MUSIC with Capped Frobenius Norm: Efficient Robust Direction-of-Arrival Estimator

Xiao Peng Li, Zhaofeng Liu, Zhang-Lei Shi, Hing Cheung So, *Fellow, IEEE*

***Abstract*—Direction-of-arrival (DOA) estimation is a frequent need in the field of array signal processing. While many conventional algorithms achieve excellent performance in Gaussian noise, they are vulnerable to impulsive noise. Although several approaches have been proposed for robust DOA estimation against gross errors, their disadvantages might limit the applicability in practice. For instance, maximum likelihood (ML) estimation based algorithms involve high computational complexity, and ℓ_p-MUSIC with $p \in (1, 2)$ requires tweaking p for handling different noises. In this work, we devise a capped Frobenius norm (CFN) for complex-valued data inspired by the truncated least squares loss function. Since the cap threshold is the boundary to differentiate the normal and outlier-contaminated entries, we propose a normalized median absolute deviation based strategy for its automatic determination. In doing so, accurate estimation is achieved in both Gaussian and impulsive noise. As the CFN is nonconvex and nonsmooth, we exploit the half-quadratic theory to simplify the resultant problem into a tractable optimization, which is then handled by alternating convex optimization with computationally-efficient closed-form solution. Furthermore, its convergence behaviors are analyzed, i.e., the objective function value is convergent, and there exists a subsequence in the variable sequence converging to a critical point. Simulation results exhibit its superior performance over several state-of-the-art algorithms in terms of estimation accuracy and resolution capability. MATLAB code is available at https://github.com/Li-X-P/Code-Robust-DOA-Estimator.**

***Index Terms*—Direction-of-arrival estimation, capped Frobenius norm, MUSIC, robust recovery, proximal block coordinate descent.**

I. Introduction

DIRECTION-of-arrival (DOA) estimation has been an important task in array signal processing, and has a wide range of applications, such as radar [1], sonar [2], wireless communications [3], and electronic reconnaissance [4]. Exploiting the second-order statistics, numerous DOA estimators have been proposed, including multiple signal classification (MUSIC) [5], [6], estimation of signal parameter via rotational invariance techniques (ESPRIT) [7], [8], principal-eigenvector utilization for modal analysis (PUMA) [9] and their variants [10]–[12]. In addition, the low-rank property is adopted to remove the noise in the received signal for performance enhancement [13]–[17]. This is because the corresponding noise-free data matrix is of low rank when the number of targets is less than those of antennas and snapshots. Low-rank matrix decomposition (LRMD) [13] and low-rank and sparse decomposition (LRSD) [14] exploit the nuclear norm to replace the rank function. Besides, gain-phase atomic norm minimization (GP-ANM) employs a new atomic norm to characterize the low-rank property [16]. Since the nuclear norm is a convex relaxation of the rank function, its solution might deviate from the optimality. To handle this issue, ℓ_p-norm with $p \in (0, 1)$, logarithm, and Laplace functions are adopted as the penalties on the singular values of the covariance matrix [17]. Moreover, compressed sensing has been applied for DOA estimation where the total-variation norm replaces ℓ_1-norm for sparsity constraint [18]. All of these algorithms are derived based on the assumption that the noise in the received signal is independent identically distributed (IID) Gaussian random variable [19], [20]. Therefore, they exhibit high estimation accuracy in Gaussian noise scenarios, but suffer performance degradation in the presence of non-Gaussian noise.

In real-life situations, non-Gaussian interference is also common, e.g., impulsive noise in radar signals [21] and communications system [22]. The probability density function (PDF) of impulsive noise has heavier tails than the Gaussian distribution. This feature is analogous to outliers in statistics since the heavy-tailed distribution results in a higher probability of occurrence for values over a few standard deviations than the Gaussian distribution. Thereby, the conventional DOA estimators might not work properly in impulsive noise.

One approach for DOA estimation in the presence of outliers is to exploit maximum likelihood (ML) estimation. For example, [23] and [24] adopt Gaussian mixture model (GMM) to model impulsive noise and hence their algorithms outperform the conventional methods in GMM noise. Besides, [25] and [26] assume that noise obeys the Cauchy distribution and thus the resultant approaches perform well in α-stable noise. However, the ML estimation based algorithms possess high computational complexity. This is because it requires dealing with a complicated nonlinear and nonconvex multi-dimensional optimization problem.

Another approach for outlier resistance is to preprocess the received data using a nonlinear function, such that the impact

This work was supported in part by the National Science Fund for Distinguished Young Scholars under Grant 61925108, in part by the Key Project of International Cooperation and Exchanges of the National Natural Science Foundation of China under Grant 62220106009, in part by the project of Shenzhen Peacock Plan Teams under Grant KQTD20210811090051046, in part by Shenzhen University 2035 Program for Excellent Research, and in part by the grant from the Research Grants Council of the Hong Kong SAR, China [Project No. CityU 11207922]. *(Corresponding author: Zhaofeng Liu).*

Xiao Peng Li is with the College of Electronics and Information Engineering, Shenzhen University, and also with State Key Laboratory of Radio Frequency Heterogeneous Integration, Shenzhen University, Shenzhen, 518060, China. (e-mail: x.p.li@szu.edu.cn).

Zhaofeng Liu and Hing Cheung So are with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China (e-mail: zhaofeliu3-c@my.cityu.edu.hk; hcso@ee.cityu.edu.hk).

Zhang-Lei Shi is with College of Science, China University of Petroleum (East China), Qingdao, 266580, China (e-mail: zlshi@upc.edu.cn).

of outliers on the DOA estimates reduces. For instance, [27] and [28] propose a zero-memory nonlinear (ZMNL) function to clip the amplitude of the observed measurements. In addition, [29] suggests utilizing four hyperbolic functions to suppress gross errors. Although they are easy to implement, and possess low computational complexity, the nonlinear mapping generally destroys the low-rank property of the signal subspace. Thereby, their performance might be degraded when the rank of the signal subspace increases [24]. To handle this issue, [30] proposes an empirical characteristic function to achieve robustness, which does not change the rank of the signal subspace.

Moreover, the fractional lower-order statistics method has been utilized for robust DOA estimation, resulting in robust covariation-based MUSIC (ROC-MUSIC) [31], fractional lower order moment based MUSIC (FLOM-MUSIC) [32], sign covariance matrix based MUSIC (SCM-MUSIC) [33], Kendall's tau covariance matrix based MUSIC (TCM-MUSIC) [33], and their variants [34], [35]. However, they are suboptimal and need large sample sizes for satisfactory performance [24], [36].

On the other hand, the low-rank property has been exploited for robust DOA estimation. Specifically, ℓ_p-MUSIC adopts the low-rank matrix factorization and ℓ_p-norm with $p \in (1, 2)$ to seek for the signal subspace [36]. Although it achieves good performance in impulsive noise, it may not satisfy the practical timing requirement since solving the ℓ_p-norm based problem has relatively high computational complexity. To handle this issue, Liu *et al.* [37] suggest exploiting robust principal component analysis (RPCA) developed by Candès *et al.* [38] to remove the anomalies in the output signal and then the reconstructed data are used for DOA estimation. Nevertheless, the resultant algorithm introduces two auxiliary parameters, such that it requires tweaking the trade-off parameters.

In this work, we aim at devising an accurate and computationally-efficient DOA estimator for Gaussian noise and impulsive noise environments. To resist anomalies, we propose a capped Frobenius norm (CFN) for complex-valued data inspired by the truncated least squares loss function (TLS) [39], [40]. The TLS has been applied for outlier estimation in computer vision where the nonconvex and nonsmooth TLS is approximated by a smooth function [41], [42]. Herein, the CFN is exploited for robust subspace estimation. Besides, the difference between the CFN and TLS is discussed in the next section. We then combine the CFN with the low-rank matrix factorization strategy to formulate the robust DOA estimation problem, resulting in a nonsmooth and nonconvex optimization. Different from the existing works, we adopt the half quadratic theory [43], [44] to convert the resultant problem into a tractable task, that is, convex optimization with respect to (w.r.t.) each individual variable. The proximal block coordinate descent method (PBCD) [45] is then adopted as the solver for the multi-variable problem, leading to alternating convex optimization with the closed-form solution. Furthermore, the convergence of the suggested method is established, including the objective function value and variable sequence.

We briefly summarize the contributions of this work as follows:

(i) We exploit the normalized median absolute deviation (MAD) to adaptively determine the cap threshold of the CFN. In doing so, the suggested algorithm achieves excellent performance in both impulsive and Gaussian noise without tweaking parameter.

(ii) We adopt the half quadratic theory to simplify the nonconvex and nonsmooth problem, leading to a Frobenius norm optimization with a regularization term. By utilizing PBCD, the multi-variable optimization is separated into three convex problems with closed-form solutions.

(iii) We analyze the convergence behavior of the devised algorithm. Specifically, the proposed method guarantees that the objective function value is convergent, and there exists a subsequence in the variable sequence converging to a critical point.

(iv) Simulation results exhibit the superior performance of our method over the state-of-the-art robust approaches in the presence of impulsive noise. Besides, in Gaussian noise scenarios, our algorithm attains comparable performance to conventional methods.

The remainder of this paper is organized as follows. We introduce the signal model, and review related works in Section II. In Section III, we present the CFN and then the robust DOA estimator is derived. Its convergence behavior and computational complexity are analyzed in Section IV. In Section V, numerical examples are included to evaluate the devised method by comparing with several existing algorithms. Finally, concluding remarks are included in Section VI.

II. Background

In this section, notations and signal model are provided, and relevant works are reviewed.

A. Notation

Italic, bold lower-case, and bold upper-case letters denote scalars, vectors, and matrices, respectively. Consider a $M \times N$ matrix $\boldsymbol{A}$, its (i, j) entry is signified by $a_{i,j}$ or $\boldsymbol{A}(i, j)$. Besides, $\boldsymbol{A}^T$ and $\boldsymbol{A}^H$ are the transpose and Hermitian transpose of $\boldsymbol{A}$, respectively. The ℓ_p-norm with $p \in (0, 2)$ is defined as $\|\boldsymbol{A}\|_p = (\sum_{i=1,j=1}^{M,N} |a_{i,j}|^p)^{1/p}$ where $|a_{i,j}|$ is the magnitude of $a_{i,j}$, while $\|\boldsymbol{A}\|_{2,1} = \sum_{j=1}^{N} \sqrt{\sum_{i=1}^{M} |a_{i,j}|^2}$ is the $\ell_{2,1}$-norm. The vectorization operator is defined as $\mathrm{vec}(\boldsymbol{A}) = [\boldsymbol{a}_{:,1}^T; \boldsymbol{a}_{:,2}^T; \ldots; \boldsymbol{a}_{:,N}^T]^T$, where $\boldsymbol{a}_{:,j}$ for $j \in [1, N]$ stands for the jth column of $\boldsymbol{A}$. Moreover, $\mathrm{E}\{\cdot\}$ and $(\cdot)^{-1}$ denote the expectation and inverse operators, respectively.

B. Signal Model

Consider an M-sensor ULA with inter-element spacing d. Note that it requires $d \leq \lambda/2$ to avoid phase ambiguity, where λ is the wavelength of the incoming signal. Assume Q far-field, uncorrelated narrow-band signals impinge on the array from distinct directions $\{\theta_1, \theta_2, ..., \theta_Q\}$. The discrete-time complex baseband signal received by the mth sensor at time instant n is modeled as

$$x_m(n) = \sum_{q=1}^{Q} s_q(n) e^{\mathrm{j}2\pi(m-1)\sin(\theta_q)d/\lambda} + e_m(n), \quad (1)$$

where $s_q(n)$ denotes the qth source signal, $\mathrm{j} = \sqrt{-1}$ is the imaginary unit, and $e_m(n)$ is the additive noise, which can be white Gaussian noise and/or non-Gaussian noise.

