

Article

Direction of Arrival Estimation in Elliptical Models via Sparse Penalized Likelihood Approach

Chen Chen ¹, Jie Zhou ^{1,*}  and Mengjiao Tang ²

¹ College of Mathematics, Sichuan University, Chengdu 610064, China; chenchen_uni@foxmail.com

² Center for Information Engineering Science Research, School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; qingqing_tmj@126.com

* Correspondence: jzhou@scu.edu.cn; Tel.: +86-28-8541-5060

Received: 4 May 2019; Accepted: 20 May 2019; Published: 22 May 2019

Abstract: In this paper, an l_1 -penalized maximum likelihood (ML) approach is developed for estimating the directions of arrival (DOAs) of source signals from the complex elliptically symmetric (CES) array outputs. This approach employs the l_1 -norm penalty to exploit the sparsity of the gridded directions, and the CES distribution setting has a merit of robustness to the uncertainty of the distribution of array output. To solve the constructed non-convex penalized ML optimization for spatially either uniform or non-uniform sensor noise, two majorization-minimization (MM) algorithms based on different majorizing functions are developed. The computational complexities of the above two algorithms are analyzed. A modified Bayesian information criterion (BIC) is provided for selecting an appropriate penalty parameter. The effectiveness and superiority of the proposed methods in producing high DOA estimation accuracy are shown in numerical experiments.

Keywords: direction of arrival (DOA); complex elliptically symmetric (CES) distributions; sparse penalized likelihood method; majorization-minimization (MM) algorithm

1. Introduction

Estimating the directions of arrival (DOAs) of a number of far-field narrow-band source signals is an important problem in signal processing. Many DOA estimation methods were proposed early on, such as multiple signal classification (MUSIC) [1], estimation of signal parameters via rotation invariance techniques (ESPRIT) [2] and their variants [3,4]. Many of them work well when having accurate estimates of the array output covariance matrix and source number. In scenarios with sufficient array snapshots and a moderately high signal-to-noise (SNR), the array output covariance matrix and source number can be accurately estimated.

Recently, some sparse DOA estimation methods are popularly proposed based on sparse constructions of the array output model. They are applicable if the number of sources is unknown and many of them are effective when with a limited number of data snapshots. In [5], the sparse DOA estimation methods are categorized into three groups: the on-grid, the off-grid and the grid-less. The on-grid method, which is widely studied and straightforward to implement, assumes that the true DOAs are on a predefined grid [6–22]. The off-grid method also uses a prior grid but does not constrain the DOA estimates to be on this grid, while it introduces more unknown parameters to be estimated and complicates the algorithm [23–28]. The grid-less method directly operates in the entire direction domain without a pre-specified grid but, currently, is developed mainly for linear array and encounters rather large computational burden [29–31]. Although the on-grid methods may induce the grid mismatch, it is still attractive due to its easy accessibility in a wide range of applications (see, e.g., [28,32,33]). To alleviate the grid mismatch, methods for selecting an appropriate prior grid are proposed in [9,34].

During the past two decades, various on-grid sparse techniques, such as linear least-square, covariance-based and maximum-likelihood (ML) type methods, were researched. The linear least-square methods [6,9,11,12], minimizing the l_2 -norm of the noise residual of the deterministic model, enforce the sparsity by constraining the l_p -norm of the signal vector for $p \in [0, 1]$ (note that l_p -norm for $p \in [0, 1]$ is defined as the same form as the one for $p \geq 1$ but is not truly a norm in mathematics). The covariance-based methods, such as the sparse iterative covariance-based estimation (SPICE) method [13,14] and its variants [15,16], and those proposed in [17–22], are derived from different covariance-fitting criteria and use the l_p -norm penalty to enforce the sparsity. The ML type methods, including the sparse Bayesian learning (SBL) methods [7,8,10] and the likelihood-based estimation of sparse parameters (LIKES) method [15,16], are deduced under the assumption that the array output signal is multivariate Gaussian distributed. The SBL methods use different prior distributions to model the sparsity. The LIKES method appears to be not explicitly utilizing the sparsity, but provides sparse estimates.

