# A Novel DOA Estimation Method for Coherently Distributed Sources in the Presence of Impulsive Noise

Ruiyan Cai*
College of Information Engineering
Dalian University
Dalian, China
Email: 34010361@qq.com

Tianshuang Qiu
Faculty of Electronic Information
and Electrical Engineering
Dalian University of Technology
Dalian, China
Email: qiutsh@dlut.edu.cn

Quan Tian
Faculty of Electronic Information
and Electrical Engineering
Dalian University of Technology
Dalian, China
Email: tianquan10@163.com

***Abstract*—Many approaches have been studied for the direction of arrival (DOA) estimation of distributed sources under the additive Gaussian noise environments, but these schemes typically perform poorly when the noise is modeled as an alpha-stable distribution. Inspired by the Hampel identifier, this paper extends the definition of correntropy that exhibits a 'robust statistics' property under the impulsive noise environments to propose a novel generalized auto-correntropy (GCO) operator. For improving the performance of GCO, an adaptive kernel size function is deduced for symmetric alpha-stable (S$\alpha$S) distributed random variables. Since the adaptive kernel size function does not require any empirical parameter of the impinging signals, it is particularly suitable for cases of practical interest. Based on the advantage of GCO operator, a distributed sources parameter estimator (DSPE) like algorithm is proposed for the DOA estimation of coherently distributed (CD) sources in the presence of impulsive noise. Comprehensive simulation results show that the proposed algorithm precedes the existing algorithms concerning estimation accuracy and probability of resolution, especially under the highly impulsive noise environments.**

***Keywords—Alpha-stable distribution; Coherently distributed sources; DOA estimation; Hampel identifier; Correntropy; DSPE.***

## I. Introduction

Numerous studies on direction of arrival (DOA) estimation have indicated that the scholars of the academe and industry have paid more attention to it [1]. However, most of them assume that the target sources are points. For these point sources, due to the wide-band sources, sensor position error, etc., the algorithm estimated steering vector does not match the steering vector of the actual sources, which will cause the model error. In [2, 3], the performance of the DOA estimation methods of the modeling error, which is modeled as random distortions on the steering vector, is analyzed.

Among the subspace-based DOA estimation algorithms, the multiple signal classification (MUSIC) [4–6] is famous for its supper resolution under the assumption of point sources. However, in practice, the angular dispersion of the sources can even reach 10°, for example, multipath propagation and mobile channel communication [7]. In this case, the source can no longer be regarded as a point because it has the central angle and the angular spread, but is treated as a spatially distributed source.

The models for distributed sources contain two types: incoherently distributed (ID) sources and coherently distributed (CD) sources. For ID sources, we consider the signals, which come from different points of the same distributed source, are uncorrelated. However, in CD sources case, the received signal components are scaled replicas and delayed from different points of the same source [8]. In CD sources case, some DOA estimation algorithms have been proposed [9–11]. Since the angular dispersion mismatches between the actual steering vector and the estimated steering vector of the sources, the performance of MUSIC is significantly reduced. To solve the central DOA estimation problem of CD sources, a generalized MUSIC algorithm, namely distributed sources parameter estimator (DSPE), is proposed in [12]. On this basis, the signal and noise subspaces are extended into the field of distributed sources. This pseudo-composition is exploited in other subspace methods, such as WPSF [13] and DISPARE [14]. Since determining the effective dimension of the pseudo signal subspace depends on the unknown parameters of the received data of the array sensors, it becomes the main difficulty of the distribution source and point source algorithm based on subspace. Moreover, These algorithms also have high computational complexity.

Most DOA estimation methods assume that noise is additive and subjects to Gaussian distribution, which has finite second-order statistics. However, studies have shown that many noises and signals, which are encountered in practical applications, have been found to be non-Gaussian that may be either natural or human-made. For instance, the lightning in the atmosphere, the influence of mountains and sea waves, the switching transients in power lines and the underwater acoustic signals [15]. These types of the signals and noises often cause significant degradation in system performance optimized under the Gaussian assumption, that is to say, if the statistical characteristics of the noise deviate from the Gaussian, serious degradation in performance occurs [16]. Since the probability density function (pdf) contains heavy tails, the alpha-stable distribution, in which the heaviness of tails is controlled by a variable parameter $\alpha$, provides a more efficient mathematical tool than the Gaussian distribution to describe the signals and noises.

For improving the robustness of the subspace-based DOA estimation methods, many methods have been proposed to instead the second-order covariance, for example, the fractional lower order moments (FLOM) [17], and the phase fractional lower order moments (PFLOM) [18, 19]. By adopting the shift-invariant kernel method, new statistics, namely correntropy, is proposed in [20]. By quantifying the statistical distribution and temporal structure the two random processes, the correntropy is able to extract more information than the second-order moment. Ref. [21] introduces the correntropy into the traditional covariance matrix; the CRCO algorithm is constructed to suppress impulsive noise. The correntropy-based unitary (COBU) MUSIC algorithm is proposed for suppressing impulsive noise and reducing computational cost [22]. In recent approaches, with $\ell_p$-norm as the fidelity criterion, a class of impulsive noise suppression algorithms is proposed. To overcome the error accumulation and estimation bias of the two-step method of time delay estimation (TDE), [23] proposes the $\ell_p$-norm minimization algorithm. Ref. [24] devises robust greedy pursuit algorithms that are based on $\ell_p$-correlation in the $\ell_p$-space. On the other hand, to overcome the drawback of many techniques for matrix completion are not robust to outliers, the iterative $\ell_p$-regression and the alternating direction method of multipliers (ADMM) are derived in [25].

In the paper, we concentrate on passive radio monitoring and signal source target location with CD sources under the impulsive noise environments. The main contributions of this work are summarized as follows:

- Inspired by the Hampel identifier, we define a generalized auto-correntropy (GCO) operator that can effectively reduce the impulsive noise of the data samples. On this basis, this paper proposes a robust central DOA estimation algorithm for CD sources, namely GCO-DSPE.
- For the GCO operator, its kernel size of the kernel function directly affects the suppression of the impulsive noise. Furthermore, we deduce an adaptive kernel size function, which sets the kernel size only through the array outputs to ensure that GCO-DSPE algorithm can achieve the best performance under the impulsive noise environments without the need for other priors.

The organization of this paper is as follows. Section II gives a brief presentation of some background on the alpha-stable distributed noise and presents the signal model of CD sources for the sensor array outputs. By extending correntropy, a GCO operator for array signal processing in alpha-stable distributed noise is defined in Section III, and we also analyze the robustness of GCO operator in impulsive noise. Further, we propose the GCO-DSPE algorithm and deduce an adaptive kernel size function for central DOA estimation. Finally, simulation results for verifying the theoretical results are given in Section IV, and Section V contains some conclusions.

## II. Signals and Model

### A. alpha-stable distribution

We can describe the alpha-stable distribution by its characteristic function as follows:

***Definition:*** The distribution function of a random variable is stable if and only if its characteristic function [16, 22] is defined by

$$\varphi(t) = \exp\left\{\mathrm{j}\mu u - \gamma|t|^{\alpha}\left[1 + \mathrm{j}\beta \operatorname{sign}(t)\omega(t,\alpha)\right]\right\} \tag{1}$$

where

$$\omega(t,\alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ \frac{2}{\pi}\log|t| & \alpha = 1 \end{cases} \tag{2}$$

$$\operatorname{sign}(t) = \begin{cases} 1 & t > 0 \\ 0 & t = 0 \\ -1 & t < 0 \end{cases} \tag{3}$$

and

$$0 < \alpha \leq 2,\ -1 \leq \beta \leq 1,\ \gamma > 0,\ -\infty < \mu < +\infty \tag{4}$$

Thus, an alpha-stable distribution is determined by the parameters: $\alpha$, $\beta$, $\gamma$ and $\mu$:

- $\alpha$ is called the characteristic exponent that measures the thickness of the tails of the distribution. If $\alpha = 2$ or $\alpha = 1$, the alpha-stable distribution is degenerated into the Gaussian distribution or the Cauchy distribution, respectively.
- $\beta$ is a symmetry parameter, which determines the degree and sign of asymmetry. If $\beta = 0$, the alpha-stable distribution is symmetric about $\mu$, then, it is named symmetric alpha-stable ($S\alpha S$).
- $\gamma$ is a dispersion parameter that is similar to the variance of the Gaussian distribution.
- $\mu$ is a location parameter. For $S\alpha S$ distributions, if $0 < \alpha \leq 1$ or $1 < \alpha \leq 2$, $\mu$ corresponds to the median or the mean, respectively.