Organizing the output signal of M sensors in the vector form yields

$$\boldsymbol{x}_n = \boldsymbol{A}\boldsymbol{s}_n + \boldsymbol{e}_n, \tag{2}$$

where $\boldsymbol{x}_n = [x_1(n), \ldots, x_M(n)]^T \in \mathbb{C}^M$ is the received signal vector, $\boldsymbol{s}_n = [s_1(n), \ldots, s_Q(n)]^T \in \mathbb{C}^Q$ is the source vector, $\boldsymbol{e}_n = [e_1(n), \ldots, e_M(n)]^T \in \mathbb{C}^M$ is the noise vector, and $\boldsymbol{A} \in \mathbb{C}^{M\times Q}$ is the array manifold matrix of the form:

$$\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_Q)]. \tag{3}$$

Here, $\boldsymbol{a}(\theta_q)$ is the steering vector:

$$\boldsymbol{a}(\theta_q) = [1, e^{\mathrm{j}2\pi \sin(\theta_q)d/\lambda}, \ldots, e^{\mathrm{j}2\pi(M-1)\sin(\theta_q)d/\lambda}]^T. \tag{4}$$

Collecting N snapshots, the received signal is

$$\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]. \tag{5}$$

In practice, the number of sources is less than those of sensors and snapshots, that is, $Q < \min(M, N)$. Thereby, the noise-free $\boldsymbol{X}$ is of low rank. Note that, in this work, the number of sources Q is assumed to be known. There are many algorithms to detect the number of sources. Interested readers are referred to [46]–[49]. It is worth mentioning that the number of sources can be accurately estimated when they are not very close. Still, source enumeration for closely-spaced sources is a challenging topic [50].

C. Related Work

The MUSIC method exploits the second-order statistics and assumes IID Gaussian noise with variance σ^2 [5]. The covariance matrix of the received signal is $\widehat{\boldsymbol{C}}_x = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^H$. Since the signal and noise are uncorrelated, the covariance matrix of the received signal is given by:

$$\boldsymbol{C}_x = \boldsymbol{A}\boldsymbol{C}_s\boldsymbol{A}^H + \sigma^2\boldsymbol{I}, \tag{6}$$

where $\boldsymbol{C}_s = \mathrm{E}\{\boldsymbol{s}_n\boldsymbol{s}_n^H\}$ is the source covariance matrix. The eignenvalue decomposition of $\boldsymbol{C}_x$ is:

$$\boldsymbol{C}_x = \boldsymbol{U}_s\boldsymbol{\Lambda}_s\boldsymbol{U}_s^H + \sigma^2\boldsymbol{U}_n\boldsymbol{U}_n^H, \tag{7}$$

where $\boldsymbol{U}_s \in \mathbb{C}^{M\times Q}$ is the signal subspace, $\boldsymbol{\Lambda}_s \in \mathbb{C}^{Q\times Q}$ is a diagonal matrix containing the corresponding eigenvalues sorted in descending order, and $\boldsymbol{U}_n \in \mathbb{C}^{M\times(N-Q)}$ is called the noise subspace. The MUSIC first computes the spatial spectrum:

$$P_{\mathrm{MUSIC}}(\theta) = \frac{1}{\boldsymbol{a}^H(\theta)(\boldsymbol{I} - \boldsymbol{U}_s\boldsymbol{U}_s^H)\boldsymbol{a}(\theta)}. \tag{8}$$

Then, the DOA estimates can be obtained via finding the peaks of the spatial spectrum. The conventional MUSIC has excellent performance in Gaussian noise scenarios. When the received signal is corrupted by impulsive noise, its performance is degraded.

To be robust against impulsive noise, ℓ_p-MUSIC combines the ℓ_p-norm with $p \in (1, 2)$ and low-rank matrix factorization strategy to estimate the signal subspace [36]:

$$\min_{\boldsymbol{Y},\boldsymbol{Z}} \|\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}\|_p^p, \tag{9}$$

where $\boldsymbol{Y} \in \mathbb{C}^{M\times Q}$ is a full column rank matrix and $\boldsymbol{Z} \in \mathbb{C}^{Q\times N}$ is a full row rank matrix. It is clear that the range space spanned by the columns of $\boldsymbol{Y}$ is the same as the subspace of $\boldsymbol{A}$, that is,

$$\mathrm{span}(\boldsymbol{A}) = \mathrm{span}(\boldsymbol{Y}). \tag{10}$$

After $\boldsymbol{Y}$ is attained, the projection matrix on the noise subspace is given by

$$\boldsymbol{P}_n = \boldsymbol{I} - \boldsymbol{Y}(\boldsymbol{Y}^H\boldsymbol{Y})^{-1}\boldsymbol{Y}^H. \tag{11}$$

The spatial spectrum is then computed via

$$P_{\ell_p-\mathrm{MUSIC}}(\theta) = \frac{1}{\boldsymbol{a}^H(\theta)\boldsymbol{P}_n\boldsymbol{a}(\theta)}. \tag{12}$$

Subsequently, DOA estimates can be obtained. Although the ℓ_p-MUSIC with $p \in (1, 2)$ attains high estimation accuracy in impulsive noise, it has two drawbacks.

(i) The ℓ_p-MUSIC might not meet the practical timing requirement since solving the ℓ_p-norm based problem has relatively high computational complexity.

(ii) The ℓ_p-MUSIC requires manually tweaking the value of p for different noise environments, e.g., $p \to 2$ for white Gaussian noise and $p \to 1$ for stronger impulsive noise.

Another strategy for robust DOA estimation is to apply the RPCA concept [37], resulting in

$$\min_{\boldsymbol{M},\boldsymbol{S}} \; J(\boldsymbol{M}) + \eta r(\boldsymbol{S}) + \frac{\gamma}{2}\|\boldsymbol{X} - \boldsymbol{M} - \boldsymbol{S}\|_F^2, \tag{13}$$

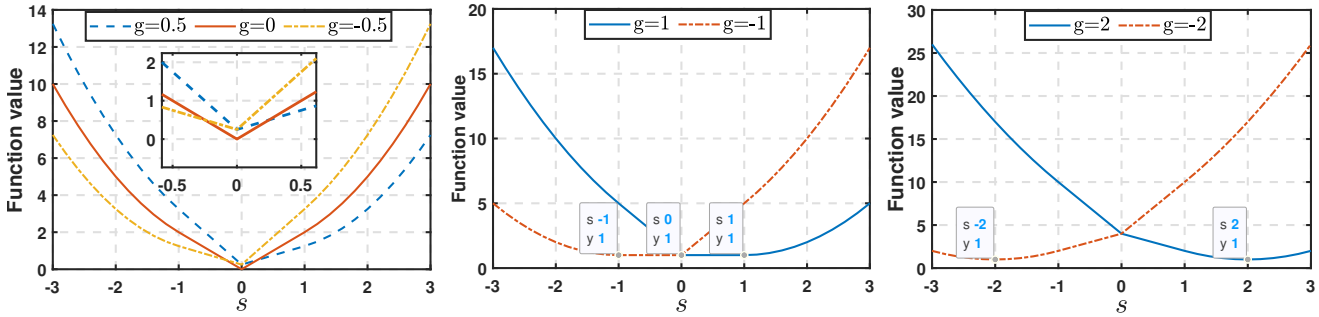
where $\boldsymbol{S}$ is the noise matrix, $\eta > 0$ and $\gamma > 0$ are the penalty parameters, while $r(\cdot)$ is a regularization term. For example, $r(\boldsymbol{S}) = \sum_{j=1}^N \sqrt{\sum_{i=1}^M |s_{i,j}|^2}$ will generate a column-wise sparse $\boldsymbol{S}$ and $r(\boldsymbol{S}) = \sum_{i=1,j=1}^{M,N} |s_{i,j}|$ yields a sparse $\boldsymbol{S}$. Meanwhile, $J(\boldsymbol{M})$ is defined as

$$J(\boldsymbol{M}) \triangleq \sum_{i=1}^{\min(M,N)} F(\delta_i(\boldsymbol{M})), \tag{14}$$

where $\delta_i(\boldsymbol{M})$ denotes the ith singular value of $\boldsymbol{M}$ and $F(\cdot)$ is a weakly convex sparseness measure [37]. It is clear that $J(\boldsymbol{M})$ is capable of seeking for a low-rank $\boldsymbol{M}$ that can be considered as a noise-free received signal. Then, $\boldsymbol{M}$ is exploited for DOA estimation. Note that, in (13), the choices of η and γ have significant impact on the rank of $\boldsymbol{M}$ and sparsity level of $\boldsymbol{S}$, which affects the estimation performance. To the best of our knowledge, there is no strategy to automatically determine these two parameters for different noise scenarios. Therefore, in real-life scenarios, (13) requires manually tuning η and γ to attain satisfactory performance.

III. Proposed Algorithm

In this section, we first introduce the CFN for complex-valued data. The suggested algorithm is then presented.



(a) $|g| < \vartheta$ (b) $|g| = \vartheta$ (c) $|g| > \vartheta$

Fig. 1: Illustration of $\varphi(s)$ in one-dimensional case, where $\vartheta = 1$.

A. Capped Frobenius Norm

Prior to introducing the CFN, we review the Frobenius norm for complex-valued entries. Give a complex matrix $\boldsymbol{A} \in \mathbb{C}^{M\times N}$, the Frobenius norm is defined as

$$\begin{aligned}\|\boldsymbol{A}\|_F &= \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}|a_{i,j}|^2} \\ &= \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\Big(\Re(a_{i,j})^2+\Im(a_{i,j})^2\Big)} \\ &= \left\|\begin{bmatrix}\Re(\boldsymbol{A})\\ \Im(\boldsymbol{A})\end{bmatrix}\right\|_F,\end{aligned} \tag{15}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ are the real and imaginary parts of a complex number.

Based on the Frobenius norm, the CFN for complex-valued matrices is defined as

$$\begin{aligned}&\|\boldsymbol{A}\|_{CF} \\ =&\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\Big(\min(\Re(a_{i,j})^2,\vartheta^2)+\min(\Im(a_{i,j})^2,\vartheta^2)\Big)} \\ =&\left\|\begin{bmatrix}\Re(\boldsymbol{A})\\ \Im(\boldsymbol{A})\end{bmatrix}\right\|_{CF}.\end{aligned} \tag{16}$$

It is easy to see that when $\vartheta \to \infty$, the CFN is equivalent to the Frobenius norm.

Besides, we review TLS [39], [40] for comparison, which has the following form:

$$\begin{aligned}\|\boldsymbol{A}\|_{TLS} &= \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\min(|a_{i,j}|^2,\vartheta^2)} \\ &= \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\min(\Re(a_{i,j})^2+\Im(a_{i,j})^2,\vartheta^2)}.\end{aligned} \tag{17}$$

Compared with the TLS, the CFN introduces the cap threshold to the real and imaginary parts. Consider one entry case, if only the real part is corrupted by an outlier, the CFN limits the real part and does not change the imaginary part. In contrast, the TLS destroys the component without anomaly.

