Signal Processing 138 (2017) 98–105

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

A novel cauchy score function based DOA estimation method under alpha-stable noise environments

Jin-feng Zhang [a,b], Tian-shuang Qiu [b,*], Peng Wang [b], Sheng-yang Luan [b]

[a] *Shenzhen Key Laboratory of Antennas and Propagation, Shenzhen University, Shenzhen, Guangdong, 518060, China*
[b] *Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, Liaoning, 116024, China*

ARTICLE INFO

Article history:
Received 21 July 2016
Revised 9 March 2017
Accepted 14 March 2017
Available online 16 March 2017

Keywords:
MUSIC
Alpha-stable distribution
Cauchy distribution
Impulsive noise
M-estimation

ABSTRACT

Based on the study of alpha-stable distribution and its score function, by combining M-estimation theory and maximum likelihood estimation theory, this paper proposes a novel Cauchy score function based cost function of the residual fitting error matrix. And then, this cost function is employed as a substitute for the ℓ_p-norm based cost function of the residual fitting error matrix which is utilized in the ℓ_p-MUSIC algorithm. To solve the cost function, alternating convex optimization is applied. And the complex-valued Newton's method with optimal step size is developed to solve the resulting convex problem. With the obtained signal subspace, the direction of arrival (DOA) estimates are retrieved by the MUSIC technique. Comprehensive simulation results demonstrate that, the proposed algorithm can achieve more robust performance than its counterpart both in terms of resolution probability and root mean square error (RMSE), especially when the generalized signal to noise ratio (GSNR) is fairly low, or the underlying noise is extremely impulsive.

1. Introduction

High resolution direction finding, as an important research branch of array signal processing, has been widely applied in many applications, such as, radar, sonar, wireless communications and so on. The well-known MUSIC [1], which is proposed under the framework of subspace theory, is one of the most classical high-resolution direction finding method. By exploiting the orthogonality between the noise subspace and array manifold, it gains the direction of arrivals (DOA) for the source signals. So far, plenty of MUSIC like algorithms have been proposed under the framework of subspace theory [2]. However, for the sake of computational efficiency, most of these algorithms have been studied under the assumption of Gaussian noise.

In fact, recent studies show that in some scenarios, sudden bursts or sharp spikes are exhibited at the array outputs which can be characterized as impulsive due to clutter sources such as mountains, forests and sea waves [3]. This impulsive component of noise has been found to be significant in many problems, including atmospheric noise and underwater problems such as sonar and submarine communication. Nikias and Shao [3] show that this impulsive noise can be well modeled by $S\alpha S$ (symmetric alpha-stable) distribution. The $S\alpha S$ distribution is defined by its characteristic function $\varphi(\omega) = \exp(-\gamma|\omega|^{\alpha})$, where α is the characteristic exponent taking values $0 < \alpha \leq 2$, and γ is the dispersion which is similar to the variance of the Gaussian distribution. Specially, Gaussian processes are $S\alpha S$ processes with $\alpha = 2$. Those $S\alpha S$ distributions with $\alpha < 2$, as we call the fractional lower-order alpha (FLOA) distributions, present heavier tails than those of Gaussian distribution and possess finite pth-order moments only for $p < \alpha < 2$. Due to the infinite variance of FLOA distribution, the conventional MUSIC algorithms which are developed based on the second order statistics degraded dramatically in their performance in impulsive noise environments.

To overcome the defect of the second order statistics based MUSIC algorithms, fractional lower order statistics (FLOS) based subspace methods have been developed for the DOA estimation. In [4–6], the covariation, one of the fractional lower order moments, and the phased fractional lower order moment have been proposed as substitutes for the covariance in the conventional MUSIC, respectively. By performing an eigenvalue decomposition (EVD) on the FLOS based matrix of the array outputs, the signal subspace and the noise subspace is partitioned. To further improve the performance of these FLOS based MUSIC algorithms in fairly low SNR (signal to noise ratio) or in highly impulsive noise scenarios, the correntropy based correlation (CRCO) has been developed in [7]. In [7], the CRCO based matrix for the array outputs is formulated and applied with MUSIC for the retrieval of DOAs. Despite the robustness of these above mentioned algorithms in impulsive noises,

* Corresponding author.
E-mail address: qiutsh@dlut.edu.cn (T.-s. Qiu).

http://dx.doi.org/10.1016/j.sigpro.2017.03.014
0165-1684/

they generally require a large number of data samples and a fairly high SNR level for a satisfactory performance.

The key step of the MUSIC method is computing the signal or noise subspace. Different from the conventional subspace decomposition on the constructed covariance or covariance like matrices which are formulated through the aforementioned techniques [4–7], in [8], a new subspace decomposition rule is proposed. Based on the minimization of the ℓ_p-norm of the residual fitting error matrix, and by solving the resulting nonconvex optimization problem, the signal subspace is directly computed. And naturally, the EVD is not required. Employing the obtained signal subspace to construct the spatial spectrum in MUSIC, the ℓ_p-MUSIC is proposed. Computer simulations have presented the effectiveness of ℓ_p-MUSIC in impulsive noise, especially with fairly few data samples.

However, through a thoroughly investigation of alpha-stable distribution, we can observe that, the ℓ_p-norm minimization based cost function is far from the optimal cost function for alpha-stable distributions. The degeneration of ℓ_p-MUSIC in extremely low generalized signal to noise ratio (GSNR) situations or highly impulsive noise environments, as we show in the simulations in Section 4, can also verify this point.

Examining the score functions of alpha-stable distributions with different characteristic exponents, we can find that they have similar features and can be selected as the nonlinear functions to eliminate outliers. Considering that the maximum likelihood (ML) estimator of a deterministic signal in an additive white Cauchy distributed ($S\alpha S$ distribution with $\alpha = 1$) noise is robust to deviations in the characteristic exponent α [9], in this paper, we choose the score function of Cauchy distribution as the nonlinear function. Based on M-estimation theory, we propose a new cost function for the search of the signal subspace. The resulting nonconvex optimization problem can also be solved by alternating convex optimization, as Ref. [8]. Finally, we employ the obtained signal subspace to construct the corresponding spatial spectrum of MUSIC for the retrieval of DOAs. As the new DOA estimator can be traced back to the employment of the score function of Cauchy distribution, we call it SC-MUSIC in this paper.

The remainder of this paper is organized as follows. In Section 2, we define the problem of interest and give a brief review of the ℓ_p-MUSIC algorithm. In Section 3, we propose the SC-MUSIC algorithm. Finally, simulation experiments are presented in Section 4, and conclusions are drawn in Section 5.

The following notations are utilized throughout this paper. The superscripts "T", "H" denote the transpose, the conjugate transpose, respectively. "$\|\cdot\|$" stands for the Frobenius norm. $\mathbf{I}_M$ is the $M \times M$ identity matrix.

2. Problem formulation and the ℓ_p-MUSIC

2.1. Array model and the conventional MUSIC

Assume Q narrow-band independent, complex signals with locations $\{\theta_1, \cdots, \theta_Q\}$ are impinging on a uniformly linear array of M sensors. The underlying noises are i.i.d. isotropic complex random processes. Using the complex envelope representation, the array output received at the mth sensor can be expressed as

$$x_m(t) = \sum_{q=1}^{Q} a(\theta_q) s_q(t) + n_m(t), \quad m = 1, 2, \cdots, M, \tag{1}$$

where $a(\theta_q) = e^{-j\frac{2\pi}{\lambda}(m-1)d\sin(\theta_q)}$ is the steering coefficient of the mth sensor toward direction θ_q, λ denotes the wavelength of the carrier and d is the distance between two sensors, $s_q(t)$ is the qth signal received at the reference sensor, and $n_m(t)$ is the underlying noise at sensor $x_m(t)$.