The alpha-stable distribution holds the generalized central limit theorem and the stability property. These properties are the main advantage that $S\alpha S$ can model for the statistical distribution of uncertainty [16]. The details of the $S\alpha S$ are shown in [26]. However, under the impulsive noise environments, the performance of the subspace-based DOA estimation methods will be significantly degraded because the covariance matrix of the $S\alpha S$ does not satisfy the boundedness [27].

### B. Signal Model and DSPE Algorithm

Consider $L$ uncorrelated narrow-band independent, complex isotropic CD sources with the locations $\{\theta_1, \theta_2, \cdots, \theta_L\}$ impinge on a uniform linear array (ULA). The ULA consists of $M$ identical omni-directional sensors. We assume that the delay spread caused by multipath propagation is smaller than the inverse bandwidth of the signal, so even if there is scattering, the narrow-band hypothesis is still valid [28]. Using the complex envelope representation, the signals vector $\boldsymbol{x}(t)$ received on the sensors can be expressed as

$$\boldsymbol{x}(t) = \sum_{i=1}^{L} \int_{\theta\in\Theta} \boldsymbol{a}(\theta) s_i(\theta, \psi_i)\, d\theta + \boldsymbol{n}(t) \tag{5}$$

where we have the following:

- $s_i(\theta, \psi_i)$ is angular signal density of the $i$-th source, $\theta \in [-\pi/2, \pi/2]$ is the direction and $\psi_i$ is the parameter vector to be estimated.

- $\Theta$ is the angular field of view.
- $\boldsymbol{n}(t) = [n_1(t), n_2(t), \cdots, n_M(t)]^T$ is the $M \times 1$ vector of additive measurement noise.
- $\boldsymbol{a}(\theta)$ is the response of the array to the source from the direction $\theta$.

In CD sources case, angular signal density can be represented as

$$s_i(\theta, \boldsymbol{\psi}_i) = \delta_i g(\theta, \boldsymbol{\psi}_i) \tag{6}$$

where $g(\theta, \boldsymbol{\psi}_i)$ denotes deterministic angular signal density, and $\delta_i$ denotes a random variable. Thus, we can rewrite the Eq.(5) as

$$\boldsymbol{x}(t) = \sum_{i=1}^{L} \delta_i \boldsymbol{b}(\boldsymbol{\psi}_i) + \boldsymbol{n}(t) \tag{7}$$

in which the generalized steering vector of the $i$-th source, $\boldsymbol{b}(\boldsymbol{\psi}_i)$, can be expressed as

$$\boldsymbol{b}(\boldsymbol{\psi}_i) = \int_{\theta\in\Theta} \boldsymbol{a}(\theta) g(\theta, \boldsymbol{\psi}_i)\, d\theta \tag{8}$$

The pdf of angular spread is generally modeled as Gaussian or uniform distribution; thus, if the pdf follows Gaussian distribution, it can be expressed as

$$g(\theta, \boldsymbol{\psi}_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(\theta-\theta_i)^2}{2\sigma_i^2}\right) \tag{9}$$

where $\boldsymbol{\psi}_i = (\theta_i, \sigma_i)$, $\theta_i$ and $\sigma_i$ are the central DOA and angular spread of the $i$-th source, respectively. The corresponding $\boldsymbol{b}(\boldsymbol{\psi}_i)$ is given by

$$\begin{aligned}\boldsymbol{b}(\boldsymbol{\psi}) = \Bigg[&1, \exp\left(-\frac{[(2\pi d/\lambda)\sigma_{\theta_0}\cos(\theta_0)]^2}{2} - \mathrm{j}\frac{2\pi d}{\lambda}\sin(\theta_0)\right), \\ &\cdots, \exp\left(-\frac{[(2\pi d/\lambda)(m-1)\sigma_{\theta_0}\cos(\theta_0)]^2}{2} - \mathrm{j}\frac{2\pi d}{\lambda}(m-1)\sin(\theta_0)\right)\Bigg]\end{aligned} \tag{10}$$

On the other hand, if the CD source follows a uniform distribution, the pdf is defined by

$$g(\theta, \boldsymbol{\psi}_i) = \begin{cases} \frac{1}{2\sigma_i} & |\theta-\theta_i| \le \sigma_i \\ 0 & |\theta-\theta_i| > \sigma_i \end{cases} \tag{11}$$

Substituting Eq.(11) into Eq.(8), the corresponding $\boldsymbol{b}(\boldsymbol{\psi}_i)$ is given by

$$\begin{aligned}\boldsymbol{b}(\boldsymbol{\psi}) = \Bigg[&1, \frac{\sin[\sigma_{\theta_0}(2d/\lambda)\cos(\theta_0)]}{\sigma_{\theta_0}(2d/\lambda)\cos(\theta_0)} \exp\left(-\mathrm{j}\frac{2\pi}{\lambda} d\sin(\theta_0)\right), \\ &\cdots, \frac{\sin[(m-1)\sigma_{\theta_0}(2d/\lambda)\cos(\theta_0)]}{(m-1)\sigma_{\theta_0}(2d/\lambda)\cos(\theta_0)} \exp\left(-\mathrm{j}\frac{2\pi}{\lambda}(m-1)d\sin(\theta_0)\right)\Bigg]\end{aligned} \tag{12}$$

Let $\boldsymbol{B} = [\boldsymbol{b}(\boldsymbol{\psi}_1), \boldsymbol{b}(\boldsymbol{\psi}_2), \cdots, \boldsymbol{b}(\boldsymbol{\psi}_L)]$ be the matrix of the column vectors, the covariance matrix of the array received signal $\boldsymbol{x}(t)$ can be expressed as

$$\begin{aligned}\boldsymbol{R} &= \mathbf{E}\left\{\boldsymbol{x}(t)\boldsymbol{x}^H(t)\right\} = \boldsymbol{B}\mathbf{E}\left\{\boldsymbol{s}(t)\boldsymbol{s}^H(t)\right\}\boldsymbol{B} + \sigma_n^2\boldsymbol{I}_M \\ &= \boldsymbol{B\Gamma B} + \sigma_n^2\boldsymbol{I}_M\end{aligned} \tag{13}$$

where $\boldsymbol{\Gamma}$ is the noise-free covariance matrix of CD sources. If the sources are uncorrelated with each other, $\boldsymbol{\Gamma}$ will be diagonal [11] and its $(i, j)$-th entry is defined as $\mathbf{E}\{\delta_i(t)\delta_j{}^*(t)\}$.

