\(\ell_p\)-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 \(\ell_p\)-MUSIC, for direction-of-arrival (DOA) estimation in impulsive noise is proposed. The \(\ell_p\)-MUSIC estimator adopts the \(\ell_p\)-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 \(\ell_p\)-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 \(\ell_p\)-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 \(\ell_p\)-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—\(\ell_p\)-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].

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 \(\alpha\)-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 \(\alpha\)-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 minimizing 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 $\ell_p$-norm ($1 \leq p < 2$) of the residual fitting error matrix. The motivation to use the $\ell_p$-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 $\ell_p$-norm minimization for robust estimation [29]. The $\ell_p$-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 $\ell_p$-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 $\ell_p$-norm minimization by alternating convex optimization (ACO). In each iteration of the ACO, the matrix $\ell_p$-norm minimization is converted into a series of vector $\ell_p$-norm minimization subproblems. The most widely used approach to vector $\ell_p$-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 $\ell_p$-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 $\ell_p$-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 $\ell_p$-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 $\ell_p$-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 $\ell_p$-norm minimization framework.

iii) The convergence of the ACO algorithm is proved.

iv) The ACO converts the original matrix $\ell_p$-norm minimization problem into a series of decoupled subproblems of convex vector $\ell_p$-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 $\ell_p$-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 $m$th ($m = 1, \ldots, M$) sensor is expressed as

$$x_m(n) = \sum_{q=1}^{Q} s_q(n)e^{j2\pi(m-1)\sin(\theta_q)/\lambda} + e_m(n)$$

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

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

$$\mathbf{x}_n = \mathbf{A}s_n + \mathbf{e}_n$$

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

$$\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_Q)]$$

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

$$\mathbf{a}(\theta) = \left[1, e^{j2\pi \sin(\theta)/\lambda}, \ldots, e^{j2\pi(M-1)\sin(\theta)/\lambda}\right]^T.$$ (4)

Given $\mathbf{X} \triangleq [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{C}^{M \times N}$ 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 independent with each other, while the non-Gaussian noises 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 $\sigma^2$. The covariance matrix of the received signal $\mathbf{C}_x = E\{\mathbf{X}_n\mathbf{X}_n^H\}$ is of full rank. Similarly, $\mathbf{S}$ is of full rank. In the absence of noise or $\mathbf{e} = 0$, we have $\text{rank}(\mathbf{X}) = Q < M$ because $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{S}) = Q$. This means that the received data matrix $\mathbf{X}$ has the low-rank decomposition

$$\mathbf{X} = \mathbf{Y} \mathbf{Z}$$

where $\mathbf{Y} \in \mathbb{C}^{M \times Q}$ is a full rank matrix and $\mathbf{Z} \in \mathbb{C}^{Q \times N}$ is a full rank matrix. It is clear that the columns of $\mathbf{Y}$ span the same subspace of the array manifold matrix $\mathbf{A}$, that is,

$$\text{range}(\mathbf{Y}) = \text{range}(\mathbf{A}).$$

III. $\ell_p$-MUSIC ALGORITHM

A. $\ell_p$-norm based Subspace Decomposition Criterion

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

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{e}$$

where $\mathbf{S} = [s_1, \ldots, s_Q] \in \mathbb{C}^{Q \times N}$ and $\mathbf{e} = [e_1, \ldots, e_N] \in \mathbb{C}^{M \times N}$ are the source and noise data matrices, respectively. The array manifold matrix $\mathbf{A} \in \mathbb{C}^{M \times Q}$ is of full column rank. Similarly, $\mathbf{S}$ is of full row rank. In the absence of noise or $\mathbf{e} = 0$, we have $\text{rank}(\mathbf{X}) = Q < M$ because $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{S}) = Q$. This means that the received data matrix $\mathbf{X}$ has the low-rank decomposition

$$\mathbf{X} = \mathbf{Y} \mathbf{Z}$$

where $\mathbf{Y} \in \mathbb{C}^{M \times Q}$ is a full rank matrix and $\mathbf{Z} \in \mathbb{C}^{Q \times N}$ is a full rank matrix. It is clear that the columns of $\mathbf{Y}$ span the same subspace of the array manifold matrix $\mathbf{A}$, that is,

$$\text{range}(\mathbf{Y}) = \text{range}(\mathbf{A}).$$

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

$$J_p(\mathbf{Y}, \mathbf{Z}) = \|\mathbf{X} - \mathbf{Y} \mathbf{Z}\|_p^2 = \sum_{m=1}^{M} \sum_{n=1}^{N} \|x_{mn} - (\mathbf{Y} \mathbf{Z})_{mn}\|^p$$