The vector form of (1) can be expressed as

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t), \tag{2}$$

where $\mathbf{A}(\boldsymbol{\theta})$ is the $M \times Q$ matrix of the array steering vectors $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_Q)]$, in which $\mathbf{a}(\theta_q) = [1, \cdots, e^{-j\frac{2\pi}{\lambda}d(M-1)\sin(\theta_q)}]^T$, $\mathbf{x}(t)$ is the $M \times 1$ vector of signals received by the array sensors $\mathbf{x}(t) = [x_1(t), \cdots, x_M(t)]^T$, $\mathbf{s}(t)$ is the $Q \times 1$ vector of the signals $\mathbf{s}(t) = [s_1(t), \cdots, s_Q(t)]^T$, and $\mathbf{n}(t)$ is the $M \times 1$ vector of the noise $\mathbf{n}(t) = [n_1(t), \cdots, n_M(t)]^T$.

The covariance matrix of the received signal $\mathbf{C_x} = E\{\mathbf{x}\mathbf{x}^H\}$ with $E\{\cdot\}$ denoting the expectation is estimated from the T data samples as

$$\hat{\mathbf{C}}_\mathbf{x} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t {\mathbf{x}_t}^H = \frac{1}{T}\mathbf{X}\mathbf{X}^H. \tag{3}$$

The EVD of $\mathbf{C_x}$ is given by

$$\mathbf{C_x} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H = \mathbf{U_s}\boldsymbol{\Lambda}_\mathbf{s}{\mathbf{U_s}}^H + \sigma^2\mathbf{U_n}{\mathbf{U_n}}^H, \tag{4}$$

where $\mathbf{U} = [\mathbf{U_s}, \mathbf{U_n}]$, in which $\mathbf{U_s}$ and $\mathbf{U_n}$ span the signal subspace and the noise subspace, respectively. $\boldsymbol{\Lambda}_\mathbf{s} = \mathrm{diag}\{\lambda_1, \cdots, \lambda_Q\}$ is a diagonal matrix containing Q non-increasing principle eigenvalues. σ^2 is the variance of the noise. The MUSIC method searches for the peaks of the following spatial spectrum

$$P_{\mathrm{MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{U_n}\mathbf{U_n}^H\mathbf{a}(\theta)}, \quad -90^\circ \le \theta \le 90^\circ. \tag{5}$$

2.2. ℓ_p-MUSIC

Assume the sample length of the received data $\mathbf{x}(t)$ is T, in the absence of noise, the received data matrix $\mathbf{X}$ has the low-rank decomposition as

$$\mathbf{X} = \mathbf{Y}\mathbf{Z}, \tag{6}$$

where $\mathbf{Y}$ is an $M \times Q$ full column rank matrix and $\mathbf{Z}$ is a $Q \times T$ full row rank matrix. Obviously, the columns of $\mathbf{Y}$ span the same subspace of the array manifold matrix, namely, the signal subspace. Considering the non-Gaussian impulsive noise situation, in ℓ_p-MUSIC, the following ℓ_p-norm ($1 \le p \le 2$) based objective function is designed to solve $\mathbf{Y}$

$$\min_{\mathbf{Y},\mathbf{Z}} J_p = ||\mathbf{X} - \mathbf{Y}\mathbf{Z}||_p^p = \sum_{m=1}^{M}\sum_{t=1}^{T} |x_{mt} - (\mathbf{Y}\mathbf{Z})_{mt}|^p, \tag{7}$$

where $\mathbf{X} - \mathbf{Y}\mathbf{Z}$ is the residual fitting error matrix, x_{mt}, and $(\mathbf{Y}\mathbf{Z})_{mt}$ denote the (m, t) entries of $\mathbf{X}$ and $\mathbf{Y}\mathbf{Z}$, respectively.

To solve the nonconvex problem of (7), an alternating convex optimization approach is proposed in [8].

Suppose the two matrices $\mathbf{Y}^{(k)}$ and $\mathbf{Z}^{(k)}$ are obtained at the kth iteration. Then $\mathbf{Z}^{(k+1)}$ and $\mathbf{Y}^{(k+1)}$ can be solved by alternately optimizing the following two sub-problems

$$\mathbf{Z}^{(k+1)} = \arg\min_{\mathbf{Z}} ||\mathbf{X} - \mathbf{Y}^{(k)}\mathbf{Z}||_p^p, \tag{8}$$

$$\mathbf{Y}^{(k+1)} = \arg\min_{\mathbf{Y}} ||\mathbf{X} - \mathbf{Y}\mathbf{Z}^{(k+1)}||_p^p. \tag{9}$$

By rewriting (9) as $\min_\mathbf{Y}\|\mathbf{X} - \mathbf{Y}\mathbf{Z}^{(k+1)}\|_p^p = \min_\mathbf{Y}\|\mathbf{X}^T - (\mathbf{Z}^{(k+1)})^T \mathbf{Y}^T\|_p^p$, it is obvious that (8) and (9) have the same structure and can be solved in a similar manner.

The ℓ_p-norm error function in (8) can be further expressed as

$$\left\|\mathbf{X} - \mathbf{Y}^{(k)}\mathbf{Z}\right\|_p^p = \sum_{t=1}^{T}\left\|\mathbf{x}_t - \mathbf{Y}^{(k)}\mathbf{z}_t\right\|_p^p, \tag{10}$$

where $\mathbf{x}_t$ and $\mathbf{z}_t$ represent the tth column vectors of $\mathbf{X}$ and $\mathbf{Z}$, respectively. Since the T summation terms in (10) are independent, obtaining the minimum of (10) can be converted into solving the following T sub-problems

$$\mathbf{z}_t{}^{(k+1)} = \arg\min_{\mathbf{z}_t} ||\mathbf{x}_t - \mathbf{Y}^{(k)}\mathbf{z}_t||_p^p, \tag{11}$$

where $\mathbf{z}_t{}^{(k+1)}$ is the tth column of $\mathbf{Z}^{(k+1)}$, that is $\mathbf{Z}^{(k+1)} = [\mathbf{z}_1{}^{(k+1)}, \cdots, \mathbf{z}_T{}^{(k+1)}]$.

Finally, the vector ℓ_p-norm minimization problem (11) is solved by the complex-valued newton's method [8].

The convergence for the iterations of $\mathbf{Y}^{(k)}$ and $\mathbf{Z}^{(k)}$ is determined by checking whether the following inequality holds

$$\frac{J_p\left(\mathbf{Y}^{(k)}, \mathbf{Z}^{(k)}\right) - J_p\left(\mathbf{Y}^{(k+1)}, \mathbf{Z}^{(k+1)}\right)}{J_p\left(\mathbf{Y}^{(k)}, \mathbf{Z}^{(k)}\right)} < \varepsilon, \tag{12}$$

for some small tolerance ε.

After obtaining $\mathbf{Y}$ through solving the optimization problem (7), the projection matrix onto the noise subspace is computed as

$$\mathbf{P}_n = \mathbf{I} - \mathbf{Y}(\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H. \tag{13}$$

Then the ℓ_p-MUSIC spatial spectrum is given by

$$P_{\ell_p-\text{MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{P}_n\mathbf{a}(\theta)}, \quad -90^\circ \le \theta \le 90^\circ. \tag{14}$$

The DOA estimates can be obtained by searching for the peaks of (14).

