ℓ_p-MUSIC: Robust Direction-of-Arrival Estimator for Impulsive Noise Environments

Wen-Jun Zeng, *Member, IEEE*, H. C. So, *Senior Member, IEEE*, and Lei Huang, *Member, IEEE*

***Abstract*—A family of algorithms, named ℓ_p-MUSIC, for direction-of-arrival (DOA) estimation in impulsive noise is proposed. The ℓ_p-MUSIC estimator adopts the ℓ_p-norm $(1 \leq p < 2)$ of the residual fitting error matrix as the objective function for subspace decomposition, rather than the Frobenius norm that is used in the conventional MUSIC method. Although the matrix ℓ_p-norm minimization based subspace decomposition will lead to a nonconvex optimization problem, two iterative algorithms are designed for achieving efficient solutions. The first algorithm is the iteratively reweighted singular value decomposition (IR-SVD), where the SVD of a reweighted data matrix is performed in each iteration. The second algorithm solves the nonconvex matrix ℓ_p-norm minimization by alternating convex optimization. Two complex-valued Newton's methods with optimal step size in each iteration are devised to solve the resulting convex problem. The convergence of the iterative procedure is also proved. Numerical results verify that the ℓ_p-MUSIC methodology outperforms the standard MUSIC scheme and several existing outlier-resistant DOA estimation approaches in terms of resolution capability and estimation accuracy.**

***Index Terms*—ℓ_p-norm minimization, complex-valued Newton's method, direction-of-arrival estimation, impulsive noise, nonconvex optimization, robust estimation, subspace method.**

I. Introduction

DIRECTION-OF-ARRIVAL (DOA) estimation of multiple emitting sources is an important issue in array processing and has various applications in radar, sonar, wireless communications and source localization [1]–[3]. MUSIC [4] is one of the most well-known high resolution DOA estimation techniques and it belongs to the subspace methodology [5]. It estimates the DOAs by exploiting the orthogonality between the noise subspace and array manifold. It has been shown that the MUSIC method is an asymptotically unbiased and efficient DOA estimator based on the Gaussian noise assumption [6]–[8].

Manuscript received August 26, 2012; revised January 01, 2013 and April 22, 2013; accepted May 07, 2013. Date of publication May 16, 2013; date of current version August 07, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Marius Pesavento. The work described in this paper was supported by a grant from the NSFC/RGC Joint Research Scheme sponsored by the Research Grants Council of Hong Kong and the National Natural Science Foundation of China (Project No.: N_CityU 104/11, 61110229/61161160564).

W.-J. Zeng and H. C. So are with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: wenjzeng@cityu.edu.hk; hcso@ee.cityu.edu.hk).

L. Huang is with Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: lhuang@hitsz.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2013.2263502

Many existing DOA estimators explicitly or implicitly assume that the ambient noise is Gaussian distributed. However, the noise in practice often exhibits non-Gaussian properties. The performance of the conventional DOA estimators may severely degrade in the presence of non-Gaussian noise. One important class of non-Gaussian noises that are frequently encountered in many practical wireless radio systems is impulsive noise, also known as burst noise [9], [10]. The probability density function (PDF) of impulsive noise has heavier tails than the Gaussian distribution. The property of impulsive noise is somewhat similar to outliers in statistics. It is because the heavy tailed distributions give higher probability of occurrence to values which exceed a few standard deviations than the Gaussian distribution. Under a nominal Gaussian noise model, these large values are unlikely to appear and can therefore be considered as outliers. There are three typical PDFs for modeling impulsive noise [11]. They are Gaussian mixture model (GMM) [12], [13] generalized Gaussian distribution (GGD) [14], and α-stable distribution [15]. A number of DOA estimation methods for various impulsive noise models have been proposed. The maximum likelihood (ML) estimation schemes are developed for GMM noise [12], [13], [16] and Cauchy distribution noise [17]. To obtain the ML estimate, we need to solve a complicated nonlinear and nonconvex multi-dimensional optimization problem. This usually leads to a heavy computational load for the ML estimator. In [12], the expectation-maximization (EM) algorithm is designed for jointly estimating the DOAs, signal waveforms and noise parameters in GMM noise. Nevertheless, the ML estimator generally requires the analytical form of the noise PDF. As a result, it is difficult to apply this optimum approach to the case where the PDF of the impulsive noise has no closed-form expression such as α-stable process with $\alpha \neq 1$ and $\alpha \neq 2$.

The conventional subspace based DOA estimation techniques exploit eigenvalue decomposition (EVD) of the covariance matrix of the received data. The DOA estimators based on the second-order sample covariance are not robust against outliers. A class of subspace based DOA estimation algorithms uses the fractional lower-order statistics such as the robust covariation in ROC-MUSIC [18], fractional lower-order moments in FLOM-MUSIC [19], sign covariance matrix in SCM-MUSIC, and Kendall's tau covariance matrix in TCM-MUSIC [20], instead of the second-order sample covariance. However, the fractional lower-order statistics based algorithms are suboptimal and require large sample sizes for a satisfactory performance [12], [18]. Swami *et al.* have proposed to apply zero-memory nonlinear (ZMNL) functions to limit the influence of outliers by clipping the amplitude of the received signal [21]–[23]. The ZMNL preprocessing achieves robust covariance estimation and provides more accurate DOA estimates than the fractional lower-order schemes [12], [13], [21], [23]. Furthermore, the data-adaptive ZMNL approach is simple and has a low computational complexity. Despite

these advantages, there is a tradeoff in the ZMNL scheme between outlier suppression and subspace preservation. The ZMNL preprocessing generally destroys the low-rank property of the signal subspace. Its performance may degrade due to the rank increase of the signal subspace [12], [13]. Similar idea using outlier-trimming has been developed in [24], where the Shapiro-Wilk *W* test for Gaussianity is used. One of its limitations is that Gaussian distributed source signal is required [24].

Another representative DOA estimation scheme resistant to impulsive noise is based on robust statistics [25]–[27]. This approach first uses a robust scheme such as the M-estimator [26], S-estimator [25], or MM-estimator [28], to estimate the covariance matrix, and then conventional subspace decomposition is exploited to obtain the DOA estimates. The success of this method depends on choosing the appropriate robust statistics. Different from the covariance based methodology which exploits the sample covariance, fractional-order moment, or any robust statistics computed from the received data, we directly compute the signal subspace without explicitly constructing the covariance. Naturally, the EVD is not required.

The key step of the MUSIC method is computing the signal or noise subspace. The subspace decomposition rule in MUSIC is equivalent to minimization of the Frobenius norm of the residual fitting error matrix. The resulting Frobenius norm minimization can be efficiently solved by singular value decomposition (SVD) of the received data matrix. The orthonormal bases of the signal and noise subspaces are given by the singular vectors associated with the principal and minor singular values, respectively. The subspace decomposition using Frobenius norm minimization is statistically optimal when the additive noise is Gaussian distributed. It, however, is no longer optimal and the performance of the conventional MUSIC schemes based on SVD will degrade in the presence of impulsive noise.

In this paper, we propose a new subspace decomposition rule via minimizing the ℓ_p-norm ($1 \leq p < 2$) of the residual fitting error matrix. The motivation to use the ℓ_p-norm with $p < 2$ is that it is less sensitive to large outliers than the square function, which corresponds to the Frobenius norm. It is worth mentioning that Mateos and Giannakis have recently suggested a matrix formulation in the ℓ_p-norm minimization for robust estimation [29]. The ℓ_p-norm minimization approach has been widely used to resist impulsive noise in wireless radio channels [30]–[32]. Here we address robust subspace decomposition and DOA estimation based on ℓ_p-norm minimization. Despite this subspace decomposition rule will result in a nonconvex optimization problem, we devise two efficient iterative algorithms to solve it. The first algorithm is called iteratively reweighted singular value decomposition (IR-SVD), where the SVD of a reweighted data matrix is performed in each iteration. The second algorithm solves the nonconvex *matrix* ℓ_p-norm minimization by alternating convex optimization (ACO). In each iteration of the ACO, the *matrix* ℓ_p-norm minimization is converted into a series of *vector* ℓ_p-norm minimization subproblems. The most widely used approach to vector ℓ_p-norm minimization is the iteratively reweighted least squares (IRLS) [25], [35]. However, the IRLS may suffer divergence when p is close to 1 and has a slow convergence rate, that is, it at most has a linear convergence rate. We propose two complex-valued Newton's algorithms, called the pseudo-Newton's method and full-Newton's method, for the resulting vector ℓ_p-norm minimization problem. The pseudo-Newton's method only exploits the partial Hessian matrix of the complex-valued variables while the full-Newton's method uses the full Hessian matrix. Optimal step size in each iteration is also derived for the two schemes. It is revealed that the IRLS can be interpreted as a special case of the pseudo-Newton's method using a fixed step size of $p/2$. The two Newton's methods are more numerically stable and converge faster than the IRLS. More specifically, the full-Newton's method converges quadratically. The ℓ_p-MUSIC method is more robust against impulsive noise than the conventional counterpart.

It is worth noting that the robust statistics based approach has also been used to compute the signal subspace. In [33], an algorithm for robust subspace decomposition based on Yohai's regression MM-estimates [34] is proposed. Our approach is different from [33] because it utilizes ℓ_p-norm minimization and is applicable to complex-valued data. Detection of the source number is crucial for subspace methods. Conventional model order selection techniques such as Akaike information criterion (AIC) and minimum description length (MDL) [36], [37] are only effective in Gaussian noise. To correctly estimate the source number in the presence of impulsive noise, several robust model order selection schemes have been developed in the literature [27], [33], [38]. In this paper, our focus is on parameter estimation and we assume that the number of sources has been accurately identified by a robust detection method.

We briefly summarize the contributions of this work as follows.

i) A new subspace decomposition rule based on minimization of the ℓ_p-norm of the residual error matrix is proposed.
ii) Two efficient iterative algorithms, the IR-SVD and ACO, are developed to solve the resulting nonconvex matrix decomposition under the ℓ_p-norm minimization framework.
iii) The convergence of the ACO algorithm is proved.
iv) The ACO converts the original matrix ℓ_p-norm minimization problem into a series of decoupled subproblems of convex vector ℓ_p-norm minimization. To solve these subproblems, the complex-valued pseudo-Newton's and full-Newton's algorithms which outperform the widely used IRLS scheme are devised. We point out that the latter is a special case of the pseudo-Newton's method with a fixed step size.

The remainder of this paper is organized as follows. In Section II, the array processing model for DOA estimation is given and the conventional MUSIC method is briefly reviewed. In Section III, we present the ℓ_p-MUSIC DOA estimator, where two iterative algorithms for robust subspace estimation, namely, IR-SVD and ACO, are also developed. Computer simulations are performed to demonstrate the effectiveness of the proposed algorithm in impulsive noise in Section IV. Finally, conclusions are drawn in Section V.

II. Problem Formulation

A. Signal Model

Consider Q far-field, narrowband sources emitting plane waves impinging on a uniform linear array (ULA) of M sensors with inter-sensor spacing d. In the subspace DOA estimation methods, it is required that the number of sources is less than the number of sensors, that is, $Q < M$. The discrete-time

complex baseband signal received by the mth $(m = 1, \ldots, M)$ sensor is expressed as

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

where $s_q(n)$ is the qth source signal with n denoting the discrete-time index, $\mathrm{j} = \sqrt{-1}$ is the imaginary unit and λ is the wavelength of the signal. Note that $d \leq \lambda/2$ is required to avoid the phase ambiguity. The θ_q is the DOA of the qth source and $e_m(n)$ denotes the non-Gaussian additive noise of the mth sensor.