where $\mathbf{X} - \mathbf{Y} \mathbf{Z}$ is the residual fitting error matrix, $x_{mn}$ and $(\mathbf{Y} \mathbf{Z})_{mn}$ denote the $(m, n)$ entries of $\mathbf{X}$ and $\mathbf{Y} \mathbf{Z}$, 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 \( z = z_R + jz_I \) 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 \( J_F(Y, Z) \) leads to the ML estimates of \( Y \) and \( Z \) if the noise is white Gaussian distributed. Although the Frobenius norm minimization of (13) with respect to (w.r.t.) \( Y \) and \( Z \) is nonconvex, its global minimum can be obtained through the SVD of \( X \) shown in (9) and the global optima of \( Y \) and \( Z \) are given by [39]

\[
Y = U_s, \quad Z = \Sigma_s V_s^H.
\] (14)

Note that the global optimum solution of (14) is not unique. For instance, \( U_s, T \) and \( T^{-1} \Sigma_s V_s^H \) are also the global optima of \( Y \) and \( Z \) for any nonsingular matrix \( T \). Despite the non-uniqueness of the optimal \( Y \), we are not interested in \( Y \) itself but the subspace spanned by its columns. The range space of the optimal \( Y \) 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 \( U_s \) 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 \( \ell_p \)-norm of the residual fitting error matrix instead of the Frobenius norm in (13). The \( \ell_p \)-norm \((1 \leq p \leq 2) \) based objective function

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

where the \( \ell_p \)-norm \( \| \cdot \|_p \) of a complex-valued matrix is defined as

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

When \( p = 2 \), (15) reduces to the Frobenius norm minimization criterion of (13). It is worth pointing out that the matrix \( \ell_p \)-norm defined in (16) is different from the widely used matrix induced norm or operator norm [39]. Minimization of the \( \ell_p \)-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 \( Y \) by minimizing \( J_p(Y, Z) \) in (15) since its columns span the signal subspace. If an estimate of \( Y \) is obtained, then its columns are adopted as a basis of the signal subspace. Unfortunately, the matrix \( \ell_p \)-norm minimization

\[
\min_{Y,Z} J_p(Y, Z) = \|X - YZ\|_p^p
\] (17)

w.r.t. \( Y \) and \( Z \) 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 \( \ell_p \)-norm minimization problem.

### B. Iteratively Reweighted SVD Algorithm

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

\[
J_p(Y, Z) = \|R\|_p^p = \sum_{m-1}^{M} \sum_{n-1}^{N} r_{mn}^p - \sum_{m-1}^{M} r_{mn}|^p \frac{1}{2} r_{mn}^\frac{1}{2}
\] (18)

where \( \circ \) denotes the element-wise multiplication and \( D \) is the weighting matrix with its \((m,n)\) entry \( d_{mn} = r_{mn}^{-1/2} \). Equation (18) means that the matrix \( \ell_p \)-norm minimization problem can be converted into a Frobenius norm minimization one. We can perform SVD to the weighted matrix \( D \circ X \) to obtain a better subspace estimation based on the \( \ell_p \)-norm minimization criterion. Note that the weighting matrix \( D \) depends on the unknowns \( Y \) and \( Z \), that is, it is a function of \( Y \) and \( Z \) which is written as \( D(Y, Z) \). Due to this reason, we cannot immediately obtain the optimal solution by performing the SVD of the weighted matrix \( D \circ X \) 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 \( k \)th 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 \( Y^{(0)} \) with a random matrix of full column rank and \( Z^{(0)} \) of full row rank.

for \( k = 0, 1, 2, \ldots \) do

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

Perform SVD:

\[
D^{(k)} \circ X = U_s^{(k)} \Sigma_s^{(k)} \left( V_s^{(k)} \right)^H + U_n^{(k)} \Sigma_n^{(k)} \left( V_n^{(k)} \right)^H.
\]

where \( U_s^{(k)} \in \mathbb{C}^{M \times Q} \) contains the first \( Q \) principal left singular vectors of \( D^{(k)} \circ X \).

Set \( Y^{(k)} \leftarrow U_s^{(k)} \) and \( Z^{(k)} \leftarrow \Sigma_s^{(k)} \left( V_s^{(k)} \right)^H \).