The ML principle is generally believed to be statistically more sound than the covariance-fitting principle [15]. The existing on-grid ML methods are developed on the Gaussian distribution assumption that are not satisfied in many signal processing practical applications. The complex elliptically symmetric (CES) distributions, containing the t -distribution, the K -distribution, the Gaussian distribution and so on, can be used to characterise the Gaussian and non-Gaussian random variables. Particularly, the heavy-tailed data, which usually appear in the field of radar and array signal processing [35,36], can be modeled by some CES distributions.

The l_p -norm penalty with $p \in [0, 1]$ is well-known for its contributions in deriving sparse solutions [37]. In particular, the convex l_1 -norm penalty, also known as the least absolute shrinkage and selection operator (LASSO) penalty, is easy to handle and widely used in various applications. Generally, a sparse and accurate estimate of the unknown sparse parameter can be expected when minimizing the negative likelihood plus an l_p -norm penalty.

For estimating the DOAs from the CES distributed array outputs, we provide an l_1 -norm penalized ML method based on a sparse reconstruction of the array output covariance matrix. The characteristics and advantages of our method are explained as follows:

1. Our method is a sparse on-grid method based on the penalized ML principle, and is designed especially for the CES random output signals. The Gaussian distribution and many heavy-tailed and light-tailed distributions are included in the class of the CES distributions, and it is worth mentioning that the heavy-tailed output signals that are non-Gaussian are common in the field of signal processing [35,36].
2. The sparsity of the unknown sparse vector is enforced by penalizing its l_1 -norm. When the penalty parameter becomes zero, the proposed method is actually the general ML DOA estimation method that is applicable to the scenarios with CES random output signals.
3. Two penalized ML optimization problems are formulated for spatially uniform and non-uniform white sensor noise, respectively. Since it is difficult to solve the two non-convex penalized ML optimizations globally, two majorization-minimization (MM) algorithms having different iterative procedures are developed for seeking the optimal solutions locally for each of them. Some discussions on the computational complexities of the above two algorithms are provided. In addition, the optimal penalty parameter is suggested.
4. The proposed methods are evaluated numerically in scenarios with Gaussian and non-Gaussian output signals. Particularly, the performance gains originated from the added l_1 -norm penalty are numerically demonstrated.

The remainder of this paper is organized as follows. Section 2 introduces a sparse CES data model of the array output. In Section 3, a sparse penalized ML method is developed to estimate the DOAs. For solving the proposed l_1 -penalized ML optimizations that are non-convex, algorithms in the MM framework are developed in Section 4. Section 5 numerically shows the performance of our method in Gaussian and non-Gaussian scenarios. Finally, some conclusions are given in Section 6.

Notations

The notation $\mathbb{C}^m(\mathbb{R}^m)$ denotes the set of all m -dimensional complex- (real-) valued vectors, and $\mathbb{C}^{m \times n}(\mathbb{R}^{m \times n})$ denotes the set of all $m \times n$ complex- (real-) valued matrices. $\mathbf{1}_p$ and $\mathbf{0}_p$ are the p -dimensional vectors with all elements equal to 1 and 0, respectively. \mathbf{I}_p is the $p \times p$ identity matrix. $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the l_1 -norm and l_2 -norm of a vector, respectively. The superscripts $(\cdot)^T$ and $(\cdot)^H$, respectively, denote the transpose and the conjugate transpose of a vector or matrix. The imaginary unit is denoted as i defined by $i^2 = -1$.

