these drawbacks, a family of robust M-shaped (FRMS) functions for censored regression (CR-FRMS) is proposed in this paper. When the system to be identified exhibits a certain degree of sparsity, the CR-FRMS algorithm cannot fully utilize the characteristics of the sparse system. Therefore, in this paper, proportionate FRMS (PFRMS) algorithm based on l_0 -norm constraint for censored regression (l_0 -CRPFRMS) is also proposed accordingly. The simulations using Gaussian white noise as the input signal and the non-Gaussian mixed noise as the background noise show that the proposed algorithm performs better than other algorithms.

Keywords A family of robust M-shaped (FRMS) functions · l_0 -Norm constraint · Proportionate algorithm · Censored regression · Adaptive filtering

✉ Haiquan Zhao
hqzhao@home.swjtu.edu.cn

Feng Zhao
xiianzhang@163.com

Wenyuan Wang
wenyuanwang@my.swjtu.edu.cn

¹ Key Laboratory of Magnetic Suspension Technology and Maglev Vehicle, Ministry of Education, Chengdu, China

² School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China

1 Introduction

Linear regression models are widely used for signal processing, communication, and many other yields, assuming that the observed data is fully available. To solve this model problem, the researchers have proposed a large number of adaptive algorithms, such as least mean square (LMS) algorithm [13], normalized LMS (NLMS) algorithm, affine projection algorithm (APA) [14, 30, 33, 42] and so on.

Unfortunately, the requirement for linear regression may not be usually met in many practical applications. In general, output data whose value exceeds the limit of the recording device cannot be observed [3, 15, 29]. In other words, the output data whose values lie in a certain range are available. This situation occurs in economics [4], statistics [11, 27], engineering applications [28, 26], and medical research [7, 31]. In some systems, due to sensors' saturation characteristics [1, 9, 43], it is not possible to collect complete data very efficiently. For example, in microphone array signal processing [2, 16], when the amplitude of speech signal exceeds a certain amplitude threshold of the microphones, it is cut off and the signal wave shape may be distorted to be flat because the signal's positive and negative peaks exceeding the threshold are lost, or censored. Actually, the censored regression can be seen as a nonlinear regression model which includes a saturated nonlinearity model and linear system [23, 38]. Since the output data of the censored regression may lose significant information, using the traditional algorithms to identify this type of model may result in biased and wrong estimates [22]. Recently, in an attempt to deal with the censored regression problem, numerous algorithms have been proposed, such as maximum likelihood (ML) methods [8], two-step estimator [15], least absolute deviation [29]. To solve online censored regression problems, Liu et al. [24] proposed the adaptive Heckman two-step algorithm (TSA) which significantly outperforms the conventional adaptive algorithms while the output data is censored.

As is well known, most of the adaptive algorithms are based on Gaussian noise environment. However, real-world signals often exhibit non-Gaussian properties. For example, in the beamforming, sub-Gaussian (light-tailed) signals are frequently encountered [17]. In some active noise control (ANC) applications, mechanical friction, vibration noise, and speech signal are super-Gaussian/impulsive (heavy-tailed) signals [39]. In the blind source separation (BSS), adaptive receivers with multiple antennas, and image denoising, the signal may be the mixed sub-Gaussian and super-Gaussian/impulsive signal [18–20, 32, 35, 40, 46].

In an attempt to improve the robustness of the algorithm in mixed non-Gaussian background noise environments, a family of robust M-shaped (FRMS) functions was applied to the adaptive algorithms [45].

Furthermore, when the system exhibits a certain degree of sparsity, the appeal algorithm cannot utilize the characteristics of the system. To deal with this issue, proportional algorithms [10] and multiple norm forms of zero-attraction algorithms, such as l_0 -norm, l_1 -norm, and l_p -norm [5, 12, 25, 34, 37, 39, 46], are applied to this type of system. In fact, the ideal sparse measure is the l_0 -norm that counts the number of nonzero components. Therefore, the l_0 -norm constraint is adopted in this paper.

In this paper, contributions are made as follows:

- (i) For the first time, a family of robust M-shaped algorithm for mixed non-Gaussian background noises under the censored regression model is proposed. The algorithm not only can effectively compensate the output signal, but also can deal with the effects of sub-Gaussian and super-Gaussian/impulsive noises.
- (ii) For the characteristics of sparse system, the l_0 -norm proportional FRMS algorithm under the censored regression model (l_0 -CRPFRMS) is also proposed for the first time.
- (iii) Simulation examples to demonstrate the performance of the proposed algorithm in non-Gaussian background noise environments.

2 Description and Preliminaries

2.1 Problem Formulation

Consider the following linear regression model

$$\hat{d}_n = \mathbf{u}_n^T \mathbf{w}_o + \eta_n + \xi_n \quad (1)$$

where \mathbf{w}_o is the unknown column vector of size, η_n denotes the background noise with zero mean and variance σ_i^2 . In this paper, sub-Gaussian and super-Gaussian noises are involved, and ξ_n represents the impulsive noise with zero mean. When the data \hat{d}_n and \mathbf{u} are completely observed, using some traditional methods, the unknown vector \mathbf{w}_o can be identified easily. However, the data \hat{d}_n may be not completely observed in some practical application. The censored output d_n can be formulated by

$$d_n = (\mathbf{u}_n^T \mathbf{w}_o + \eta_n)_+ = (\hat{d}_n)_+ \quad (2)$$

where $(\hat{d}_n)_+ = \max\{0, \hat{d}_n\}$.

Remark 1 In this paper, although the left-censored to 0 is only applied to the algorithm, the right-censored and both-sides-censored cases can also be applied in the algorithm, and the processing method is similar. In particular, if the output \hat{d}_n is left-censored or right-censored to a constant c , the censored output d_n can be expressed as

$$d_n = \max\{c, \hat{d}_n\} = (\hat{d}_n - c)_+ + c \quad (3)$$

and

$$d_n = \min\{c, \hat{d}_n\} = c - (c - \hat{d}_n)_+ \quad (4)$$

respectively.

2.2 Review of FRMS Algorithm

In [23], the authors divide the existing error nonlinear adaptive algorithms based on LMS into three categories, V-shaped, Λ -shaped, and M-shaped algorithms. After comparing the advantages and disadvantages of each algorithm, the FRMS function was proposed, where its weighted function is

$$f(e_n) = \frac{|e_n|^p}{\zeta + |e_n|^{p+1}}, p > 0 \quad (5)$$

where $\zeta > 0$ is the parameter. For instance, when $\zeta \rightarrow 0$, it will be a Λ -shaped algorithm. (Λ -shaped algorithm is applicable to the super-Gaussian noise than that for sub-Gaussian noise environment.) When $\zeta \rightarrow \infty$, it is a V-shaped algorithm. (V-shaped algorithm is more suitable for sub-Gaussian noise.) In the proposed robust M-shaped function, its denominator is with an order higher than the numerator. Using the gradient descent, this algorithm essentially solves the stochastic cost function

$$J(e_n) = \int_0^{e_n} \frac{x|x|^p}{\zeta + |x|^{p+1}} dx \quad (6)$$

which is positive semi-definite for $\zeta > 0; p > 0$. Thus, the proposed robust M-shaped algorithm will not suffer from the local minimum problem.