By arranging the output of the M sensors in a vector $\boldsymbol{x}_n = [x_1(n), \ldots, x_M(n)]^T$ with the superscript $(\cdot)^T$ denoting the transpose, the matrix formulation of (1) can be written as

$$\boldsymbol{x}_n = \boldsymbol{A}\boldsymbol{s}_n + \boldsymbol{e}_n \quad (2)$$

where $\boldsymbol{s}_n = [s_1(n), \ldots, s_Q(n)]^T$ is the source vector, $\boldsymbol{e}_n = [e_1(n), \ldots, e_M(n)]^T$ is the noise vector, and $\boldsymbol{A}$ is the array manifold matrix which is of the form:

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

with $\boldsymbol{a}(\theta)$ being the steering vector:

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

Given $\boldsymbol{X} \triangleq [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N] \in \mathbb{C}^{M\times N}$ with N being the sample size, our goal is to estimate the DOAs of the Q sources. The number of sources Q is assumed known or has been determined by an outlier-resistant source enumeration method [27], [33], [38]. It is assumed that the sources are zero-mean and mutually independently with each other, while the non-Gaussian noises $\{e_m(t)\}_{m=1}^M$ are temporally and spatially white, and statistically independent of the sources. These assumptions are quite mild for practical applications.

B. Conventional MUSIC Method

The MUSIC method proposed by Schmidt [4] for DOA estimation of multiple sources is based on subspace decomposition. It exploits the second-order statistics and assumes independent and identically distributed Gaussian noises with variance σ^2. The covariance matrix of the received signal $\boldsymbol{C}_x = E\{\boldsymbol{x}_n\boldsymbol{x}_n^H\}$ with $E\{\cdot\}$ denoting the expectation is estimated from the N snapshots as

$$\widehat{\boldsymbol{C}}_x = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n\boldsymbol{x}_n^H = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^H \quad (5)$$

where the superscript $(\cdot)^H$ denotes the Hermitian transpose. Since there is no correlation between the signals and noises, the covariance matrix of the received signal is

$$\boldsymbol{C}_x = \boldsymbol{A}\boldsymbol{C}_s\boldsymbol{A}^H + \sigma^2\boldsymbol{I} \quad (6)$$

where $\boldsymbol{C}_s = E\{\boldsymbol{s}_n\boldsymbol{s}_n^H\}$ is the source covariance matrix and $\boldsymbol{I}$ denotes the identity matrix. The EVD of $\boldsymbol{C}_x$ is given by

$$\boldsymbol{C}_x = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H = \boldsymbol{U}_s\boldsymbol{\Lambda}_s\boldsymbol{U}_s^H + \sigma^2\boldsymbol{U}_n\boldsymbol{U}_n^H \quad (7)$$

where $\boldsymbol{U} = [\boldsymbol{U}_s, \boldsymbol{U}_n]$, $\boldsymbol{\Lambda}_s = \mathrm{diag}\{\lambda_1, \ldots, \lambda_Q\}$ is a diagonal matrix containing the Q non-increasing principal eigenvalues and $\boldsymbol{U}_s$ contains the corresponding orthonormal eigenvectors. While $\boldsymbol{U}_n$ contains the $(M-Q)$ orthonormal eigenvectors associated with the eigenvalue σ^2. The range space spanned by $\boldsymbol{U}_s$ is known as signal subspace and its orthogonal complement, spanned by $\boldsymbol{U}_n$, is called noise subspace. The MUSIC method searches for the peaks of the following spatial spectrum

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

as the DOA estimates.

We can also calculate the signal and noise subspaces by the SVD of the received data matrix $\boldsymbol{X}$ to avoid computing the covariance matrix $\boldsymbol{C}_x$. The SVD of $\boldsymbol{X}$ is written as

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H = \boldsymbol{U}_s\boldsymbol{\Sigma}_s\boldsymbol{V}_s^H + \boldsymbol{U}_n\boldsymbol{\Sigma}_n\boldsymbol{V}_n^H \quad (9)$$

where the columns of $\boldsymbol{U} \in \mathbb{C}^{M\times M}$ and $\boldsymbol{V} \in \mathbb{C}^{N\times N}$ are the left and right singular vectors, respectively, and $\boldsymbol{\Sigma}$ is a diagonal matrix whose diagonal elements are the singular values. Note that the left singular vectors of $\boldsymbol{X}$ are equal to the eigenvectors of $\widehat{\boldsymbol{C}}_x$. Hence the columns of matrices $\boldsymbol{U}_s \in \mathbb{C}^{M\times Q}$ and $\boldsymbol{U}_n$ in (9) are the orthonormal base vectors of the signal and noise subspaces, respectively.

The performance of the MUSIC method is good in Gaussian noise at moderate signal-to-noise ratio (SNR) conditions. However, its accuracy degrades severely in impulsive noise. In the next section, the ℓ_p-MUSIC algorithm which is robust against impulsive noise is developed.

III. ℓ_p-MUSIC Algorithm

A. Matrix ℓ_p-norm based Subspace Decomposition Criterion

The received data matrix $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{E} \quad (10)$$

where $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N] \in \mathbb{C}^{Q\times N}$ and $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N] \in \mathbb{C}^{M\times N}$ are the source and noise data matrices, respectively. The array manifold matrix $\boldsymbol{A} \in \mathbb{C}^{M\times Q}$ is of full column rank. Similarly, $\boldsymbol{S}$ is of full row rank. In the absence of noise or $\boldsymbol{E} = \boldsymbol{0}$, we have $\mathrm{rank}(\boldsymbol{X}) = Q < M$ because $\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{S}) = Q$. This means that the received data matrix $\boldsymbol{X}$ has the low-rank decomposition

$$\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{Z} \quad (11)$$

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 columns of $\boldsymbol{Y}$ span the same subspace of the array manifold matrix $\boldsymbol{A}$, that is,

$$\mathrm{range}(\boldsymbol{Y}) = \mathrm{range}(\boldsymbol{A}). \quad (12)$$

Since the received data are always noisy in practice, the decomposition of (11) is only an approximation. We first consider minimizing the square error function

$$J_F(\boldsymbol{Y}, \boldsymbol{Z}) = \|\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}\|_F^2 = \sum_{m=1}^{M}\sum_{n=1}^{N}|x_{mn} - (\boldsymbol{Y}\boldsymbol{Z})_{mn}|^2 \quad (13)$$

where $\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}$ is the residual fitting error matrix, x_{mn} and $(\boldsymbol{Y}\boldsymbol{Z})_{mn}$ denote the (m, n) entries of $\boldsymbol{X}$ and $\boldsymbol{Y}\boldsymbol{Z}$, respectively,

and $\|\cdot\|_F$ is the Frobenius norm. Throughout the paper, $|\cdot|$ represents the absolute value of a real number or the modulus of a complex number. In particular, for a complex number $z = z_R + \mathrm{j}z_I$ with the real part z_R and imaginary part z_I, its modulus is defined as $|z| = \sqrt{z_R^2 + z_I^2}$.

It is obvious that minimization of $J_F(\boldsymbol{Y}, \boldsymbol{Z})$ leads to the ML estimates of $\boldsymbol{Y}$ and $\boldsymbol{Z}$ if the noise is white Gaussian distributed. Although the Frobenius norm minimization of (13) with respect to (w.r.t.) $\boldsymbol{Y}$ and $\boldsymbol{Z}$ is nonconvex, its global minimum can be obtained through the SVD of $\boldsymbol{X}$ shown in (9) and the global optima of $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are given by [39]

$$\boldsymbol{Y} = \boldsymbol{U}_s,\ \boldsymbol{Z} = \boldsymbol{\Sigma}_s \boldsymbol{V}_s^H. \tag{14}$$

Note that the global optimum solution of (14) is not unique. For instance, $\boldsymbol{U}_s\boldsymbol{T}$ and $\boldsymbol{T}^{-1}\boldsymbol{\Sigma}_s\boldsymbol{V}_s^H$ are also the global optima of $\boldsymbol{Y}$ and $\boldsymbol{Z}$ for any nonsingular matrix $\boldsymbol{T}$. Despite the nonuniqueness of the optimal $\boldsymbol{Y}$, we are not interested in $\boldsymbol{Y}$ itself but the subspace spanned by its columns. The range space of the optimal $\boldsymbol{Y}$ is unique. Now it is clear that the Frobenius norm minimization criterion will result in the conventional subspace decomposition which corresponds to MUSIC. The conventional MUSIC method takes the columns of $\boldsymbol{U}_s$ as an orthogonal basis of the estimated signal subspace. However, when the PDF of the noise is non-Gaussian, the performance of (13) will degrade.

To design a subspace based DOA estimator that is more robust against non-Gaussian impulsive noise, we propose to use the ℓ_p-norm of the residual fitting error matrix instead of the Frobenius norm in (13). The ℓ_p-norm $(1 \le p \le 2)$ based objective function

$$J_p(\boldsymbol{Y}, \boldsymbol{Z}) = \|\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}\|_p^p = \sum_{m=1}^{M}\sum_{n=1}^{N} |x_{mn} - (\boldsymbol{Y}\boldsymbol{Z})_{mn}|^p \tag{15}$$

where the ℓ_p-norm $\|\cdot\|_p$ of a complex-valued matrix is defined as

$$\|\boldsymbol{X}\|_p = \left(\sum_{m=1}^{M}\sum_{n=1}^{N} |x_{mn}|^p\right)^{1/p}. \tag{16}$$

When $p = 2$, (15) reduces to the Frobenius norm minimization criterion of (13). It is worth pointing out that the matrix ℓ_p-norm defined in (16) is different from the widely used matrix induced norm or operator norm [39]. Minimization of the ℓ_p-norm error function in (15) with $p < 2$ is a more suitable criterion in the presence of impulsive noise.

It is of interest to estimate $\boldsymbol{Y}$ by minimizing $J_p(\boldsymbol{Y}, \boldsymbol{Z})$ in (15) since its columns span the signal subspace. If an estimate of $\boldsymbol{Y}$ is obtained, then its columns are adopted as a basis of the signal subspace. Unfortunately, the matrix ℓ_p-norm minimization

$$\min_{\boldsymbol{Y},\boldsymbol{Z}} J_p(\boldsymbol{Y}, \boldsymbol{Z}) = \|\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}\|_p^p \tag{17}$$

w.r.t. $\boldsymbol{Y}$ and $\boldsymbol{Z}$ is a nonconvex optimization problem. It is well known that the nonconvex optimization is much more intractable than the convex counterpart [40]. In the following, we develop two efficient iterative algorithms to solve this nonconvex ℓ_p-norm minimization problem.