For a vector $\mathbf{x} = [x_1, \dots, x_p]^T \in \mathbb{C}^p$, define element-wise square root function $\text{sqrt}(\mathbf{x}) = [\sqrt{|x_1|}, \dots, \sqrt{|x_p|}]^T$ and element-wise division operation $\mathbf{x} \oslash \mathbf{y} = [x_1/y_1, \dots, x_p/y_p]^T$ for a vector $\mathbf{y} = [y_1, \dots, y_p]^T \in \mathbb{C}^p$ with non-zero elements. $\max(\mathbf{x}) = \max\{x_1, \dots, x_p\}$, $\mathbf{x}_{[a:b]} = [x_a, \dots, x_b]^T$ for $1 \leq a \leq b \leq p$, $\mathbf{x} > 0$ (≥ 0) means $x_i > 0$ (≥ 0) for $i = 1, \dots, p$, $[\mathbf{x}; \mathbf{z}]$ denotes the stacked vector of the column vectors \mathbf{x} and \mathbf{z} . $\text{Diag}(\mathbf{x})$ denotes a square diagonal matrix with the elements of vector \mathbf{x} on the main diagonal, and $(\text{Diag}(\mathbf{x}))^{-1}$ denotes the diagonal matrix with main diagonal elements $1/x_i$ ($i = 1, \dots, p$) by taking $1/0 = \infty$.

For a square matrix $\mathbf{X} \in \mathbb{C}^{p \times p}$, $\mathbf{X}_{i,j}$ denotes the (i, j) th entry of \mathbf{X} , $\mathbf{X} > 0$ (≥ 0) means that \mathbf{X} is Hermitian and positive definite (semidefinite), $\text{tr}(\mathbf{X})$ denotes the trace of \mathbf{X} , and $\text{diag}(\mathbf{X})$ denotes a column vector of the main diagonal elements of \mathbf{X} .

2. Problem Formulation

Consider the problem of estimating the DOAs of k_0 narrow-band signals impinging on an array of m sensors.

Given a set of grid points

$$\{\theta_1, \dots, \theta_k\}, \quad (1)$$

we assume that the true k_0 ($k_0 \ll k$) DOAs, respectively, denoted as ξ_1, \dots, ξ_{k_0} , take values in it. The array output measurement at the instant t , denoted as $\mathbf{x}(t) \in \mathbb{C}^m$, can be modeled as

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

where

- $\boldsymbol{\theta} = [\theta_1, \dots, \theta_k]^T$,
- $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_k)] \in \mathbb{C}^{m \times k}$ is the known array manifold matrix with $\mathbf{a}(\theta_i)$ being the steering vector corresponding to θ_i , $i = 1, \dots, k$;
- $\mathbf{s}(t) = [s_1(t), \dots, s_k(t)]^T \in \mathbb{C}^k$ is the source signal vector at the time instant t , in which $s_i(t)$ is the unknown random signal from a possible source at θ_i and then is zero if θ_i is not in the true DOA set $\{\xi_1, \dots, \xi_{k_0}\}$, $i = 1, \dots, k$;
- $\mathbf{v}(t) = [v_1(t), \dots, v_m(t)]^T \in \mathbb{C}^m$ is the noise vector impinging on the sensor array at the time instant t .

Some necessary statistical assumptions are made as follows:

- The possible source signals $s_1(t), \dots, s_k(t)$ are uncorrelated and zero-mean at any time instant t .
- The noise components $v_1(t), \dots, v_m(t)$ are uncorrelated, zero-mean, and independent of $s_1(t), \dots, s_k(t)$ for any time instant t .
- The n ($n > m$) snapshots $\mathbf{x}(1), \dots, \mathbf{x}(n)$ of sensor array signals are independent and identically distributed from a CES distribution.

Note that the zero-mean assumptions above are common in the signal processing literature [5,35]. The CES distribution setting of the array output $\mathbf{x}(t)$ enables us to effectively process the Gaussian, heavy-tailed or light-tailed array snapshots, because the class of CES distributions [38] includes the Gaussian distribution, the t -distribution, the K -distribution and so on.