Since the CFN is nonsmooth and nonconvex, it is difficult to optimize the resultant problem. We propose a strategy to convert this challenging problem into a tractable optimization via Lemma 1.

Lemma 1. *Given a* ***real-valued*** *matrix $\boldsymbol{G} \in \mathbb{R}^{M\times N}$. Let $f_\vartheta(\boldsymbol{G}) = \|\boldsymbol{G}\|_{CF}^2$, then $\min_{\boldsymbol{G}} f_\vartheta(\boldsymbol{G})$ is equivalent to*

$$\min_{\boldsymbol{G},\boldsymbol{S}} \Big(\|\boldsymbol{G}-\boldsymbol{S}\|_F^2 + h_\vartheta(\boldsymbol{S})\Big), \tag{18}$$

where $h_\vartheta(\boldsymbol{S})$ is an entry-wise operator, defined as

$$h_\vartheta(\boldsymbol{S}(i,j)) = \begin{cases}-(\vartheta-|s_{i,j}|)^2+\vartheta^2, & |s_{i,j}|<\vartheta,\\ \vartheta^2 & |s_{i,j}|\geq\vartheta.\end{cases} \tag{19}$$

Proof: See Appendix A.

Now we know that optimizing the CFN based problem is simplified as the Frobenius norm optimization with a regularization term. It is clear that its subproblem w.r.t. $\boldsymbol{G}$ is a linear least squares problem with closed-form solution. We then introduce the following lemma to derive the solution of its subproblem w.r.t. $\boldsymbol{S}$.

Lemma 2. *For the subproblem of* (18) *w.r.t. $\boldsymbol{S}$:*

$$\min_{\boldsymbol{S}} \varphi(\boldsymbol{S}) = \min_{\boldsymbol{S}} \Big(\|\boldsymbol{G}-\boldsymbol{S}\|_F^2 + h_\vartheta(\boldsymbol{S})\Big). \tag{20}$$

Its optimal solution is

$$\boldsymbol{S}^*(i,j) = T_\vartheta(g_{i,j}) = \begin{cases}0, & |g_{i,j}|<\vartheta,\\ g_{i,j}, & |g_{i,j}|\geq\vartheta.\end{cases} \tag{21}$$

In addition, the subgradient of $\varphi(\boldsymbol{S})$ at minimizer $\boldsymbol{S}^*$ is

$$\frac{\partial\varphi(\boldsymbol{S})}{\partial s_{i,j}^*} = \begin{cases}0\in[-(g_{i,j}+\vartheta),\vartheta-g_{i,j}], & |g_{i,j}|<\vartheta,\\ 0, & |g_{i,j}|\geq\vartheta.\end{cases} \tag{22}$$

Proof: See Appendix B.

Fig. 1 shows the solution of $\min_s \varphi(s)$ in one-dimensional case where $\vartheta = 1$. We see that when $|g| < \vartheta$, the unique minimizer is $s^* = 0$. For $|g| = \vartheta$, there are multiple minimizers, and one of them is $s^* = g$. When $|g| > \vartheta$, the unique minimizer is $s^* = g$.

Moreover, the solution to (20) is affected by the selection of ϑ. Herein, we propose a strategy to adaptively determine it. If $\boldsymbol{G}$ is considered as the fitting error with the mean being assumed 0, $-\vartheta < g_{i,j} < \vartheta$ is considered as a confidence interval to identify outliers. To obtain a robust confidence

interval in the presence of outliers, we suggest adopting the normalized MAD [51] to calculate the standard deviation:

$$\sigma = 1.4826 \times \mathrm{Med}(|\mathrm{vec}(\boldsymbol{G}) - \mathrm{Med}(\mathrm{vec}(\boldsymbol{G}))|), \tag{23}$$

where $\mathrm{Med}(\cdot)$ is the sample median operator. Then, ϑ is updated via

$$\vartheta = \zeta \times \sigma, \tag{24}$$

where $\zeta > 0$ controls the range of confidence interval. The selection of ζ is discussed in Section V.

B. CFN-MUSIC

1) Signal Subspace Estimation: To achieve robust DOA estimation, we suggest combining the CFN and low-rank factorization to search for the signal subspace, resulting in

$$\min_{\boldsymbol{Y},\boldsymbol{Z}} \|\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}\|_{CF}^2. \tag{25}$$

In accordance to the definition of the CFN, (25) is equivalent to the following real-valued problem:

$$\min_{\boldsymbol{Y},\boldsymbol{Z}} \left\| \begin{bmatrix} \Re(\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}) \\ \Im(\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}) \end{bmatrix} \right\|_{CF}^2. \tag{26}$$

Employing Lemma 1, (26) is converted into the Frobenius norm optimization with a regularization term:

$$\begin{aligned} \min_{\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{S}} \mathcal{L}_\vartheta(\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{S}) = \min_{\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{S}} \left\| \begin{bmatrix} \Re(\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}) \\ \Im(\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}) \end{bmatrix} - \begin{bmatrix} \Re(\boldsymbol{S}) \\ \Im(\boldsymbol{S}) \end{bmatrix} \right\|_F^2 \\ + h_\vartheta \left(\begin{bmatrix} \Re(\boldsymbol{S}) \\ \Im(\boldsymbol{S}) \end{bmatrix} \right). \end{aligned} \tag{27}$$

It is clear that (27) is a multi-variable nonconvex optimization problem. We then exploit the PBCD concept to deal with (27), resulting in the following iterative procedure:

$$\boldsymbol{S}^{k+1} = \arg\min_{\boldsymbol{S}} \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}), \tag{28a}$$

$$\boldsymbol{Z}^{k+1} = \arg\min_{\boldsymbol{Z}} \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k, \boldsymbol{Z}, \boldsymbol{S}^{k+1}) + \mu \left\| \boldsymbol{Z} - \boldsymbol{Z}^k \right\|_F^2, \tag{28b}$$

$$\boldsymbol{Y}^{k+1} = \arg\min_{\boldsymbol{Y}} \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1}) + \mu \left\| \boldsymbol{Y} - \boldsymbol{Y}^k \right\|_F^2, \tag{28c}$$

where $\mu > 0$ is the proximal parameter. It is seen that the PBCD alternately updates one of the variables while fixing the remaining variables at each iteration.

We first focus on tackling (28a) and reformulate it:

$$\widetilde{\boldsymbol{S}}^{k+1} = \arg\min_{\widetilde{\boldsymbol{S}}} \left(\|\widetilde{\boldsymbol{G}}^k - \widetilde{\boldsymbol{S}}\|_F^2 + h_{\vartheta^{k+1}}(\widetilde{\boldsymbol{S}}) \right), \tag{29}$$

where $\widetilde{\boldsymbol{G}}^k = \begin{bmatrix} \Re(\boldsymbol{X} - \boldsymbol{Y}^k\boldsymbol{Z}^k) \\ \Im(\boldsymbol{X} - \boldsymbol{Y}^k\boldsymbol{Z}^k) \end{bmatrix} \in \mathbb{R}^{2M \times N}$ and $\widetilde{\boldsymbol{S}}^k = \begin{bmatrix} \Re(\boldsymbol{S}) \\ \Im(\boldsymbol{S}) \end{bmatrix} \in \mathbb{R}^{2M \times N}$. Based on Lemma 2, an optimal solution of (29) is computed as

$$\widetilde{\boldsymbol{S}}^{k+1} = T_{\vartheta^{k+1}}(\widetilde{\boldsymbol{G}}^k). \tag{30}$$

Algorithm 1 CFN-MUSIC

Input: Received data matrix $\boldsymbol{X} \in \mathbb{R}^{M \times N}$, target number Q, $\mu = 0.1$, and $\vartheta^0 = 100$.
Initialize: Randomize $\boldsymbol{Y} \in \mathbb{R}^{M \times Q}$ and $\boldsymbol{Z} \in \mathbb{R}^{Q \times N}$
1. Signal subspace estimation
for $k = 1, 2, \ldots, K_{\max}$ **do**
1) Update ϑ^{k+1} via (32)
2) Compute $\widetilde{\boldsymbol{S}}^{k+1}$ via (30)
3) Update $\boldsymbol{S}^{k+1}$ via (31)
4) Compute $\boldsymbol{Z}^{k+1}$ via (34)
5) Compute $\boldsymbol{Y}^{k+1}$ via (36)
Stop if stopping criterion is met.
end for
2. DOA estimation
6) Compute projection matrix on signal subspace via (37)
7) Compute spatial spectrum via (38)
8) Find $\widetilde{\theta}_q$ with $q \in [1, Q]$ via peak search of spatial spectrum
Output: $\widetilde{\theta}_q$ with $q \in [1, Q]$ and spatial spectrum

After obtaining $\widetilde{\boldsymbol{S}}^{k+1}$, $\boldsymbol{S}^{k+1}$ is updated by:

$$\boldsymbol{S}^{k+1} = \widetilde{\boldsymbol{S}}^{k+1}(1:M,:) + \mathrm{j}\widetilde{\boldsymbol{S}}^{k+1}(M+1:2M,:), \tag{31}$$

where $\widetilde{\boldsymbol{S}}^{k+1}(1:M,:)$ extracts the rows from the 1st to Mth to obtain a matrix.

Note that we require updating ϑ prior to handling (28a). To ensure convergence of the proposed algorithm, we suggest computing ϑ^{k+1} via:

$$\vartheta^{k+1} = \min(\zeta \times \sigma^{k+1}, \vartheta^k). \tag{32}$$

That is, the value of ϑ^k is monotonically nonincreasing during iterations.

Subsequently, we deal with (28b). Apparently, it is equivalent to the following complex-valued problem:

$$\boldsymbol{Z}^{k+1} = \arg\min_{\boldsymbol{Z}} \|\boldsymbol{X} - \boldsymbol{Y}^k\boldsymbol{Z} - \boldsymbol{S}^{k+1}\|_F^2 + \mu\|\boldsymbol{Z} - \boldsymbol{Z}^k\|_F^2. \tag{33}$$

Its solution is

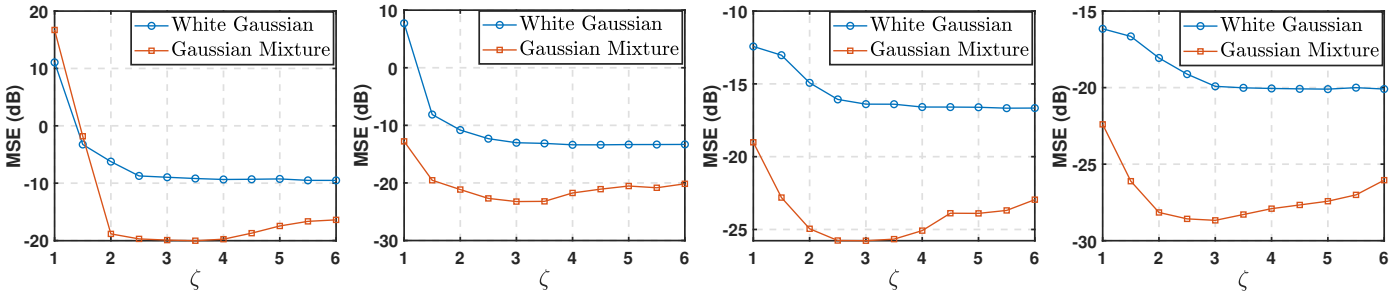
$$\boldsymbol{Z}^{k+1} = \left((\boldsymbol{Y}^k)^H\boldsymbol{Y}^k + \mu\boldsymbol{I} \right)^{-1} \left((\boldsymbol{Y}^k)^H(\boldsymbol{X} - \boldsymbol{S}^{k+1}) + \mu\boldsymbol{Z}^k \right). \tag{34}$$

Finally, we handle (28c). Similar to (28b), it is reformulated as a complex-valued problem:

$$\boldsymbol{Y}^{k+1} = \arg\min_{\boldsymbol{Y}} \|\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}^{k+1} - \boldsymbol{S}^{k+1}\|_F^2 + \mu\|\boldsymbol{Y} - \boldsymbol{Y}^k\|_F^2, \tag{35}$$

whose solution is

$$\boldsymbol{Y}^{k+1} = \left((\boldsymbol{X} - \boldsymbol{S}^{k+1})(\boldsymbol{Z}^{k+1})^H + \mu\boldsymbol{Y}^k \right) \left(\boldsymbol{Z}^{k+1}(\boldsymbol{Z}^{k+1})^H + \mu\boldsymbol{I} \right)^{-1}. \tag{36}$$



(a) SNR = -3 (b) SNR = 0 (c) SNR = 3 (b) SNR = 6

Fig. 2: Impact of ζ for different SNRs.