B. Iteratively Reweighted SVD Algorithm

Denoting the residual error matrix as $\boldsymbol{R} = \boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}$ with the (m, n) entry r_{mn}, the matrix ℓ_p-norm of (17) can be expressed as

$$\begin{aligned} J_p(\boldsymbol{Y}, \boldsymbol{Z}) &= \|\boldsymbol{R}\|_p^p = \sum_{m=1}^{M}\sum_{n=1}^{N} |r_{mn}|^p = \sum_{m,n} |r_{mn}|^{p-2}|r_{mn}|^2 \\ &= \|\boldsymbol{D} \odot \boldsymbol{R}\|_F^2 = \|\boldsymbol{D} \odot \boldsymbol{X} - \boldsymbol{D} \odot (\boldsymbol{Y}\boldsymbol{Z})\|_F^2 \end{aligned} \tag{18}$$

where $\odot$ denotes the element-wise multiplication and $\boldsymbol{D}$ is the weighting matrix with its (m, n) entry $d_{mn} = |r_{mn}|^{(p-2)/2}$. Equation (18) means that the matrix ℓ_p-norm minimization problem can be converted into a Frobenius norm minimization one. We can perform SVD to the weighted matrix $\boldsymbol{D} \odot \boldsymbol{X}$ to obtain a better subspace estimation based on the ℓ_p-norm minimization criterion. Note that the weighting matrix $\boldsymbol{D}$ depends on the unknowns $\boldsymbol{Y}$ and $\boldsymbol{Z}$, that is, it is a function of $\boldsymbol{Y}$ and $\boldsymbol{Z}$ which is written as $\boldsymbol{D}(\boldsymbol{Y}, \boldsymbol{Z})$. Due to this reason, we cannot immediately obtain the optimal solution by performing the SVD of the weighted matrix $\boldsymbol{D} \odot \boldsymbol{X}$ only once. An iterative procedure must be adopted, which is shown in Algorithm 1, where the superscript $(\cdot)^{(k)}$ is used to denote the result at the kth iteration. In each iteration, the SVD of a reweighted matrix is performed. Therefore we refer to this algorithm as iteratively reweighted SVD (IR-SVD).

Algorithm 1: IR-SVD

Initialize $\boldsymbol{Y}^{(0)}$ with a random matrix of full column rank and $\boldsymbol{Z}^{(0)}$ of full row rank.

for $k = 0, 1, 2, \ldots$ **do**

Compute the residual matrix $\boldsymbol{R}^{(k)} = \boldsymbol{X} - \boldsymbol{Y}^{(k)}\boldsymbol{Z}^{(k)}$ and weighting matrix $\boldsymbol{D}^{(k)} = |\boldsymbol{R}^{(k)}|^{(p-2)/2}$.

Perform SVD:

$$\boldsymbol{D}^{(k)} \odot \boldsymbol{X} = \boldsymbol{U}_s^{(k)}\boldsymbol{\Sigma}_s^{(k)}\left(\boldsymbol{V}_s^{(k)}\right)^H + \boldsymbol{U}_n^{(k)}\boldsymbol{\Sigma}_n^{(k)}\left(\boldsymbol{V}_n^{(k)}\right)^H$$

where $\boldsymbol{U}_s^{(k)} \in \mathbb{C}^{M\times Q}$ contains the first Q principal left singular vectors of $\boldsymbol{D}^{(k)} \odot \boldsymbol{X}$.

Set $\boldsymbol{Y}^{(k)} \leftarrow \boldsymbol{U}_s^{(k)}$ and $\boldsymbol{Z}^{(k)} \leftarrow \boldsymbol{\Sigma}_s^{(k)}\left(\boldsymbol{V}_s^{(k)}\right)^H$.

end for

Remark 1: It should be pointed out that the IR-SVD algorithm does not always converge when p is much smaller than 2. It rapidly decreases the objective function $J_p(\boldsymbol{Y}, \boldsymbol{Z})$ to a low level at the first few iterations. The IR-SVD procedure will converge toward to this low level value if it converges. This is the general convergence behavior of the IR-SVD algorithm. If it does not converge, it oscillates around this low level value. Although its convergence is not guaranteed, it is a simple and effective approach to achieve a satisfactory subspace estimation. If it oscillates around a lower level value, this means that the objective function has attained a lower value. Then the algorithm is terminated and the subspace estimate can be obtained from the minimum point among all the iterations.

Remark 2: The IR-SVD algorithm is applicable to the case of $0 < p < 1$. After obtaining $\boldsymbol{Y}$ through IR-SVD, the projection matrix onto the noise subspace is computed as

$$\boldsymbol{P}_n = \boldsymbol{I} - \boldsymbol{Y}\left(\boldsymbol{Y}^H\boldsymbol{Y}\right)^{-1}\boldsymbol{Y}^H. \tag{19}$$

Then the ℓ_p-MUSIC spatial spectrum of is given by

$$P_{\ell_p-\mathrm{MUSIC}}(\theta) = \frac{1}{\boldsymbol{a}^H(\theta)\boldsymbol{P}_n\boldsymbol{a}(\theta)}. \tag{20}$$

The DOA estimates can be obtained by searching for the peaks of (20). Certainly, in the ℓ_p-MUSIC, we can also use the root finding technique, namely, root-MUSIC [41] to compute the DOAs instead of the spectrum search.

C. Alternating Convex Optimization Approach

Despite the simplicity of the IR-SVD algorithm, it may not converge. In this section, we propose an ACO approach, whose convergence is guaranteed, to efficiently solve the nonconvex problem of (17). Note that if we fix one of the two matrices, that is, either $\boldsymbol{Y}$ or $\boldsymbol{Z}$, the error function $J_p(\boldsymbol{Y},\boldsymbol{Z})$ is convex w.r.t. the other matrix alone and the global minimum point can be more easily found. Therefore we use an iterative procedure that minimizes the error function alternately over the two matrices $\boldsymbol{Y}$ and $\boldsymbol{Z}$. Suppose we have obtained the two matrices $\boldsymbol{Y}^{(k)}$ and $\boldsymbol{Z}^{(k)}$ at the kth iteration. Then $\boldsymbol{Z}^{(k+1)}$ and $\boldsymbol{Y}^{(k+1)}$ can be solved by alternately optimizing the following two subproblems:

$$\boldsymbol{Z}^{(k+1)} = \arg\min_{\boldsymbol{Z}} \left\|\boldsymbol{X} - \boldsymbol{Y}^{(k)}\boldsymbol{Z}\right\|_p^p \tag{21}$$

$$\boldsymbol{Y}^{(k+1)} = \arg\min_{\boldsymbol{Y}} \left\|\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}^{(k+1)}\right\|_p^p. \tag{22}$$

Both (21) and (22) are convex. This observation motivates us to use an alternating convex programming approach to solve the original nonconvex optimization problem of (15). Since (22) can be rewritten as

$$\min_{\boldsymbol{Y}} \left\|\boldsymbol{X}-\boldsymbol{Y}\boldsymbol{Z}^{(k+1)}\right\|_p^p = \min_{\boldsymbol{Y}} \left\|\boldsymbol{X}^T - \left(\boldsymbol{Z}^{(k+1)}\right)^T\boldsymbol{Y}^T\right\|_p^p \tag{23}$$

it is obvious that (21) and (22) have the same structure and can be solved in a similar manner. Therefore we only need to discuss the solver for (21). The problem of (22) is handled in the same way.

The ℓ_p-norm error function in (21) can be further expressed as

$$\left\|\boldsymbol{X} - \boldsymbol{Y}^{(k)}\boldsymbol{Z}\right\|_p^p = \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{Y}^{(k)}\boldsymbol{z}_n\|_p^p. \tag{24}$$

Recall that $\boldsymbol{x}_n \in \mathbb{C}^M$ and $\boldsymbol{z}_n \in \mathbb{C}^Q$ are the nth columns of $\boldsymbol{X}$ and $\boldsymbol{Z}$, respectively. Since the N summation terms in (24) are independent, we can obtain the minimum of (21) by separately solving the following N subproblems

$$\boldsymbol{z}_n^{(k+1)} = \arg\min_{\boldsymbol{z}_n\in\mathbb{C}^Q} \|\boldsymbol{x}_n - \boldsymbol{Y}^{(k)}\boldsymbol{z}_n\|_p^p,\ n = 1,\ldots,N \tag{25}$$

where $\boldsymbol{z}_n^{(k+1)}$ is the nth column of $\boldsymbol{Z}^{(k+1)}$, that is,

$$\boldsymbol{Z}^{(k+1)} = \left[\boldsymbol{z}_1^{(k+1)},\ldots,\boldsymbol{z}_N^{(k+1)}\right]. \tag{26}$$

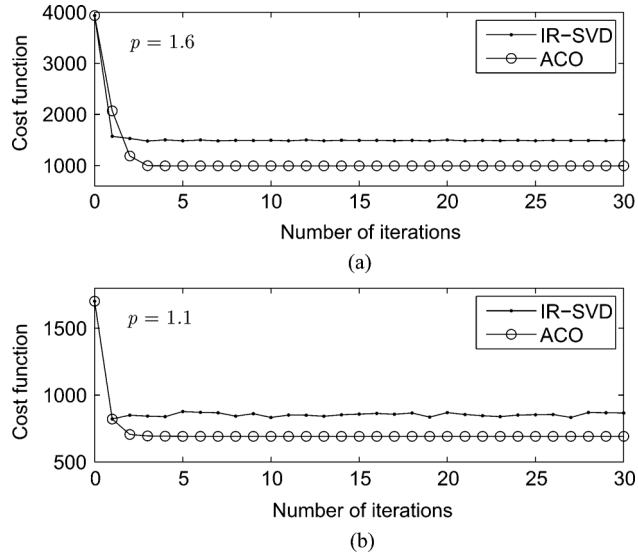


Fig. 1. Convergence behavior of IR-SVD and ACO algorithms. (a) $p = 1.6$; (b) $p = 1.1$.

Now the *matrix* ℓ_p-norm minimization of (21) is converted into N subproblems of *vector* ℓ_p-norm minimization. The last remaining problem is to solve the vector ℓ_p-norm minimization of (25), which will be elaborated in the next subsection.

The ACO convergence for subspace decomposition is illustrated in the following proposition.

Proposition 1: The ℓ_p-norm minimization based subspace estimation algorithm monotonically non-increases the error value as defined in (15), thus it converges in the limit.

Proof: Denote the residual error value at the kth iteration of the alternating algorithm as $J_p\left(\boldsymbol{Y}^{(k)},\boldsymbol{Z}^{(k)}\right)$. From (21) and (22), it follows that

$$\begin{aligned} J_p\left(\boldsymbol{Y}^{(k+1)},\boldsymbol{Z}^{(k+1)}\right) &\le J_p\left(\boldsymbol{Y}^{(k)},\boldsymbol{Z}^{(k+1)}\right) \\ &\le J_p\left(\boldsymbol{Y}^{(k)},\boldsymbol{Z}^{(k)}\right). \end{aligned} \tag{27}$$

This means that the error function $J_p(\boldsymbol{Y},\boldsymbol{Z})$ does not increase at each iteration. In addition, the error function $J_p(\boldsymbol{Y},\boldsymbol{Z})$ is bounded from below by 0. Therefore the iterative ACO procedure converges. □

The relative reduction of the error value can be used to examine the convergence. Specifically, the convergence is determined by checking whether the following inequality holds:

$$\frac{J_p\left(\boldsymbol{Y}^{(k)},\boldsymbol{Z}^{(k)}\right) - J_p\left(\boldsymbol{Y}^{(k+1)},\boldsymbol{Z}^{(k+1)}\right)}{J_p\left(\boldsymbol{Y}^{(k)},\boldsymbol{Z}^{(k)}\right)} < \varepsilon \tag{28}$$

for some small tolerance ε. In the simulations, we choose $\varepsilon = 10^{-7}$.