By performing singular value decomposition (SVD), the signal and noise subspaces of the covariance matrix can be obtained

$$\boldsymbol{R} = \boldsymbol{U}_s\boldsymbol{\Sigma}_s\boldsymbol{U}_s^H + \boldsymbol{U}_n\boldsymbol{\Sigma}_n\boldsymbol{U}_n^H \tag{14}$$

We can find that for CD sources the signal and noise subspaces are spanned by the eigenvectors of the signal covariance matrix $\boldsymbol{U}_s$ corresponding to the $L$ largest eigenvalues and noise covariance matrix $\boldsymbol{U}_n$ corresponding to the $M-L$ smallest eigenvalues, respectively. Thus, the dimension of $\boldsymbol{\Sigma}_s$ or the rank of $\boldsymbol{\Gamma}$ can be used to estimate the number of sources.

With the deterministic angular signal density $g(\theta, \boldsymbol{\psi}_i)$, the reciprocal of the spatial spectrum can be expressed as

$$f(\boldsymbol{\psi}) = \frac{1}{\int_{\theta\in\Theta}\int_{\theta'\in\Theta} g^*(\theta, \boldsymbol{\psi})\,\boldsymbol{a}^H(\theta)\boldsymbol{U}_n\boldsymbol{U}_n^H\boldsymbol{a}(\theta')g^*(\theta', \boldsymbol{\psi})\, d\theta d\theta'} \tag{15}$$

the unknown parameters of CD sources can be obtained by

$$\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \frac{1}{\|\boldsymbol{b}^H(\boldsymbol{\psi})\boldsymbol{U}_n\|^2} \tag{16}$$

This is called DSPE algorithm. Moreover, in practical applications, $\boldsymbol{R}$ can be constructed by employed $N$ received data samples

$$\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}(t)\,\boldsymbol{x}^H(t) \tag{17}$$

In this paper, to simplify the derivation of the expressions in the equations, we focus on the central DOA estimation, omitting the angular spread parameters.

## III. Subspace-based Estimation

### A. Hampel identifier and outliers suppression

In practice, the Hampel identifier is effective to detect and suppress outliers [29–31], so it is often employed to identify whether the received data contains by impulsive noise components. We rewrite the data received by the $i$-th sensor of ULA:

$$\boldsymbol{x}_i = [x_i(1), x_i(2), \cdots, x_i(N)], \quad i = 1, 2, \cdots, M \tag{18}$$

where $N$ is the number of snapshots. Since the received data containing the impulsive components have a large magnitude, the modulus of the received data can be used to identify whether the received data are the outliers. Without loss of generality, denoting the modulus of the $i$-th data by $|x_i(j)|$, we can formulate the Hampel identifier as

$$\rho[x_i(j)] = \begin{cases} x_i(j) & ||x_i(j)| - \bar{\mu}_i(j)| < \varepsilon \\ 0 & ||x_i(j)| - \bar{\mu}_i(j)| \ge \varepsilon \end{cases} \tag{19}$$

where $\bar{\mu}_i(j)$ is the median of $\boldsymbol{x}_i(j)$, and $\varepsilon$ is the discrimination threshold.

The median absolute deviation $\bar{\sigma}_i(j)$ is defined by [32]

$$\bar{\sigma}_i(j) = 1.4826 \cdot \mathrm{med}\{\boldsymbol{U}(|x_i(j)|)\} \tag{20}$$

where med$\{\cdot\}$ stands for statistical mean, $\boldsymbol{U}(|x_i(j)|) = \{||x_i(j-W)| - \bar{\mu}_i(j)|, \cdots, ||x_i(j+W)| - \bar{\mu}_i(j)|\}$, and $\bar{\mu}_i(j)$ can be expressed as

$$\bar{\mu}_i(j) = \mathrm{med}\{|x_i(j-W)|, \cdots, |x_i(j)|, \cdots, |x_i(j+W)|\} \tag{21}$$

in which $W$ is the length of the estimation windows. Then, the suppressed impulsive noise data can be obtained by

$$\tilde{\boldsymbol{x}}_i = \{\rho[x_i(1)], \rho[x_i(2)], \cdots, \rho[x_i(N)]\} \tag{22}$$

In Eq. (19), $\varepsilon$ is always defined by $\varepsilon = \varpi \cdot \bar{\sigma}_i(j)$, where the threshold $\varpi$ is usually set between 2 and 5.

### B. Correntropy

We assume the received data only contain noises, Eq. (19) can detect the outliers well and effectively suppress them. However, for the mixture of both signals and noises received from the sensor array, when the noises contain impulsive components, the Hampel identifier will make the received data to zero. In this case, not only the outliers are eliminated, but also the effective signal components are trimmed, this will directly lead to a decrease in the accuracy of the parameter estimation. To suppress the impulsive components while retaining the effective information in the signals, we quote correntropy.

Correntropy, which extends the autocorrelation function to the non-linear systems analysis [33] and is defined as a generalized correlation function, not only quantifies the size and shape of the data samples but also gives information about the statistical distribution in the feature space [20]. On the other hand, correntropy includes the higher-order moments of the pdf of the random processes and is much easier to estimate than the higher-order moments from data samples [20]. It has also been extensively employed in some fields such as DOA estimation [34, 35] and face recognition [36, 37].

Assume $X$ and $Y$ are arbitrary random variables, correntropy is defined by

$$\boldsymbol{V}_\sigma(X, Y) = \mathbf{E}\left[\kappa_\sigma(X - Y)\right] = \int_\Omega \kappa_\sigma(x - y) d f_\sigma(x, y) \tag{23}$$

in which $\kappa_\sigma(\cdot)$ denotes a translation-invariant kernel, $f_\sigma(\cdot, \cdot)$ is the joint distribution function. In practical situations, there is not a know joint pdf, given a set of data $\{(x_1, y_1), \cdots, (x_N, y_N)\}$, we can obtain the sample estimator of correntropy by

$$\hat{\boldsymbol{V}}_\sigma(X, Y) = \frac{1}{N} \sum_{n=1}^{N} \kappa_\sigma(x_n - y_n) \tag{24}$$

In general, the correntropy uses the Gaussian kernel as the kernel function, it can be defined by

$$\kappa_\sigma(x_n - y_n) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_n - y_n)^2}{2\sigma^2}\right), \quad n = 1, 2, \cdots, N \tag{25}$$

where $\sigma > 0$ is the kernel size. The Silverman's rule, kurtosis of signals or maximum-likelihood (ML) can be used to determine the kernel size.

By using the Gaussian kernel function, we can also construct a Taylor series of correntropy:

$$\boldsymbol{V}_\sigma(X, Y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} \mathbf{E}\left[\frac{(X - Y)^{2n}}{\sigma^{2n}}\right] \tag{26}$$

which illustrates that all the even-order moments of the random variable $X - Y$ are included in the correntropy. Furthermore, it is effective for non-Gaussian array signal processing. From Eq. (26), another observation is that the correntropy contains the conventional covariance function and the kernel size $\sigma$ determines the suppression effect of impulsive noise [38]. Further, compared with conventional moments, the correntropy, which can be estimated directly from data samples, has lower computational complexity.

### C. Generalized auto-correntropy operator

In Eq. (19), $||x_i(j)| - \bar{\mu}_i|$ is used as criteria to identify whether noise contains outliers. In Eq. (23), $x_i - y_i$ indicates the difference between two random variables $x_i$ and $y_i$. The conditions of both $x_i$ and $y_i$, which contain impulsive noise, are shown in TABLE I. For the sake of discussion, we assume that both $x_i$ and $y_i$ are positive-valued.