2) DOA Estimation: After we obtain the signal subspace $\boldsymbol{Y}^{k+1}$, denoted by $\boldsymbol{Y}$ for concise expression, the projection matrix on the signal subspace is given by

$$\boldsymbol{P}_s = \boldsymbol{Y}(\boldsymbol{Y}^H\boldsymbol{Y})^{-1}\boldsymbol{Y}^H. \tag{37}$$

The spatial spectrum is then computed via

$$P_{\text{CFN-MUSIC}}(\theta) = \frac{1}{\boldsymbol{a}^H(\theta)(\boldsymbol{I}-\boldsymbol{P}_s)\boldsymbol{a}(\theta)}. \tag{38}$$

Finally, the DOA estimates can be found via seeking for the Q peaks of the spatial spectrum.

The proposed approach is termed **c**apped **F**robenius **n**orm based MUSIC (CFN-MUSIC) whose steps are summarized in Algorithm 1. There are two termination conditions. One is to reach the maximum iteration number. The other one is that CFN-MUSIC stops when it is approximately convergent [52], defined as

$$\frac{\|\boldsymbol{Y}^{k+1}\boldsymbol{Z}^{k+1}-\boldsymbol{Y}^k\boldsymbol{Z}^k\|_F^2}{MN} \le 10^{-8}. \tag{39}$$

IV. Theoretical Analysis

This section analyzes the convergence behavior and computational complexity of CFN-MUSIC.

A. Convergence Behavior

The convergence of the objective function value is established in Theorem 1.

Theorem 1. *Let $\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)$ be the objective function value generated by Algorithm 1, then it satisfies the following properties:*

(i) $\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)$ is nonincreasing as all variables update.

(ii) $\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)$ is lower bounded.

Therefore, $\{\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)\}_{k\in\mathbb{N}}$ is convergent.

Proof: See Appendix C.

We then analyze the variable sequence behavior in Theorem 2.

Theorem 2. *Let $(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)$ be the sequence generated by Algorithm 1. For any initialization with finite $\|\boldsymbol{Y}^1\|_F$, $\|\boldsymbol{Z}^1\|_F$ and $\mathcal{L}_{\vartheta^1}(\boldsymbol{Y}^1,\boldsymbol{Z}^1,\boldsymbol{S}^1)$, we have the following:*

(i) The sequence $(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)$ is bounded.

(ii) There exists a subsequence $(\boldsymbol{Y}^{k_i},\boldsymbol{Z}^{k_i},\boldsymbol{S}^{k_i})$ converging to an accumulation point $(\boldsymbol{Y}^*,\boldsymbol{Z}^*,\boldsymbol{S}^*)$.

(iii) The accumulation point $(\boldsymbol{Y}^*,\boldsymbol{Z}^*,\boldsymbol{S}^*)$ is a critical point.

Therefore, the estimated variable sequence has a subsequence converging to a critical point.

Proof: See Appendix D.

B. Computational Complexity

In this subsection, we study the computational complexity of CFN-MUSIC. The complexity of updating $\boldsymbol{S}^k$ is $\mathcal{O}(MNQ)$, while computing $\boldsymbol{Z}^k$ and $\boldsymbol{Y}^k$ has a complexity of $\mathcal{O}(MNQ^2)$. Therefore, the computational complexity in the signal subspace estimation is $\mathcal{O}(K_{\max}MNQ^2)$. In DOA estimation, the complexity for calculating the projection matrix is $\mathcal{O}(M^2Q)$. Then, the spatial spectrum computation has a complexity of $\mathcal{O}(M^2T_\theta)$ where T_θ is the number of searches which depends on the resolution and range. Thereby, the computational complexity in DOA estimation is $\mathcal{O}(M^2T_\theta)$. As a result, the overall complexity is $\mathcal{O}(K_{\max}MNQ^2+M^2T_\theta)$.

Table I tabulates the computational complexities of three robust estimators where K_{inner} is the number of inner iterations in the ℓ_p-MUSIC. For LRMA-ADMM, $M+Q>Q^2$ holds when the number of targets is much less than antenna number. It is observed that our method has the lowest computational complexity among them.

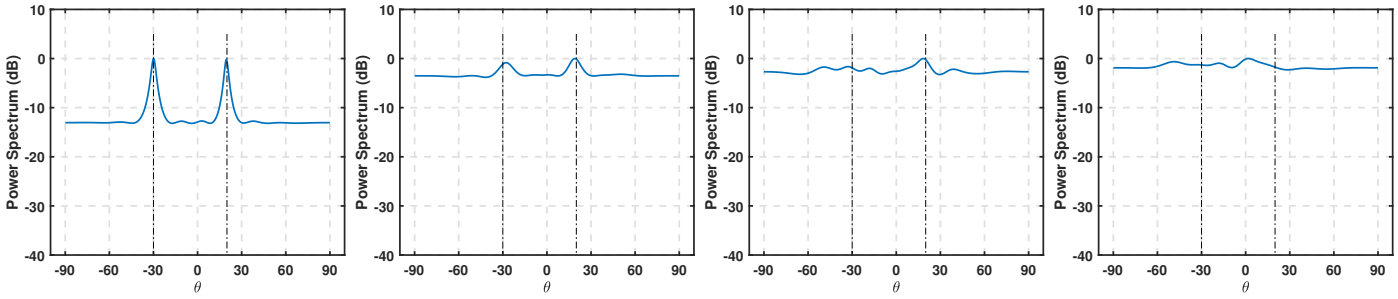
TABLE I: Complexity comparison of robust algorithms

Method	Computational complexity
CFN-MUSIC	$\mathcal{O}(K_{\max}MNQ^2+M^2T_\theta)$
ℓ_p-MUSIC	$\mathcal{O}(K_{\max}K_{\text{inner}}MNQ^2+M^2T_\theta)$
LRMA-ADMM	$\mathcal{O}(K_{\max}MN(M+Q)+M^2T_\theta)$

V. Simulation

In our simulations, the signals are Gaussian distributed with equal power, and the ULA is comprised of $M=10$ sensors with the inter-element spacing of $d=0.5\lambda$. In addition, the number of snapshots is $N=128$, and unless stated otherwise, the DOAs of the signals are $\theta_1=-30^\circ$ and $\theta_2=20^\circ$.

We consider two types of noise, namely, zero-mean white Gaussian noise and two-component GMM noise. The GMM



(a) Proposed (b) ℓ_p-MUSIC (c) LRMA-ADMM (b) MUSIC

Fig. 3: Reconstructed spatial spectrum of two signals by different algorithms in -9dB GMM noise.

noise is a popular impulsive noise [53], and its PDF is given by

$$p_v(v) = \frac{c_1}{\pi\sigma_1^2}\exp\left(-\frac{|v|^2}{\sigma_1^2}\right) + \frac{c_2}{\pi\sigma_2^2}\exp\left(-\frac{|v|^2}{\sigma_2^2}\right), \quad (40)$$

where $0 < c_i < 1$ with $c_1 + c_2 = 1$ while σ_1^2 and σ_2^2 are variances. To simulate the impulsive noise, it requires $\sigma_2^2 \gg \sigma_1^2$ and $c_2 < c_1$. It means that sparse and high power noise samples with variance σ_2^2 and occurrence probability c_2 are mixed in Gaussian background noise with small variance σ_1^2. We set $\sigma_2^2 = 100\sigma_1^2$ and $c_2 = 0.1$. The signal-to-noise ratio (SNR) of GMM noise in dB is defined as

$$\text{SNR} = 10\log_{10}\Big(\frac{\|\boldsymbol{X}\|_F^2}{MN\sigma_v^2}\Big), \quad (41)$$

where $\sigma_v^2 = \sum_{i=1}^2 c_i\sigma_i^2$ is the total noise variance.

A. Performance Metrics

The performance is measured by the mean square error (MSE) in dB, defined as

$$\text{MSE} = 10\log_{10}\Big(\frac{1}{M_c}\sum_{m=1}^{M_c}\sum_{q=1}^{Q}\frac{(\widehat{\theta}_{q,m} - \theta_q)^2}{Q}\Big), \quad (42)$$

where $M_c = 100$ is the number of Monte Carlo trials and $\widehat{\theta}_{q,m}$ is the qth DOA estimate at the mth trial.

Another performance metric is the probability of resolution. For each individual source, if the magnitude of the DOA estimation error is less than $0.5°$, then the corresponding source is considered to be successfully resolved.

B. Investigation of ζ

Since the proposed algorithm has one tunable parameter, viz. ζ, we first investigate its impact on estimation performance. The results are plotted in Fig. 2 where four SNR levels from -3dB to 6dB are considered. It is seen that, in all noise levels, the values of MSE decrease with boosting ζ in Gaussian noise, while MSEs first reduce and then increase with ζ in impulsive noise. This is because a small ζ generates a narrow confidence interval, leading to many entries being considered as outlier-contaminated elements. In Gaussian noise scenarios, all observed entries are not corrupted by anomalies and thus a bigger ζ results in better performance. Under impulsive noise, a small ζ leads to many entries to be mistaken as outliers, while a very large ζ cannot identify all anomaly-contaminated entries. To achieve satisfactory performance in both types of noise, we select $\zeta = 3$ for the following simulations.

Moreover, we provide an intuitive interpretation why ζ decreases as the ratio of the number of anomalies and that of the noisy components increases. Based on (41), we attain

$$\sigma_1^2 = \frac{\|\boldsymbol{X}\|_F^2}{MN(99c_2+1)10^{\text{SNR}/10}}, \quad (43)$$

that is, σ_1^2 decreases as c_2 increases. This means that when the ratio of anomalies to noisy components is increased, the variance of Gaussian component reduces. With an appropriate sigma, the 3-sigma rule has been adopted for outlier detection [54], [55]. Specifically, the samples that fall outside 3-sigma interval are likely to be outliers as 99.7% of samples fall within 3-sigma interval for normal distribution. For unimodal distributions, the probability of being within the interval is at least 95% by the Vysochanskij–Petunin inequality [56]. Therefore, for GMM noise, the interval $[-3\sigma_1, 3\sigma_1]$ can be considered as a baseline to identify outliers. It is reported that the normalized MAD provides a robust measurement for standard deviation estimation [51]. Therefore, given a SNR, the estimated standard deviation σ of GMM noise via the normalized MAD remains relatively stable with various c_2. If ζ is chosen as $\zeta\sigma \approx 3\sigma_1$, $[-\zeta\sigma, \zeta\sigma]$ can be deemed an appropriate interval to differentiate anomalies in GMM noise. We have analyzed that when c_2 becomes large, σ_1 reduces, whereas σ remains relatively stable. Consequently, ζ decreases as c_2 increases.