3 Proposed New Algorithm

3.1 Proposed CR-FRMS Algorithm

Since the observations in this paper are assumed to be censored, the FRMS algorithm would obtain a biased estimate which is called sample selection bias [34, 44]. In other words, when $\hat{d}_n < 0$, the data \hat{d}_n is missing, which leads to the bias and the inequality $E[d_n | \mathbf{u}_n^T \mathbf{w}_o] \neq \mathbf{u}_n^T \mathbf{w}_o$. To compensate the bias, firstly, it is necessary to correct the sample selection bias using FRMS in censored observations, which is inspired by the Heckman two-stage approach [15]. Due to the left-censored property of d_n , only positive values of d_n can be correctly obtained. Recalling (1) and noting that the background noise η_n and impulsive noise ξ_n are both zero-mean signals, the expectation of d_n under the condition $d_n > 0$ can be expressed by

$$\begin{aligned} & E[d_n | \mathbf{u}_n^T \mathbf{w}_o, d_n > 0] \\ &= E[\hat{d}_n | \mathbf{u}_n^T \mathbf{w}_o, \hat{d}_n > 0] \\ &= \mathbf{u}_n^T \mathbf{w}_o + E[\eta_n + \xi_n | \eta_n + \xi_n > -\mathbf{u}_n^T \mathbf{w}_o]. \end{aligned} \quad (7)$$

Since the impulsive noise occurs with a very low probability, the following approximation is reasonable,

$$E[\eta_n + \xi_n | \eta_n + \xi_n > -\mathbf{u}_n^T \mathbf{w}_o] \approx E[\eta_n | \eta_n > -\mathbf{u}_n^T \mathbf{w}_o] \quad (8)$$

Table 1 Probability density function and distribution function of three different noises

Noise type	Probability density function	Cumulative distribution function
Gaussian	$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\Phi(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$
Uniform	$\varphi(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{else} \end{cases}$	$\Phi(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$
Laplace	$\varphi(x) = \frac{1}{2\lambda} e^{-\frac{ x-\mu }{\lambda}}$	$\Phi(x) = \int_{-\infty}^x \frac{1}{2\lambda} e^{-\frac{ y-\mu }{\lambda}} dy$

Before calculating the last term of (5), the following lemma is introduced.

Lemma *The condition expectation $E[x|x > -c]$ satisfies*

$$E[x|x > -c] = \frac{\varphi(c)}{\Phi(c)} = \Omega(c). \tag{9}$$

Proof See “Appendix” for detail.

Using the lemma, we have

$$E[\eta_n | \eta_n > -\mathbf{u}_n^T \mathbf{w}_o] = \sigma_n \Omega(\mathbf{u}_n^T \boldsymbol{\alpha}) \tag{10}$$

with $\varphi(\cdot)$ and $\Phi(\cdot)$ given in Table 1. In addition, the vector $\boldsymbol{\alpha}$ is given by

$$\boldsymbol{\alpha} = \frac{\mathbf{w}_o}{\sigma_i} \tag{11}$$

Then, using (9) and the probability theory yields

$$\begin{aligned} E[d_n | \mathbf{u}_n^T \mathbf{w}_o] &= Pr(d_n > 0) E[d_n | \mathbf{u}_n^T \mathbf{w}_o, d_n > 0] \\ &= \Phi(\mathbf{u}_n^T \boldsymbol{\alpha}) \mathbf{u}_n^T \mathbf{w}_o + \sigma_i \varphi(\mathbf{u}_n^T \boldsymbol{\alpha}) \end{aligned} \tag{12}$$

where the second equation comes from the fact that the probability of $d_n > 0$ is equal to $\Phi(\mathbf{u}_n^T \boldsymbol{\alpha})$, i.e., $Pr(d_n > 0) = \Phi(\mathbf{u}_n^T \boldsymbol{\alpha})$. According to (12), the censored regression model (2) can be expressed as:

$$d_n = \Phi(\mathbf{u}_n^T \boldsymbol{\alpha}) \mathbf{u}_n^T \mathbf{w}_o + \sigma_i \varphi(\mathbf{u}_n^T \boldsymbol{\alpha}) + v_n \tag{13}$$

where v_n is the random variable with zero mean, i.e.,

$$E[v_n | \mathbf{u}_n^T \mathbf{w}_o] = 0. \tag{14}$$

Since the algorithm based on MSE criterion cannot estimate \mathbf{w}_o correctly, the cost function based on FRMS is adopted

$$\zeta_{\text{FRMS}} = E \left(\frac{|\varepsilon_n(\mathbf{w}, \sigma_n, \boldsymbol{\alpha})|^p}{\zeta + |\varepsilon_n(\mathbf{w}, \sigma_n, \boldsymbol{\alpha})|^{p+1}} \right) \tag{15}$$

where

$$\varepsilon_n(\mathbf{w}, \sigma_n, \boldsymbol{\alpha}) = d_n - \Phi(\mathbf{u}_n^T \boldsymbol{\alpha}) \mathbf{u}_n^T \mathbf{w} + \sigma_i \varphi(\mathbf{u}_n^T \boldsymbol{\alpha}). \tag{16}$$

Obviously, to estimate \mathbf{w}_o , the estimation for $\boldsymbol{\alpha}$ and σ_n , should also be available. In the sequel, an indicator variable a_n is introduced to estimate $\boldsymbol{\alpha}$,

$$a_n = \begin{cases} 1, & \text{if } d_n > 0 \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

The probabilities of a_n is expressed as

$$Pr(a_n = 1) = \Phi(\mathbf{u}_n^T \boldsymbol{\alpha}) \tag{18}$$

$$Pr(a_n = 0) = \Phi(-\mathbf{u}_n^T \boldsymbol{\alpha}). \tag{19}$$

Then, the following optimization problem is considered to estimate $\boldsymbol{\alpha}$ [22],

$$\hat{\boldsymbol{\alpha}} = \arg \max_{\tilde{\boldsymbol{\alpha}}} \Gamma(\tilde{\boldsymbol{\alpha}}) = \arg \max_{\tilde{\boldsymbol{\alpha}}} E[\Gamma_n(\tilde{\boldsymbol{\alpha}})] \tag{20}$$

where

$$\Gamma(\tilde{\boldsymbol{\alpha}}) = E[\Gamma_n(\tilde{\boldsymbol{\alpha}})] \tag{21}$$

and

$$\Gamma_n(\tilde{\boldsymbol{\alpha}}) = \log(Pr(d_n | \mathbf{u}_n, \boldsymbol{\alpha})) \tag{22}$$

with

$$Pr(d_n | \mathbf{u}_n, \boldsymbol{\alpha}) = \Phi(\mathbf{u}_n^T \boldsymbol{\alpha})^{a_n} \Phi(-\mathbf{u}_n^T \boldsymbol{\alpha})^{1-a_n}. \tag{23}$$

In the sequel, using the steepest ascent principle yields [22]

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_n &= \hat{\boldsymbol{\alpha}}_{n-1} + \mu \frac{\partial \Gamma_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \Big|_{\hat{\boldsymbol{\alpha}}_{n-1}} \\ &= \hat{\boldsymbol{\alpha}}_{n-1} + \mu [a_n \mathcal{Q}(\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_{n-1}) \mathbf{u}_n - (1 - a_n) \mathcal{Q}(-\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_{n-1}) \mathbf{u}_n]. \end{aligned} \tag{24}$$