The key of ℓ_p-MUSIC is the utilization of the ℓ_p-norm based objective function which is more robust against non-Gaussian impulsive noise than the L2 norm based objective function. However, the ℓ_p-norm based objective function is far from optimal for $S\alpha S$ distributions, especially for highly impulsive situations. The degeneration of ℓ_p-MUSIC in extremely low GSNR situations or highly impulsive noise environments, as we show in the simulations in Section 4, can also verify this point. In order to improve the performance in severe impulsive noise situations, in the following Section, by examining the score functions of $S\alpha S$ distributions, we propose a new cost function as a substitute for the ℓ_p-norm based objective function in (7).

3. SC-MUSIC

3.1. Cauchy score function based cost function

Consider the problem of estimating some unknown parameters of a deterministic signal received in additive impulsive noise, where $\{x_i\}$ $(i = 1, \cdots, N)$ are the received data, and ξ is the parameter to be estimated. The M-estimator for ξ is given by

$$\hat{\xi} = \arg\min_{\xi} \frac{1}{N}\sum_{i=1}^{N} \rho(e_i; \xi), \tag{15}$$

where e_i $(1 \le i \le N)$ are the errors produced by the system during supervised learning and $\rho(\cdot)$ is the M-estimate function [9]. Specific to our problem, the information of ξ is fully embedded in the errors $e_i (1 \le i \le N)$. Therefore, in the following content, we simplify $\rho(e_i; \xi)$ as $\rho(e)$. It is easily to find that, if $\rho(e)$ is chosen as $-\log f(e)$, the M-estimator gives the ordinary maximum likelihood estimator (MLE), where $f(e)$ is the probability density function (PDF) of $\{e_i\}$ $(i = 1, \cdots, N)$. However, it is hard to directly employ the ML method in $S\alpha S$ situations. The reason is that, for $S\alpha S$ distributions, there generally does not exist the closed form of PDF, except for some specific distributions (e.g. Gaussian distribution for α=2, and Cauchy distribution for $\alpha = 1$).

Differentiating (15) yields a set of estimating equations for the parameter of interest ξ

$$\sum_{i=1}^{N} \rho'(e_i) = 0. \tag{16}$$

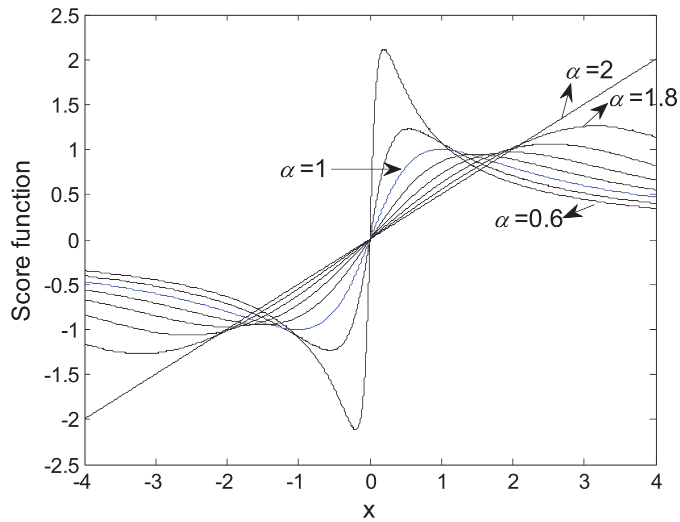


Fig. 1. Score functions of $S\alpha S$ distributions for different characteristic exponent α's. (blue line: $\alpha = 1$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Associating (16) with ML estimation, we have that

$$\sum_{i=1}^{N} -\frac{f'(e_i)}{f(e_i)} = 0, \tag{17}$$

where $\varphi(\cdot) = -\frac{f'(\cdot)}{f(\cdot)}$ is the score function of the distribution.

Based on the above analysis, in the following content, our intention is to seek the suitable M-estimation function $\rho(e)$ for the robust estimation of the parameter ξ by a thorough investigation of the score functions of $S\alpha S$ distributions. An integral expression for the zero-centered, unit dispersion, $S\alpha S$ density is given by

$$f_\alpha(x) = \frac{\alpha}{(1-\alpha)\pi}|x|^{\frac{1}{\alpha-1}} \int_0^{\pi/2} \upsilon(\vartheta) e^{-\upsilon(\vartheta)|x|^{\frac{-\alpha}{1-\alpha}}} d\vartheta, \quad \alpha \neq 1, \tag{18}$$

where $\upsilon(\vartheta) = \frac{1}{(sin\alpha\vartheta)^{\alpha/\alpha-1}} \cos[(\alpha-1)\vartheta](cos\vartheta)^{1/\alpha-1}$. And for $\alpha = 1$, $f_1(x) = \frac{1}{\pi(1+x^2)}$, which corresponds to the Cauchy distribution. It is worthy to note that, for $\alpha = 2$, $f_2(x) = \frac{1}{2\sqrt{\pi}}e^{-x^2/4}$, which corresponds to the Gaussian distribution with zero mean and the variance 2.

In order to find a cost function in the M-estimator (15) that yields robust performance in the $S\alpha S$ noise case, we examine the behavior of score functions of $S\alpha S$ distributions with different characteristic exponent α's. Fig. 1 shows the numerical evaluations of the score functions $\varphi(x)$ as α varies from 0.6 to 2 with the step size 0.2, respectively.

Excluding the Gaussian case, these score functions are all odd and bounded. In addition, it can be shown that for large values of x, all score functions behave similarly, i.e., $\lim_{x\to\infty}\varphi(x) \propto (1/x)$ for every $\alpha < 2$. As analyzed, from the viewpoint of ML estimation, the best choice of the nonlinear M-estimate cost function for an M-estimator is only decided by its score function. From the relationship between the score function and the best M-estimate cost function we can derive that, if the underlying noise is Gaussian distributed, the best M-estimate cost function is $\rho(e) = \frac{e^2}{2}$, which corresponds to the conventional LS (least square) cost function. And if the underlying noise is Cauchy distributed, the best M-estimate cost function is $\rho(e) = log(1 + e^2)$. Fig. 2 depicts the contours of the distance from $\mathbf{e}$ to the origin in a 2-D space of these two M-estimate cost functions as well as the employed cost function in l_p-MUSIC, i.e. $\rho(e) = |e|^{1.1}$.

From Fig. 2 we can observe that, for the LS M-estimate cost function $\rho(e) = \frac{e^2}{2}$, the quadratic increase function has the net

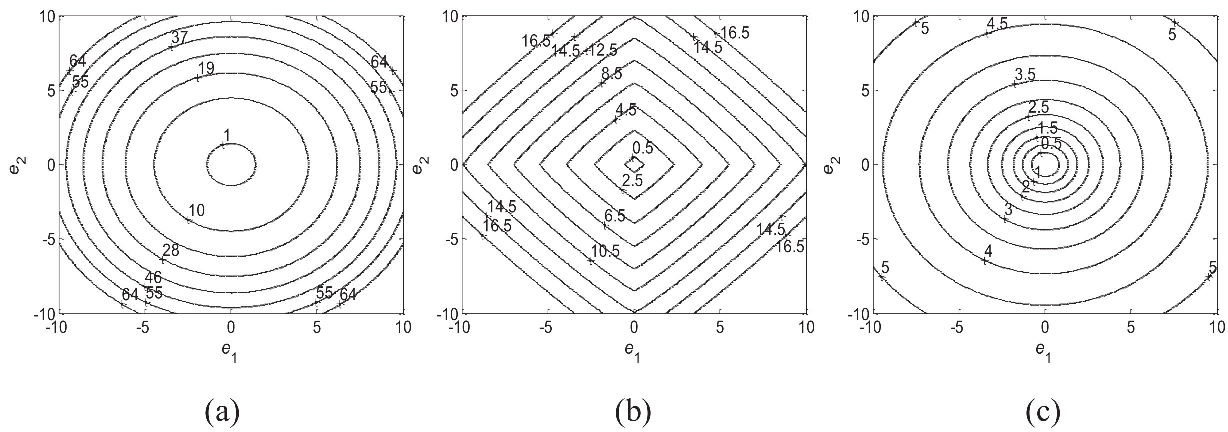


Fig. 2. Contours of the distance from **e** to the origin in a 2-D space of three M-estimate cost functions. (a) $\rho(e)=\frac{e^2}{2}$; (b) $\rho(e)=|e|^{1.1}$; (c) $\rho(e)=log(1+e^2)$.