TABLE I: Comparison of two variables

| Variable | Outliers | | | |
|---|---|---|---|---|
| | Condition I | Condition II | Condition III | Condition IV |
| $x_i$ | Yes | Yes | No | No |
| $y_i$ | Yes | No | Yes | No |

In TABLE I, Condition II and Condition III illustrate that only one random variable in $x_i$ and $y_i$ contains impulsive components, therefore, the difference between $x_i$ and $y_i$ is large, then correntropy can effectively suppress the impulsive noise; in Condition IV, both random variables $x_i$ and $y_i$ do not contain impulsive components. In this case, the parameter estimation can be performed under the Gaussian assumption. Condition I displays that both random variables $x_i$ and $y_i$ contain impulsive components, then the difference between $x_i$ and $y_i$ is small, that is, from Eq. (25), the exponential part of the Gaussian kernel $\exp[-(x_i - y_i)^2/(2\sigma^2)]$ will approach to 1. Consequently, from the view of impulsive noise suppression, correntropy may not work well in Condition I.

Based on these analyses, we expand the correntropy as generalized auto-correntropy (GCO) operator. Consider an i.i.d. $S\alpha S$ random variable $X$, the GCO is defined by

$$\boldsymbol{G}_\sigma(X) = \mathbf{E}\left[\kappa_\sigma\left(|X| - \mu_X\right)\right] \tag{27}$$

where $\mu_X$ can be expressed as

$$\mu_X = \frac{1}{N} \|X\|_p^p \tag{28}$$

In $\ell_p$-space, the $\ell_p$-norm of $X$ with $0 < p < 2$ can be expressed as

$$\|X\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} \tag{29}$$

Note that $\|X\|_p$ is a norm only when $p \geq 1$. For $0 < p < 1$, its norm interpretation is not appropriate because the triangle inequality $\|X\|_p + \|Y\|_p \geq \|X + Y\|_p$ is not held. Nevertheless, the $\ell_p$-norm is still referred for $0 < p < 1$ in this paper.

The correntropy is constructed by using the Gaussian kernel [38], but, in fact, not only the Gaussian kernel but also other symmetric kernels can be used to do. According to our previous research, the robustness of the Gaussian kernel to suppress impulsive noise is inferior to the exponential kernel; therefore, we employ the exponential kernel to construct the GCO. The exponential kernel is defined by

$$\kappa_\sigma(x) = \exp\left(-\frac{|x|}{2\sigma^2}\right), \quad i = 1, 2, \cdots, N \tag{30}$$

The GCO that is defined in Eq. (27) can be reformulated by employing the exponential kernel as

$$\boldsymbol{G}_\sigma(X) = \mathbf{E}\left[\exp\left(-\frac{\left||X| - \mu_X\right|}{2\sigma^2}\right)\right] \tag{31}$$

Consider that the $i$-th array sensor receives sample data $\boldsymbol{x}_i$, which involves signals and impulsive noises, the suppression of impulsive noise of GCO is shown in Fig. 1. As can be

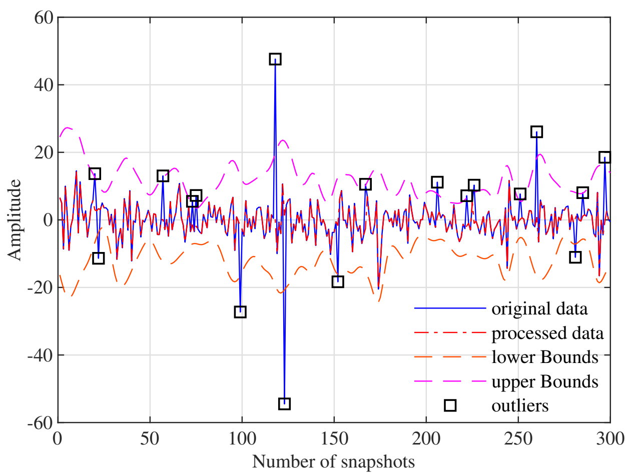



Fig. 1: **The suppression of impulsive noise**

seen from Fig. 1, the GCO operator can suppress the impulsive components of the noise sufficiently, and effectively preserve the signals. If the Hampel identifier is employed, the signal at the location where the impulsive noise occurs will be set to 0. Therefore, by ensuring that the valid data are not lost, the GCO operator is superior to the Hampel identifier in impulsive noise immunity.

The $S\alpha S$ distribution without finite variance results in a significant degradation in the performance of conventional second-order statistic-based DOA estimation algorithms [39]. In the paper, we propose a novel central DOA estimation algorithm for CD sources by using the GCO operator as an adaptive factor to suppress outliers in $S\alpha S$ processes with $1 < \alpha \leq 2$ and make it an excellent substitute for the estimation matrix in DSPE algorithm.

***Definition:*** Consider two i.i.d. $S\alpha S$ random variables $X$ and $Y$ corresponding to $1 < \alpha \leq 2$, the GCO-based estimation matrix by employing the exponential kernel is defined by

$$\begin{aligned}\boldsymbol{R}_c &= \mathbf{E}\left[\kappa_\sigma\left(|X| - \mu_X\right)\kappa_\sigma\left(|Y| - \mu_Y\right)XY\right]\\ &= \mathbf{E}\left[\exp\left(-\frac{\left||X| - \mu_X\right|}{2\sigma_1^2}\right)\exp\left(-\frac{\left||Y| - \mu_Y\right|}{2\sigma_2^2}\right)XY\right]\end{aligned} \tag{32}$$

We find that two exponential multipliers

$$\omega_1 = \exp\left[-\left||X| - \mu_X\right| / \left(2\sigma_1^2\right)\right] \tag{33a}$$

$$\omega_2 = \exp\left[-\left||Y| - \mu_Y\right| / \left(2\sigma_2^2\right)\right] \tag{33b}$$

are involved in correlation function. Eq. (32) can be rewritten in a compact form:

$$\boldsymbol{R}_c = \mathbf{E}\left\{\left[\exp\left(-\frac{\left||X| - \mu_X\right|}{2\sigma_1^2}\right)X\right]\cdot\left[\exp\left(-\frac{\left||Y| - \mu_Y\right|}{2\sigma_2^2}\right)Y\right]\right\} \tag{34}$$

is bounded.

From the view of data processing, multipliers $\omega_1$ and $\omega_2$ can be seen as "*preprocessing*" for random variables $X$ and $Y$, respectively. By the weighting processing, the impulsive noises are effectively eliminated.

### *D. Adaptive kernel size function*

Correntropy behaves as a 'robust statistics' like the M-estimator. Further, in 2D space, the correntropy induced metric (CIM) that is defined by $\mathrm{CIM}(X, Y) = \sqrt{\kappa_\sigma(0) - \boldsymbol{V}_\sigma(X, Y)}$ is induced from the correntropy. Apparently, due to the inhomogeneity of CIM, it can induce multiple norms. If the points are close, CIM behaves like the 2-norm; if the points get further apart, CIM behaves like to the 1-norm; and if the points are far apart, CIM behaves like the 0-norm, that is to say, the correntropy is a robust measure of local similarity. Another important observation is that the kernel size can completely control the size of CIM under the impulsive noise environments [38].

From the perspective of central DOA estimation accuracy, if kernel size is smaller, overfull noise components are suppressed, which cases signal subspace is extended into noise subspace, the orthogonality of noise subspace and signal subspace cannot be guaranteed; further, if the kernel size is large, some impulsive noises are remained in the outputs after preprocessing by GCO operator, which will destroy the semi-positive definite property of the covariance matrix.

Based on these analyses, we define an adaptive kernel size function according to the following criteria: 1) the kernel size is positive and has the upper and lower bounds; 2) the kernel size is monotonic with changes of impulsive noise, which ensures that both the signal and noise subspaces are entirely preserved. In [21] and [40], some kernel sizes are defined, however, obtaining these kernel sizes require some prior knowledge that is not easy to be getting in practical applications. Thus, we address to define an adaptive kernel size function, which depends only on the received data of the sensor array and can achieve the best performance of central DOA estimation.