Then, the estimation for σ_n is considered. Using the decent method and the cost function $f(\varepsilon_n(\mathbf{w}, \sigma_n, \boldsymbol{\alpha}))$ with respected to $\sigma_{n,i-1}$, we have

$$\begin{aligned} \hat{\sigma}_{i,n} &= \hat{\sigma}_{i,n-1} - \frac{\mu}{2} \frac{\partial f(\varepsilon_n(\mathbf{w}, \sigma_n, \boldsymbol{\alpha}))}{\partial \sigma_i} \Big|_{\hat{\sigma}_{i,n-1}} \\ &= \hat{\sigma}_{i,n-1} + \mu \varphi(\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_n) f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_n)) \varepsilon_n(\mathbf{w}, \sigma_{n,i-1}, \hat{\boldsymbol{\alpha}}_n) \end{aligned} \tag{25}$$

where

$$f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_{n-1})) = \frac{|d_n - \Phi(\mathbf{u}_n^T \boldsymbol{\alpha}) \mathbf{u}_n^T \mathbf{w} + \sigma_i \varphi(\mathbf{u}_n^T \boldsymbol{\alpha})|^p}{\zeta + |d_n - \Phi(\mathbf{u}_n^T \boldsymbol{\alpha}) \mathbf{u}_n^T \mathbf{w} + \sigma_i \varphi(\mathbf{u}_n^T \boldsymbol{\alpha})|^{p+1}}. \quad (26)$$

Similarly, the weight update formula can be obtained by the gradient descent method

$$\begin{aligned} \mathbf{w}_n &= \mathbf{w}_{n-1} - \mu \frac{\partial \zeta_{FRMS}(\mathbf{w})}{\partial \mathbf{w}} \Big|_{\mathbf{w}_{n-1}} \\ &= \mathbf{w}_{n-1} + \mu \Phi_n(\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_n) \mathbf{u}_n^T f(\varepsilon_n(\mathbf{w}, \sigma_{i,n}, \hat{\boldsymbol{\alpha}}_n)) \varepsilon_n(\mathbf{w}, \sigma_{i,n}, \hat{\boldsymbol{\alpha}}_n) \end{aligned} \quad (27)$$

where

$$\zeta_{FRMS}(\mathbf{w}) = f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_{n-1})). \quad (28)$$

3.2 Proposed l_0 -CRPFRMS algorithm

In many practical applications, such as digital TV transmission channels and echo paths, the systems that need to be identified are sparse. However, the algorithm in Sect. 3.1 does not make full use of the characteristics of the sparse system. Therefore, this section proposes a l_0 -norm proportional FRMS algorithm based on the censored regression model (l_0 -CRPFRMS).

Sparse system is defined whose impulse response contains many near-zero coefficients and few large ones. In [36], the author first proposed a proportional idea, the (PNLMS) algorithm. Its iteration formula is as follows:

$$\hat{y}(k) = \sum_{n=0}^{N-1} \hat{h}_n(k) x(k-n) \quad (29)$$

$$e(k) = y(k) - \hat{y}(k) \quad (30)$$

$$l_\infty(k) = \max\{|\hat{h}_0(k)|, \dots, |\hat{h}_{N-1}(k)|\} \quad (31)$$

$$l'_\infty(k) = \max\{\delta, l_\infty(k)\} \quad (32)$$

$$g_n(k) = \max\{\rho l'_\infty(k), |\hat{h}_n(k)|\} \quad (33)$$

$$\bar{g}(k) = \frac{1}{N} \sum_{n=0}^{N-1} g_n(k) \quad (34)$$

$$\hat{\sigma}_x^2(k) = \frac{1}{N} \sum_{n=0}^{N-1} x^2(k-n) \quad (35)$$

$$\hat{h}_n(k+1) = \hat{h}_n(k) + \frac{\mu}{N} \frac{g_n(k)}{\bar{g}(k)} \frac{e(k)x(k-n)}{\hat{\sigma}_x^2(k)}. \quad (36)$$

where $x(\cdot)$ is input signal, $\hat{y}(\cdot)$ is output signal and $y(\cdot)$ is expected signal, and $\hat{h}_n(\cdot)$ is tap weight. The parameters ρ and δ affect small-signal regularization. In (32), ρ prevents $\hat{h}_n(k + 1)$ from stalling when it is much smaller than the largest coefficient and δ regularizes the updating when all coefficients are zero at initialization. It can be seen from (31–34, 36) that when the tap weight is closer to zero, the $g_n(k)/\bar{g}(k)$ item will become smaller and smaller. On the contrary, when the tap weight is far from zero, more energy will be gained in iteration. In this way, the aim of proportional algorithm is achieved, that is, the convergence speed of the algorithm is accelerated without changing the steady-state mean square deviation (MSD).

In order to further accelerate the convergence speed of the algorithm better, this section adds another method for sparse systems, i.e., l_0 -norm constraint [12, 34], on the premise of adding proportional algorithm. As the name implies, it is to add a $\gamma\|\mathbf{w}\|_0$ to the original cost function. In this section, the cost function is changed into

$$\vartheta(n) = \int_0^{e_n} \frac{x|x|^p}{\varsigma + |x|^{p+1}} dx + \gamma\|\mathbf{w}_n\|_0. \tag{37}$$

where $\gamma > 0$ is a factor to balance the new penalty and the estimation error. In [12], in order to reduce the computational complexity, the author firstly approximates $\|\mathbf{w}_n\|_0$ to $\sum_i^{L-1} (1 - e^{-\beta|w_n(i)|})$, then uses gradient descent method to derive the cost function, and then uses the Taylor formula to perform a first-order expansion on the zero attracting term in the iterative formula, and finally obtains the weight update formula with lower computational complexity,

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \text{gradient correction} + \text{zero attraction} \tag{38}$$

where zero attraction means $\kappa f_\beta(\mathbf{w}_n) \cdot \kappa = \mu\gamma$ is a positive constant and $f_\beta(\cdot)$ is defined as

$$f_\beta(x) = \begin{cases} \beta^2 x + \beta, & -\frac{1}{\beta} \leq x \leq 0 \\ \beta^2 x - \beta, & 0 \leq x \leq \frac{1}{\beta} \\ 0, & \text{elsewhere} \end{cases}. \tag{39}$$

The parameter β is set to 5 in this paper. From (39), it can be seen that when the coefficients are in the range of $(-1/\beta, 1/\beta)$, they will be constantly attracted to zero, and when the coefficients are not in the range, there will be no additional attraction, which will improve the convergence speed of those coefficients close to zero, and the overall convergence speed will be accelerated.