effect of amplifying the contribution of samples that are far away from the mean value of the error distribution. This interprets why Gaussian distributed residuals provide optimality for the LS procedure while other data distributions will make the LS non-optimal, in particular if the error distribution has outliers. On the other side, by means of the attenuation on the large error terms, M-estimate cost functions $\rho(e)=|e|^{1.1}$and $\rho(e)=log(1+e^2)$ show adaptability to outliers. We can also observe from Fig. 2 that, the attenuation on the large error terms of $\rho(e)=log(1+e^2)$ is much stronger than that of $\rho(e)=|e|^{1.1}$, which indicates that $\rho(e)=log(1+e^2)$ is more suitable for handling impulsive noise environments. In addition, $\rho(e)=log(1+e^2)$ is continuous and differentiable everywhere indicating it is more suitable for developing gradient descent algorithms.

Combining Figs. 1 and 2 we can also deduce that, the smaller the characteristic exponent α is, the stronger the attenuation for the large error terms should be. Suppose the underlying noise is $S\alpha S$ distributed but the characteristic exponent α is neither 1 nor 2, the best M-estimation cost function cannot be obtained because the closed form of PDF functions do not exist for these distributions. Hence, choosing the score function of Cauchy distribution to handle the impulsive noise is straightforward and logical. The reasons we choose the score function of Cauchy distribution for the nonlinearity in M-estimate also lie in that, (1) It is simple and easy to be implemented. (2) The underlying impulsive noise in real applications can basically be well modeled by $S\alpha S$ distributions with the characteristic exponent $\alpha \in$ [12], while the extremely impulsive noise environments with $\alpha < 1$ is hardly encountered in real applications [3].

Back to the estimate for the parameter ξ in (15), based on the above analysis, we can formulate the following M-estimator as

$$\hat{\xi}=\arg\min_{\xi}\frac{1}{N}\sum_{i=1}^{N}log(1+e_i^2). \tag{19}$$

Obviously, the M-estimator (19) which is developed on the score function of Cauchy distribution is optimal when the noise is Cauchy distributed. Actually, as shown in Ref. [10], the ML estimator of a deterministic signal in an additive white Cauchy distributed noise is robust to deviations in the characteristic exponent α. Therefore, we choose the estimator (19) to handle the $S\alpha S$ noise situations with different characteristic exponents. Based on this estimator, we develop the SC-MUSIC method to gain more robust DOA estimation in impulsive noise environments.

3.2. SC-MUSIC method

As described in Section 2.1, the data received at the sensor array is expressed as the matrix $\mathbf{X}$, in the absence of noise, $\mathbf{X}$ has the low-rank decomposition $\mathbf{X}=\mathbf{YZ}$, where $\mathbf{Y}$ is an $M\times Q$ full column rank matrix and $\mathbf{Z}$ is a $Q\times T$ full row rank matrix, in which, M, Q, T represent the number of the array sensors, the number of the incoming signals, and the sample length of the received data, respectively.

For the case of $S\alpha S$ distributed noises, the residual fitting error matrix is defined as

$$\mathbf{X}-\mathbf{YZ}=[\mathbf{r}_1,\cdots,\mathbf{r}_T], \tag{20}$$

where $\mathbf{r}_t=\mathbf{x}_t-\mathbf{Y}\mathbf{z}_t$, in which, $\mathbf{x}_t$ and $\mathbf{z}_t$denote the tth vectors of $\mathbf{X}$ and $\mathbf{Y}$, respectively [8].

Based on M-estimate theory, we can define the following objective function

$$\min_{\mathbf{Y},\mathbf{Z}}J_\rho=\min_{\mathbf{Y},\mathbf{Z}}\frac{1}{T}\sum_{t=1}^{T}\rho(\mathbf{r}_t),\quad t=1,\cdots,T, \tag{21}$$

to estimate $\mathbf{Y}$ and $\mathbf{Z}$. Where $\rho(\cdot)$ is an M-estimate function. Moreover, if we execute transpose on the residual fitting error matrix, i.e. $(\mathbf{X}-\mathbf{YZ})^T=(\mathbf{X}^T-\mathbf{Z}^T\mathbf{Y}^T)=[\mathbf{q}_1,\cdots,\mathbf{q}_M]$, where $\{\mathbf{q}_j\}_{j=1}^{M}$ are the column vectors of $(\mathbf{X}-\mathbf{YZ})^T$, we can also define the following objective function as

$$\min_{\mathbf{Y},\mathbf{Z}}J_\rho=\min_{\mathbf{Y},\mathbf{Z}}\frac{1}{M}\sum_{m=1}^{M}\rho(\mathbf{q}_m),\quad m=1,\cdots,M, \tag{22}$$

to obtain an estimate for $\mathbf{Y}$ and $\mathbf{Z}$.

The following alternating convex optimization approach can be utilized to solve (21) and (22)

$$\mathbf{Z}^{(k+1)}=\arg\min_{\mathbf{Z}}\frac{1}{T}\sum_{t=1}^{T}\rho(\mathbf{x}_t-\mathbf{Y}^{(k)}\mathbf{z}_t),\quad t=1,\cdots,T, \tag{23}$$

$$\mathbf{Y}^{(k+1)}=\arg\min_{\mathbf{Y}}\frac{1}{M}\sum_{m=1}^{M}\rho(\mathbf{x}_{(m)}^T-(\mathbf{Z}^{(k+1)})^T\mathbf{y}_{(m)}^T),\quad m=1,\cdots,M, \tag{24}$$

where $\mathbf{Z}^{(k+1)}$ and $\mathbf{Y}^{(k+1)}$ are the obtained $(k+1)$th iteration results, $\mathbf{x}_{(m)}$ and $\mathbf{y}_{(m)}$ denotes the mth row vectors of $\mathbf{X}$ and $\mathbf{Y}$, respectively. By the formulation of (23) and (24), we can observe that $\mathbf{Y}$ and $\mathbf{Z}$ can be solved in a similar manner. Therefore, in the following content, we only discuss the solution for (23). The solution for (24) can be handled in the same manner and it is hence omitted in this paper.