#### 1. Definition

Since the array outputs $\boldsymbol{x}(t)$, which is contaminated by noises, are the only data we can handle with, we defined an adaptive kernel size function by

$$\sigma = \frac{\pi}{1 + \exp(-H)} \tag{35}$$

and we expand the local entropy of information theory [40–42], in Eq. (35), $H$ is defined by:

$$H = -\sum_{i=1}^{N}\sum_{j=1}^{M} p_{i,j}\log(p_{i,j}) \tag{36}$$

where

$$p_{i,j} = \frac{|\boldsymbol{x}_i(t)|}{\sum_i^N \sum_j^M |\boldsymbol{x}_i(t)|} = \frac{|\boldsymbol{x}_i(t)|}{\ell_1[\boldsymbol{x}(t)]} \tag{37}$$

in which, $\ell_1[\cdot]$ denotes the $\ell_1$-norm. It is easy to find that Eq.(35) only involves receiving data of sensor array but not using any empirical parameter or prior knowledge; therefore, the adaptive kernel function is suitable for practical scenarios, particularly.

### 2. Properties

Some important properties of the adaptive kernel size function are presented next. These are obtained after straightforward derivations and will, therefore, not be proved here.

- P1: The adaptive kernel function and its inverse function are monotonic.
- P2: The adaptive kernel function is positive and bounded: $\pi/2 \le \sigma < \pi$. It reaches its minimum if and only if $H = 0$.
- P3: The adaptive kernel function is differentiable, real-valued function, which is defined for all real-valued inputs and has a non-negative derivative at each point.

The comparison experiments between fixed and adaptive kernel sizes will be involved in Section IV-A.

### *E. Subspace methods under GCO framework*

In this section, a novel GCO-based impulsive noise suppression algorithm that can be employed with DSPE is proposed to achieve the estimation of the central DOAs.

Let $\boldsymbol{C}$ be an $M \times M$ estimation matrix, its $(i,j)$-th entry $C_{ij}$ is defined by

$$C_{ij} = \mathbf{E}\left[\exp\left(-\frac{\left||x_i| - \mu_{x_i}\right|}{2\sigma_1^2}\right) x_i \exp\left(-\frac{\left||x_j{}^*| - \mu_{x_j}\right|}{2\sigma_2^2}\right) x_j{}^*\right] \tag{38}$$

where $x_i$ and $x_j$ are the $i$-th and $j$-th components of the vector $\boldsymbol{x}(t)$, respectively.

We extend the DSPE algorithm from the Gaussian distribution noise to the S$\alpha$S distribution noise environments by applying GCO operator, and this algorithm is named as GCO-DSPE. The implementation steps are shown in **Algorithm 1**.

## IV. Simulations

In our simulations, consider that two independent quadrature phase-shift keying (QPSK) CD sources impinge on a ULA with $M = 8$ sensors. The underlying noises are modeled as an additive complex isotropic $S\alpha S$ distribution with $1 \le \alpha \le 2$. However, since there is no finite variance in the alpha-stable family with $1 \le \alpha < 2$, we utilized the generalized signal-to-noise ratio (GSNR) to evaluate the rate of the signal power over noise dispersion by

$$\mathrm{GSNR} = 10\log\frac{\mathbf{E}\left[|s(t)|^2\right]}{\gamma} \tag{40}$$

in which $\gamma$ is the dispersion parameter that is defined in Eq. (1).

In order to evaluate the performance of the algorithms, we use the following two criteria: (1) the probability of resolution; (2) the root mean square error (RMSE). When the following criterion is held, we define that two sources can be successfully resolved.

$$\Delta(\theta_1,\theta_2) = \boldsymbol{r}(\theta_e) - \frac{[\boldsymbol{r}(\theta_1) + \boldsymbol{r}(\theta_2)]}{2} > 0 \tag{41}$$

where $\theta_1$ and $\theta_2$ are the arrival angles of the two sources and $\theta_e = (1/2)(\theta_1 + \theta_2)$. $\boldsymbol{r}(\theta)$ denotes the reciprocal of the spatial spectrum $\boldsymbol{P}(\theta)$. The probability of resolution is defined as the ratio of the number of successful runs to the total number of Monte Carlo runs, and the RMSE of those successful runs is defined by

$$\mathrm{RMSE} = \frac{1}{2}\left[\sqrt{\frac{1}{Q}\sum_{q=1}^{Q}\left(\hat{\theta}_1(q) - \theta_1\right)^2} + \sqrt{\frac{1}{Q}\sum_{q=1}^{Q}(\hat{\theta}_2(q) - \theta_2)^2}\right] \tag{42}$$

where $Q$ is the number of the successful runs and $\hat{\theta}_1$ and $\hat{\theta}_2$ are the successful estimation of $\theta_2$ and $\theta_2$, respectively.

In these experiments, we execute 300 Monte Carlo runs to evaluate the probability of resolution and RMSE of the central DOA estimation. We examine the performance of the algorithms as the functions of several parameters, namely, the kernel size, the characteristic exponent $\alpha$, the GSNR, the number of snapshots, respectively.

In addition to conventional DSPE, we also apply the impulsive noise suppression algorithms, i.e., FLOM, PFLOM and CRCO, on the signal model of CD sources and compare them

---

**Algorithm 1** GCO-DSPE

**Input:** Observation data
**Output:** Central DOA estimation

1: Compute the adaptive kernel function $\sigma$ based on Eq. (35);
2: Construct the $M \times M$ estimation matrix $\hat{\boldsymbol{C}}$;

$$\hat{C}_{ij} = \frac{1}{N}\sum_{t=1}^{N}\left[\exp\left(-\frac{\left||x_i(t)| - \mu_{x_i}\right|}{2\sigma_1^2}\right)x_i \times\exp\left(-\frac{\left||x_j^*(t)| - \mu_{x_j}\right|}{2\sigma_2^2}\right)x_j^*(t)\right] \tag{39}$$

3: Apply SVD to $\hat{\boldsymbol{C}}$, and obtain the $M\times(M-L)$ noise subspace $\hat{\boldsymbol{U}} = [u_{L+1}, u_{L+2}, \cdots, u_M]$ that corresponds the smallest $M-L$ singular values of $\hat{\boldsymbol{C}}$;
4: Construct the spatial spectrum function $\boldsymbol{P}_{\text{GCO}}(\boldsymbol{\theta})$ of GCO-DSPE based on Eq. (16);
5: Search the $L$ local peaks of $\boldsymbol{P}_{\text{GCO}}(\boldsymbol{\theta})$ and estimate central DOA of CD sources.

---

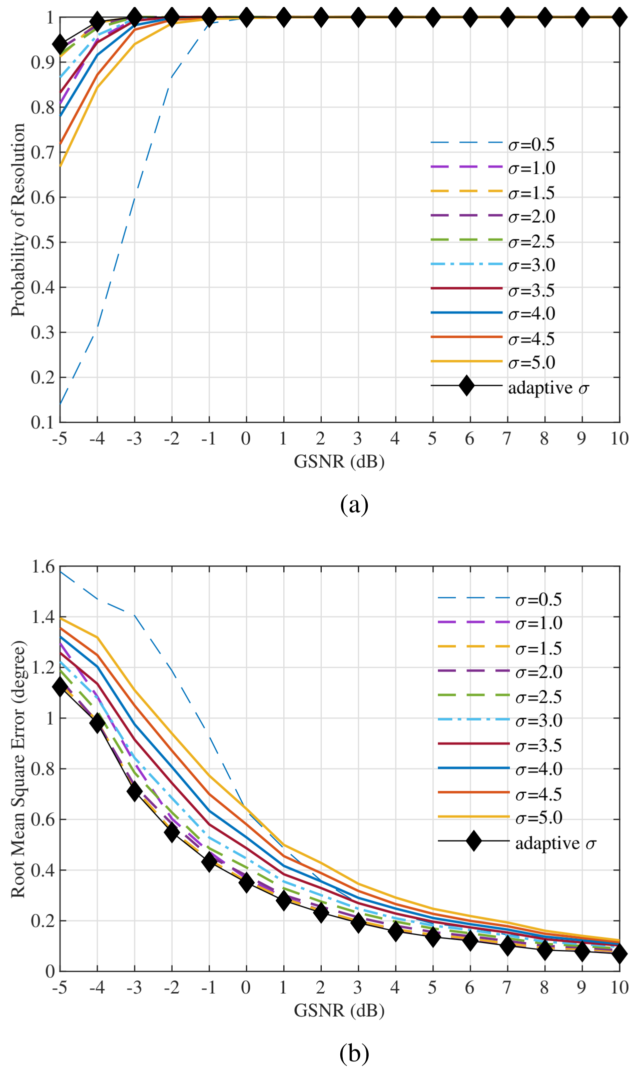



Fig. 2: **Kernel size analysis: (a) Probability of resolution; (b) Root mean square error.**