As in Sect. 3.1, primarily, it is necessary to process the output signal to get Eq. (13). Then, the error can be expressed as Eq. (16). Next, it is also needed to process α and then use the steepest descent method to get Eq. (24). Again, the next step is to estimate the parameter σ_n . That is, the cost function based on l_0 -CRPFRMS is derived and obtained.

$$\hat{\sigma}_{i,n} = \hat{\sigma}_{i,n-1} - \frac{\mu}{2} \frac{\partial \vartheta_{l_0\text{-CRPFRMS}}(\mathbf{w}, \sigma_{i,n-1}, \alpha)}{\partial \sigma_i} \Big|_{\hat{\sigma}_{i,n-1}}$$

$$= \hat{\sigma}_{i,n-1} + \mu\varphi(\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_n) f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_n)) \varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_n). \tag{40}$$

Similarly, the weight update formula can be obtained as follows

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mu G_{n-1} \Phi_n(\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_n) u_n^T f(\varepsilon_n(\mathbf{w}, \sigma_{i,n}, \hat{\boldsymbol{\alpha}}_n)) \varepsilon_n(\mathbf{w}, \sigma_{i,n}, \hat{\boldsymbol{\alpha}}_n) - \kappa f_{\beta}(\mathbf{w}_{n-1}) \tag{41}$$

where

$$\mathbf{w}_n = (w_n(0), \dots, w_n(L - 1)) \tag{42}$$

$$\mathbf{G}_{n-1} = \text{diag}\{g_{n-1}(0), \dots, g_{n-1}(L - 1)\}. \tag{43}$$

The diagonal elements of \mathbf{G}_{n-1} are calculated as follows:

$$g_{n-1}(l) = \frac{\theta_{n-1}(l)}{\sum_{i=0}^{L-1} \theta_{n-1}(i)}, \quad 0 \leq l \leq L - 1 \tag{44}$$

$$\theta_{n-1}(l) = \max\{\rho \max[|w_{n-1}(0)|, \dots, |w_{n-1}(L - 1)|], |w_{n-1}(l)|\} \tag{45}$$

where $\rho = 5/L$.

3.3 The Convergence Analysis of l_0 -CRPFRMS Algorithm

The steady-state performance of l_0 -CRPFRMS algorithm is analyzed in the following part. To make the analysis tractable, the following assumptions are given, which are commonly used in the analysis of adaptive filtering algorithm [3, 13].

Assumption 1 The noise η_n is independent of the input signal. The impulsive noise ξ_n doesn't occur.

Assumption 2 The weight error vector $\hat{\mathbf{w}}_n = \mathbf{w}_o - \mathbf{w}_n$ is independent of the input signal.

Obviously, (24) can be rewritten

$$\hat{\boldsymbol{\alpha}}_n = \hat{\boldsymbol{\alpha}}_{n-1} + \mu \Gamma'_n(\hat{\boldsymbol{\alpha}}_{n-1}). \tag{46}$$

Then, the Taylor expansion of $\Gamma'_n(\hat{\boldsymbol{\alpha}}_{n-1})$ at the point $\hat{\boldsymbol{\alpha}}_{n-1} =$ is given by,

$$\begin{aligned} \Gamma'_n(\hat{\alpha}_{n-1}) &= \Gamma'_n(\hat{\alpha}) + \Gamma''_n(\hat{\alpha}_{n-1})(\hat{\alpha}_{n-1} - \hat{\alpha}) \\ &= \Gamma'_n(\hat{\alpha}) - \mathbf{Z}_n(\hat{\alpha}_{n-1} - \hat{\alpha}) \end{aligned} \tag{47}$$

where $\mathbf{Z}_n = \Gamma''_n(\hat{\alpha}_{n-1})$. Letting $\tilde{\alpha}_{n-1} = \alpha - \hat{\alpha}_{n-1}$, subtracting both sides of (46) by α and using (47) yields

$$\tilde{\alpha}_n = \tilde{\alpha}_{n-1} - \mu \Gamma'_n(\hat{\alpha}_{n-1}) = (\mathbf{I}_M + \mu \mathbf{Z}_n) \tilde{\alpha}_{n-1} - \mu \Gamma'_n(\alpha). \tag{48}$$

Using the Assumptions A1–A2 and taking expectation both sides of (4) leads to

$$E[\tilde{\alpha}_n] = (\mathbf{I}_M + \mu E[\mathbf{Z}_n])E[\tilde{\alpha}_{n-1}] - \mu E[\Gamma'_n(\alpha)]. \tag{49}$$

According to [24], $E[\mathbf{Z}_n]$ is the hessian matrix of $\Gamma_n(\hat{\alpha}_{n-1})$ and is negative definite. In the sequel, taking the expectation $\Gamma_n(\alpha)$ with only v_n results in

$$E_v[a_n] = Pr(a_n = 1) = \Phi(u_n^T \tilde{\alpha}). \tag{50}$$

Using (50) and Assumptions A1–A2 yields (see [24] for detail)

$$E[\Gamma'_n(\alpha)] = 0. \tag{51}$$

Substituting (52) into (50) yields

$$E[\tilde{\alpha}_n] = (\mathbf{I}_M + \mu E[\mathbf{Z}_n])E[\tilde{\alpha}_{n-1}]. \tag{52}$$

To guarantee the stability in the mean sense, the matrix should be stable. Note that the negative-definite has negative eigenvalues, and hence, the step size should be selected according to

$$0 < \mu < -\frac{2}{\lambda_{\min}(E[\mathbf{Z}_n])}. \tag{53}$$

Under this condition, we have

$$E[\tilde{\alpha}_\infty] = \alpha. \tag{54}$$

In the other words, Eq. (24) is an unbiased estimate of α if the proposed algorithm is stable. Then, (40) and (41) can be rewritten, respectively,

$$\begin{aligned} \hat{\sigma}_{i,n} &= \hat{\sigma}_{i,n-1} - \frac{\mu}{2} \frac{\partial \vartheta_{l_0\text{-CRPFRMS}}(\mathbf{w}, \sigma_{i,n-1}, \hat{\alpha}_n)}{\partial \sigma_i} \Big|_{\hat{\sigma}_{i,n-1}} \\ &= \hat{\sigma}_{i,n-1} + \mu \varphi(\mathbf{u}_n^T \hat{\alpha}_n) f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\alpha}_n)) \varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\alpha}_n) \end{aligned} \tag{55}$$

$$\begin{aligned} \mathbf{w}_n &= \mathbf{w}_{n-1} + \mu G_{n-1} \Phi_n(\mathbf{u}_n^T \hat{\alpha}_n) u_n^T f(\varepsilon_n(\mathbf{w}, \sigma_{i,n}, \hat{\alpha}_n)) \varepsilon_n(\mathbf{w}, \sigma_{i,n}, \hat{\alpha}_n) \\ &\quad - \kappa f_\beta(\mathbf{w}_{n-1}) \end{aligned}$$

$$\approx \mathbf{w}_{n-1} + \mu G_{n-1} \Phi_n(\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_n) u_n f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_n)) \varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_n) - \kappa f_\beta(\mathbf{w}_{n-1}). \tag{56}$$

Combining (55) and (56) gives

$$\hat{\boldsymbol{\theta}}_n = \begin{bmatrix} \mathbf{w}_n \\ \hat{\sigma}_{i,n} \end{bmatrix} = \begin{bmatrix} \mathbf{w}_{n-1} \\ \hat{\sigma}_{i,n-1} \end{bmatrix} + \mu f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_n)) \varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_n) \bar{\mathbf{G}}_{n-1} \hat{\mathbf{h}}_n - \begin{bmatrix} 0 \\ \kappa f_\beta(\mathbf{w}_{n-1}) \end{bmatrix} \tag{57}$$

where

$$\bar{\mathbf{G}}_{n-1} = \begin{bmatrix} \mathbf{G}_{n-1} & 0 \\ 0 & 1 \end{bmatrix} \tag{58}$$

$$\hat{\mathbf{h}}_n = \begin{bmatrix} \Phi_n(\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_n) u_n \\ \varphi(\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_n) \end{bmatrix}. \tag{59}$$