Based on the independent assumptions of signals and noises, the minimum optimization of (23) can be obtained by solving the following T sub-problems separately

$$\mathbf{z}_t^{(k+1)} = \arg\min_{\mathbf{z}_t} \rho(\mathbf{x}_t - \mathbf{Y}^{(k)}\mathbf{z}_t), \quad t = 1, \cdots, T, \tag{25}$$

where $\mathbf{z}_t^{(k+1)}$ is the tth column vector of $\mathbf{Z}^{(k+1)}$, that is, $\mathbf{Z}^{(k+1)} = [\mathbf{z}_1{}^{(k+1)}, \cdots, \mathbf{z}_T{}^{(k+1)}]$. For notation simplicity, the superscripts and subscripts in (25) are omitted, then, (25) can be re-expressed as

$$\mathbf{z}^{(k+1)} = \arg\min_{\mathbf{z}} \rho(\mathbf{x} - \mathbf{Y}\mathbf{z}). \tag{26}$$

Defining the residual fitting error vector $\mathbf{r} = \mathbf{x} - \mathbf{Y}\mathbf{z} = [r_1, \cdots, r_M]^T$, obviously, $r_1, \cdots, r_M$ can be modeled by $S\alpha S$ distribution. Based on the discussion in Section 3.1, we can construct the following objective function to solve $\mathbf{z}$

$$\begin{aligned}\mathbf{z}^{(k+1)} &= \arg\min_{\mathbf{z}} \rho(\mathbf{r}) \\ &= \arg\min_{\mathbf{z}} \sum_{i=1}^{M} log(1 + |r_i|^2),\end{aligned} \tag{27}$$

where r_i is the ith entry of the residual fitting error vector $\mathbf{r} = \mathbf{x} - \mathbf{Y}\mathbf{z}$. It is easily to deduce that, the estimator of $\mathbf{z}$ developed on the objective function (27) is optimal when the noise is Cauchy distributed. However, as analyzed in Section 3.1, this estimator can also achieve robust results even if the characteristic exponent α is different from 1. The simulations provided in Section 4 can also verify this point.

We adopt complex-valued pseudo-Newton's method to solve (27). Defining $f(\mathbf{z}) = \sum_{i=1}^{M} log(1 + ||r_i||_2^2)$, the partial derivative of f w.r.t the complex quantity $\mathbf{z}$ can be expressed as

$$g(\mathbf{z}) = \frac{\partial f}{\partial \mathbf{z}} = \left(\frac{\partial \mathbf{r}}{\partial \mathbf{z}}\right)^T \frac{\partial f}{\partial \mathbf{r}^*} = -\mathbf{Y}^H \frac{\partial f}{\partial \mathbf{r}^*} = -\mathbf{Y}^H\mathbf{W}\mathbf{r}, \tag{28}$$

where the diagonal matrix $\mathbf{W} = \mathrm{diag}\{1/(1 + |r_1|^2), \cdots, 1/(1 + |r_M|^2)\}$. And then, the $Q \times Q$ partial Hessian matrix of f w.r.t $\mathbf{z}$, denoted by $\mathbf{H}_{\mathbf{z}^*\mathbf{z}}$ is

$$\mathbf{H}_{\mathbf{z}^*\mathbf{z}} = \frac{\partial^2 f}{\partial \mathbf{z}^* \partial \mathbf{z}^T} = \left(\frac{\partial \mathbf{r}}{\partial \mathbf{z}^H}\right)^T \mathbf{H}_{\mathbf{r}^*\mathbf{r}} \frac{\partial \mathbf{r}}{\partial \mathbf{z}^T} = \mathbf{Y}^H\mathbf{H}_{\mathbf{r}^*\mathbf{r}}\mathbf{Y}, \tag{29}$$

where $\mathbf{H}_{\mathbf{r}^*\mathbf{r}}$ denotes the $M \times M$ partial Hessian matrix of f w.r.t $\mathbf{r}$, and has the form

$$\mathbf{H}_{\mathbf{r}^*\mathbf{r}} = \frac{\partial^2 f}{\partial \mathbf{r}^* \partial \mathbf{r}^T} = \mathrm{diag}\{1/(1 + |r_1|^2)^2, \cdots, 1/(1 + |r_M|^2)^2\}.$$

Consequently, the sequence $\{\mathbf{z}^{(k)}\}(k = 0, 1, \cdots)$ can be iteratively obtained as

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \mu_k \Delta\mathbf{z}^{(k)}, \tag{30}$$

where $\Delta\mathbf{z}^{(k)} = -\mathbf{H}_{\mathbf{z}^*\mathbf{z}}^{-1}\mathbf{g}(\mathbf{z}^{(k)})$, and μ_k is a positive step size. For a given Newton direction $\Delta\mathbf{z}^{(k)}$, the optimal step size μ_k is given by solving the line search

$$\mu_k = \arg\min_{\mu_k \geq 0} \sum_{i=1}^{M} log(1 + |r_i|^2). \tag{31}$$

This is a one-dimensional optimization problem and can be solved by the existing line search techniques, such as Golden section search or tangential method [11].

At last, the convergence of our method can be determined by checking whether the following inequality hold:

$$\frac{J_\rho(\mathbf{Y}^{(k)}, \mathbf{Z}^{(k)}) - J_\rho(\mathbf{Y}^{(k+1)}, \mathbf{Z}^{(k+1)})}{J_\rho(\mathbf{Y}^{(k)}, \mathbf{Z}^{(k)})} < \varepsilon, \tag{32}$$

for some small tolerance ε.

The proposed SC-MUSIC algorithm can be summarized as follows:

Given the received data matrix $\mathbf{X}$, the number of incoming sources Q and the tolerance $\varepsilon = 10^{-6}$.

S1. Initialize $\mathbf{Y}^{(0)}$ with a random matrix of full column rank.

S2. For $k = 1, 2, \cdots$, do until converge

S2.1 Initialize $\{\mathbf{z}_t^{(0)}\}(t = 1, \cdots, T)$ with the least-squares solutions, that is, $\mathbf{z}_t^{(0)} = (\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H\mathbf{x}_t$, $t = 1, \cdots, T$.

S2.2 Utilize (30) to update $\{\mathbf{z}_t^{(k)}\}$ $(t = 1, \cdots, T)$.

S2.3 Calculate $\mathbf{Z}^{(k)} = [\mathbf{z}_1{}^{(k)}, \cdots, \mathbf{z}_T{}^{(k)}]$.

S2.4 Obtain $\mathbf{Y}^{(k+1)} = [\mathbf{y}_1^{(k+1)}, \cdots, \mathbf{y}_M^{(k+1)}]$ by a similar manner as $\mathbf{Z}^{(k)}$.

S2.5 Check the stopping condition in (32).

end For.

S3. Compute the projection matrix onto the noise $\mathbf{P_n} = \mathbf{I} - \mathbf{Y}(\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H$.

S4. Compute the spatial spectrum of (33) and search for its peaks as the DOA estimates.

$$P_{\mathrm{SC-MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{P_n}\mathbf{a}(\theta)}, \quad -90^\circ \leq \theta \leq 90^\circ. \tag{33}$$

3.3. Computational complexity of SC-MUSIC

Let's review the updating equation of $\{\mathbf{z}_t^{(k)}\}$ $(t = 1, \cdots, T)$ in the kth iteration in SC-MUSIC:

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \mu_k \Delta\mathbf{z}^{(k)}, \tag{34}$$

where $\Delta\mathbf{z}^{(k)} = -\mathbf{H}_{\mathbf{z}^*\mathbf{z}}^{-1}\mathbf{g}(\mathbf{z}^{(k)})$ with $g(\mathbf{z}) = -\mathbf{Y}^H\mathbf{W}\mathbf{r}$ and $\mathbf{H}_{\mathbf{z}^*\mathbf{z}} = \mathbf{Y}^H\mathbf{H}_{\mathbf{r}^*\mathbf{r}}\mathbf{Y}$, in which, $\mathbf{H}_{\mathbf{r}^*\mathbf{r}} = \mathrm{diag}\{1/(1 + |r_1|^2)^2, \cdots, 1/(1 + |r_M|^2)^2\}$. It can be derived that the complexity of matrix multiplication $\mathbf{Y}^H\mathbf{W}\mathbf{r}$ and $\mathbf{Y}^H\mathbf{H}_{\mathbf{r}^*\mathbf{r}}\mathbf{Y}$ is $O(MQ)$ and $O(MQ^2)$, respectively. While, the complexity for computing $\Delta\mathbf{z}^{(k)} = -\mathbf{H}_{\mathbf{z}^*\mathbf{z}}^{-1}\mathbf{g}(\mathbf{z}^{(k)})$ is $O(Q^3)$. Thus, the complexity of solving $\{\mathbf{z}_t^{(k)}\}$ is $O(MQ^2)$ due to $M > Q$. Clearly, the numerical complexity of SC-MUSIC is $O(TMQ^2)$ in each iteration.