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 \( J_p(Y, Z) \) 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 $\mathbf{Y}$ through IR-SVD, the projection matrix onto the noise subspace is computed as

$$
P_n = \mathbf{I} - \mathbf{Y} \left( \mathbf{Y}^H \mathbf{Y} \right)^{-1} \mathbf{Y}^H.
$$

(19)

Then the $\ell_p$-MUSIC spatial spectrum of is given by

$$
P_{\ell_p,\text{MUSIC}}(\mathbf{\theta}) = \frac{1}{a^H(\mathbf{\theta}) P_n a(\mathbf{\theta})}.
$$

(20)

The DOA estimates can be obtained by searching for the peaks of (20). Certainly, in the $\ell_p$-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 $\mathbf{Y}$ or $\mathbf{Z}$, the error function $J_p(\mathbf{Y}, \mathbf{Z})$ 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 $\mathbf{Y}$ and $\mathbf{Z}$. Suppose we have obtained the two matrices $\mathbf{Y}^{(k)}$ and $\mathbf{Z}^{(k)}$ at the $k$th iteration. Then $\mathbf{Z}^{(k+1)}$ and $\mathbf{Y}^{(k+1)}$ can be solved by alternately optimizing the following two subproblems:

$$
\mathbf{Z}^{(k+1)} = \arg \min_{\mathbf{Z}} \left\| \mathbf{X} - \mathbf{Y}^{(k)} \mathbf{Z} \right\|^p_p,
$$

(21)

$$
\mathbf{Y}^{(k+1)} = \arg \min_{\mathbf{Y}} \left\| \mathbf{X} - \mathbf{Y} \mathbf{Z}^{(k+1)} \right\|^p_p.
$$

(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

$$
\min_{\mathbf{Y}} \left\| \mathbf{X} - \mathbf{Y} \mathbf{Z}^{(k+1)} \right\|^p_p = \min_{\mathbf{Y}} \left\| \mathbf{X}^T - \left( \mathbf{Z}^{(k+1)} \right)^T \mathbf{Y}^T \right\|^p_p
$$

(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 $\ell_p$-norm error function in (21) can be further expressed as

$$
\left\| \mathbf{X} - \mathbf{Y}^{(k)} \mathbf{Z} \right\|^p_p = \sum_{n=1}^N \left\| \mathbf{z}_n - \mathbf{Y}^{(k)} \mathbf{z}_n \right\|^p_p.
$$

(24)

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

$$
\mathbf{z}_n^{(k+1)} = \arg \min_{\mathbf{z}_n} \left\| \mathbf{z}_n - \mathbf{Y}^{(k)} \mathbf{z}_n \right\|^p_p, \quad n = 1, \ldots, N
$$

(25)

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

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

(26)

Now the matrix $\ell_p$-norm minimization of (21) is converted into $N$ subproblems of vector $\ell_p$-norm minimization. The last remaining problem is to solve the vector $\ell_p$-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 $\ell_p$-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 $k$th iteration of the alternating algorithm as $J_p \left( \mathbf{Y}^{(k)}, \mathbf{Z}^{(k)} \right)$. From (21) and (22), it follows that

$$
J_p \left( \mathbf{Y}^{(k+1)}, \mathbf{Z}^{(k+1)} \right) \leq J_p \left( \mathbf{Y}^{(k)}, \mathbf{Z}^{(k+1)} \right) \leq J_p \left( \mathbf{Y}^{(k)}, \mathbf{Z}^{(k)} \right).
$$

(27)

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

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:

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

(28)

for some small tolerance $\varepsilon$. 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
but also it more rapidly decreases the objective function value
than the latter. It is noticed that the ACO procedure does not nec-
essarily converge to the global minimum point. The point that
it converges to depends on the initial value
. We can ini-
tialize 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 spans the same range space of
Here the normalized subspace distance between and is
used to evaluate the quality of subspace estimate, which is de-

defined as

\[
d_p(Y, A) = \frac{\| P_Y - P_A \|}{\| P_A \|} (29)
\]

where and are the projection matrices onto the column spaces of and
respectively. Again, we can take different values of . The definition of (29) aligns
with [39] for but here we allow other values of . The
subspace distance will become zero if \(
\text{range}(Y) = \text{range}(A)\)
exactly holds. A smaller subspace distance means a better sub-
space 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 guar-
anteed.

D. Complex-Valued Newton’s Methods for Vector \(\ell_p\)-Norm

Minimization

This subsection addresses the solver for the last remaining
problem, namely, the complex-valued vector \(\ell_p\)-norm mini-
mization 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

\[
\min_z f(z) = \| z - Yz \|^p_p (30)
\]

where \(Y \in \mathbb{C}^{M \times Q}, z \in \mathbb{C}^M\) and \(z \in \mathbb{C}^Q\).