Then, the desired signal d_n can be written as

$$d_n = \mathbf{h}_n^T \boldsymbol{\theta}_{\text{opt}} + v_n \tag{60}$$

where

$$\boldsymbol{\theta}_{\text{opt}} = \begin{bmatrix} \mathbf{w}_o \\ \sigma_i \end{bmatrix} \tag{61}$$

$$\mathbf{h}_n = \begin{bmatrix} \Phi_n(\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_n) u_n \\ \varphi(\mathbf{u}_n^T \hat{\boldsymbol{\alpha}}_n) \end{bmatrix}. \tag{62}$$

Using (60), (16) implies

$$\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_{n-1}) \approx \varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \boldsymbol{\alpha}) = d_n - \mathbf{h}_n^T \hat{\boldsymbol{\theta}}_{n-1} = \mathbf{h}_n^T \tilde{\boldsymbol{\theta}}_{n-1} + v_n \tag{63}$$

where

$$\tilde{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_{\text{opt}} - \hat{\boldsymbol{\theta}}_n. \tag{64}$$

Inserting (64) into (57) yields

$$\tilde{\boldsymbol{\theta}}_n = (\mathbf{I} - \mu f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_n)) \mathbf{H}_n) \tilde{\boldsymbol{\theta}}_{n-1} - \mu f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}_n)) v_n \bar{\mathbf{G}}_{n-1} \mathbf{h}_n + \begin{bmatrix} 0 \\ \kappa f_\beta(\mathbf{w}_{n-1}) \end{bmatrix} \tag{65}$$

where

$$\mathbf{H}_n = \bar{\mathbf{G}}_{n-1} \mathbf{h}_n \mathbf{h}_n^T. \tag{66}$$

Then, taking the expectation of both sides of (65)

$$\begin{aligned}
 E[\tilde{\boldsymbol{\theta}}_n] &= (\mathbf{I} - \mu E[f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}))\mathbf{H}_n])E[\tilde{\boldsymbol{\theta}}_{n-1}] \\
 &\quad - \mu f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}))E[v_n]E[\tilde{\mathbf{G}}_{n-1}]E[\mathbf{h}_n] + \begin{bmatrix} 0 \\ E[\kappa f_{\beta}(\mathbf{w}_{n-1})] \end{bmatrix} \\
 &= (\mathbf{I} - \mu \mathfrak{R}_n)E[\tilde{\boldsymbol{\theta}}_{n-1}] + \begin{bmatrix} 0 \\ E[\kappa f_{\beta}(\mathbf{w}_{n-1})] \end{bmatrix} \\
 &\approx (\mathbf{I} - \mu \mathfrak{R}_n)E[\tilde{\boldsymbol{\theta}}_{n-1}]
 \end{aligned} \tag{67}$$

where

$$\mathfrak{R}_n = E[f(\varepsilon_n(\mathbf{w}, \sigma_{i,n-1}, \hat{\boldsymbol{\alpha}}))\mathbf{H}_n]. \tag{68}$$

In order to ensure the convergence of the algorithm, this should make $-1 < (\mathbf{I} - \mu \mathfrak{R}_n) < 1$. Therefore, the range of μ is

$$0 < \mu < \frac{2}{\lambda_{\max}(\mathfrak{R}_n)}. \tag{69}$$

Combining (53) and (69), we can obtain

$$0 < \mu < \min\left(-\frac{2}{\lambda_{\min}(E[\mathbf{Z}_n])}, \frac{2}{\lambda_{\max}(\mathfrak{R}_n)}\right). \tag{70}$$

4 Simulation

4.1 Verify the Superiority of CR-FRMS

Firstly, the simulations in context of the system identification are carried out to illustrate the advantage of the proposed algorithm. The length of the unknown system generated randomly is 8 taps. It is assumed that the length of the unknown system is same as that of adaptive filter. Figure 1a, b depicts the performance of the CR-FRMS, FRMS [24], TSA [11], and LMS [13] algorithms under two background noise distributions, respectively, where $p = 2$ and $\kappa = 8 \times 10^{-6}$. The step sizes of four different algorithms are set to 0.03 in two different mixed noise environments. Figure 1a shows the simulation results in a mixture of sub-Gaussian and super-Gaussian noises, and Fig. 1b shows the simulation results in a mixture of sub-Gaussian noise and impulsive noise [uniform and Laplace distribution are independent and identically distributed (i.i.d.) over time with zero mean]. Bernoulli–Gaussian (BG) process [41] is used frequently for modeling the impulsive noise ξ_n , formulated as $\xi_n = \tau_n v_n$, where τ_n is a Bernoulli process with the probability of $p\{\tau_n = 1\} = 0.01$ and v_n is i.i.d. zero-mean Gaussian sequence with variance $\sigma = 10000$. For a fair comparison of proposed algorithms, parameters are set so that the algorithms have a comparable initial convergence speed. As observed in

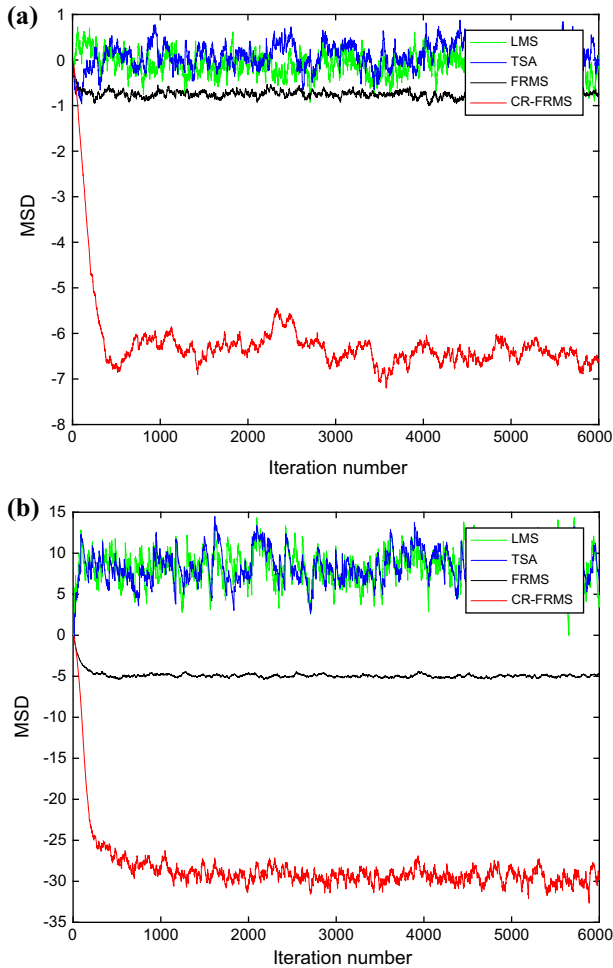


Fig. 1 MSD curves for the proposed algorithm **a** the mixed uniform and Laplace distributed noises, where $\zeta = 0.5$. **b** The mixed uniform distributed and impulsive noises, where $\zeta = 0.005$