4. Simulations

In the simulations, we assume two independent QPSK (quadrature phase shift keying) signals of the same power impinging on a uniformly linear array of sensors which space a half-wavelength apart. The underlying noise is assumed to be complex isotropic $S\alpha S$ distributed. For $\alpha = 2$, it is simplified to be Gaussian distributed. For $\alpha < 2$, due to infinite variance, the generalized signal to noise ratio (GSNR) defined by

$$\mathrm{GSNR} = 10\log\left(\frac{1}{\gamma T}\sum_{t=1}^{T} |s(t)|^2\right), \tag{35}$$

is utilized to evaluate the ratio of the signal power over noise dispersion.

After obtaining the signal subspace estimate $\mathbf{Y}$, the root-MUSIC [12] method is utilized.to retrieve the DOA estimates.

4.1. Cramer–Rao bound

Kozick and Sadler have derived a general expression of the CRB of the DOA parameters $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_Q]^T$ for non-Gaussian noise [13], which can be shown as

$$\mathrm{CRB}^{-1}(\theta) = \frac{1}{I_c}\sum_{t=1}^{T} \Re\left\{\mathbf{S}^H(t)\mathbf{D}^H\left[\mathbf{I} - \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\right]\mathbf{D}\mathbf{S}(t)\right\}, \tag{36}$$

where $\mathbf{S}(t) = diag\{s_1(t), \cdots s_Q(t)\}$ is a diagonal matrix, $\mathbf{D} = [\frac{\partial \mathbf{a}(\theta_1)}{\partial \theta_1}, \cdots, \frac{\partial \mathbf{a}(\theta_Q)}{\partial \theta_Q}]$ is the differential of the array steering matrix

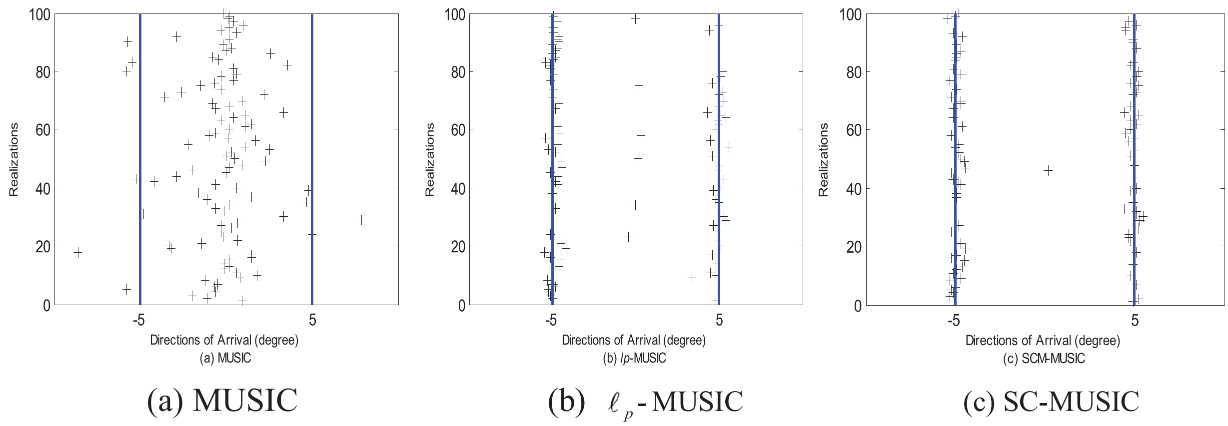


Fig. 3. 50 realizations of MUSIC, ℓ_p-MUSIC, and SC-MUSIC. (blue solid line: real DOAs). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

A, T is the number of data samples, and $\Re(\cdot)$ is the real part of a complex number. The scaling factor I_c can be computed by

$$I_c = \pi \int_0^{\infty} \frac{[f'(\upsilon)]^2}{f(\upsilon)} \upsilon \mathrm{d}\upsilon, \tag{37}$$

where $\upsilon = |e|$ is the modulus of the complex non-Gaussian variable e, and $f'(\upsilon)$ represents the derivative of PDF $f(\upsilon)$. For $S\alpha S$ distribution, due to the lack of the closed form of PDF except for the cases $\alpha = 1$ and $\alpha = 2$, the scaling factor I_c is analytically tractable only for Gaussian distribution ($\alpha = 1$) and Cauchy distribution ($\alpha = 2$). As suggested by [12], the CRB scaling factor for $1 < \alpha < 2$ can be approximated by linear interpolation of the value at $\alpha = 1$ and $\alpha = 2$.

Given the PDF $f(\upsilon)$ of the complex Gaussian distributed variable e

$$f(\upsilon) = \frac{1}{4\pi\gamma} \exp\left(-\upsilon^2/(4\gamma)\right), \tag{38}$$

where γ is the dispersion, and the PDF $f(\upsilon)$ of the complex Cauchy distributed variable e

$$f(\upsilon) = \frac{\gamma}{2\pi\left(\upsilon^2 + \gamma^2\right)^{3/2}}, \tag{39}$$

we can derive that the scaling factors for Gaussian distribution and Cauchy distribution are $I_c = \frac{1}{2\gamma}$ and $I_c = \frac{3}{5\gamma^2}$, respectively.

4.2. Experiment 1

In the first experiment, the estimation capability of SC-MUSIC is examined and compared with the traditional MUSIC and the ℓ_p-MUSIC. The number of the array sensors is 7. A fairly strong impulsive noise environment with $\alpha = 1.3$ is employed, the GSNR is set at 8 dB, and the number of data samples available at each sensor is 100. Two directions of arrival are $\theta_1 = -5°$, and $\theta_2 = 5°$, respectively. Fig. 3 shows 50 realizations of the DOA estimates of MUSIC, ℓ_p-MUSIC, and SC-MUSIC, respectively. As shown in Fig. 3, for such a given severe impulsive noise environment, the traditional MUSIC fails to distinguish the two incoming signals in most runs of the 50 realizations. By contrast, ℓ_p-MUSIC and SC-MUSIC present a much better adaptability to the impulsive noise situation. However, as we can observe from Fig. 3 that, the estimation results of our proposed SC-MUSIC present a better gathering to the real DOAs which indicates a more accurate DOA estimation.