1) Gradient and Hessian of \(\ell_p\)-norm: By defining the
residual vector

\[
r = Yz - x = [r_1, \ldots, r_M]^T, (31)
\]

the objective function \(f(z)\) can be expressed as

\[
f(z) = f(r) = \|r\|^p_p = \sum_{i=1}^M |r_i|^p. (32)
\]

It is not difficult to derive the partial derivative w.r.t. the com-
plex quantity \(r_i\) as

\[
\frac{\partial f}{\partial r_i^*} = \frac{1}{2} \left( \frac{\partial f}{\partial \text{Re}(r_i)} \right) + \frac{1}{2} \left( \frac{\partial f}{\partial \text{Im}(r_i)} \right) = -\frac{p}{2} r_i^{p-2} r_i^*, \quad i = 1, \ldots, M (33)
\]

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

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

with \(|r|^p = [r_1^{p-2}, \ldots, r_M^{p-2}]^T\). Define a diagonal matrix

\[
W = \text{diag} \left\{ |r_1|^{p-2}, \ldots, |r_M|^{p-2} \right\}. (35)
\]

Equation (34) can be rewritten as

\[
\frac{\partial f}{\partial r^*} = \frac{p}{2} W^2 r. (36)
\]

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

We compute the second-order partial derivative as

\[
\frac{\partial^2 f}{\partial r_i^* \partial r_j} = \begin{cases} p^2 |r_i|^p /4, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} (37)
\]
Hence the $M \times M$ partial Hessian matrix of $f$ w.r.t. $r$, denoted by $H_{r^r}$, is diagonal and has the form:

$$H_{r^r} = \frac{\partial^2 f}{\partial r^r \partial r^r} = \frac{p^2}{4} W.$$ (38)

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

$$\frac{\partial r}{\partial z^*} = Y^*.$$ (39)

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

$$g(z) = \frac{\partial f}{\partial z^*} = \left( \frac{\partial r}{\partial z^*} \right)^T \frac{\partial f}{\partial r^r} = \frac{p^2}{4} Y^H W(Y z - x).$$ (40)

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

$$H_{z^z} = \frac{\partial^2 f}{\partial z^* \partial z^*} = \left( \frac{\partial r}{\partial z^*} \right)^T H_{r^r} \frac{\partial r}{\partial z^*} = -Y^H H_{r^r} Y = \frac{p^2}{4} Y^H W(z) Y.$$ (41)

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

$$H = \begin{bmatrix} H_{z^z} & H_{z^z^*} \\ H_{z^z^*} & H_{z^z} \end{bmatrix}.$$ (42)

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

$$H_{z^z^*} = \frac{\partial^2 f}{\partial z^* \partial z^*} = Y^H H_{r^r} Y^*$$ (43)

where

$$H_{r^r} = \frac{\partial^2 f}{\partial r^r \partial r^r} = \frac{p(p-2)}{4} \text{diag} \{ r_1^p, r_2^p, \ldots, r_M^p \}$$ (44)

is a diagonal matrix, $H_{zz} = H_{z^z}^*$, and $H_{zz^*} = H_{z^z^*}^*$. It is noticed that the two off-diagonal block matrices $H_{z^z^*}^*$ and $H_{zz}$ 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 $\{z^{(k)}\}$ $\{k = 0, 1, \ldots\}$ through the following iteration

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

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

$$\Delta z^{(k)} = -H_{z^z^*}^{-1} g(z^{(k)})$$ (46)

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

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

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

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

Proposition 2: The widely used IRLS approach to vector $\ell_p$-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 $\Delta z^{(k)}$, the optimal step size is given by solving the line search

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

Denoting the residual vector in the $k$th iteration as

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

the objective function w.r.t. the step size $\mu$ can be written as

$$\min_{\mu \geq 0} f(\mu) = \left\| r^{(k)} + \mu Y \Delta z^{(k)} \right\|_p.$$ (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 $\mu$ is guaranteed since $f(\mu)$ is unimodal w.r.t. $\mu$ if $p > 1$. 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

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

and the updating rule is

$$\begin{bmatrix} z^{(k+1)} \\ (z^{(k+1)})^* \end{bmatrix} = \begin{bmatrix} z^{(k)} \\ (z^{(k)})^* \end{bmatrix} + \mu_k \begin{bmatrix} \Delta z^{(k)} \\ (\Delta z^{(k)})^* \end{bmatrix}$$ (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 $\mathbf{z}^{(0)} = \hat{\mathbf{z}}_{\text{LS}} = (\mathbf{Y}^H \mathbf{Y})^{-1} \mathbf{Y}^H \mathbf{x}$.