Fig. 1 a, b, the proposed CR-FRMS algorithm is superior to other existing algorithms. In addition, the calculation formula of MSD is defined as

$$\text{MSD} = 10 \log_{10} \|\mathbf{w}_o - \mathbf{w}_n\|^2 \quad (71)$$

The second experiment is to test the convergence performance of CR-FRMS with different step sizes. Figure 2a, b illustrates the MSD learning curves with different step sizes of input signals for Gaussian white noise. As expected, when the fixed step size is applied in the CR-FRMS algorithm, there is a trade-off between the steady-state MSD and the rate of convergence. That is, a small step size corresponds to the lower steady-state error, although it slows down the convergence speed. In contrast, a large

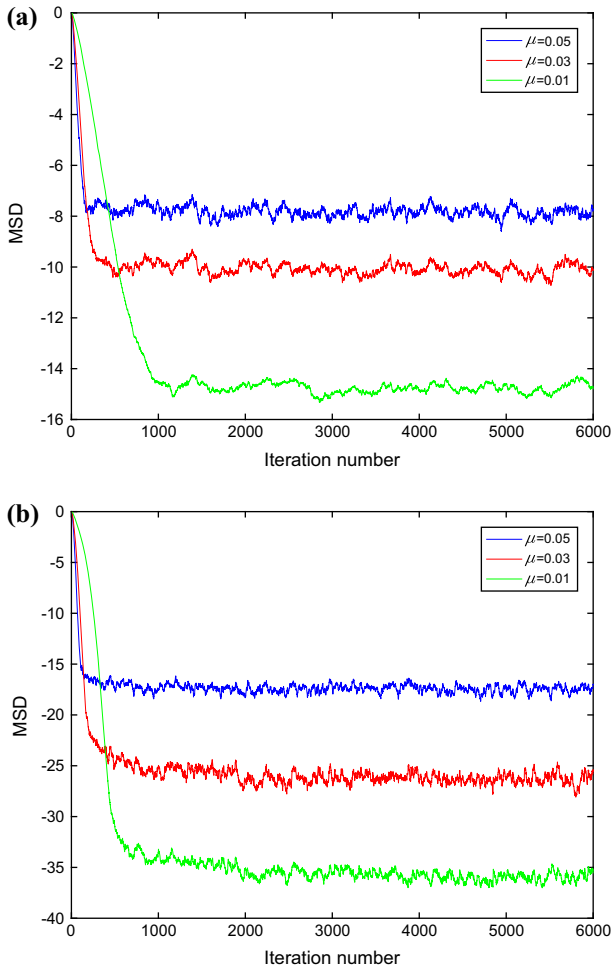


Fig. 2 Comparison of different step sizes; **a** the mixed noise environment of sub-Gaussian and super-Gaussian. **b** The mixed noise environment of sub-Gaussian and impulsive noise

step size which is located in the stable range provides the higher the convergence speed, while it achieves large steady-state error.

The third experiment is to test the convergence performance of CR-FRMS with different parameters p . Suppose that the unknown system has 8 coefficients. The driven signal is white Gaussian with unit variance. The filter length is 8. The step size of these algorithms is fixed to 3×10^{-2} , while p is set to different values. After a hundred times run, it can be seen from Fig. 3a that in the case of uniform distribution and Laplace distribution variances of 1 and 9, respectively, the larger the p , the smaller the MSD, but the effect is not significant. However, as can be seen from Fig. 3b, in the case of mixed background noise with impulsive noise and sub-Gaussian noise

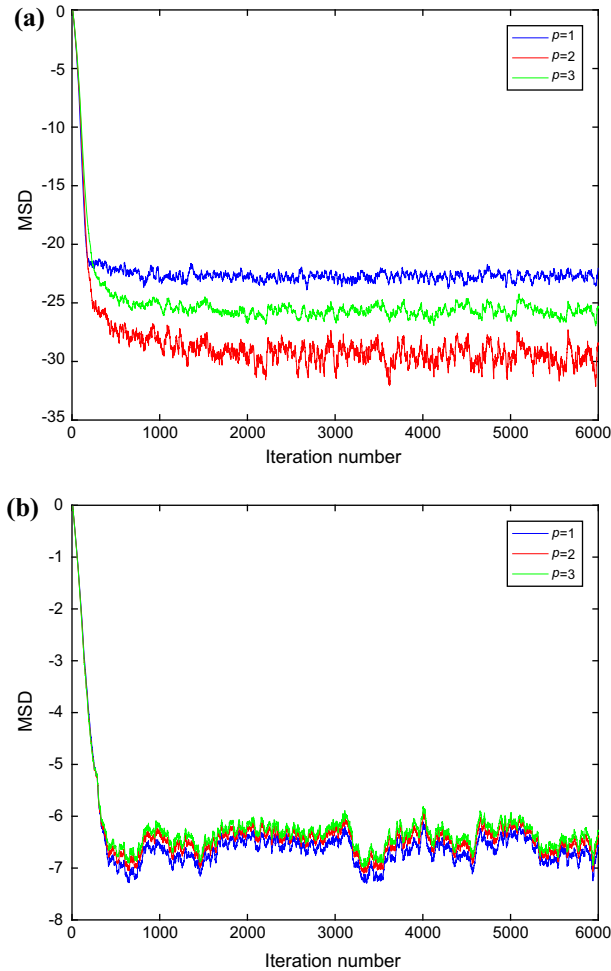


Fig. 3 Background noise is different p **a** the mixed noise environment of sub-Gaussian and super-Gaussian. **b** The mixed noise environment of sub-Gaussian and impulsive noise

(uniform distribution) with a signal-to-noise ratio (SNR) of 30, $p = 2$ obtains a better MSD when taking 1 or 3 with respect to p .

4.2 Verify the Superiority of l_0 -CRPFRMS

The proposed l_0 -CRPFRMS is compared with the algorithms CR-FRMS, PNLMS [9] and l_0 -LMS [5] in two different background noise environments. The first one is a mixed noise with a uniform distribution with variance $\sigma_u = 2.5 \times 10^{-5}$ and a Laplace distribution with variance $\sigma_L = 0.01$. The second is still mixed noise with a sub-Gaussian noise (uniform distribution) with signal-to-noise ratio of 30 dB and the impulsive noise. Besides, in two different mixed background noise environments,