In order to further examine the estimation accuracy of the proposed algorithm, in the following simulations, we assume two independent QPSK signals of the same power impinging on a uniformly linear array of 7 sensors. The robustness of SC-MUSIC and five other MUSIC algorithms which are developed to deal with the DOA estimation in $S\alpha S$ noise environments, including ℓ_p-MUSIC [8], ROC-MUSIC [4], FLOM-MUSIC [5], PFLOM-MUSIC [6], and CRCO-MUSIC [7], is compared. In each simulation, five hundred Monte-Carlo trials are run for each algorithm, and two quantities, namely, the probability of resolution and the RMSE (root mean square error) are evaluated for comparison. Among which, the two signals are considered to be resolvable if the following resolution criterion holds:

$$\Lambda(\theta_1, \theta_2) \stackrel{\mathrm{d}}{=} S(\theta_m) - \frac{1}{2}\{S(\theta_1) + S(\theta_2)\} > 0, \tag{40}$$

where θ_1 and θ_2 are the two arrival angles for the two signals, $\theta_m = (\theta_1 + \theta_2)/2$ is the mid-range between them, and the null spectrum $S(\theta) \stackrel{\mathrm{d}}{=} 1/P(\theta)$ is defined as the reciprocal of the MUSIC spectrum [4]. The RMSE of the DOA estimates was calculated by taking into consideration only the Monte-Carlo runs for which the two sources are resolved, that is,

$$\mathrm{RMSE} = \sqrt{\frac{1}{K_s Q} \sum_{k=1}^{K_s} \sum_{q=1}^{Q} \left(\hat{\theta}_{q,k} - \theta_q\right)^2}, \tag{41}$$

where K_s represents the number of the Monte-Carlo runs of success resolution, θ_q represents the real DOA of the qth signal, and $\hat{\theta}_{q,k}$ represents the estimate to θ_q in the kth Monte-Carlo run.

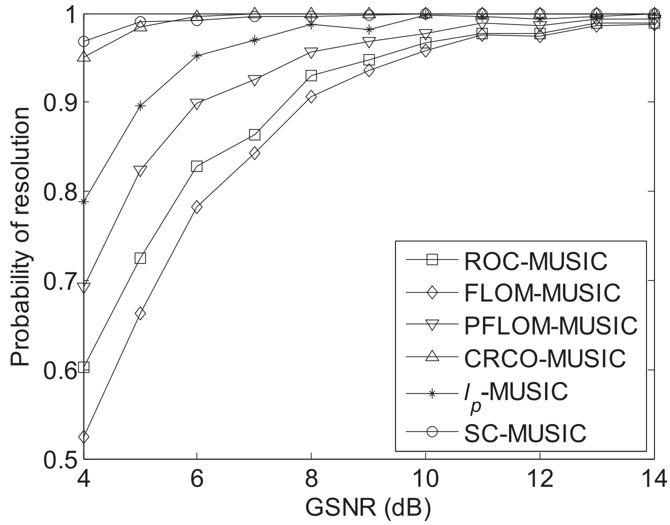
In our simulations, as recommended in Ref. [4–6], $p = 1.1$ is employed in ℓ_p-MUSIC, ROC-MUSIC, FLOM-MUSIC, and PFLOM-MUSIC. For CRCO-MUSIC, the parameter μ and the kernel σ are selected in the same way as in Ref. [7]. The performance of these algorithms is reviewed as a function of three parameters, namely, the GSNR, the characteristic exponent α, and the number of data samples.

4.3. Experiment 2 effect of GSNR

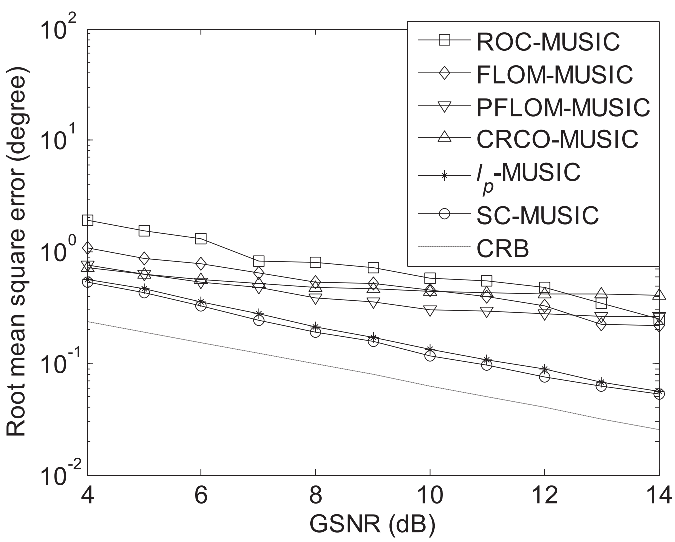
In this experiment, the underlying noise is set to be moderate impulsive with the characteristic exponent $\alpha = 1.5$. The number of the data samples is $T = 100$. The performance of these six algorithms as a function of GSNR is reviewed in this experiment and the results are depicted in Fig. 4.

It can be observed from Fig. 4(a) and (b) that, as the GSNR increases, for each algorithm, the resolution probability improves and the RMSE decreases as expected. However, the RMSEs of the ℓ_p-MUSIC and the proposed SC-MUSIC demonstrate a distinct improvement over the other four algorithms.

Another observation is that, although as shown in Fig. 4(b), the ℓ_p-MUSIC and the proposed SC-MUSIC present equivalent RMSE re-

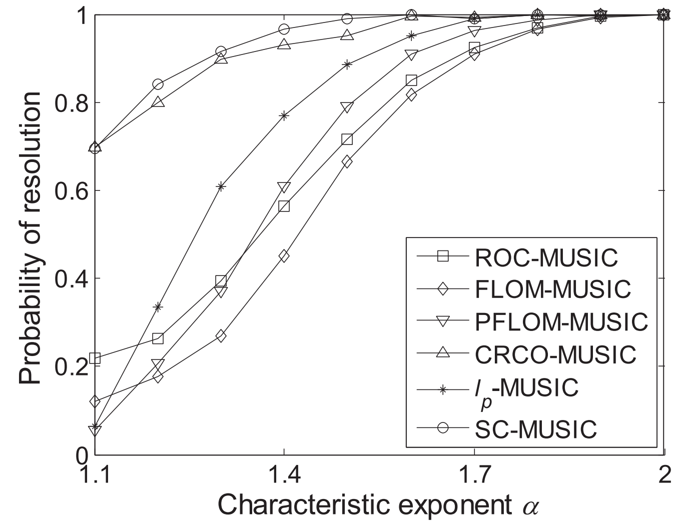


(a)

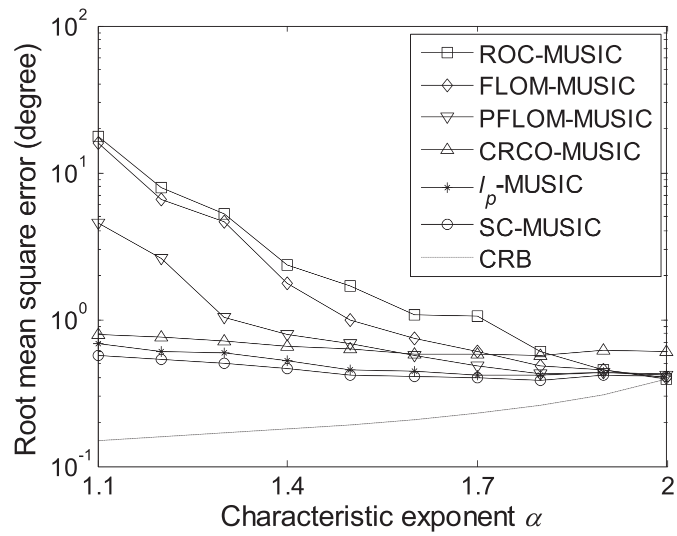


(b)

Fig. 4. Performance comparison of the algorithms as a function of the GSNR ($\alpha = 1.5, T = 100$). (a) Probability of resolution. (b) Root mean square error.