Using the experimental setting and data in Section IV-B, Fig. 1 plots the convergence behaviors of the IR-SVD and ACO algorithms. In Fig. 1(a), the IR-SVD algorithm converges for $p = 1.6$ while it slightly oscillates for $p = 1.1$ in Fig. 1(b). Fig. 2 further shows the relative error of the ACO scheme, which converges in several tens of iterations with the specific tolerance of $\varepsilon = 10^{-7}$.

Remark 3: It can be seen that the ACO algorithm is superior to the IR-SVD method. Not only its convergence is guaranteed

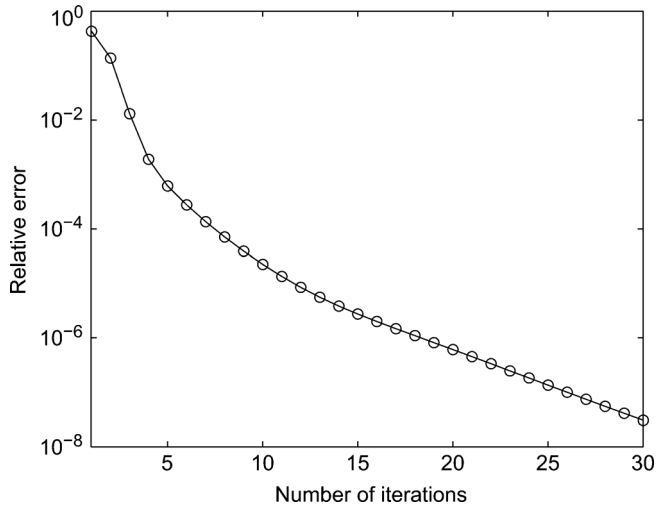


Fig. 2. Relative error versus number of iterations of ACO algorithm. It converges in several tens of iterations with a tolerance of 10^{-7}.

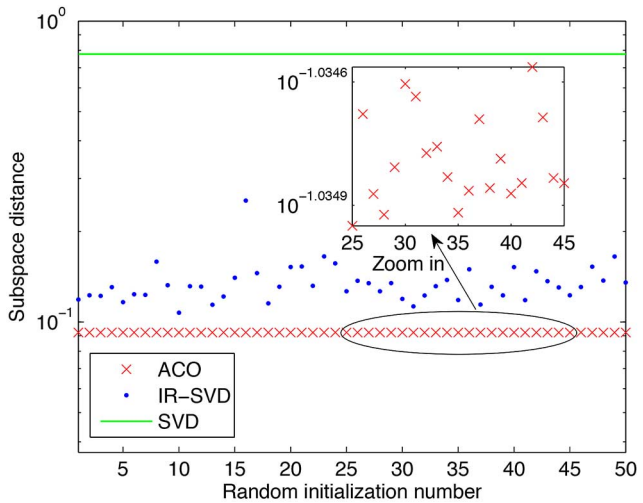


Fig. 3. Subspace distance of the estimated signal subspace and the true array manifold matrix $\boldsymbol{A}$ using random initialization.

but also it more rapidly decreases the objective function value than the latter. It is noticed that the ACO procedure does not necessarily converge to the global minimum point. The point that it converges to depends on the initial value $\boldsymbol{Y}^{(0)}$. We can initialize $\boldsymbol{Y}$ using random matrix of full column rank or the signal subspace obtained from the SVD or IR-SVD. It is expected that the estimated signal subspace $\boldsymbol{Y}$ spans the same range space of $\boldsymbol{A}$. Here the normalized *subspace distance* between $\boldsymbol{Y}$ and $\boldsymbol{A}$ is used to evaluate the quality of subspace estimate, which is defined as

$$d_p(\boldsymbol{Y},\boldsymbol{A}) = \|\boldsymbol{P}_{\boldsymbol{Y}} - \boldsymbol{P}_{\boldsymbol{A}}\|_p / \|\boldsymbol{P}_{\boldsymbol{A}}\|_p \tag{29}$$

where $\boldsymbol{P}_{\boldsymbol{Y}} = \boldsymbol{Y}\left(\boldsymbol{Y}^H\boldsymbol{Y}\right)^{-1}\boldsymbol{Y}^H$ and $\boldsymbol{P}_{\boldsymbol{A}}$ are the projection matrices onto the column spaces of $\boldsymbol{Y}$ and $\boldsymbol{A}$, respectively. Again, we can take different values of p. The definition of (29) aligns with [39] for $p = 2$ but here we allow other values of p. The subspace distance will become zero if $\mathrm{range}(\boldsymbol{Y}) = \mathrm{range}(\boldsymbol{A})$ exactly holds. A smaller subspace distance means a better subspace estimate. Fig. 3 plots the subspace distance of 50 random initializations with $p = 1.1$. It can be seen that the IR-SVD scheme is more sensitive to initial value while the results of ACO vary slightly with different initial values. Furthermore, the ACO approach can provide an accurate subspace estimate for DOA estimation although the global optimum is not guaranteed.

D. Complex-Valued Newton's Methods for Vector ℓ_p-Norm Minimization

This subsection addresses the solver for the last remaining problem, namely, the complex-valued vector ℓ_p-norm minimization of (25). Existing approaches to this problem including the most widely used IRLS method [25], [35]. However, the IRLS scheme may diverge when p is close to 1 or larger than 3. We devise two new complex-valued Newton's methods, called pseudo-Newton's and full-Newton's methods, for this problem. The two schemes adopt optimal step size in each iteration. It is revealed that the IRLS algorithm can be interpreted as a special case of the pseudo-Newton's method using a fixed step size of $p/2$, which is suboptimal.

For notation simplicity, the superscripts and subscripts in (25) are omitted. We consider the following problem

$$\min_{\boldsymbol{z}} f(\boldsymbol{z}) = \|\boldsymbol{x} - \boldsymbol{Y}\boldsymbol{z}\|_p^p \tag{30}$$

where $\boldsymbol{Y} \in \mathbb{C}^{M\times Q}$, $\boldsymbol{x} \in \mathbb{C}^M$ and $\boldsymbol{z} \in \mathbb{C}^Q$.

1) Gradient and Hessian of ℓ_p-norm: By defining the residual vector

$$\boldsymbol{r} = \boldsymbol{Y}\boldsymbol{z} - \boldsymbol{x} = [r_1, \ldots, r_M]^T, \tag{31}$$

the objective function $f(\boldsymbol{z})$ can be expressed as

$$f(\boldsymbol{z}) = f(\boldsymbol{r}) = \|\boldsymbol{r}\|_p^p = \sum_{i=1}^{M} |r_i|^p. \tag{32}$$

It is not difficult to derive the partial derivative w.r.t. the complex quantity r_i as

$$\begin{aligned}\frac{\partial f}{\partial r_i^*} &= \frac{1}{2}\left(\frac{\partial f}{\partial \mathrm{Re}(r_i)} + \mathrm{j}\frac{\partial f}{\partial \mathrm{Im}(r_i)}\right)\\ &= \frac{p}{2}|r_i|^{p-2}r_i,\ i = 1, \ldots, M\end{aligned} \tag{33}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ are the real and imaginary parts of a complex number, respectively. The gradient of $f(\boldsymbol{r})$ is given by

$$\frac{\partial f}{\partial \boldsymbol{r}^*} = \left[\frac{\partial f}{\partial r_1^*}, \ldots, \frac{\partial f}{\partial r_M^*}\right]^T = \frac{p}{2}|\boldsymbol{r}|^{p-2} \odot \boldsymbol{r} \tag{34}$$

with $|\boldsymbol{r}|^{p-2} = \left[|r_1|^{p-2}, \ldots, |r_M|^{p-2}\right]^T$. Define a diagonal matrix

$$\boldsymbol{W} = \mathrm{diag}\left\{|r_1|^{p-2}, \ldots, |r_M|^{p-2}\right\}. \tag{35}$$

Equation (34) can be rewritten as

$$\frac{\partial f}{\partial \boldsymbol{r}^*} = \frac{p}{2}\boldsymbol{W}\boldsymbol{r}. \tag{36}$$

Note that sometimes we use $\boldsymbol{W}(\boldsymbol{z})$ to emphasize that $\boldsymbol{W}$ is a function of $\boldsymbol{z}$. Clearly, $\boldsymbol{W}(\boldsymbol{z})$ is positive definite.

We compute the second-order partial derivative as

$$\frac{\partial^2 f}{\partial r_i^* \partial r_j} = \begin{cases} p^2|r_i|^{p-2}/4, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases} \tag{37}$$

Hence the $M \times M$ *partial Hessian* matrix of f w.r.t. $\boldsymbol{r}$, denoted by $\boldsymbol{H}_{\boldsymbol{r}^*\boldsymbol{r}}$, is diagonal and has the form:

$$\boldsymbol{H}_{\boldsymbol{r}^*\boldsymbol{r}} = \frac{\partial^2 f}{\partial \boldsymbol{r}^* \partial \boldsymbol{r}^T} = \frac{p^2}{4}\boldsymbol{W}. \tag{38}$$

The $M \times Q$ Jacobian matrix of $\boldsymbol{r}(\boldsymbol{z})$ w.r.t. $\boldsymbol{z}^*$ is given by

$$\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{z}^H} = \boldsymbol{Y}^*. \tag{39}$$

Then the gradient of $f(\boldsymbol{z})$ w.r.t. $\boldsymbol{z}$ is computed as

$$\begin{aligned}\boldsymbol{g}(\boldsymbol{z}) = \frac{\partial f}{\partial \boldsymbol{z}^*} &= \left(\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{z}^H}\right)^T \frac{\partial f}{\partial \boldsymbol{r}^*} = \boldsymbol{Y}^H \frac{\partial f}{\partial \boldsymbol{r}^*} \\ &= \frac{p}{2}\boldsymbol{Y}^H \boldsymbol{W}(\boldsymbol{Y}\boldsymbol{z} - \boldsymbol{x}).\end{aligned} \tag{40}$$

The $Q \times Q$ leading partial Hessian matrix of f w.r.t. $\boldsymbol{z}$ is

$$\begin{aligned}\boldsymbol{H}_{\boldsymbol{z}^*\boldsymbol{z}} = \frac{\partial^2 f}{\partial \boldsymbol{z}^* \partial \boldsymbol{z}^T} &= \left(\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{z}^H}\right)^T \boldsymbol{H}_{\boldsymbol{r}^*\boldsymbol{r}} \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{z}^T} \\ &= \boldsymbol{Y}^H \boldsymbol{H}_{\boldsymbol{r}^*\boldsymbol{r}} \boldsymbol{Y} = \frac{p^2}{4}\boldsymbol{Y}^H \boldsymbol{W}(\boldsymbol{z})\boldsymbol{Y}.\end{aligned} \tag{41}$$