When $c_2 = 0$, the GMM noise reduces to white Gaussian noise. The normalized MAD is able to provide an accurate standard deviation estimate for Gaussian noise, indicating that the interval $[-3\sigma, 3\sigma]$ can cover 99.7% samples of zero-mean white Gaussian noise. Therefore, we set $\zeta = 3$ to balance the estimation performance between GMM and Gaussian noise scenarios, although $\zeta = 3$ generates a little bit of performance loss for Gaussian noise.

C. Performance Comparison

The CFN-MUSIC is compared with TLS-MUSIC, MUSIC [5], PUMA [9], EPUMA [12], ESPRIT [7], root-MUSIC [57], ℓ_p-MUSIC [36], LRMA-ADMM [37], and fast improved covariance matrix reconstruction approach

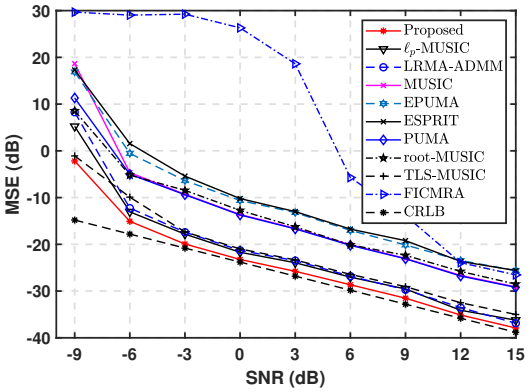


Fig. 4: MSE versus SNR in GMM noise.

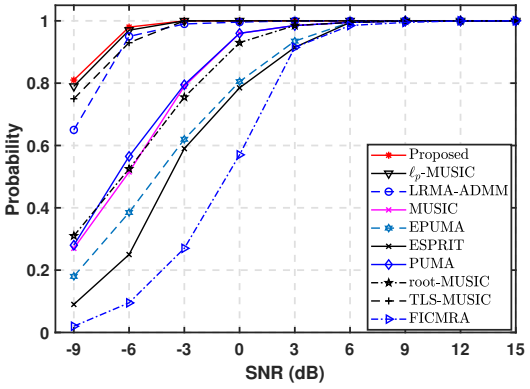


Fig. 5: Probability of resolution versus SNR in GMM noise.

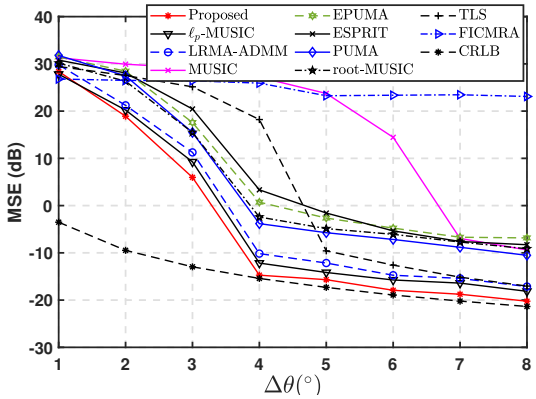


Fig. 6: MSE versus $\Delta\theta(^\circ)$ in 3dB GMM noise

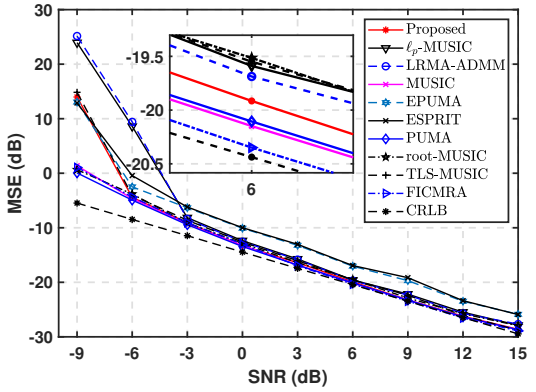


Fig. 7: MSE versus SNR in zero-mean white Gaussian noise.

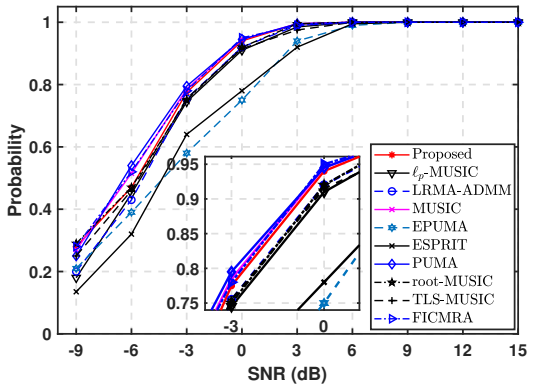


Fig. 8: Probability of resolution versus SNR in zero-mean white Gaussian noise.

(FICMRA) [17]. For the ℓ_p-MUSIC, p is set as 1.1 which is the suggested value in [36]. Moreover, we include the Cramér–Rao lower bound (CRLB) as a performance benchmark. The CRLB in Gaussian noise is presented in [58], while CRLB for GMM noise is provided in [36].

We first compare the spatial spectrum of CFN-MUSIC, ℓ_p-MUSIC, LRMA-ADMM, and MUSIC since these four algorithms are able to compute the spatial spectrum. The results are plotted in Fig. 3 where the output signal is corrupted by -9dB GMM noise. The corresponding MSEs of CFN-MUSIC, ℓ_p-MUSIC, LRMA-ADMM, and MUSIC are -10.5dB, 4.8dB, 7.4dB, and 25.3dB, respectively. It is clear that the CFN-MUSIC achieves the smallest estimation error among them. In addition, the computation times of CFN-MUSIC, ℓ_p-MUSIC, LRMA-ADMM, and MUSIC are 0.13s, 5.80s, 0.16s and 0.11s. It is known that the proposed algorithm is faster than the existing robust methods. Note that since PUMA, EPUMA, ESPRIT, and root-MUSIC do not need to compute the spatial spectrum, they are faster than these four approaches.

1) Performance in GMM Noise: We then compare all methods in GMM noise. Fig. 4 shows MSE versus SNR $\in [-9, 15]$dB. It is seen that the CFN-MUSIC, TLS-MUSIC, ℓ_p-MUSIC, and LRMA-ADMM attain smaller MSEs than MUSIC, PUMA, EPUMA, ESPRIT, root-MUSIC, and FICMRA. Besides, the MSEs of CFN-MUSIC are smaller than those of TLS-MUSIC, ℓ_p-MUSIC, and LRMA-ADMM in all SNRs. Fig 5 plots the probability of resolution versus SNR. Similarly, CFN-MUSIC, TLS-MUSIC, ℓ_p-MUSIC, and LRMA-ADMM achieve higher probabilities than the other five algorithms in the range of -9dB to 6 dB. In addition, CFN-MUSIC and ℓ_p-MUSIC have comparable probabilities that are slightly higher than those of TLS-MUSIC and LRMA-ADMM in low SNRs. When SNR is larger than 9dB, all methods exhibit good performance. Fig. 6 shows the estimation performance versus source separation, where DOAs are $\{20^\circ, 20^\circ + \Delta\theta\}$. It is seen that FICMRA attains the best performance with $\Delta\theta = 1^\circ$ since it is a high-resolution DOA estimation method. In addition, the

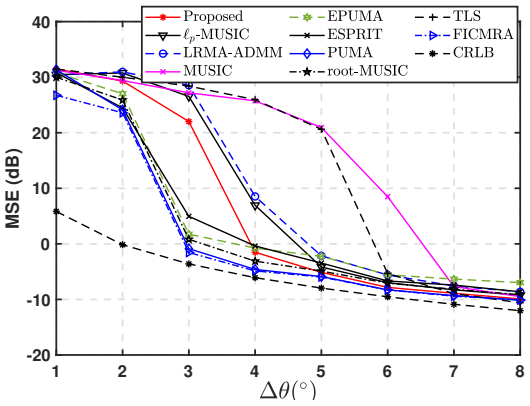


Fig. 9: MSE versus $\Delta\theta(^\circ)$ in 3dB white Gaussian noise.

CFN-MUSIC and ℓ_p-MUSIC yield comparable MSEs, which are superior to the remaining methods. When $\Delta\theta \in [2^\circ, 8^\circ]$, the CFN-MUSIC has the lowest estimation error.

2) Performance in White Gaussian Noise: We further compare all methods in white Gaussian noise. Note that the parameters of CFN-MUSIC, TLS-MUSIC, ℓ_p-MUSIC, and LRMA-ADMM do not change. The results are shown in Figs. 7 and 8. It is seen that the PUMA, MUSIC, CFN-MUSIC, and FICMRA obtain better performance than TLS-MUSIC, root-MUSIC, ℓ_p-MUSIC, LRMA-ADMM, ESPRIT, and EPUMA. That is, the performance of ℓ_p-MUSIC and LRMA-ADMM degrades in white Gaussian noise. Besides, CFN-MUSIC attains comparable performance to MUSIC and PUMA, indicating that our method does not degrade in white Gaussian noise. A similar conclusion is found in probability of resolution versus SNR, plotted in Fig. 8. The estimation performance versus $\Delta\theta$ is shown in Fig. 9. We see that the CFN-MUSIC obtains the best performance among the four robust algorithms. When $\Delta\theta \in [1^\circ, 4^\circ]$, the MSEs of the CFN-MUSIC are slightly higher than those of the conventional approaches, except for MUSIC. When $\Delta\theta \in [5^\circ, 8^\circ]$, the CFN-MUSIC is on par with PUMA and FICMRA.

VI. Conclusion

In this article, we have devised an efficient DOA estimator using the CFN and low-rank matrix factorization. The proposed algorithm, termed CFN-MUSIC, has one auxiliary parameter that is considered as a threshold to differentiate normal and outlier-contaminated elements. Thereby, we exploit the robust statistics concept to design a strategy with which the threshold is adaptively determined. Besides, we adopt the half-quadratic theory to simplify the nonconvex and nonsmooth problem caused by the CFN. The intractable problem is converted into the Frobenius norm optimization with a regularization term, which is handled via alternating convex optimization with closed-form solution. The convergence behavior of CFN-MUSIC has been analyzed, that is, the objective function is convergent while there exists a subsequence in the variable sequence converging to a critical point. Simulation results demonstrate that the CFN-MUSIC achieves better performance in GMM noise in terms of estimation accuracy and the probability of resolution. Besides, in Gaussian noise, the performance of the CFN-MUSIC is comparable to that of MUSIC and PUMA.