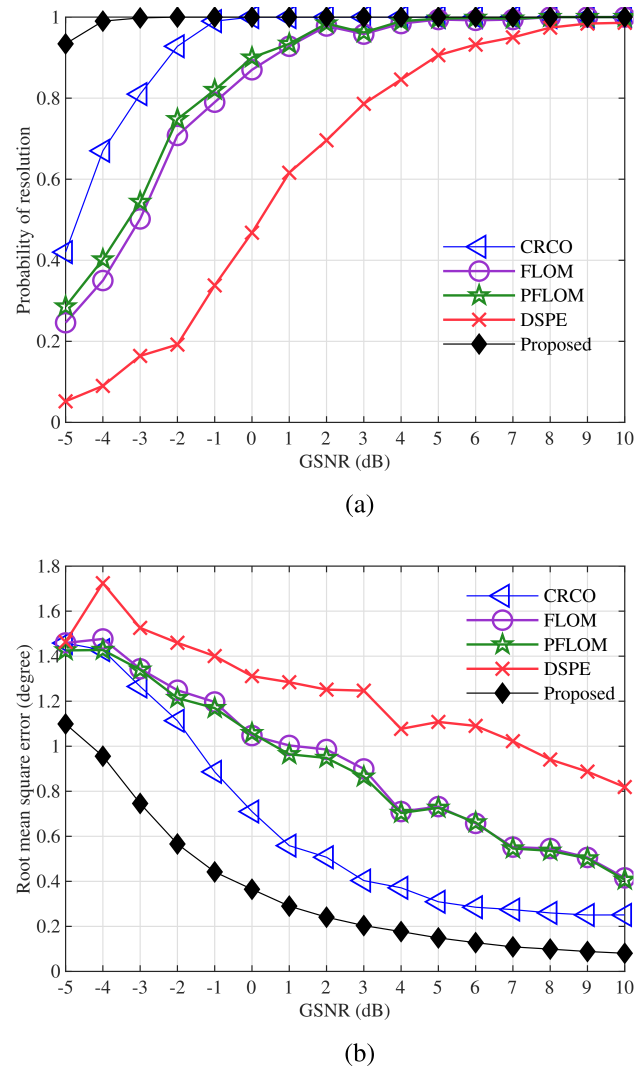



Fig. 3: **Performance comparison of the algorithms as a function of the GSNR at $\alpha = 1.5$: (a) Probability of resolution; (b) Root mean square error.**

with the GCO-DSPE. The directions of the two impinging CD sources $\theta_1 = 5°$ and $\theta_2 = 15°$ are uncorrelated with each other unless otherwise specified. Angular spread parameter $\Delta = 2°$ is identical for the two impinging CD sources.

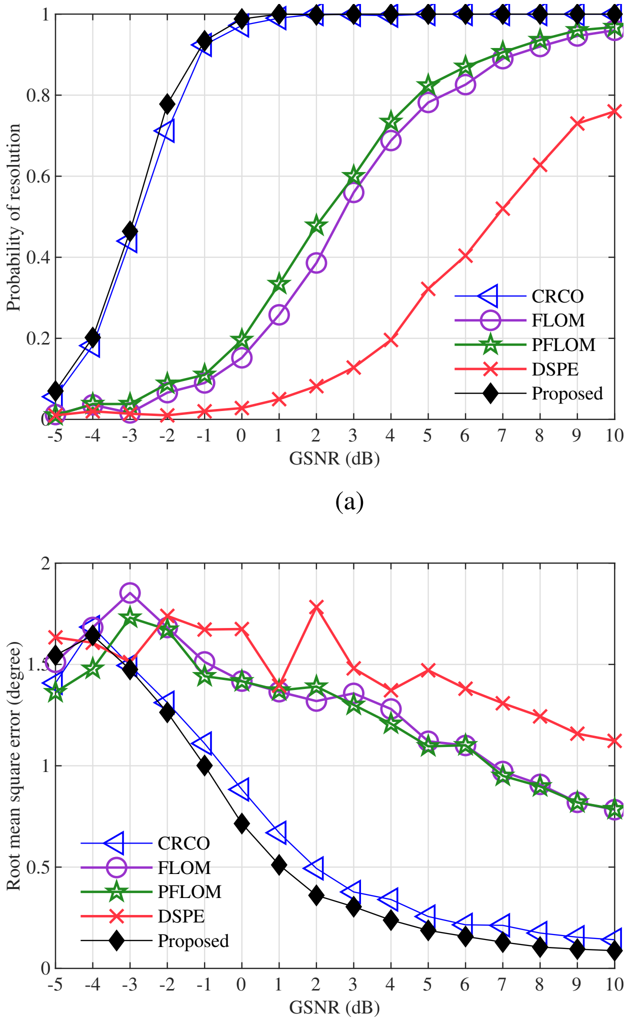
### *A. Kernel size analysis*

In this simulation, we analyze the influence of different kernel sizes on the accuracy of central DOA estimation. In Fig. 2, the kernel sizes are chosen as $\{0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0\}$. Further, we also compare the adaptive kernel size in this section. As can be seen from Fig. 2, when kernel size $\sigma = 0.5$, the probability of resolution and RMSE are the worst of the various kernel sizes. With fixing $\sigma = 1.5$ as a reference, whether kernel size is reduced or increased, the performance of central DOA estimation will decrease. When the kernel size becomes smaller, the performance of DOA estimation declines faster. However, both the RMSE and probability of resolution, the adaptive kernel size achieves the best performance.

### *B. The effect of the GSNR*

In this section, Fig. 3 and Fig. 4 illustrate the performance comparison of the algorithms with respect to the GSNR of the two impinging CD sources. The number of snapshots is fixed to 500. In Fig. 3, the additive impulsive noise is modeled as the $S\alpha S$ distribution with the characteristic exponent $\alpha = 1.5$. We can find that with increasing GSNR, the performance of involved algorithms is improved. Further, the CRCO and proposed algorithms are superiors to the FLOM, PFLOM and conventional DSPE. For low GSNR, the GCO-DSPE has the highest success probability than others. Compared with other algorithms, RMSE of GCO-DSPE is the smallest.