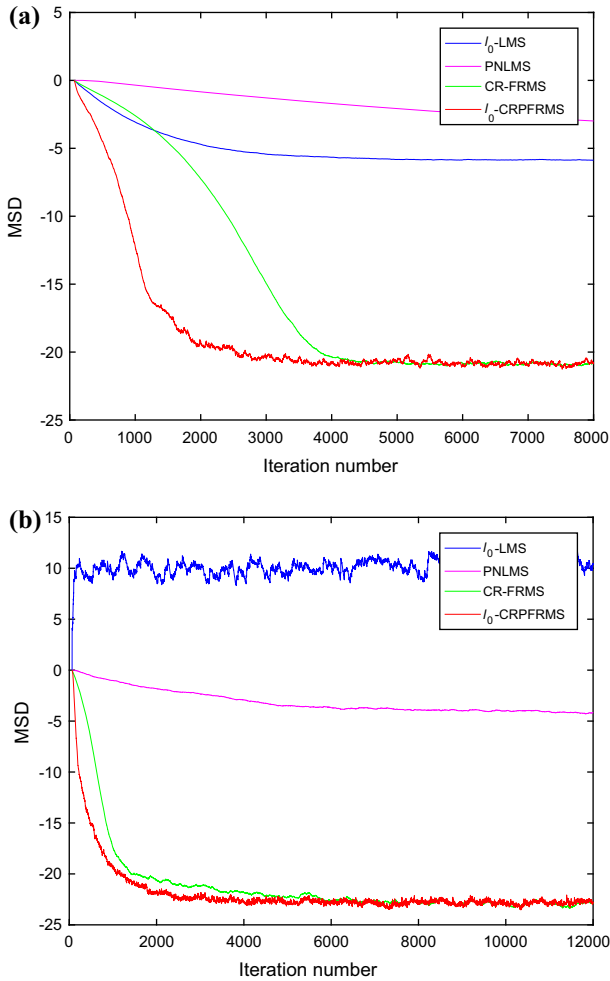


Fig. 4 MSD curves for the proposed algorithm **a** the mixed sub-Gaussian and super-Gaussian noises, where $\zeta = 1 \times 10^{-6}$, $\mu = 0.001$, $\kappa = 2 \times 10^{-6}$. **b** The mixed sub-Gaussian and impulsive noises, where $\zeta = 0.001$, $\mu = 0.005$, $\kappa = 1 \times 10^{-6}$

$p = 2$. Suppose that the unknown system has 64 coefficients, in which two of them are nonzero ones (their locations and values are randomly selected). The input signal is white Gaussian with unit variance. The filter length is 64. After fifty independent operations, their MSD curves are shown in Fig. 4a, b. It is evidently recognized that l_0 -CRPFMS algorithm converges faster than its ancestor.

Comparing different algorithms is also needed to do experimental research on different values of parameter κ . Theoretically, when κ is larger, the update weight will be closer to zero faster, that is, the convergence speed of the algorithm will be accelerated. However, the MSD of the algorithm will be increased. In this paper, the range of values of κ is relatively small. Therefore, as can be seen from Fig. 5a, b, the

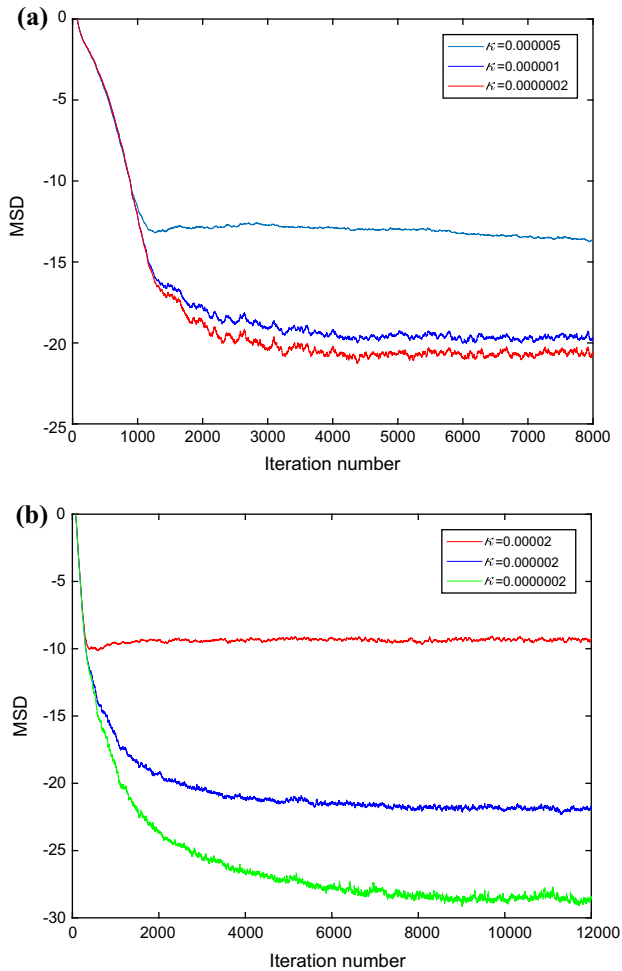


Fig. 5 Comparison of different κ in **a** the mixed sub-Gaussian and super-Gaussian noises, **b** mixed sub-Gaussian and impulsive noises

convergence speed of the algorithm is almost the same when κ is different, but MSD has obvious difference, that is, the smaller the κ , the smaller the MSD.

5 Conclusion

In this paper, two algorithms based on censored regression, namely CR-FRMS and l_0 -CRPFRMS, are proposed under two different mixed background noises. CR-FRMS show superiority to LMS, TSA and FRMS in terms of MSD. Meanwhile, in the case where the unknown system is a sparse system, l_0 -CRPFRMS exhibits a faster convergence speed than CR-FRMS without changing the MSD. Since the two algo-

rithms show different advantages, CR-FRMS has lower computational complexity and l_0 -CRPFRMS converges fast. Therefore, in real life, we must combine the actual requirements and choose a reasonable algorithm.

Acknowledgments This work was partially supported by National Science Foundation of P.R. China (Grant Nos. 61871461, 61571374 and 61433011), Sichuan Science and Technology Program (Grant No. 19YYJC0681).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

For a random x with standard distribution, i.e., the cumulative distribution function satisfies the condition in Table 1, the conditional density function of conditional random variable denoted by is given by

$$\begin{aligned} p(x|x > c) &= \frac{dPr(V < x|x > c)}{dx} \\ &= \frac{d}{dv} \frac{Pr(V < x \text{ and } x > c)}{Pr(x > c)} \\ &= \frac{d}{dx} \frac{\int_b^x \varphi(u)du}{Pr(x > c)} = \frac{\varphi(x)}{1 - \Phi(c)} \end{aligned}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ is expressed in Table 1. Using the probability theory, the conditional expectation $E[x|x > c]$ can be computed by

$$\begin{aligned} E[x|x > c] &= \int_c^\infty xf(x|x > c)dv \\ &= \int_c^\infty \frac{x\varphi(x)}{1 - \Phi(x)}dx = -\frac{1}{1 - \Phi(c)}F(x)|_c^\infty \\ &= \frac{\varphi(c)}{1 - \Phi(c)} = \frac{\varphi(c)}{\Phi(-c)} = \Psi(-c) \end{aligned}$$

References

1. B. Allik, C. Miller, M.J. Piovoso, R. Zurakowski, Estimation of saturated data using the Tobit Kalman filter, in Proceedings of the American Control Conference (ACC), Portland, OR, USA, June (2014), pp. 4151–4156
2. J. Benesty, J. Chen, Y. Huang, *Microphone Array Signal Processing*, Ser. Springer Topics in Signal Processing, 2008th edn. (Springer, New York, NY, 2008)
3. M. Bottai, J. Zhang, Laplace regression with censored data. *Biom. J.* **52**(4), 487–503 (2010)
4. A.C. Cameron, P.K. Trivedi, *Microeconometrics: Methods and Applications* (Cambridge University Press, New York, NY, 2005)