Note that the partial Hessian $\boldsymbol{H}_{\boldsymbol{z}^*\boldsymbol{z}}$ is positive definite because $\boldsymbol{W}(\boldsymbol{z})$ is positive definite. The pseudo-Newton's method only uses the $Q \times Q$ partial Hessian $\boldsymbol{H}_{\boldsymbol{z}^*\boldsymbol{z}}$ whereas the full-Newton's method exploits the following $2Q \times 2Q$ full Hessian matrix

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_{\boldsymbol{z}^*\boldsymbol{z}} & \boldsymbol{H}_{\boldsymbol{z}^*\boldsymbol{z}^*} \\ \boldsymbol{H}_{\boldsymbol{z}\boldsymbol{z}} & \boldsymbol{H}_{\boldsymbol{z}\boldsymbol{z}^*} \end{bmatrix}. \tag{42}$$

The full Hessian matrix is positive definite when $p > 1$. The other three partial Hessian matrices are given by

$$\boldsymbol{H}_{\boldsymbol{z}^*\boldsymbol{z}^*} = \frac{\partial^2 f}{\partial \boldsymbol{z}^* \partial \boldsymbol{z}^H} = \boldsymbol{Y}^H \boldsymbol{H}_{\boldsymbol{r}^*\boldsymbol{r}^*} \boldsymbol{Y}^* \tag{43}$$

where

$$\begin{aligned}\boldsymbol{H}_{\boldsymbol{r}^*\boldsymbol{r}^*} &= \frac{\partial^2 f}{\partial \boldsymbol{r}^* \partial \boldsymbol{r}^H} \\ &= \frac{p(p-2)}{4}\mathrm{diag}\left\{|r_1|^{p-4}r_1^2, \ldots, |r_M|^{p-4}r_M^2\right\}\end{aligned} \tag{44}$$

is a diagonal matrix, $\boldsymbol{H}_{\boldsymbol{z}\boldsymbol{z}} = \boldsymbol{H}^*_{\boldsymbol{z}^*\boldsymbol{z}^*}$, and $\boldsymbol{H}_{\boldsymbol{z}\boldsymbol{z}^*} = \boldsymbol{H}^*_{\boldsymbol{z}^*\boldsymbol{z}}$. It is noticed that the two off-diagonal block matrices $\boldsymbol{H}^*_{\boldsymbol{z}^*\boldsymbol{z}^*}$ and $\boldsymbol{H}_{\boldsymbol{z}\boldsymbol{z}}$ become zero if $p = 2$. In this case, these two partial Hessian matrices contain no information. When $p \neq 2$, these two matrices do not vanish and contain useful information for optimization.

2) Newton's Method with Optimal Step Size: The pseudo-Newton's method generates a sequence $\{\boldsymbol{z}^{(k)}\}$ $(k = 0, 1, \ldots)$ through the following iteration

$$\boldsymbol{z}^{(k+1)} = \boldsymbol{z}^{(k)} + \mu_k \Delta \boldsymbol{z}^{(k)} \tag{45}$$

to find the minimum of $f(\boldsymbol{z})$, where $\mu_k \geq 0$ is a positive step size, and

$$\Delta \boldsymbol{z}^{(k)} = -\boldsymbol{H}^{-1}_{\boldsymbol{z}^*\boldsymbol{z}} \boldsymbol{g}(\boldsymbol{z}^{(k)}) \tag{46}$$

is the search direction or *pseudo-Newton step* in the kth iteration. According to (40) and (41), the pseudo-Newton step is computed as

$$\Delta \boldsymbol{z}^{(k)} = -\frac{2}{p}\left(\boldsymbol{Y}^H \boldsymbol{W}(\boldsymbol{z}^{(k)})\boldsymbol{Y}\right)^{-1} \boldsymbol{Y}^H \boldsymbol{W}(\boldsymbol{z}^{(k)})(\boldsymbol{Y}\boldsymbol{z}^{(k)} - \boldsymbol{x}). \tag{47}$$

Selection of step size is an important issue. In conventional Newton's method, the fixed step size $\mu_k = 1$ is adopted, which is clearly not an optimal choice. If a fixed step size $\mu_k = p/2$ is used, then (45) can be simplified to a fixed-point iteration

$$\boldsymbol{z}^{(k+1)} = \left(\boldsymbol{Y}^H \boldsymbol{W}(\boldsymbol{z}^{(k)})\boldsymbol{Y}\right)^{-1} \boldsymbol{Y}^H \boldsymbol{W}(\boldsymbol{z}^{(k)})\boldsymbol{x} \tag{48}$$

which is reduced to the widely used IRLS algorithm [25], [35] with $\boldsymbol{W}(\boldsymbol{z}^{(k)})$ being the weighting matrix. The relation between the IRLS and pseudo-Newton's methods for ℓ_p-norm minimization is clear now and described in the following proposition.

Proposition 2: The widely used IRLS approach to vector ℓ_p-norm minimization is a special case of the pseudo-Newton's method using a fixed step size of $p/2$.

However, the fixed step size strategy is not optimal. We consider the variable step size. For a given Newton direction $\Delta \boldsymbol{z}^{(k)}$, the optimal step size is given by solving the line search

$$\mu_k = \arg\min_{\mu \geq 0} \left\| \boldsymbol{Y}\left(\boldsymbol{z}^{(k)} + \mu \Delta \boldsymbol{z}^{(k)}\right) - \boldsymbol{x} \right\|_p^p. \tag{49}$$

Denoting the residual vector in the kth iteration as

$$\boldsymbol{r}^{(k)} = \boldsymbol{Y}\boldsymbol{z}^{(k)} - \boldsymbol{x} \tag{50}$$

the objective function w.r.t. the step size μ can be written as

$$\min_{\mu \geq 0} f(\mu) = \left\| \boldsymbol{r}^{(k)} + \mu \boldsymbol{Y} \Delta \boldsymbol{z}^{(k)} \right\|_p^p. \tag{51}$$

This is a one-dimensional optimization problem and can be easily solved by the existing line search techniques, such as Golden section search or tangential method [42]. The global optimality of μ is guaranteed since $f(\mu)$ is unimodal w.r.t. μ if $p \geq 1$. Unlike the pseudo-Newton's and IRLS algorithms that only use the leading partial Hessian matrix, the full-Newton's method employs the full $2Q \times 2Q$ Hessian matrix to compute the search direction

$$\begin{bmatrix} \Delta \boldsymbol{z}^{(k)} \\ (\Delta \boldsymbol{z}^{(k)})^* \end{bmatrix} = -\begin{bmatrix} \boldsymbol{H}_{\boldsymbol{z}^*\boldsymbol{z}} & \boldsymbol{H}_{\boldsymbol{z}^*\boldsymbol{z}^*} \\ \boldsymbol{H}_{\boldsymbol{z}\boldsymbol{z}} & \boldsymbol{H}_{\boldsymbol{z}\boldsymbol{z}^*} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{g}(\boldsymbol{z}^{(k)}) \\ \boldsymbol{g}^*(\boldsymbol{z}^{(k)}) \end{bmatrix} \tag{52}$$

and the updating rule is

$$\begin{bmatrix} \boldsymbol{z}^{(k+1)} \\ (\boldsymbol{z}^{(k+1)})^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{z}^{(k)} \\ (\boldsymbol{z}^{(k)})^* \end{bmatrix} + \mu_k \begin{bmatrix} \Delta \boldsymbol{z}^{(k)} \\ (\Delta \boldsymbol{z}^{(k)})^* \end{bmatrix} \tag{53}$$

where the optimal step size is determined according to (49). Equations (52) and (53) give a true Newton's method since it utilizes all second-order derivatives. The initial value of the IRLS

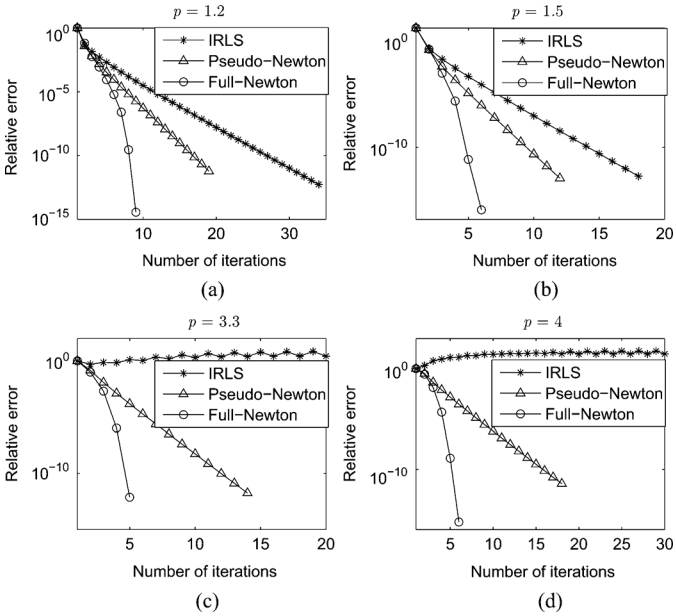


Fig. 4. Convergence rate versus number of iterations of IRLS and two Newton's methods with optimal step sizes for $p = 1.2$, 1.5, 3.3, and 4.

and two Newton's methods can be taken as the least-squares (LS) solution $\boldsymbol{z}^{(0)} = \widehat{\boldsymbol{z}}_{\rm LS} = \left(\boldsymbol{Y}^H\boldsymbol{Y}\right)^{-1}\boldsymbol{Y}^H\boldsymbol{x}$.

Note that it is not necessary to directly compute the inverse of the Hessian matrices of (46) and (52) when computing the pseudo-Newton step and Newton step. For example, we can use some efficient algorithms such as the conjugate gradient (CG) method [43] for solving the linear equation $\boldsymbol{H}_{\boldsymbol{z}^*\boldsymbol{z}}\Delta\boldsymbol{z}^{(k)} = -\boldsymbol{g}(\boldsymbol{z}^{(k)})$ to obtain pseudo-Newton step $\Delta\boldsymbol{z}^{(k)}$.

The convergence rates of the IRLS and two Newton's methods with optimal step sizes for a variety of values of p are compared. We take four values of p for illustration, i.e., $p = 1.2$, 1.5, 3.3, and 4. In this numerical example, we randomly generate the coefficient matrix $\boldsymbol{Y} \in \mathbb{C}^{50\times 20}$ and vector $\boldsymbol{x} \in \mathbb{C}^{50}$. We are primarily interested in the behavior, as a function of the number of iterations, of the relative error $|f(\boldsymbol{z}^{(k)}) - f(\boldsymbol{z}^\star)|/f(\boldsymbol{z}^\star)$, where $f(\boldsymbol{z}^\star)$ is the global minimum. We can calculate this global minimum exactly (in practice up to computer round-off precision) with a finite number of steps using the proposed Newton's method or any optimization software package in advance. Fig. 4 shows the convergence rates of the three methods. It is observed that the IRLS algorithm does not converge for $p = 3.3$ and $p = 4$ while the two Newton's methods converge in all cases. When the IRLS algorithm converges, it has a linear convergence rate. The pseudo-Newton's method also has a linear convergence rate but it converges faster than the IRLS scheme. The full-Newton's method has a quadratic convergence rate and converges very fast. It only needs several iterations for convergence.