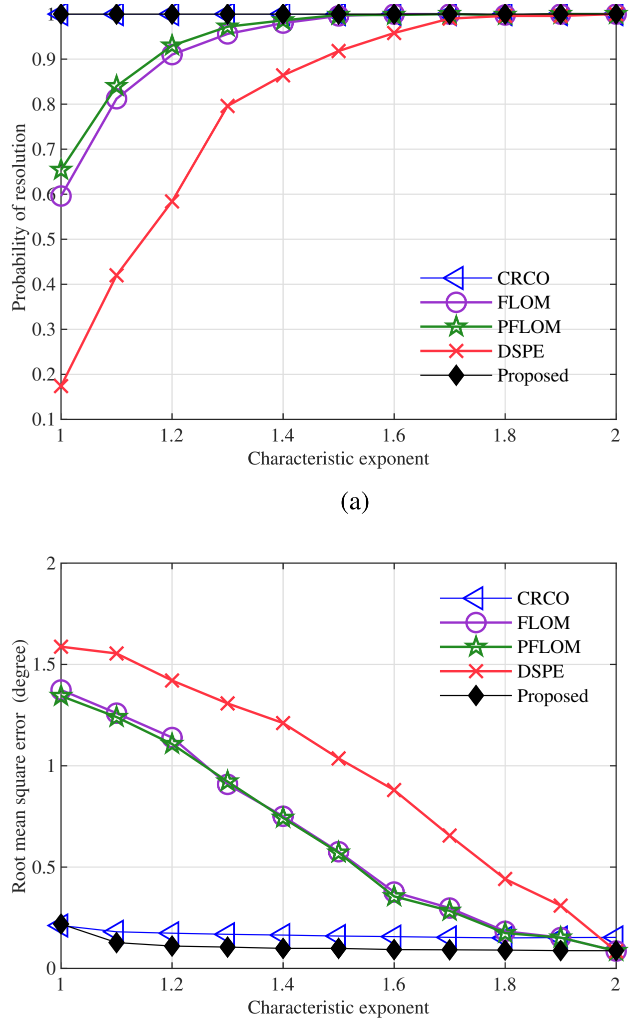
A very highly impulsive noise condition with $\alpha = 1.1$ is used in this experiment, and we present the central DOA estimation in Fig. 4. As the characteristic exponent $\alpha$ decreases from 1.5 to 1.1, the impulsive components in the noise increase significantly, which leads to a reduction in the performance of




(a)

(b)

Fig. 4: **Performance comparison of the algorithms as a function of the GSNR at $\alpha$ = 1.1: (a) Probability of resolution; (b) Root mean square error.**




(a)

(b)

Fig. 5: **Performance comparison of the algorithms as a function of the characteristic exponent: (a) Probability of resolution; (b) Root mean square error.**

all algorithms. When the GSNR increases, the performance of the conventional DSPE, FLOM and PFLOM is slow to improve. However, the performance of the GCO-DSPE is still the best.

### *C. The effect of the characteristic exponent $\alpha$*

In the experiment, the performance comparison of the algorithms with different characteristic exponent $\alpha$, which varies from extremely strong impulsive conditions with $\alpha = 1.0$ (Cauchy noise) to impulsive-free conditions with $\alpha = 2.0$ (Gaussian noise), are displayed in Fig. 5. In order to directly compare the performance of different central DOA estimation algorithms, the GSNR is set to 8 dB and the number of snapshots is $N = 500$.

As expected, when the characteristic exponent $\alpha$ is reduced from 2.0 to 1.0, the performance of all algorithms decreases. From Fig. 5(a) we can find that the change in the characteristic exponent $\alpha$ has little influence on the probability of resolution the GCO-DSPE. Another interesting observation is that the probability of resolution of the GCO-DSPE is always equal to 1 along with the characteristic exponent $\alpha$ varies from 1 to 2. As shown in Fig. 5(b), when the characteristic exponent $\alpha$ is greater than 1.1, its change has little effect on the RMSE of the GCO-DSPE, which indicates that the GCO-DSPE is highly robust to changes in the characteristic exponent $\alpha$. The simulation results show that the GCO-DSPE has better performance than other algorithms under very highly impulsive noise environments.

### *D. The effect of the number of snapshots*

The effect of the number of snapshots on the performance of central DOA estimation is studied in this section, involving the probability of resolution and RMSE. The GSNR is kept almost at 8 dB. The characteristic exponent $\alpha$ is set to 1.5

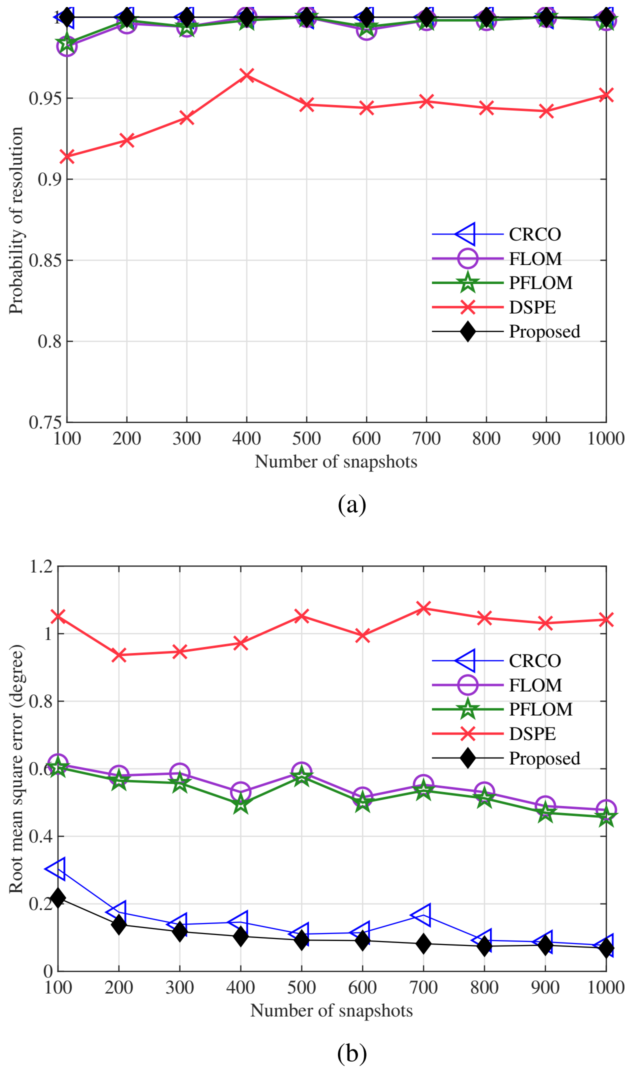



Fig. 6: **Performance comparison of the algorithms as a function of the number of snapshots: (a) Probability of resolution; (b) Root mean square error.**

corresponding to the impulsive noise. It can be observed from Fig. 6(a) that although the CRCO and PFLOM are superior to the FLOM and conventional DSPE algorithms, however, the GCO-DSPE superiors to others concerning the probability of resolution. The performance of conventional DSPE is very poor.

Fig. 6(b) illustrates that with the number of snapshots increasing from 100 to 1000, the RMSEs decrease, however, the magnitude of the decrease is not obvious. That is to say, under the condition that the probability of impulsive noise is determined, increasing the information quantity is not obvious for improving the robustness of these algorithms.

## V. Conclusions

Inspired by Hampel identifier, we have presented a novel operator referred to as the GCO for central DOA estimation of CD sources in the impulsive noise conditions. In order to make the proposed algorithm more suitable for the real scenarios, we also deduce an adaptive kernel size function, which only involves received data from the sensor array but not relying on any empirical parameter or prior knowledge. The GCO-based estimation matrix is formulated and applied with DSPE to obtain central DOA estimation. Comprehensive simulations verify the advantage of GCO-DSPE over other impulsive noise suppression algorithms, i.e. CRCO, FLOM and PFLOM, for central DOA estimation of CD sources in quite highly impulsive noises and very low GSNR situations.

## Acknowledgment

This work is partly supported by the National Natural Science Foundation of China under Grant 81241059, Grant 61172108, Grant 61671105, and Grant 61139001.