(a)



(b)

Fig. 5. Performance comparison of the algorithms as a function of characteristic exponent α (GNSR = 5 dB, $T = 100$). (a) Probability of resolution. (b) Root mean square error.

sults, the probability resolution of the SC-MUSIC is much higher than that of the ℓ_p-MUSIC, especially for the low GSNR values, indicating that the proposed SC-MUSIC can achieve a better estimation accuracy than the ℓ_p-MUSIC does.

4.4. Experiment 3 *effect of characteristic exponent α*

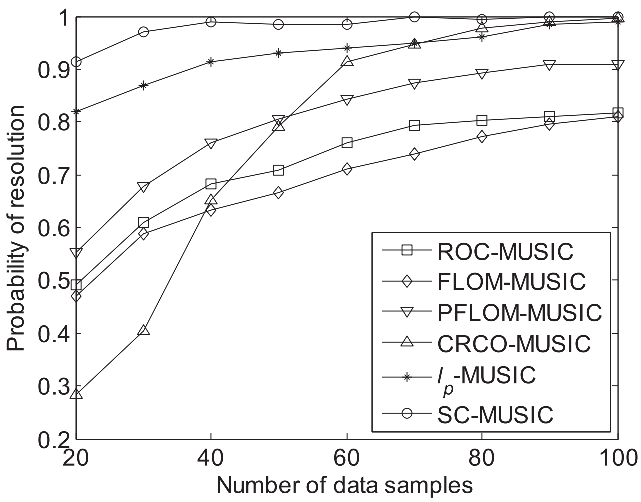
In this experiment, the GSNR is set at 5 dB. And the number of the data samples is $T = 100$. Fig. 5 demonstrates the performance comparison of these six algorithms along with the characteristic exponent α varies from 1.1 to 2, which indicates a transition from highly impulsive noise situations to Gaussian noise environments.

From Fig. 5(a) and (b) we can observe that, the proposed SC-MUSIC demonstrates the most superior performance in both resolution probability and RMSE for each characteristic exponent α. Moreover, as shown in Fig. 5, the smaller the characteristic exponent α gets, the more significant performance improvement the SC-MUSIC achieves, which indicates that the proposed SC-MUSIC is more suitable for highly impulsive noise environments than the other algorithms.

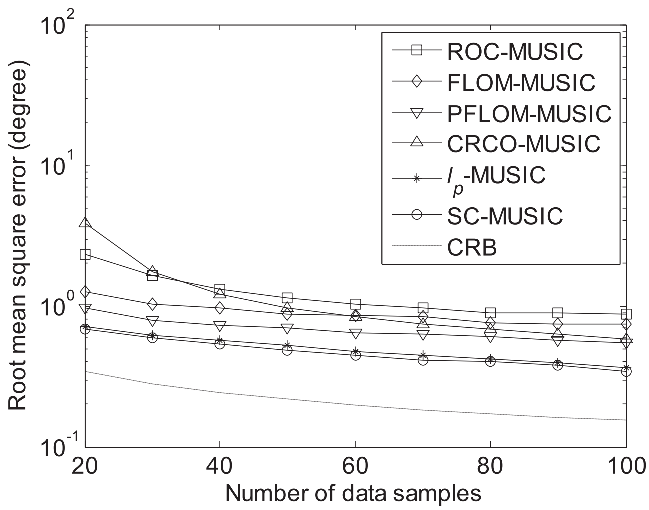
4.5. Experiment 4 *effect of number of snapshots*

In this experiment, the characteristic exponent α of the underlying noise is set as 1.5, and the GSNR is set as 6 dB. Fig. 6 demonstrates the performance comparison of these six algorithms as the number of the data samples increase from 20 to 100 under a step size 10. It can be seen from Fig. 6 that, when the available data samples are few, the FLOS based MUSIC algorithms, including the ROC-MUSIC, the FLOM-MUSIC, and the PFLOM-MUSIC, degrade in the performance. In contrast, the ℓ_p-MUSIC and the proposed SC-MUSIC gain better performance in both resolution probability and RMSE. In addition, the SC-MUSIC achieves the best performance among these algorithms.

Another interesting observation of Fig. 6 is the behavior of the CRCO-MUSIC, that is, when the data samples available are few, the performance of the CRCO-MUSIC degrade dramatically, which indicates that the CRCO-MUSIC is not suitable for the DOA estimation with few data samples.



(a)



(b)

Fig. 6. Performance comparison of the algorithms as a function of number of snapshots ($\alpha = 1.5$, GSNR = 6 dB). (a) Probability of resolution. (b) Root mean square error.

5. Conclusions

In this paper, by investigating the score functions of $S\alpha S$ distribution, a novel Cauchy score function based cost function of the residual fitting error matrix is proposed and employed as a substitute for the ℓ_p-norm based cost function in the ℓ_p-MUSIC algorithm. Alternating convex optimization is applied for the solution of the cost function. And the complex-valued Newton's method with optimal step size is employed to solve the resulting convex problem. With the estimated signal subspace, the DOA estimates are obtained by the MUSIC technique. Simulation comparisons demonstrate that, with less data samples, the new proposed algorithm can achieve more robust DOA estimates than its counterpart, especially in fairly low GSNR or extremely strong impulsive noise environments.

Acknowledgments

The authors acknowledge the support from National Natural Science Foundation of China under Grants 61501301, 61172108, 61139001, 81241059, 61261046, and the support from Natural Science Foundation of SZU under grant 201413, and would like to thank the anonymous reviewers for their useful comments and suggestions that significantly improved the paper.

References

[1] R.O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (1986) 276–280.
[2] D. Johnson, D. Dudgeon, Array Signal Processing: Concepts and Techniques, Prentice-Hall, Englewood Cliffs. NJ, 1993.
[3] C.L. Nikias, M. Shao, Signal Processing with α-Stable Distribution and Applications, John Wiley & Sons, 1995.
[4] P. Tsakalides, C.L. Nikias, Robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments, IEEE Trans. Signal Process. 44 (1996) 1623–1633.
[5] T.H. Liu, J.M. Mendel, A subspace-based direction finding algorithm using fractional lower order statistics, IEEE Trans. Signal Process. 49 (2001) 1605–1613.
[6] H. Belkacemi, S. Marcos, Robust subspace-based algorithms for joint angle/Doppler estimation in non-Gaussian clutter, Signal Process. 87 (2007) 1547–1558.
[7] J.F. Zhang, T.S. Qiu, A.M. Song, H. Tang, A novel correntropy based DOA estimation algorithm in impulsive noise environments, Signal Process. 104 (2014) 346–357.
[8] W.J. Zeng, H.C. So, L. Huang, ℓ_p-MUSIC: Robust direction-of-arrival estimator for impulsive noise environments, IEEE Trans. Signal Process. 61 (17) (2013) 4296–4308.
[9] P.J. Huber, Robust Statistics, Wiley, New York, 1981.
[10] J. Fiedmann, H. Messer, J.F. Cardoso, Robust parameter estimation of a deterministic signal in impulsive noise, IEEE Trans. Signal Process. 48 (4) (2000) 935–942.
[11] D. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, MA, USA, 1999.
[12] B.D. Rao, K.V.S. Hari, Performance analysis of root-MUSIC, IEEE Transa. Acoust. Speech Signal Process. 37 (12) (1990) 1939–1949.
[13] R.J. Kozick, B.M. Sadler, Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures, IEEE Trans. Signal Process. 48 (12) (2000) 3520–3535.