Computational Complexity: We first analyze the complexities of the Newton's and IRLS methods. The complexity of matrix multiplication $\boldsymbol{Y}^H\boldsymbol{W}(\boldsymbol{z})\boldsymbol{Y}$ is $\mathcal{O}(MQ^2)$ because $\boldsymbol{W}(\boldsymbol{z})$ is diagonal. This is the complexity of computing the partial and full Hessian matrices. The complexity for solving the partial and full Newton steps of (46) and (52) is $\mathcal{O}(Q^3)$. Hence the complexity of the Newton's methods is $\mathcal{O}(MQ^2)$ in each iteration due to $M > Q$, which is the same as that of the IRLS algorithm.

It is clear that the complexity of the first convex optimization problem in (21) is $\mathcal{O}(NMQ^2)$, which is also the complexity of the second convex optimization problem in (22). Therefore the complexity of the ACO approach in each iteration is $\mathcal{O}(NMQ^2)$.

E. Summary of ℓ_p-MUSIC Algorithm

The steps of the ℓ_p-MUSIC algorithm for robust DOA estimation using ACO are summarized in Algorithm 2.

Algorithm 2: ℓ-MUSIC

Input: $M \times N$ received data matrix $\boldsymbol{X}$, number of sources Q and tolerance ε.
Output: DOA estimates.

1. Initialize $\boldsymbol{Y}^{(0)}$ with a random matrix of full column rank.
2. **for** $k = 1, 2, \ldots,$ **do** until converge
   1) Calculate
   $$\boldsymbol{Z}^{(k+1)} = \arg\min_{\boldsymbol{Z}\in\mathbb{C}^{Q\times N}} \left\|\boldsymbol{X} - \boldsymbol{Y}^{(k)}\boldsymbol{Z}\right\|_p^p$$
   2) Use the Newton's method to solve the following N subproblems:
   $$\boldsymbol{z}_n^{(k+1)} = \arg\min_{\boldsymbol{z}_n\in\mathbb{C}^Q} \|\boldsymbol{x}_n - \boldsymbol{Y}^{(k)}\boldsymbol{z}_n\|_p^p, \quad n = 1, \ldots, N.$$
   3) Solve the second minimization problem
   $$\boldsymbol{Y}^{(k+1)} = \arg\min_{\boldsymbol{Y}\in\mathbb{C}^{M\times Q}} \left\|\boldsymbol{X} - \boldsymbol{Y}\boldsymbol{Z}^{(k+1)}\right\|_p^p$$
   in the same way.
   4) Check the stopping condition in (28).
   end for
3. Compute the projection matrix onto the noise subspace $\boldsymbol{P}_n = \boldsymbol{I} - \boldsymbol{Y}\left(\boldsymbol{Y}^H\boldsymbol{Y}\right)^{-1}\boldsymbol{Y}^H$, where $\boldsymbol{Y}$ is obtained from Step 2.
4. Compute the spatial spectrum of (20) and search for its peaks as the DOA estimates.

Remark 4: We can also use the IR-SVD procedure shown in Algorithm 1 instead of the ACO in Step 2. In Step 4, the root-MUSIC can be adopted to avoid spectrum search. Moreover, once the robust subspace estimate is obtained in Step 3, other subspace based DOA estimation techniques such as ESPRIT [44], can be applied. Selection of p is important for the ℓ_p-MUSIC algorithm. Clearly, the optimal p is related to the PDF of the noise. However, since the noise PDF is often unavailable in practical applications, it is difficult to obtain the optimal p. For GGD noise, the optimal p equals the shape parameter β in (55) since $p = \beta$ leads to ML estimate of the signal subspace. For α-stable and GMM noises, determining the optimal p is difficult and remains an open problem, even though the noise PDF is known. In practice, we aim at an appropriate rather than the optimal value of p because a proper but suboptimal p can also give good performance. From the simulation results, it is seen that $p = 1.1$ works well for a wide class of impulsive noises. Hence the value near $p = 1$ is an appropriate choice for achieving robustness. In addition, a rule-of-thumb on

selection of p is that the more impulsive the noise is, the smaller p is preferred.

IV. Simulation Results

The performance of the ℓ_p-MUSIC, conventional MUSIC [4], ROC-MUSIC [18], FLOM-MUSIC [19], ZMNL [21], [23], MM-estimator [27], [28], and EM [12] algorithms, as well as the Cramer-Rao bound (CRB), are compared. The ZMNL first uses a Gaussian-tailed ZMNL (GZMNL) function to clip outliers. After this data preprocessing, the conventional MUSIC is applied. Therefore this method is referred to as GZMNL-MUSIC for short. The method in [27] first robustly estimates the covariance matrix by MM-estimator [28], and then employs MUSIC for DOA estimation. ROC-MUSIC and FLOM-MUSIC use fractional lower-order moments instead of the second-order sample covariance matrix. For the purpose of fair comparison, the fractional order used in ROC-MUSIC and FLOM-MUSIC is set as the same in the ℓ_p-MUSIC. Although the EM algorithm in [12] is originally designed for GMM noise, it is claimed that it performs well over a wide range of noise environments and is robust to PDF mismatch. We consider a ULA with inter-sensor spacing of $d = \lambda/2$. The emitting sources are two independent quadrature phase-shift keying (QPSK) signals with equal power. In all simulations, we use the ACO rather than the IR-SVD since ACO obtains a better estimate of the signal subspace.

A. Impulsive Noise Model and Cramer-Rao Bound

The PDF of impulsive noise exhibits heavier tail than that of the Gaussian distribution. Three widely used PDF models for impulsive noise, i.e., GMM, GGD, and α-stable distribution, are considered.

1) GMM: The PDF of the two-term circular (complex symmetric) Gaussian mixture noise $e(n)$ is

$$p_e(e) = \sum_{i=1}^{2} \frac{c_i}{\pi\sigma_i^2} \exp\left(-\frac{|e|^2}{\sigma_i^2}\right) \tag{54}$$

where $0 \le c_i \le 1$ and σ_i^2 are the probability and variance of the ith term, respectively, with $c_1 + c_2 = 1$. If $\sigma_2^2 \gg \sigma_1^2$ and $c_2 < c_1$ are selected, large noise samples of variance σ_2^2 occurring with a smaller probability c_2 can be viewed as outliers embedded in Gaussian background noise of variance σ_1^2. Therefore the GMM can well model the phenomenon in the presence of both Gaussian thermal noise and impulsive noise. The total variance of $e(n)$ is $\sigma_e^2 = \sum_i c_i\sigma_i^2$ and the SNR is defined as $\mathrm{SNR} = \mathrm{E}\left\{|s(n)|^2\right\}/\sigma_e^2$ with $\mathrm{E}\left\{|s(n)|^2\right\}$ representing the average power of the source signal. In our simulation, we set $\sigma_2^2 = 100\sigma_1^2$ and $c_2 = 0.1$. Hence there are 10% noise samples that are considered as outliers.

2) GGD: The PDF of the circular zero-mean GGD with variance σ_e^2 is

$$p_e(e) = \frac{\beta\Gamma(4/\beta)}{2\pi\sigma_e^2\Gamma^2(2/\beta)} \exp\left(-\frac{|e|^\beta}{c\sigma_e^\beta}\right) \tag{55}$$

where $\beta > 0$ is the shape parameter, $\Gamma(\cdot)$ is the Gamma function and $c = (\Gamma(2/\beta)/\Gamma(4/\beta))^{\beta/2}$. When $\beta = 2$, the GGD reduces to the circular Gaussian distribution. The $\beta > 2$ models short-tailed noise while $\beta < 2$ models heavy-tailed one. Especially, $\beta = 1$ corresponds to Laplacian noise. The smaller the value of β, the more impulsive the noise is. We take $\beta = 0.4$ in the simulation.

3) α-stable Process: Here we adopt the symmetric $\alpha-\mathrm{stable}$ ($\mathrm{S}\alpha\mathrm{S}$) distribution with zero-location, whose characteristic function is expressed as

$$\varphi(\omega) = \exp\left(-\gamma^\alpha|\omega|^\alpha\right) \tag{56}$$

where $0 < \alpha \le 2$ is called the characteristic exponent that describes the tail of the distribution and $\gamma > 0$ is the scale. When $\alpha = 2$, the α-stable distribution reduces to the Gaussian distribution and γ^2 is similar to the variance of the Gaussian distribution. When $\alpha = 1$, the $\alpha-\mathrm{stable}$ distribution becomes the Cauchy distribution. When $\alpha < 2$, α-stable noise shows heavy tails and hence is impulsive. The smaller the value of α, the more impulsive the noise is. Although the characteristic function of the $\mathrm{S}\alpha\mathrm{S}$ noise is of closed-form, the PDF does not have a closed-form expression, except for $\alpha = 1$ and $\alpha = 2$ [15]. Since the second-order and higher-order moments of the α-stable distribution are infinite for $\alpha < 2$, the commonly used SNR is meaningless. To quantify the relative strength between signal and noise, a generalized SNR (GSNR)

$$\mathrm{GSNR} = \mathrm{E}\left\{|s(n)|^2\right\}/\gamma^\alpha \tag{57}$$

is adopted.

4) CRB for Non-Gaussian Noise: Kozick and Sadler have derived a general expression of the CRB of the DOA parameters $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_Q]^T$ for non-Gaussian noise [12], [45], which is shown as

$$\mathrm{CRB}(\boldsymbol{\theta}) = \frac{1}{I_c}\mathrm{diag}\left\{\sum_{n=1}^{N} \mathrm{Re}\left(\boldsymbol{S}_n^H\boldsymbol{\Xi}^H(\boldsymbol{\theta})\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{\Xi}(\boldsymbol{\theta})\boldsymbol{S}_n\right)\right\} \tag{58}$$

where $\boldsymbol{S}_n = \mathrm{diag}\{s_1(n), \ldots, s_Q(n)\}$ is a diagonal matrix, $\boldsymbol{\Xi}(\boldsymbol{\theta}) = [\boldsymbol{\xi}(\theta_1), \ldots, \boldsymbol{\xi}(\theta_Q)]$ with $\boldsymbol{\xi}(\theta) = \mathrm{d}\boldsymbol{a}(\theta)/\mathrm{d}\theta$, $\boldsymbol{P}_{\boldsymbol{A}}^{\perp} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}$ is the projection onto the orthogonal complementary space of $\boldsymbol{A}$, and

$$I_c = \pi\int_0^\infty \frac{(p_e'(\rho))^2}{p_e(\rho)}\rho\mathrm{d}\rho \tag{59}$$

with $\rho = |e|$ being the modulus of the complex variable e and $p_e'(\rho)$ the derivative of $p_e(\rho)$. The noise PDF affects the CRB only through the scalar I_c. The CRBs of GMM, GGD and $\mathrm{S}\alpha\mathrm{S}$ noises can be numerically computed using (58) and(59).