## References


[1] Y. Shen and M. Z. Win, “Fundamental limits of wideband localization – Part I: A general framework,” *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 4956–4980, 2010.

[2] B. Friedlander and A. J. Weiss, “Effects of model errors on waveform estimation using the music algorithm,” *IEEE Transactions on Signal Processing*, vol. 42, no. 1, pp. 147–155, 1994.

[3] A. Ferreol, P. Larzabal, and M. Viberg, “Statistical analysis of the music algorithm in the presence of modeling errors, taking into account the resolution probability,” *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4156–4166, 2010.

[4] R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[5] Z. H. Chen, Y. T. Ding, S. W. Ren, and Z. M. Chen, “A novel noncircular music algorithm based on the concept of the difference and sum coarray,” *Sensors*, vol. 18, no. 2, p. 17, 2018.

[6] K. Todros and A. O. Hero, “Robust multiple signal classification via probability measure transformation,” *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1156–1170, 2015.

[7] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, “A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments,” *IEEE Transactions on Vehicular Technology*, vol. 49, no. 2, pp. 437–447, 2000.

[8] P. Zetterberg, “Mobile cellular communications with base station antenna arrays: Spectrum efficiency, algorithms and propagation models,” 1997.

[9] L. T. Wan, G. J. Han, L. Shu, and N. X. Feng, “The Critical Patients Localization Algorithm Using Sparse Representation for Mixed Signals in Emergency Healthcare System,” *IEEE Systems Journal*, vol. 12, no. 1, pp. 52–63, 2018.

[10] Z. Zheng, G. J. Li, and Y. L. Teng, “2D DOA Estimator for multiple coherently distributed sources using modified propagator,” *Circuits Systems and Signal Processing*, vol. 31, no. 1, pp. 255–270, 2012.

[11] S. Shahbazpanahi, S. Valaee, and M. H. Bastani, “Distributed source localization using ESPRIT algorithm,”

*IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2169–2178, 2001.

[12] S. Valaee, B. Champagne, and P. Kabal, “Parametric location of distributed sources,” *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2144–2153, 1995.

[13] M. Bengtsson, “Antenna array signal processing for high rank data models,” Thesis, Signaler, sensorer och system, 2000.

[14] Y. Meng, P. Stoica, and K. M. Wong, “Estimation of the directions of arrival of spatially dispersed signals in array processing,” *IEE Proceedings-Radar Sonar and Navigation*, vol. 143, no. 1, pp. 1–9, 1996.

[15] G. A. Tsihrintzis and C. L. Nikias, “Performance of optimum and suboptimum receivers in the presence of impulsive noise modeled as an alpha-stable process,” *IEEE Transactions on Communications*, vol. 43, no. 2-4, pp. 904–914, 1995.

[16] C. L. Nikias and M. Shao, *Signal processing with alpha-stable distributions and applications*. Wiley-Interscience, 1995.

[17] T. H. Liu and J. M. Mendel, “A subspace-based direction finding algorithm using fractional lower order statistics,” *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1605–1613, 2001.

[18] H. Belkacemi and S. Marcos, “Robust subspace-based algorithms for joint angle/Doppler estimation in non-Gaussian clutter,” *Signal Processing*, vol. 87, no. 7, pp. 1547–1558, 2007.

[19] Y. Bian and B. Mercer, “Interferometric SAR Extended Coherence Calculation Based on Fractional Lower Order Statistics,” *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 841–845, 2010.

[20] I. Santamaria, P. P. Pokharel, and J. C. Principe, “Generalized correlation function: Definition, properties, and application to blind equalization,” *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, 2006.

[21] J. F. Zhang, T. S. Qiu, A. M. Song, and H. Tang, “A novel correntropy based DOA estimation algorithm in impulsive noise environments,” *Signal Processing*, vol. 104, pp. 346–357, 2014.

[22] Q. Tian, T. Qiu, J. Ma, J. Li, and R. Li, “A simplified doa estimation method based on correntropy in the presence of impulsive noise,” *IEEE Access*, vol. 6, pp. 67 010–67 025, 2018.

[23] W. J. Zeng, H. C. So, and A. M. Zoubir, “An $\ell_p$-norm minimization approach to time delay estimation in impulsive noise,” *Digital Signal Processing*, vol. 23, no. 4, pp. 1247–1254, 2013.

[24] W.-J. Zeng, H. C. So, and X. Jiang, “Outlier-robust greedy pursuit algorithms in $\ell_p$-space for sparse approximation,” *IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 60–75, 2016.

[25] W. J. Zeng and H. C. So, “Outlier-Robust Matrix Completion via $\ell_p$-Minimization,” *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1125–1140, 2018.

[26] E. Pitman and J. Pitman, “A direct approach to the stable distributions,” *Advances in Applied Probability*, vol. 48, no. A, pp. 261–282, 2016.

[27] L. Lu, H. Zhao, and B. Chen, “Robust adaptive algorithm for smart antenna system with $\alpha$-stable noise,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2017.

[28] M. Ghogho, O. Besson, and A. Swami, “Estimation of directions of arrival of multiple scattered sources,” *IEEE Transactions on Signal Processing*, vol. 49, no. 11, pp. 2467–2480, 2001.

[29] J. Astola and P. Kuosmanen, *Fundamentals of nonlinear digital filtering*. CRC press, 1997, vol. 8.

[30] R. K. Pearson, “Outliers in process modeling and identification,” *IEEE Transactions on control systems technology*, vol. 10, no. 1, pp. 55–63, 2002.

[31] L. Davies and U. Gather, “The identification of multiple outliers,” *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 782–792, 1993.

[32] P. J. Huber, “Robust statistics,” 1981.

[33] J. P. Guimaraes, A. I. Fontes, J. B. Rego, A. d. M. Martins, and J. C. Principe, “Complex correntropy function: properties, and application to a channel equalization problem,” *Expert Systems with Applications*, vol. 107, pp. 173–181, 2018.

[34] J. F. Zhang, T. S. Qiu, P. Wang, and S. Y. Luan, “A novel cauchy score function based DOA estimation method under alpha-stable noise environments,” *Signal Processing*, vol. 138, pp. 98–105, 2017.

[35] B. Wang, Y. M. D. Zhang, and W. Wang, “Robust DOA Estimation in the presence of miscalibrated sensors,” *IEEE Signal Processing Letters*, vol. 24, no. 7, pp. 1073–1077, 2017.

[36] R. He, W.-S. Zheng, B.-G. Hu, and X.-W. Kong, “Two-stage nonnegative sparse representation for large-scale face recognition,” *IEEE transactions on neural networks and learning systems*, vol. 24, no. 1, pp. 35–46, 2013.

[37] R. He, W.-S. Zheng, and B.-G. Hu, “Maximum correntropy criterion for robust face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1561–1576, 2011.

[38] W. F. Liu, P. P. Pokharel, and J. C. Principe, “Correntropy: properties and applications in non-gaussian signal processing,” *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.

[39] W. J. Zeng, H. C. So, and L. Huang, “$\ell_p$-MUSIC: robust direction-of-arrival estimator for impulsive noise environments,” *IEEE Transactions on Signal Processing*, vol. 61, no. 17, pp. 4296–4308, 2013.

[40] T. Dai, W. Z. Lu, W. Wang, J. L. Wang, and S. T. Xia, “Entropy-based bilateral filtering with a new range kernel,” *Signal Processing*, vol. 137, pp. 223–234, 2017.

[41] C. E. Shannon, “A mathematical theory of communication,” *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[42] J. N. Kapur, P. K. Sahoo, and A. K. Wong, “A new method for gray-level picture thresholding using the entropy of the histogram,” *Computer vision, graphics, and image processing*, vol. 29, no. 3, pp. 273–285, 1985.