Appendix A
Proof of Lemma 1

Since the half-quadratic theory is adopted, we first introduce it using the following lemma:

Lemma 3. *[43]: Given $f(g)$ and $h(s)$, if $f(g)$ makes $\phi(g) = g^2 - f(g)$ convex, and $h(s)$ generates a convex function $\psi(s) = s^2 + h(s)$, we then have*

$$f(g) = \inf_s \ \Big((g-s)^2 + h(s)\Big), \ g \in (-\infty, +\infty), \tag{44a}$$

$$h(s) = \sup_g \ \Big(-(g-s)^2 + f(g)\Big), \ s \in (-\infty, +\infty). \tag{44b}$$

Based on the half-quadratic theory, we derive the equivalent problem of $\min_{\boldsymbol{G}} f_\vartheta(\boldsymbol{G})$. It is clear that $\min_{\boldsymbol{G}} f_\vartheta(\boldsymbol{G})$ is equivalent to

$$\begin{aligned}\min_{\boldsymbol{G}} f_\vartheta(\boldsymbol{G}) &= \min_{\boldsymbol{G}} \sum_{i=1,j=1}^{M,N} \min(g_{i,j}^2, \vartheta^2) \\ &= \sum_{i=1,j=1}^{M,N} \min_{g_{i,j}} \min(g_{i,j}^2, \vartheta^2).\end{aligned} \tag{45}$$

That is, if each entry attains minimum, the function value achieves minimum. Therefore, we focus on one entry, say, g, without loss of generality. In accordance to Lemma 3, we know that $f_\vartheta(g)$ results in convex $\phi(g) = g^2 - \min(g^2, \vartheta^2)$ with $g \in (-\infty, +\infty)$. It is possible to make an assumption that there exists $h_\vartheta(s)$ to render $\psi(s) = s^2 + h_\vartheta(s)$ convex. Thus, we have

$$\begin{aligned} h_\vartheta(s) &= \begin{cases} \sup\limits_g \ (-(g-s)^2 + g^2), & |g| < \vartheta, \\ \sup\limits_g \ (-(g-s)^2 + \vartheta^2), & |g| \geq \vartheta, \end{cases} \\ &= \begin{cases} -(\vartheta - |s|)^2 + \vartheta^2, & |s| < \vartheta, \\ \vartheta^2, & |s| \geq \vartheta. \end{cases} \end{aligned} \tag{46}$$

We then plug (46) into $\psi(s) = s^2 + h_\vartheta(s)$, and know that the resultant $\psi(s)$ is convex. Therefore, $f_\vartheta(g)$ and $h_\vartheta(s)$ make Lemma 1 hold, resulting in

$$f_\vartheta(g) = \inf_s \ \Big((g-s)^2 + h_\vartheta(s)\Big). \tag{47}$$

Since $l(s) = (g-s)^2 + h_\vartheta(s)$ with $s \in (-\infty, +\infty)$ is continuous, we have

$$\begin{aligned}\min_g \ f_\vartheta(g) &= \min_g \inf_s \ \Big((g-s)^2 + h_\vartheta(s)\Big) \\ &= \min_{g,s} \ \Big((g-s)^2 + h_\vartheta(s)\Big).\end{aligned} \tag{48}$$

Therefore, $\min_{\boldsymbol{G}} f_\vartheta(\boldsymbol{G})$ is equivalent to

$$\begin{aligned}\min_{\boldsymbol{G}} f_\vartheta(\boldsymbol{G}) &= \sum_{i=1,j=1}^{M,N} \min_{g_{i,j},s_{i,j}} \Big((g_{i,j}-s_{i,j})^2 + h_\vartheta(s_{i,j})\Big) \\ &= \min_{\boldsymbol{G},\boldsymbol{S}} \sum_{i=1,j=1}^{M,N} \Big((g_{i,j}-s_{i,j})^2 + h_\vartheta(s_{i,j})\Big) \\ &= \min_{\boldsymbol{G},\boldsymbol{S}} \Big(\|\boldsymbol{G}-\boldsymbol{S}\|_F^2 + h_\vartheta(\boldsymbol{S})\Big). \end{aligned} \tag{49}$$

The proof is complete. ■

APPENDIX B
PROOF OF LEMMA 2

$$\begin{aligned}&\min_{\boldsymbol{S}} \Big(\|\boldsymbol{G}-\boldsymbol{S}\|_F^2 + h_\vartheta(\boldsymbol{S})\Big) \\ =&\min_{\boldsymbol{S}} \sum_{i=1,j=1}^{M,N} \Big((g_{i,j}-s_{i,j})^2 + h_\vartheta(s_{i,j})\Big) \\ =& \sum_{i=1,j=1}^{M,N} \min_{s_{i,j}} \Big((g_{i,j}-s_{i,j})^2 + h_\vartheta(s_{i,j})\Big) \end{aligned} \tag{50}$$

Similar to the proof of Lemma 1, we focus on one entry, say, s, resulting in

$$\begin{aligned} s^* &= \arg\min_s \ \varphi(s) \\ &= \begin{cases} \arg\min_s \ g^2 - 2gs + 2\vartheta|s|, & |s|<\vartheta, \\ \arg\min_s \ (g-s)^2 + \vartheta^2 & |s|\geq\vartheta, \end{cases} \\ &= \begin{cases} \arg\min_s \ g^2 - 2s(g-\vartheta), & 0\leq s<\vartheta, \\ \arg\min_s \ g^2 - 2s(g+\vartheta), & -\vartheta<s<0, \\ \arg\min_s \ (g-s)^2 + \vartheta^2, & |s|\geq\vartheta, \end{cases} \\ &= \begin{cases} 0, & |g|<\vartheta, \\ g, & |g|\geq\vartheta. \end{cases} \end{aligned} \tag{51}$$

Although $h_\vartheta(s)$ is nonconvex, $\varphi(s)$ is convex since $\psi(s) = s^2 + h_\vartheta(s)$ is convex according to Lemma 3. As a result, (51) is the optimal solution.

Therefore, the optimal solution of (50) is

$$\boldsymbol{S}^*(i,j) = T_\vartheta(g_{i,j}) = \begin{cases} 0, & |g_{i,j}|<\vartheta, \\ g_{i,j}, & |g_{i,j}|\geq\vartheta. \end{cases} \tag{52}$$

Moreover, for $s^* = g_{i,j}$ with $|g_{i,j}|\geq\vartheta$, the subgradient is $2(g_{i,j}-s^*)=0$. For $s^*=0$ such that $|s^*|<\vartheta$ and $|g_{i,j}|<\vartheta$, the subgradient is $[-(g_{i,j}+\vartheta), \vartheta - g_{i,j}]$, such that $0\in[-(g_{i,j}+\vartheta), \vartheta-g_{i,j}]$. The proof is complete. ■

APPENDIX C
PROOF OF THEOREM 1

A. Proof of Property (i)

It is clear that the objective function value is affected by four variables, namely, ϑ^k, $\boldsymbol{S}^k$, $\boldsymbol{Z}^k$ and $\boldsymbol{Y}^k$. In $\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)$, only the term of $h_{\vartheta^k}(\boldsymbol{S}^k)$ is related to ϑ^k and thus we compute the partial derivative of $h_{\vartheta^k}(\boldsymbol{S}^k)$ w.r.t. ϑ as

$$\frac{\partial h_\vartheta(s_{i,j}^k)}{\partial\vartheta} = \begin{cases} 2|s_{i,j}^k|, & |s_{i,j}^k|<\vartheta^k, \\ 2\vartheta^k, & |s_{i,j}^k|\geq\vartheta^k, \end{cases} \tag{53}$$

where $\vartheta^k\geq 0$ and thus we have

$$\frac{\partial\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)}{\partial\vartheta} = \sum_{i=1,j=1}^{M,N} \frac{\partial h_{\vartheta^k}(s_{i,j}^k)}{\partial\vartheta} \geq 0, \tag{54}$$

which indicates that $\mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k) \leq \mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)$ with $\vartheta^{k+1}\leq\vartheta^k$. From (32), it is known that $\vartheta^{k+1}\leq\vartheta^k$ must hold and thus we attain

$$\mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k) \leq \mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k). \tag{55}$$

From Lemma 2, it is known that $\boldsymbol{S}^{k+1}$ is an optimal solution to $\min_{\boldsymbol{S}} \ \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S})$, resulting in

$$\mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^{k+1}) \leq \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k). \tag{56}$$

Furthermore, $\boldsymbol{Z}^{k+1}$ is the optimal solution to (28b), leading to

$$\begin{aligned} &\mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^{k+1},\boldsymbol{S}^{k+1}) + \mu\|\boldsymbol{Z}^{k+1}-\boldsymbol{Z}^k\|_F^2 \\ &\qquad \leq \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^{k+1}) \\ \Leftrightarrow\ &\mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^{k+1},\boldsymbol{S}^{k+1}) - \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^{k+1}) \\ &\qquad \leq -\mu\|\boldsymbol{Z}^{k+1}-\boldsymbol{Z}^k\|_F^2. \end{aligned} \tag{57}$$

Similarly, for the update of $\boldsymbol{Y}^k$, we have

$$\begin{aligned} &\mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^{k+1},\boldsymbol{Z}^{k+1},\boldsymbol{S}^{k+1}) + \mu\|\boldsymbol{Y}^{k+1}-\boldsymbol{Y}^k\|_F^2 \\ &\qquad \leq \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^{k+1},\boldsymbol{S}^{k+1}) \\ \Leftrightarrow\ &\mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^{k+1},\boldsymbol{Z}^{k+1},\boldsymbol{S}^{k+1}) - \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k,\boldsymbol{Z}^{k+1},\boldsymbol{S}^{k+1}) \\ &\qquad \leq -\mu\|\boldsymbol{Y}^{k+1}-\boldsymbol{Y}^k\|_F^2. \end{aligned} \tag{58}$$

Combining (55), (56) and (58) yields

$$\begin{aligned} &\mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^{k+1},\boldsymbol{Z}^{k+1},\boldsymbol{S}^{k+1}) - \mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k) \\ &\qquad \leq -\mu\Big(\|\boldsymbol{Z}^{k+1}-\boldsymbol{Z}^k\|_F^2 + \|\boldsymbol{Y}^{k+1}-\boldsymbol{Y}^k\|_F^2\Big). \end{aligned} \tag{59}$$

This leads to

$$\mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^{k+1},\boldsymbol{Z}^{k+1},\boldsymbol{S}^{k+1}) \leq \mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k). \tag{60}$$

Therefore, $\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)$ is monotonically nonincreasing.

B. Proof of Property (ii)

Based on the definition of $\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)$, it is easy to know that $\lim_{k\to\infty}\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)\geq 0$. Thereby, the objective function value is lower bounded.

In accordance to Properties (i) and (ii), we obtain the conclusion that $\{\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k,\boldsymbol{Z}^k,\boldsymbol{S}^k)\}_{k\in\mathbb{N}}$ is convergent. The proof is complete. ■

APPENDIX D
PROOF OF THEOREM 2

A. Proof of Property (i)

Before analyzing the boundedness of $(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k)$, we introduce a function:

$$\widetilde{\mathcal{L}}_{\vartheta^k}(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{S}^k) = \mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k) + \mu\Big(\|\boldsymbol{Z}^k - \boldsymbol{Z}\|_F^2 + \|\boldsymbol{Y}^k - \boldsymbol{Y}\|_F^2\Big). \tag{61}$$

It is easy to attain the following equalities:

$$\begin{aligned}\widetilde{\mathcal{L}}_{\vartheta^k}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k) &= \mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k),\\ \widetilde{\mathcal{L}}_{\vartheta^{k+1}}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k) &= \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k),\\ \widetilde{\mathcal{L}}_{\vartheta^{k+1}}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^{k+1}) &= \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^{k+1}).\end{aligned} \tag{62}$$

Given any initialization such that $\mathcal{L}_{\vartheta^1}(\boldsymbol{Y}^1, \boldsymbol{Z}^1, \boldsymbol{S}^1)$ is finite, then from (56) and (58), we have

$$\begin{aligned}\widetilde{\mathcal{L}}_{\vartheta^{k+1}}(\boldsymbol{Y}^k, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1}) &\le \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^{k+1}),\\ \widetilde{\mathcal{L}}_{\vartheta^{k+1}}(\boldsymbol{Y}^{k+1}, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1}) &\le \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^k, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1}).\end{aligned} \tag{63}$$

From Theorem 1, it is known that $\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k) \le \mathcal{L}_{\vartheta^1}(\boldsymbol{Y}^1, \boldsymbol{Z}^1, \boldsymbol{S}^1)$ and hence $\widetilde{\mathcal{L}}_{\vartheta^k}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k)$ is upper bounded. Moreover, it is clear that $\widetilde{\mathcal{L}}_{\vartheta^k}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k)$ is lower bounded.