B. Demonstration of Robustness against Impulsive Noise

In the first simulation, the robustness of the six subspace-based DOA estimators in $\mathrm{S}\alpha\mathrm{S}$ noise are compared. The DOAs of the two closely-spaced signals are $\theta_1 = -5^\circ$ and $\theta_2 = 5^\circ$. The number of sensors and snapshots are $M = 7$ and $N = 100$. Fig. 5 shows the spatial spectra of five independent trials in $\mathrm{S}\alpha\mathrm{S}$ noise with $\alpha = 1.4$ and $\mathrm{GSNR} = 5$ dB. Since the EM algorithm [12] does not output the spatial spectrum, it is not included in this simulation example. We take $p = 1.1$ for the ℓ_p-MUSIC, ROC-MUSIC and FLOM-MUSIC methods. The performance of the conventional MUSIC estimator severely degrades in α-stable noise because it fails in distinguishing the two closely-spaced sources. The ℓ_p-MUSIC method possesses higher resolution and is more robust against impulsive noise. The GZMNL-MUSIC and MM-MUSIC schemes are also robust to outliers and have a similar performance of the ℓ_p-MUSIC. Although the ROC-MUSIC and FLOM-MUSIC are better than the conventional MUSIC, they are inferior to the ℓ_p-MUSIC algorithm.

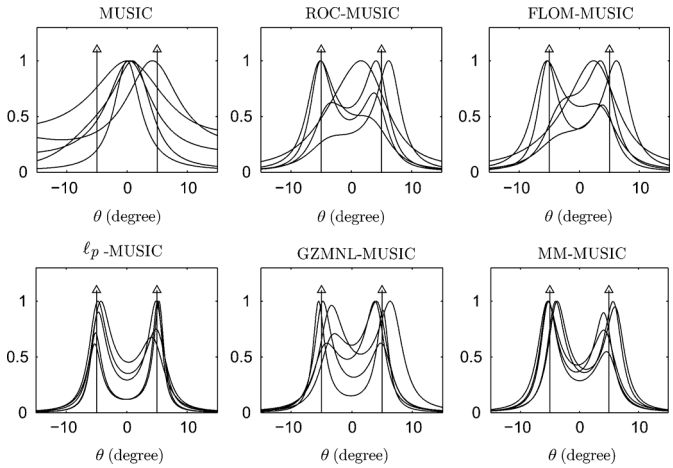


Fig. 5. Spatial spectra of five independent trials in $\mathrm{S}\alpha\mathrm{S}$ noise with $\alpha = 1.4$ and GSNR $= 5$ dB. The six DOA estimators use the same data in each trial. The vertical lines show the true DOAs.

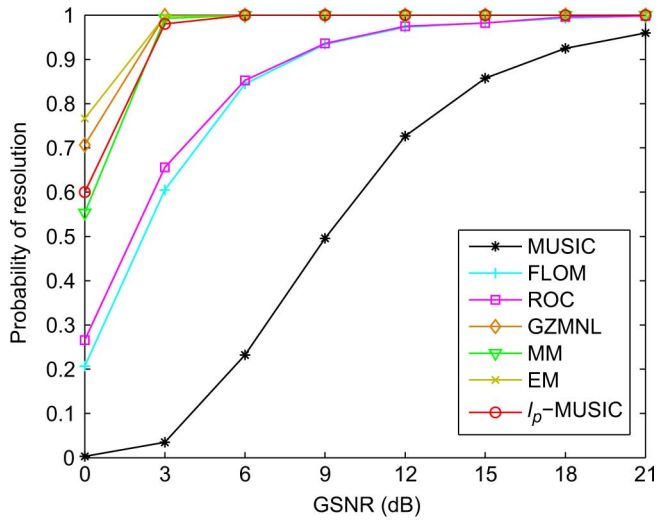


Fig. 6. Probability of resolution in $\mathrm{S}\alpha\mathrm{S}$ noise versus GSNR with $\alpha = 1.4$.

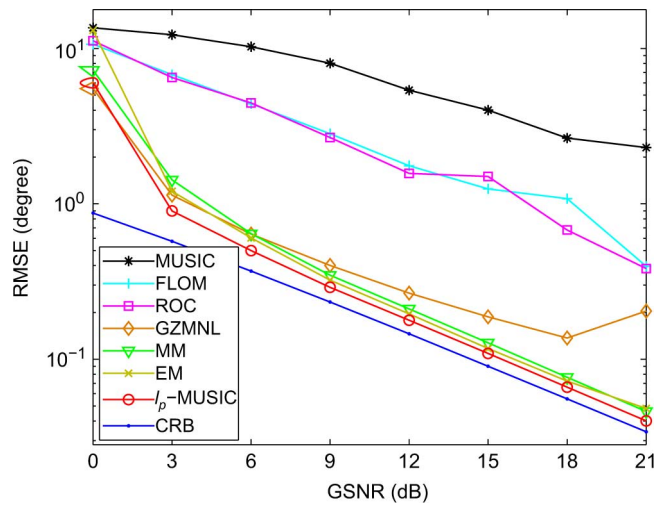


Fig. 7. RMSE of DOA estimate of the first source in $\mathrm{S}\alpha\mathrm{S}$ noise versus GSNR with $\alpha = 1.4$.

C. Statistical Performance Comparison

Monte Carlo trials have been carried out to evaluate the performance of the DOA estimation algorithms. The value of $p = 1.1$ is taken for the ℓ_p-MUSIC, ROC-MUSIC and FLOM-MUSIC methods. The DOAs of the two incoming signals are $\theta_1 = -8°$ and $\theta_2 = 8°$. The numbers of sensors and snapshots are $M = 6$ and $N = 100$. Two statistical performance measures are used. The first is the root mean square error (RMSE)

$$\mathrm{RMSE}\left(\widehat{\theta}_q\right) = \sqrt{\frac{1}{M_c}\sum_{i=1}^{M_c}\left(\widehat{\theta}_{q,i} - \theta_q\right)^2} \tag{60}$$

where M_c is the number of Monte Carlo trials and $\widehat{\theta}_{q,i}$ is the DOA estimate of the qth source in the ith trial. In the following, the root-MUSIC [41] is applied to calculate the DOA parameters after the signal subspace is obtained. Note that the root-MUSIC always gives two roots from which the DOA estimates of the two sources are obtained. If a DOA estimator cannot distinguish the two sources, the error of the DOA estimate given by the root technique will be large. When calculating the RMSE, all the DOA estimates, including those of the unresolved trials, are included. We adopt the probability of resolution as another performance measure. If the DOA estimate errors of both sources are less than 5°, we call it successes in resolving two sources. Clearly, the probability of resolution indicates the resolution capability of a DOA estimator.

All the three types of impulsive noise are considered. For $\mathrm{S}\alpha\mathrm{S}$ noise, performance versus GSNR and α is investigated. At each GSNR and α, $M_c = 2000$ Monte Carlo trials are performed. Figs. 6 and 7 show the probability of resolution and the RMSE of the DOA estimate of the first source versus GSNR with $\alpha = 1.4$, respectively. The RMSE of the second source is similar to that of the first source and hence we omit it. Figs. 8 and 9 show the results versus $\alpha \in [1.1, 2]$ at GSNR = 6 dB. For GMM and GGD noises, performance versus SNR is studied. For each SNR, 2000 Monte Carlo trials are carried out. Fig. 10 plots the RMSE versus SNR in two-term GMM noise, where there are 10% outliers whose variance is 100 times of the variance of the background noise. Fig. 11 illustrates the RMSE versus SNR in GGD noise with shape parameter $\beta = 0.4$. In addition, the CRBs of different non-Gaussian noises are also plotted for comparison. Note that the probabilities of resolution for GMM and GGD noises are not included here because all robust DOA estimators perform similarly and has high angular resolution.

It is seen from Figs. 6–11 that the conventional MUSIC is not robust in the presence of impulsive noise. The ℓ_p-norm minimization with $p < 2$ leads to an improved performance. In Gaussian noise or $\alpha = 2$, the performance of the ℓ_p-MUSIC estimator is comparable to other methods. The ℓ_p-MUSIC has the best performance in most cases: its RMSEs are close to the CRB. The EM algorithm has performance close to the CRB and is slightly superior to ℓ_p-MUSIC only in GMM noise. Its RMSE cannot attain the CRB and it is inferior to the proposed method in GGD noise and α-stable noise. Still, it provides good performance in these two impulsive noises, which verifies that it is robust to PDF mismatch. MM-MUSIC and GZMNL-MUSIC also show good robustness to outlier but GZMNL-MUSIC suffers a performance saturation as GSNR or SNR increases. The GZMNL scheme generally destroys the low-rank structure of the signal subspace, which leads to a performance saturation or even degradation [12], [13].

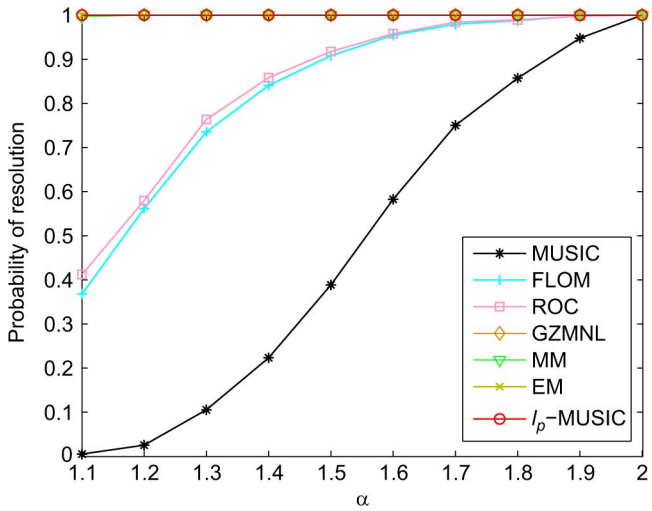


Fig. 8. Probability of resolution in SαS noise versus α at GSNR = 6 dB.

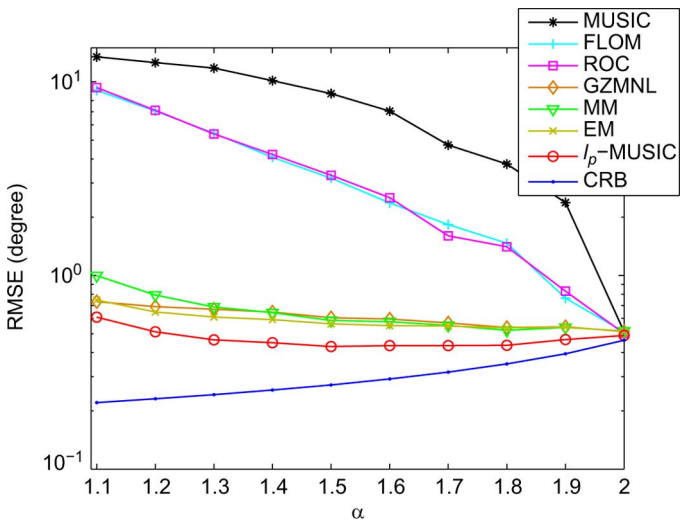


Fig. 9. RMSE of DOA estimate of the first source versus α.