5. Y. Chen, Y. Gu, A. Hero, Sparse LMS for system identification, in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE (2009), pp. 3125–3128
6. H.C. Cho, S.W. Park, K.S. Lee, N.H. Kim, Anovel multiple-channel active noise control approach with neural secondary-path model for interior acoustic noise attenuation of railway train systems. *IET Sig. Process.* **6**(8), 772–780 (2012)
7. D. Collett, *Modelling Survival Data in Medical Research*, 2nd edn. (Chapman and Hall, London, 2003)
8. J. Cook, J. McDonald, Partially adaptive estimation of interval censored regression models. *Comput. Econom.* **42**(1), 119–131 (2013)
9. D. Ding, Z. Wang, B. Shen, H. Shu, H_∞ State estimation for discrete time complex networks with randomly occurring sensor saturations and randomly varying sensor delays. *IEEE Trans. Neural Netw.* **23**(5), 725–736 (2012)
10. D.L. Dunweiler, Proportionate normalized least mean square adaptation in echo cancelers. *IEEE Trans. Speech Audio Process.* **8**, 508–518 (2000)
11. E. Ertin, Gaussian process models for censored sensor readings, in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP'07)* (2007), pp. 665–669
12. Y. Gu, J. Jin, S. Mei, l_0 -Norm constraint LMS algorithm for sparse system identification. *IEEE Signal Process. Lett.* **16**(9), 774–777 (2009)
13. S.S. Haykin, *Adaptive Filter Theory* (Pearson Education, Chennai, 2008)
14. Z. He, *Adaptive Signal Processing* (Science Press, Beijing, 2003)
15. J.J. Heckman, The common structure of statistical models of truncation, sample selection and limited dependent variables. *Ann. Econ. Soc. Meas.* **5**(4), 311–801 (1976)
16. O. Hoshuyama, A. Sugiyama, A. Hirano, A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Trans. Signal Process.* **47**(10), 2677–2684 (1999)
17. X. Jiang, W.J. Zeng, A. Yasotharan, H.C. So, T. Kirubarajan, Robust beamforming by linear programming. *IEEE Trans. Signal Process.* **62**(7), 1834–1849 (2014)
18. J. Jin, B. Yang, K. Liang, X. Wang, General image denoising framework based on compressive sensing theory. *Comput. Graph.* **38**, 382–391 (2014)
19. M. Kawamoto, K. Matsuoka, N. Ohnishi, A method of blind separation for convolved non-stationary signals. *Neurocomputing* **22**(1), 157–171 (1998)
20. T.W. Lee, M. Girolami, T. Sejnowski, Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources. *Neural Comput.* **11**(2), 409–433 (1999)
21. T.W. Lee, M.S. Lewicki, Unsupervised image classification, segmentation, and enhancement using ICA mixture models. *IEEE Trans. Image Process.* **11**(3), 270–279 (2002)
22. Z. Liu, C. Li, Censored regression with noisy input. *IEEE Trans. Signal Process.* **63**(19), 5071–5082 (2015)
23. Z. Liu, C. Li, Recursive least squares for censored regression. *IEEE Trans. Signal Process.* **65**(6), 1565–1579 (2015)
24. Z. Liu, C. Li, Y. Liu, Distributed censored regression over networks. *IEEE Trans. Signal Process.* **63**(20), 5437–5449 (2015)
25. W. Ma, B. Chen, H. Zhao, Sparse least logarithmic absolute difference algorithm with correntropy-induced metric penalty. *Circuits Syst. Sig. Process.* **35**(3), 1077–1089 (2016)
26. S. Maleki, G. Leus, Censored truncated sequential spectrum sensing for cognitive radio networks. *IEEE J. Sel. Areas Commun.* **31**(3), 364–378 (2013)
27. N.R. Mann, *Methods for Statistical Analysis of Reliability and Life Data* (Wiley, New York, NY, 1975)
28. E.J. Msechu, G.B. Giannakis, Distributed measurement censoring for estimation with wireless sensor networks, in *Proceedings of the Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, San Francisco, CA, USA (2011), pp. 176–180
29. J.L. Powell, Least absolute deviations estimation for the censored regression model. *J. Econom.* **25**(3), 303–325 (1984)
30. A.H. Sayed, *Fundamentals of Adaptive Filtering* (Wiley, Hoboken, NJ, 2003)
31. M.R. Segal, Regression trees for censored data. *Biometrics* **44**(1), 35–47 (1988)
32. L.N. Sharma, S. Dandapat, A. Mahanta, ECG signal denoising using higher order statistics in wavelet subbands. *Biomed. Signal Process. Control* **5**(3), 214–222 (2010)
33. F. Shen, *Adaptive Signal Processing* (Xi'an University of Electronic Science and Technology Press, Xi'an, 2001)

34. G. Su, J. Jin, Y. Gu, J. Wang, Performance analysis of l_0 -norm constraint least mean square algorithm. *Ann. Econ. Soc. Meas.* **60**(5), 2223–2235 (2012)
35. G.A. Tsihrintzis, C.L. Nikias, Data-adaptive algorithms for signal detection in sub-Gaussian impulsive interference. *IEEE Trans. Signal Process.* **45**(7), 1873–1878 (1997)
36. W. Wang, H. Zhao, B. Chen, Bias compensated zero attracting normalized least mean square adaptive filter and its performance analysis. *Sig. Process.* **143**, 94–105 (2017)
37. W. Wang, H. Zhao, B. Chen, Robust adaptive volterra filter under maximum correntropy criteria in impulsive environments. *Circuits Syst. Sig. Process.* **36**(10), 4097–4117 (2017)
38. W. Wang, H. Zhao, K. Doğançay, Y. Yu, L. Lu, Z. Zheng, Robust adaptive filtering algorithm based on maximum correntropy criteria for censored regression. *Sig. Process.* **160**, 88–98 (2019)
39. L. Weruaga, S. Jimaa, Exact NLMS algorithm with, ℓ_p -norm constraint. *Sig. Process. Lett. IEEE* **22**(3), 366–370 (2014)
40. W. Ying, Y. Jiang, Y. Liu, P. Li, A blind receiver with multiple antennas in impulsive noise modeled as the sub-Gaussian distribution via the MCMC algorithm. *IEEE Trans. Veh. Technol.* **62**(7), 3492–3497 (2013)
41. Y. Yu, H. Zhao, B. Chen, Z. He, Two improved normalized subband adaptive filter algorithms with good robustness against impulsive interferences. *Circuits Syst. Sig. Process.* **35**(12), 4607–4619 (2016)
42. X. Zhang, Z. Bao, *Communication Signal Processing* (Publishing of Defense Industry, Beijing, 2000)
43. H. Zhang, H. Yan, F. Yang, Q. Chen, Distributed average filtering for sensor networks with sensor saturation. *IET Control Theory Appl.* **7**(6), 887–893 (2013)
44. S. Zhang, J. Zhang, New steady-state analysis results of variable step-size LMS algorithm with different noise distributions. *IEEE Signal Process. Lett.* **21**(6), 653657 (2014)
45. S. Zhang, W. Zheng, J. Zhang, H. Han, A family of robust M-shaped error weighted least mean square algorithms: performance analysis and echo cancelation application. *IEEE Access* **5**, 14716–14727 (2017)
46. F. Zhong, J. Zhang, Linear discriminant analysis based on L1-norm maximization. *IEEE Trans. Image Process.* **22**(8), 3018–3027 (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.