Note that $\boldsymbol{Z}^k$ and $\boldsymbol{Y}^k$ are updated by (28b) and (28c). Given $\|\boldsymbol{Z}^1\|_F < \infty$ and $\|\boldsymbol{Y}^1\|_F < \infty$, if there exists $\|\boldsymbol{Z}^k\|_F \to \infty$ or $\|\boldsymbol{Y}^k\|_F \to \infty$, we must have $\widetilde{\mathcal{L}}_{\vartheta^k}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k) \to \infty$, which leads to a contradiction. Thereby, $\{\boldsymbol{Z}^k\}$ and $\{\boldsymbol{Y}^k\}$ are bounded.

Furthermore, bounded $\{\boldsymbol{Z}^k\}$ and $\{\boldsymbol{Y}^k\}$ yield bounded $\{\boldsymbol{G}^k\}$. It is easy to know that if $\|\boldsymbol{S}^k\|_F \to \infty$, $\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k) \to \infty$, contradicting Theorem 1. Thereby, $\{\boldsymbol{S}^k\}$ is also bounded.

As a result, $(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k)$ is bounded.

B. Proof of Property (ii)

For (59), by induction on k, we get

$$\begin{aligned}&\sum_{k=1}^{K}\mu\Big(\|\boldsymbol{Z}^{k+1} - \boldsymbol{Z}^k\|_F^2 + \|\boldsymbol{Y}^{k+1} - \boldsymbol{Y}^k\|_F^2\Big)\\ &\le \mathcal{L}_{\vartheta^1}(\boldsymbol{Y}^1, \boldsymbol{Z}^1, \boldsymbol{S}^1) - \mathcal{L}_{\vartheta^{k+1}}(\boldsymbol{Y}^{k+1}, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1}).\end{aligned} \tag{64}$$

In addition, since $\mathcal{L}_{\vartheta^k}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k)$ is convergent, we have

$$\lim_{K\to\infty}\sum_{k=1}^{K}\mu\Big(\|\boldsymbol{Z}^{k+1} - \boldsymbol{Z}^k\|_F^2 + \|\boldsymbol{Y}^{k+1} - \boldsymbol{Y}^k\|_F^2\Big) < \infty, \tag{65}$$

which leads to

$$\lim_{K\to\infty}\sum_{k=1}^{K}\|\boldsymbol{Z}^{k+1} - \boldsymbol{Z}_1^k\|_F^2 < \infty, \tag{66a}$$

$$\lim_{K\to\infty}\sum_{k=1}^{K}\|\boldsymbol{Y}^{k+1} - \boldsymbol{Y}^k\|_F^2 < \infty. \tag{66b}$$

Thereby, we get

$$\lim_{k\to\infty}\|\boldsymbol{Z}_1^{k+1} - \boldsymbol{Z}_1^k\|_F = 0, \tag{67a}$$

$$\lim_{k\to\infty}\|\boldsymbol{Y}^{k+1} - \boldsymbol{Y}^k\|_F = 0. \tag{67b}$$

Furthermore, since $\boldsymbol{G}^k = \boldsymbol{X} - \boldsymbol{Y}^k\boldsymbol{Z}^k$, in accordance to (67), we obtain

$$\lim_{k\to\infty}\|\boldsymbol{G}^{k+1} - \boldsymbol{G}^k\|_F = 0. \tag{68}$$

Based on (23) and (24), we have

$$\lim_{k\to\infty}|\vartheta^{k+1} - \vartheta^k| = 0. \tag{69}$$

In accordance to Lemma (2), we get

$$\lim_{k\to\infty}\|\boldsymbol{S}^{k+1} - \boldsymbol{S}^k\|_F = 0. \tag{70}$$

Combined with Property (i), we obtain the conclusion that there exists a subsequence of $(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k)$ to be convergent:

$$\lim_{k_i\to\infty}(\boldsymbol{Y}^{k_i}, \boldsymbol{Z}^{k_i}, \boldsymbol{S}^{k_i}) = (\boldsymbol{Y}^*, \boldsymbol{Z}^*, \boldsymbol{S}^*), \tag{71}$$

where $(\boldsymbol{Y}^*, \boldsymbol{Z}^*, \boldsymbol{S}^*)$ is an accumulation point. The proof is complete. ■

C. Property (iii)

Prior to the proof, we introduce the critical point definition in Lemma 4.

Lemma 4. *[59]: Given a function $\delta(x)$, then x^* is a critical point if x^* meets one of the following statements:*

(i) If $\delta(x)$ is differentiable, $\nabla\delta(x^*) = 0$.
(ii) If $\delta(x)$ is not differentiable, $0 \in \partial\delta(x^*)$ where $\partial\varphi(x^*)$ is the subgradient.

We denote $\mathcal{L}_\vartheta(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{S})$ as $\mathcal{L}(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{S})$ for concise expression. From (28a) to (28c), we have

$$\frac{\partial\mathcal{L}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^{k+1})}{\partial\boldsymbol{S}} \ni \boldsymbol{0}, \tag{72a}$$

$$\frac{\partial\mathcal{L}(\boldsymbol{Y}^k, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1})}{\partial\boldsymbol{Z}} + 2\mu(\boldsymbol{Z}^{k+1} - \boldsymbol{Z}^k) = \boldsymbol{0}, \tag{72b}$$

$$\frac{\partial\mathcal{L}(\boldsymbol{Y}^{k+1}, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1})}{\partial\boldsymbol{Y}} + 2\mu(\boldsymbol{Y}^{k+1} - \boldsymbol{Y}^k) = \boldsymbol{0}, \tag{72c}$$

We then rewrite (72b)–(72c) as

$$\frac{\partial\mathcal{L}(\boldsymbol{Y}^{k+1}, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1})}{\partial\boldsymbol{Z}} - \frac{\partial\mathcal{L}(\boldsymbol{Y}^k, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1})}{\partial\boldsymbol{Z}} - 2\mu(\boldsymbol{Z}^{k+1} - \boldsymbol{Z}^k) = \frac{\partial\mathcal{L}(\boldsymbol{Y}^{k+1}, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1})}{\partial\boldsymbol{Z}} \tag{73a}$$

$$-2\mu(\boldsymbol{Y}^{k+1} - \boldsymbol{Y}^k) = \frac{\partial\mathcal{L}(\boldsymbol{Y}^{k+1}, \boldsymbol{Z}^{k+1}, \boldsymbol{S}^{k+1})}{\partial\boldsymbol{Y}} \tag{73b}$$

It is clear that the derivatives of $\mathcal{L}(\boldsymbol{Y}^k, \boldsymbol{Z}^k, \boldsymbol{S}^k)$ w.r.t. $\boldsymbol{Z}$ and $\boldsymbol{Y}$ are continuous. Since $\lim_{k_i\to\infty}(\boldsymbol{Y}^{k_i}, \boldsymbol{Z}^{k_i}, \boldsymbol{S}^{k_i}) = (\boldsymbol{Y}^*, \boldsymbol{Z}^*, \boldsymbol{S}^*)$, we have

$$\begin{aligned}\boldsymbol{0} &\in \lim_{k_i\to\infty} \frac{\partial \mathcal{L}(\boldsymbol{Y}^{k_i}, \boldsymbol{Z}^{k_i}, \boldsymbol{S}^{k_i+1})}{\partial \boldsymbol{S}} \\ &= \lim_{k_i\to\infty} \frac{\partial \mathcal{L}(\boldsymbol{Y}^{k_i+1}, \boldsymbol{Z}^{k_i+1}, \boldsymbol{S}^{k_i+1})}{\partial \boldsymbol{S}}, \qquad (74a)\\ \lim_{k_i\to\infty} \Big(&\frac{\partial \mathcal{L}(\boldsymbol{Y}^{k_i+1}, \boldsymbol{Z}^{k_i+1}, \boldsymbol{S}^{k_i+1})}{\partial \boldsymbol{Z}} - \frac{\partial \mathcal{L}(\boldsymbol{Y}^{k_i}, \boldsymbol{Z}^{k_i+1}, \boldsymbol{S}^{k_i+1})}{\partial \boldsymbol{Z}} \\ &-2\mu(\boldsymbol{Z}^{k_i+1} - \boldsymbol{Z}^{k_i}) \Big) = \boldsymbol{0}, \qquad (74b)\\ \lim_{k_i\to\infty} &-2\mu(\boldsymbol{Y}^{k_i+1} - \boldsymbol{Y}^{k_i}) = \boldsymbol{0}, \qquad (74c)\end{aligned}$$

which indicates

$$\begin{aligned}\frac{\partial \mathcal{L}(\boldsymbol{Y}^*, \boldsymbol{Z}^*, \boldsymbol{S}^*)}{\partial \boldsymbol{S}} &\ni \boldsymbol{0}, \qquad (75a)\\ \frac{\partial \mathcal{L}(\boldsymbol{Y}^*, \boldsymbol{Z}^*, \boldsymbol{S}^*)}{\partial \boldsymbol{Z}} &= \boldsymbol{0}, \qquad (75b)\\ \frac{\partial \mathcal{L}(\boldsymbol{Y}^*, \boldsymbol{Z}^*, \boldsymbol{S}^*)}{\partial \boldsymbol{Y}} &= \boldsymbol{0}. \qquad (75c)\end{aligned}$$

Therefore, $(\boldsymbol{Y}^*, \boldsymbol{Z}^*, \boldsymbol{S}^*)$ is a critical point.

In conclusion, there exists a subsequence $(\boldsymbol{Y}^{k_i}, \boldsymbol{Z}^{k_i}, \boldsymbol{S}^{k_i})$ converging a critical point. The proof is complete. ■

REFERENCES

[1] A. Khabbazibasmenj, A. Hassanien, S. A. Vorobyov, and M. W. Morency, "Efficient transmit beamspace design for search-free based DOA estimation in MIMO radar," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1490–1500, Mar. 2014.

[2] R. Rajagopal and P. Rao, "Generalised algorithm for DOA estimation in a passive sonar," *IEE Proc. F, Radar Signal Process.*, vol. 140, no. 1, pp. 12–20, Feb. 1993.

[3] J. Wang, H. Xu, G. J. Leus, and G. A. Vandenbosch, "Experimental assessment of the coarray concept for DOA estimation in wireless communications," *IEEE Trans. Antennas Propag.*, vol. 66, no. 6, pp. 3064–3075, Jun. 2018.

[4] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[5] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[6] W. Li and W. Liao, "Stable super-resolution limit and smallest singular value of restricted Fourier matrices," *Appl. Comput. Harmon. Anal.*, vol. 51, pp. 118–156, Mar. 2021.

[7] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[8] W. Li, W. Liao, and A. Fannjiang, "Super-resolution limit of the ESPRIT algorithm," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4593–4608, Jul. 2020.