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 $\mathbf{H}_{\mathbf{x}^*} \Delta \mathbf{z}^{(k)} = -g(\mathbf{z}^{(k)})$ to obtain pseudo-Newton step $\Delta \mathbf{z}^{(k)}$.

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 $\mathbf{Y} \in \mathbb{C}^{30 \times 20}$ and vector $\mathbf{x} \in \mathbb{C}^{50}$. We are primarily interested in the behavior, as a function of the number of iterations, of the relative error $\epsilon$, where $\mathbf{z}^{(k)}$ 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 $\mathbf{Y}^H \mathbf{W}(\mathbf{z}) \mathbf{Y}$ is $O(MQ^2)$ because $\mathbf{W}(\mathbf{z})$ 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 $O(Q^3)$. Hence the complexity of the Newton’s methods is $O(MQ^2)$ 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 $O(NMQ^2)$, which is also the complexity of the second convex optimization problem in (22). Therefore the complexity of the ACO approach in each iteration is $O(NMQ^2)$.

**E. Summary of $\ell_p$-MUSIC Algorithm**

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

**Algorithm 2:** $\ell_p$-MUSIC

**Input:** $M \times N$ received data matrix $\mathbf{X}$, number of sources $Q$ and tolerance $\epsilon$.

**Output:** DOA estimates.

1. Initialize $\mathbf{Y}^{(0)}$: with a random matrix of full column rank.
2. for $k = 1, 2, \ldots$, do until converge
   1) Calculate $\mathbf{Z}^{(k+1)} = \arg \min_{\mathbf{z} \in \mathbb{C}^{Q \times N}} \| \mathbf{X} - \mathbf{Y}^{(k)} \mathbf{Z} \|^p_p$
   2) Use the Newton’s method to solve the following $N$ subproblems:
      $\mathbf{z}^{(k+1)} = \arg \min_{\mathbf{z} \in \mathbb{C}^{Q \times N}} \| \mathbf{x} - \mathbf{Y}^{(k)} \mathbf{z} \|^p_p$
   3) Solve the second minimization problem
      $\mathbf{Y}^{(k+1)} = \arg \min_{\mathbf{Y} \in \mathbb{C}^{M \times Q}} \| \mathbf{X} - \mathbf{Y} \mathbf{Z}^{(k+1)} \|^p_p$

3. Compute the projection matrix onto the noise subspace $\mathbf{P}_n = I - \mathbf{Y} (\mathbf{Y}^H \mathbf{Y})^{-1} \mathbf{Y}^H$, where $\mathbf{Y}$ 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 $\ell_p$-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 $\beta$ in (55) since $p = \beta$ leads to ML estimate of the signal subspace. For $\alpha$-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 \( \ell_p \)-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 as in the \( \ell_p \)-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 intersensor spacing of \( d = \lambda / 2 \). The emitting sources are two in the \( A \)-space.

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 \( \alpha \)-stable distribution, are considered.

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

\[
p_{\ell}(\ell) = \sum_{i=1}^{2} c_{i} \pi \sigma_{i}^{2} \exp \left( -\frac{|\ell|^{2}}{\sigma_{i}^{2}} \right)
\]

where \( 0 \leq c_{i} \leq 1 \) and \( \sigma_{i}^{2} \) are the probability and variance of the \( i \)th 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 \( \sigma_{2}^{2} \) occurring with a smaller probability \( c_{2} \) can be viewed as outliers embedded in Gaussian background noise of variance \( \sigma_{1}^{2} \). Therefore the GMM can well model the phenomenon in the presence of both Gaussian thermal noise and impulsive noise. The total variance of \( n(\ell) \) is \( \sigma_{\ell}^{2} = \sum_{i=1}^{2} c_{i} \sigma_{i}^{2} \) and the SNR is defined as SNR = \( E\{\|s(\ell)\|^2\} / \sigma_{\ell}^{2} \) with \( E\{\|s(\ell)\|^2\} \) 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 \( \sigma_{\ell}^{2} \) is

\[
p_{\ell}(\ell) = \frac{\beta \Gamma(4/\beta)}{2\pi \sigma_{\ell}^{2} \Gamma(2/\beta)} \exp \left( -\frac{|\ell|^{2}}{\sigma_{\ell}^{2}} \right)
\]

where \( \beta > 0 \) is the shape parameter, \( \Gamma(\cdot) \) is the Gamma function and \( \epsilon = (\Gamma(2/\beta)/\Gamma(4/\beta))^{1/2} \). 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 \( \beta \), the more impulsive the noise is. We take \( \beta = 0.4 \) in the simulation.