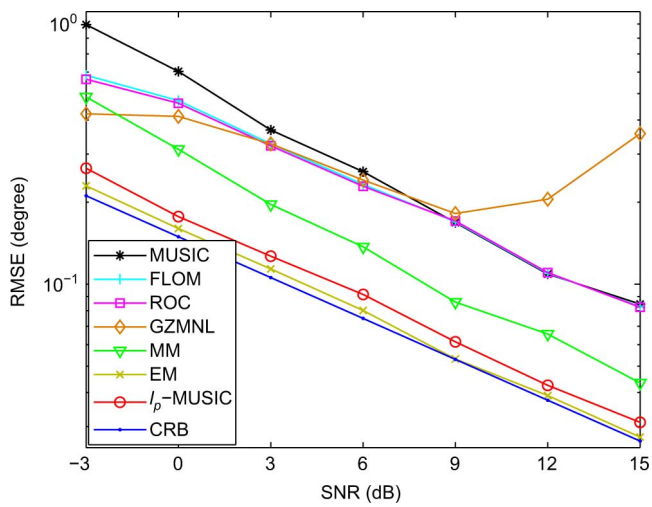


Fig. 10. RMSE of DOA estimate of the first source in GMM noise versus SNR.

Although the ROC-MUSIC and FLOM-MUSIC are superior to the conventional MUSIC, they are inferior to the ℓ_p-MUSIC,

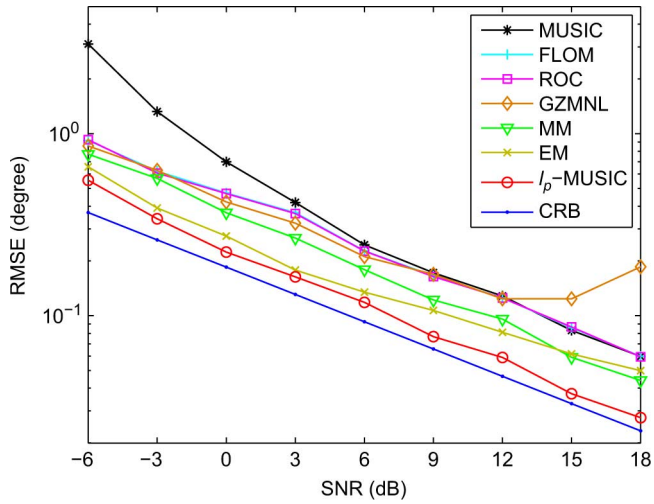


Fig. 11. RMSE of DOA estimate of the first source in GGD noise versus SNR with $\beta = 0.4$.

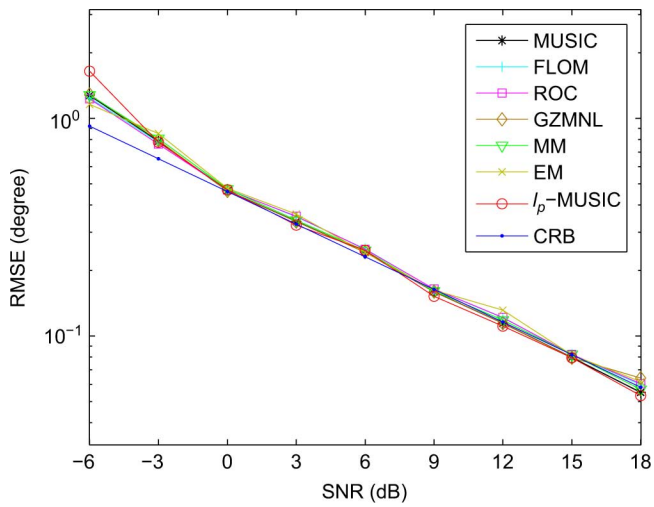


Fig. 12. RMSE of DOA estimate of the first source in Gaussian noise versus SNR.

MM-MUSIC and GZMNL-MUSIC. It has been analyzed that the performance of the fractional lower-order moment based algorithms is not satisfactory if the sample size is not large enough [12], [18].

In the last simulation, we investigate the performance of the robust DOA estimators under the Gaussian noise. The RMSE is shown in Fig. 12. It is demonstrated that all the methods have similar performance and the ℓ_p-MUSIC also has a high efficiency in the nominal Gaussian noise model.

V. Conclusion

We have proposed a family of algorithms called ℓ_p-MUSIC for DOA estimation in the presence of impulsive noise. The ℓ_p-MUSIC method replaces the Frobenius norm by the ℓ_p-norm of the residual error matrix, which results in a different subspace decomposition rule. Two efficient iterative algorithms, namely, IR-SVD and ACO, are developed to compute the signal subspace under the matrix ℓ_p-norm minimization framework. The ACO converts the original nonconvex subspace decomposition problem into a number of decoupled subproblems of convex vector ℓ_p-norm minimization and two complex-valued

Newton's methods have been developed to solve them. Interestingly, the widely used IRLS algorithm is a special case of the pseudo-Newton's method. Simulations results demonstrate that the ℓ_p-MUSIC estimator outperforms the conventional MUSIC and several existing robust subspace DOA methods in terms of robustness, resolution capability and estimation accuracy.

ACKNOWLEDGMENT

The authors are grateful to the three anonymous reviewers and the associate editor for their useful comments.

REFERENCES

[1] J. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 67–94, Jul. 1996.

[2] W.-J. Zeng and X.-L Li, "High-resolution multiple wideband and non-stationary source localization with unknown number of sources," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3125–3136, Jun. 2010.

[3] H. C. So, "Source localization: Algorithms and analysis," in *Handbook of Position Location: Theory, Practice and Advances*, S. A. Zekavat and M. Buehrer, Eds. New York, NY, USA: Wiley-IEEE Press, 2011.

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

[5] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1110–1121, May 1991.

[6] 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.

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

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

[9] K. L. Blackard, T. S. Rappaport, and C. W. Bostian, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Sele. Areas Commun.*, vol. 11, no. 7, pp. 991–1001, Sep. 1993.

[10] P. L. Brockett, M. Hinich, and G. R. Wilson, "Nonlinear and non-Gaussian ocean noise," *J. Acoust. Soc. Amer.*, vol. 82, no. 4, pp. 1386–1394, Oct. 1987.

[11] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1129–1149, May 1999.

[12] 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.

[13] R. J. Kozick and B. M. Sadler, "Robust subspace estimation in non-Gaussian noise," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Istanbul, Jun. 05–09, 2000, vol. 6, pp. 3818–3821.

[14] M. Novey, T. Adali, and A. Roy, "A complex generalized Gaussian distribution—Characterization, generation, and estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1427–1433, Mar. 2010.

[15] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. New York, NY, USA: Wiley, 1995.

[16] D. D. Lee and R. L. Kashyap, "Robust maximum likelihood bearing estimation in contaminated Gaussian noise," *IEEE Trans. Signal Process.*, vol. 40, pp. 1983–1986, Aug. 1992.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

[23] A. Swami and B. Sadler, "TDE, DOA and related parameter estimation problems in impulsive noise," in *Proc. IEEE Signal Process. Workshop Higher-Order Statist.*, Banff, Alta, Jul. 21–23, 1997, pp. 273–277.

[24] C.-H. Lim, S. C.-M. See, A. M. Zoubir, and B. P. Ng, "Robust adaptive trimming for high-resolution direction finding," *IEEE Signal Process. Lett.*, vol. 16, no. 7, pp. 580–583, Jul. 2009.

[25] P. J. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 2005.

[26] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, Jul. 2012.

[27] M. Muma, Y. Cheng, F. Roemer, M. Haardt, and A. M. Zoubir, "Robust source number enumeration for r-dimensional arrays in case of brief sensor failures," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 25–30, 2012, pp. 3709–3712, .

[28] M. Salibian-Barrera, S. Van Aelst, and G. Willems, "Principal components analysis based on multivariate MM estimators with fast and robust bootstrap," *J. Amer. Stat. Assoc.*, vol. 101, no. 475, pp. 1198–1211, Sep. 2006.

[29] G. Mateos and G. B. Giannakis, "Robust PCA as bilinear decomposition with outlier sparsity regularization," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5176–5190, Oct. 2012.

[30] F. Moghimi, A. Nasri, and R. Schober, "LP-norm spectrum sensing for cognitive radio networks impaired by non-Gaussian noise," in *Proc. GLOBECOM*, Honolulu, HI, USA, Nov. 30–Dec. 4, 2009, pp. 1–6, .

[31] A. Navia-Vazquez and J. Arenas-Garcia, "Combination of recursive least p-norm algorithms for robust adaptive filtering in alpha-stable noise," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1478–1482, Mar. 2012.

[32] S. Lee and J. Dickerson, "Least LP-norm interference suppression for DS/CDMA systems in non-Gaussian impulsive channels," in *Proc. Int. Conf. Commun. (ICC)*, Vancouver, BC, Canada, Jun. 6–10, 1999, vol. 2, pp. 907–911.

[33] R. A. Maronna and V. J. Yohai, "Robust low-rank approximation of data matrices with elementwise contamination," *Technometrics*, vol. 50, no. 3, pp. 295–304, Aug. 2008.

[34] V. J. Yohai, "High-breakdown point and high-efficiency robust estimates for regression," *Ann. Statist.*, vol. 15, no. 2, pp. 642–656, Jun. 1987.

[35] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gunturk, "Iteratively re-weighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.

[36] S. Valaee and P. Kabal, "An information theoretic approach to source enumeration in array signal processing," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1171–1178, 2004.

[37] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 3, pp. 387–392, Apr. 1985.

[38] Z. Lu, Y. Chakhchoukh, and A. M. Zoubir, "Source number estimation in impulsive noise environments using bootstrap techniques and robust statistics," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 22–27, 2011, pp. 2712–2715.

[39] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[41] B. D. Rao and K. V. S. Hari, "Performance analysis of root-MUSIC," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 1939–1949, Dec. 1989.

[42] D. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

[43] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Nat. Bur. Stand.*, vol. 49, no. 6, pp. 409–436, Dec. 1952.

[44] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, Jul. 1989.

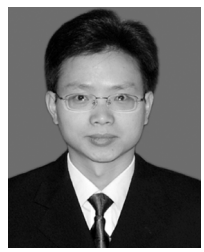
[45] B. M. Sadler, R. J. Kozick, and T. Moore, "Performance analysis for direction finding in non-Gaussian noise," *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5, pp. 2857–2860, 1999.

Wen-Jun Zeng (S'10–M'11) received the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2008.

From 2006 to 2009, he was a Research Assistant with Tsinghua University. From 2009 to 2011, he was a faculty member with the Department of Communication Engineering, Xiamen University, Xiamen, China. He is now a Senior Research Associate with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong. His research interests lie in the area of mathematical signal processing, including convex optimization, array processing, sparse approximation, and inverse problem, with applications to wireless radio and underwater acoustic communications.

H. C. So (S'90–M'95–SM'07) 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 worked as 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 an Associate Professor. His research interests include statistical signal processing, fast and adaptive algorithms, signal detection, parameter estimation, and source localization.

Dr. So has been on the editorial boards of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, *Signal Processing*, *Digital Signal Processing*, and *ISRN Applied Mathematics*, as well as a member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society.



Lei Huang (M'07) was born in Guangdong, China. He received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively.

From 2005 to 2006, he was a Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC. From 2009 to 2010, he was a Research Fellow with the Department of Electronic Engineering, City University of Hong Kong and a Research Associate with the Department of Electronic Engineering, The Chinese University of Hong Kong. Since 2011, he has joined the Department of Electronic and Information Engineering, Harbin Institute of Technology Shenzhen Graduate School, where he is currently a Professor. His research interests include spectral estimation, array signal processing, statistical signal processing, and their applications in radar and wireless communication systems. He currently is an editorial board member of Digital Signal Processing.