[9] C. Qian, L. Huang, M. Cao, J. Xie, and H. C. So, "PUMA: An improved realization of MODE for DOA estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 5, pp. 2128–2139, Oct. 2017.

[10] A. B. Gershman and M. Haardt, "Improving the performance of unitary ESPRIT via pseudo-noise resampling," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2305–2308, Aug. 1999.

[11] C. Qian, L. Huang, and H. C. So, "Improved unitary root-MUSIC for DOA estimation based on pseudo-noise resampling," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 140–144, Feb. 2013.

[12] C. Qian, L. Huang, N. D. Sidiropoulos, and H. C. So, "Enhanced PUMA for direction-of-arrival estimation and its performance analysis," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4127–4137, Aug. 2016.

[13] B. Liao, C. Guo, and H. C. So, "Direction-of-arrival estimation in nonuniform noise via low-rank matrix decomposition," in *Proc. 22nd Int. Conf. Digit. Signal Process.*, London, U.K., Aug. 2017, pp. 1–4.

[14] H. Huang and A. M. Zoubir, "Low-rank and sparse decomposition for joint DOA estimation and contaminated sensors detection with sparsely contaminated arrays," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Toronto, Ontario, Canada, Jun. 2021, pp. 4615–4619.

[15] Q. Liu, X. Li, and H. Cao, "Two-dimensional localization: Low-rank matrix completion with random sampling in massive MIMO system," *IEEE Syst. J.*, vol. 15, no. 3, pp. 3628–3631, Sep. 2020.

[16] P. Chen, Z. Chen, Z. Cao, and X. Wang, "A new atomic norm for DOA estimation with gain-phase errors," *IEEE Trans. Signal Process.*, vol. 68, pp. 4293–4306, Jul. 2020.

[17] X. Wu, W.-P. Zhu, and J. Yan, "A high-resolution DOA estimation method with a family of nonconvex penalties," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4925–4938, Jun. 2018.

[18] M. Barzegar, G. Caire, A. Flinth, S. Haghighatshoar, G. Kutyniok, and G. Wunder, "Estimation of angles of arrival through superresolution–A soft recovery approach for general antenna geometries," *arXiv preprint arXiv:1711.03996*, 2017.

[19] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 36, no. 4, pp. 532–544, Apr. 1988.

[20] H. Abeida and J.-P. Delmas, "Efficiency of subspace-based DOA estimators," *Signal Process.*, vol. 87, no. 9, pp. 2075–2084, Sep. 2007.

[21] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2206–2217, Jun. 2008.

[22] N. Rožić, P. Banelli, D. Begušić, and J. Radić, "Multiple-threshold estimators for impulsive noise suppression in multicarrier communications," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1619–1633, Mar. 2018.

[23] D. D. Lee and R. L. Kashyap, "Robust maximum likelihood bearing estimation in contaminated Gaussian noise," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 1983–1986, Aug. 1992.

[24] R. J. Kozick and B. M. Sadler, "Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3520–3535, Dec. 2000.

[25] P. Tsakalides and C. L. Nikias, "Maximum likelihood localization of sources in noise modeled as a stable process," *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2700–2713, Nov. 1995.

[26] G. Tzagkarakis, J. P. Nolan, and P. Tsakalides, "Robust nonlinear compressive sampling using symmetric alpha-stable distributions," *Signal Process.*, vol. 182, p. 107944, May 2021.

[27] A. Swami, "On some parameter estimation problems in alpha-stable processes," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5, Munich, Germany, Apr. 1997, pp. 3541–3544.

[28] A. Swami and B. M. Sadler, "On some detection and estimation problems in heavy-tailed noise," *Signal Process.*, vol. 82, no. 12, pp. 1829–1846, Dec. 2002.

[29] S. Luan, M. Zhao, Y. Gao, Z. Zhang, and T. Qiu, "Generalized covariance for non-Gaussian signal processing and GC-MUSIC under alpha-stable distributed noise," *Digital Signal Process.*, vol. 110, p. 102923, Mar. 2021.

[30] M. Asghari, M. Zareinejad, S. M. Rezaei, and H. Amindavar, "ECF-MUSIC: An empirical characteristic function based direction of arrival (DOA) estimation in the presence of impulsive noise," *Digital Signal Process.*, vol. 123, p. 103440, Apr. 2022.

[31] P. Tsakalides and C. L. Nikias, "The robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 44, no. 7, pp. 1623–1633, Jul. 1996.

[32] T.-H. Liu and J. M. Mendel, "A subspace-based direction finding algorithm using fractional lower order statistics," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1605–1613, Aug. 2001.

[33] S. Visuri, H. Oja, and V. Koivunen, "Subspace-based direction-of-arrival estimation using nonparametric statistics," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2060–2073, Sep. 2001.

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

[35] S. Li, R. He, B. Lin, and F. Sun, "DOA estimation based on sparse representation of the fractional lower order statistics in impulsive noise," *IEEE/CAA J. Autom. Sin.*, vol. 5, no. 4, pp. 860–868, Jul. 2018.

[36] W.-J. Zeng, H. C. So, and L. Huang, "ℓ_p-MUSIC: Robust direction-of-arrival estimator for impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4296–4308, Sep. 2013.

[37] Q. Liu, Y. Gu, and H. C. So, "DOA estimation in impulsive noise via low-rank matrix approximation and weakly convex optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3603–3616, Dec. 2019.

[38] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, May 2011.

[39] A. Blake and A. Zisserman, *Visual Reconstruction*. MIT press, 1987.

[40] X. P. Li, Z.-Y. Wang, Z.-L. Shi, H. C. So, and N. D. Sidiropoulos, "Robust tensor completion via capped Frobenius norm," *IEEE Trans. Neural Netw. Learn. Syst.*, Jan. 2023, Early Access.

[41] H. Yang, J. Shi, and L. Carlone, "TEASER: Fast and certifiable point cloud registration," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 314–333, Apr. 2020.

[42] P. Antonante, V. Tzoumas, H. Yang, and L. Carlone, "Outlier-robust estimation: Hardness, minimally tuned algorithms, and applications," *IEEE Trans. Robot.*, vol. 38, no. 1, pp. 281–301, Feb. 2021.

[43] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.

[44] H. T. Shen, Y. Zhu, W. Zheng, and X. Zhu, "Half-quadratic minimization for unsupervised feature selection on incomplete data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3122–3135, Jul. 2020.

[45] F. Wen, R. Ying, P. Liu, and T.-K. Truong, "Nonconvex regularized robust PCA using the proximal block coordinate descent algorithm," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5402–5416, Oct. 2019.

[46] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech, Signal Process*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[47] M. Wax, "Detection and localization of multiple sources via the stochastic signals model," *IEEE Trans. Signal Process.*, vol. 39, no. 11, pp. 2450–2456, Nov. 1991.

[48] L. Huang, C. Qian, H. C. So, and J. Fang, "Source enumeration for large array using shrinkage-based detectors with small samples," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 344–357, Jan. 2015.

[49] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, Oct. 2009.

[50] W. Liao and A. Fannjiang, "MUSIC for single-snapshot spectral estimation: Stability and super-resolution," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 1, pp. 33–67, Jan. 2016.

[51] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge Univ. Press, 2018.

[52] X. P. Li and H. C. So, "Robust low-rank tensor completion based on tensor ring rank via $\ell_{p,\epsilon}$-norm," *IEEE Trans. Signal Process.*, vol. 69, pp. 3685–3698, May 2021.

[53] X. P. Li, Z.-L. Shi, Q. Liu, and H. C. So, "Fast robust matrix completion via entry-wise ℓ_0-norm minimization," *IEEE Trans. Cybern.*, 2022, Early Access.

[54] B. Xiong and Z. Yin, "A universal denoising framework with a new impulse detector and nonlocal means," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1663–1675, Apr. 2011.

[55] V. Jakhetiya, K. Gu, T. Singhal, S. C. Guntuku, Z. Xia, and W. Lin, "A highly efficient blind image quality assessment metric of 3-D synthesized images using outlier detection," *IEEE Trans. Ind. Inf.*, vol. 15, no. 7, pp. 4120–4128, Jul. 2018.

[56] D. Vysochanskij and Y. I. Petunin, "Justification of the 3σ rule for unimodal distributions," *Theory Probab. Math. Statist.*, vol. 21, no. 25-36, 1980.

[57] B. D. Rao and K. S. Hari, "Performance analysis of root-MUSIC," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.

[58] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér–Rao bound," *IEEE Trans. Acoust. Speech. Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[59] Q. Yao, J. T. Kwok, T. Wang, and T.-Y. Liu, "Large-scale low-rank matrix learning with nonconvex regularizers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2628–2643, Nov. 2018.

Xiao Peng Li received the B.Eng. degree as an outstanding graduate in Electronic Science and Technology from Yanshan University, Qinhuangdao, China, in 2015, and the M.Sc. degree with Distinction in Electronic Information Engineering and the Ph.D. degree in Electrical Engineering from the City University of Hong Kong, Hong Kong SAR, China, in 2018 and 2022, respectively. He was a Research Assistant with the Department of Information Engineering, Shenzhen University, Shenzhen, China from 2018 to 2019, and a Postdoctoral Fellow with the Department of Electrical Engineering, City University of Hong Kong from 2022 to 2023. He is currently an Assistant Professor with the College of Electronics and Information Engineering, Shenzhen University. His research interests include robust signal processing, optimization methods, machine learning, sparse recovery, matrix processing and tensor processing with applications in target estimation, image recovery, video restoration, hyperspectral unmixing, and stock market analysis.

Zhaofeng Liu received the B.S. and M.S. degrees both in School of Electronics and Information Technology from Sun Yat-sen University, Guangzhou, China, in 2018 and 2020, respectively. He is currently pursuing the Ph.D. degree in Department of Electrical Engineering, City University of Hong Kong. His research interests include chaotic communication, multi-carrier modulation, robust signal processing, and MIMO radar.

Zhang-Lei Shi received the Ph.D. degree from the Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China, in 2021. He is currently a Lecturer with the College of Science, China University of Petroleum (East China), Qingdao, China. His current research interests include neural networks and machine learning.

Hing Cheung So (S'90–M'95–SM'07-F'15) was born in Hong Kong. He received the B.Eng. degree from the City University of Hong Kong and the Ph.D. degree from The Chinese University of Hong Kong, both in electronic engineering, in 1990 and 1995, respectively. From 1990 to 1991, he was an Electronic Engineer with the Research and Development Division, Everex Systems Engineering Ltd., Hong Kong. During 1995–1996, he was a Postdoctoral Fellow with The Chinese University of Hong Kong. From 1996 to 1999, he was a Research Assistant Professor with the Department of Electronic Engineering, City University of Hong Kong, where he is currently a Professor. His research interests include detection and estimation, fast and adaptive algorithms, multidimensional harmonic retrieval, robust signal processing, source localization, and sparse approximation.

He has been on the editorial boards of IEEE Signal Processing Magazine (2014–2017), IEEE Transactions on Signal Processing (2010–2014), Signal Processing (2010–), and Digital Signal Processing (2011–). He was also Lead Guest Editor for IEEE Journal of Selected Topics in Signal Processing, special issue on "Advances in Time/Frequency Modulated Array Signal Processing" in 2017. In addition, he was an elected member in Signal Processing Theory and Methods Technical Committee (2011–2016) of the IEEE Signal Processing Society where he was chair in the awards subcommittee (2015–2016).