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

\[
\phi(\omega) = \exp \left( -\gamma^\alpha |\omega|^\alpha \right)
\]

where \( 0 < \alpha \leq 2 \) is called the characteristic exponent that describes the tail of the distribution and \( \gamma > 0 \) is the scale. When \( \alpha = 2 \), the \( \alpha \)-stable distribution reduces to the Gaussian distribution and \( \gamma^2 \) is similar to the variance of the Gaussian distribution. When \( \alpha = 1 \), the \( \alpha \)-stable distribution becomes the Cauchy distribution. When \( \alpha < 2 \), \( \alpha \)-stable noise shows heavy tails and hence is impulsive. The smaller the value of \( \alpha \), the more impulsive the noise is. Although the characteristic function of the S\( \alpha \)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 \( \alpha \)-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) is used.

\[
\text{GSNR} = E\{\|s(\ell)\|^2\} / \gamma^\alpha
\]

is adopted.

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

\[
\text{CRB}(\theta) = \frac{1}{n} \text{diag} \left\{ \sum_{n=1}^{N} \text{Re} \left( S_n^H \Xi(\mu(\theta)) P_A^* \Xi(\mu(\theta^*)) S_n \right) \right\}
\]

where \( S_n = \text{diag}\{s_1(n), \ldots, s_Q(n)\} \) is a diagonal matrix, \( \Xi(\mu) = \{\xi(\mu_1), \ldots, \xi(\mu_Q)\} \) with \( \xi(\mu) = \text{diag}(\mu)/\mu^2 \), \( P_A = I - P_A^* \) is the projection onto the orthogonal complementary space of \( A \), and

\[
I_c = \pi \int_{0}^{\infty} \frac{p_c(\rho)^2}{p_e(\rho)} \rho d\rho
\]

with \( p = e \) being the modulus of the complex variable \( e \) and \( p'_c(\rho) \) the derivative of \( p_c(\rho) \). The noise PDF affects the CRB with the PDF mismatch. We consider a ULA with intersensor spacing of \( d = \lambda / 2 \). In all simulations, we use the ACO rather than the IR-SVD since ACO obtains a better estimate of the signal subspace.
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 $\ell_p$-MUSIC, ROC-MUSIC and FLOM-MUSIC methods. The DOAs of the two incoming signals are $\theta_1 = -8^\circ$ and $\theta_2 = 8^\circ$. 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)

$$\text{RMSE} \left( \hat{\theta}_q \right) = \sqrt{\frac{1}{M_s} \sum_{i=1}^{M_s} \left( \hat{\theta}_{q,i} - \theta_q \right)^2}$$

(60)

where $M_s$ is the number of Monte Carlo trials and $\hat{\theta}_{q,i}$ is the DOA estimate of the $q$th source in the $i$th 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 $5^\circ$, 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 $\alpha S\alpha$ noise, performance versus GSNR and $\alpha$ is investigated. At each GSNR and $\alpha$, $M_s = 2000$ 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 $\ell_p$-norm minimization with $p < 2$ leads to an improved performance. In Gaussian noise or $\alpha$, the performance of the $\ell_p$-MUSIC estimator is comparable to other methods. The $\ell_p$-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 $\ell_p$-MUSIC only in GMM noise. Its RMSE cannot attain the CRB and it is inferior to the proposed method in GGD noise and $\alpha$-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].
Although the ROC-MUSIC and FLOM-MUSIC are superior to the conventional MUSIC, they are inferior to the $\ell_p$-MUSIC, 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 $\ell_p$-MUSIC also has a high efficiency in the nominal Gaussian noise model.

V. CONCLUSION

We have proposed a family of algorithms called $\ell_p$-MUSIC for DOA estimation in the presence of impulsive noise. The $\ell_p$-MUSIC method replaces the Frobenius norm by the $\ell_p$-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 $\ell_p$-norm minimization framework. The ACO converts the original nonconvex subspace decomposition problem into a number of decoupled subproblems of convex vector $\ell_p$-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 $f_r$-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


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.

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.

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.