For the simplicity of the notations, we denote $\mathbf{A}(\boldsymbol{\theta})$ as \mathbf{A} and $\mathbf{a}(\theta_i)$ as \mathbf{a}_i for $i = 1, \dots, k$ in the following. Under the above assumptions, we can find that the array output $\mathbf{x}(t)$ in Equation (2) at any time instant t has mean zero and covariance matrix

$$\mathbf{R} = \mathbb{E}[\mathbf{x}(t)\mathbf{x}(t)^H] = \mathbf{A}\mathbf{P}\mathbf{A}^H + \mathbf{V}, \quad (3)$$

where

$$\mathbf{P} = \mathbb{E}[\mathbf{s}(t)\mathbf{s}(t)^H] = \text{Diag}(\mathbf{p}) \quad (4)$$

is the (unknown) source signal covariance matrix with signal power vector $\mathbf{p} = [p_1, \dots, p_k]^T$, and

$$\mathbf{V} = \mathbb{E}[\mathbf{v}(t)\mathbf{v}(t)^H] = \text{Diag}(\boldsymbol{\sigma}) \quad (5)$$

is the (unknown) noise covariance matrix with the noise power vector $\boldsymbol{\sigma} = [\sigma_1^2, \dots, \sigma_m^2]^T$. The matrix \mathbf{R} can be rewritten as

$$\mathbf{R} = [\mathbf{A}, \mathbf{I}_m] \text{Diag}([\mathbf{p}; \boldsymbol{\sigma}]) [\mathbf{A}, \mathbf{I}_m]^H. \quad (6)$$

For any $i = 1, \dots, k$, the signal power $p_i = 0$ if θ_i is not in the set of the true DOAs. Therefore, the true DOAs can be identified by the locations of nonzero elements of the power vector \mathbf{p} . In the following, the DOA estimation problem is formulated as a problem of estimating the locations of nonzero elements of the power vector \mathbf{p} .

3. Sparse DOA Estimation

The array output $\mathbf{x}(t)$ in Equation (2) is CES distributed with mean zero and covariance matrix \mathbf{R} , and then the normalized random vector

$$\mathbf{y}(t) \triangleq \mathbf{x}(t) / \|\mathbf{x}(t)\|_2, \quad (7)$$

which actually refers to the angle of the array output vector $\mathbf{x}(t)$, has a complex angular central Gaussian (ACG) distribution [35] with the probability density function

$$p(\mathbf{y}(t); \mathbf{R}) \propto \det(\mathbf{R})^{-1} (\mathbf{y}(t)^H \mathbf{R}^{-1} \mathbf{y}(t))^{-m}. \quad (8)$$

Denote

$$L_0(\mathbf{R}) = \log \det(\mathbf{R}) + \frac{m}{n} \sum_{t=1}^n \log(\mathbf{x}(t)^H \mathbf{R}^{-1} \mathbf{x}(t)), \quad (9)$$

and then the negative log-likelihood of $\mathbf{y}(t) = \mathbf{x}(t) / \|\mathbf{x}(t)\|_2$, $t = 1, \dots, n$, becomes

$$nL_0(\mathbf{R}) - m \sum_{t=1}^n \log(\|\mathbf{x}(t)\|_2^2) + c_1, \quad (10)$$

where c_1 is a constant.

3.1. Spatially Non-Uniform White Noise

In the case of spatially non-uniform white sensor noise, not all noise variances $\sigma_i^2, i = 1, \dots, m$, are equal. Assuming $\sigma_m^2 > 0$, we denote

$$\mathbf{r} = [\mathbf{p}; \sigma_{[1:m-1]}] / \sigma_m^2 \quad (11)$$

$$\mathbf{W} = \mathbf{B} \text{Diag}(\mathbf{r}) \mathbf{B}^H + \mathbf{J}_m, \quad (12)$$

where $\mathbf{B} \in \mathbb{C}^{m \times (k+m-1)}$ is the first $(k+m-1)$ columns of $[\mathbf{A}, \mathbf{I}_m]$ and \mathbf{J}_m is an $m \times m$ matrix with the (m, m) -entry being 1 and the other entries being 0. Since

$$\mathbf{W} = \mathbf{R} / \sigma_m^2, \quad (13)$$

$$L_0(a\mathbf{R}) = L_0(\mathbf{R}), \forall a > 0 \quad (14)$$

and that the locations of nonzero elements of the sparse vector $\mathbf{r}_{[1:k]}$ identify the true DOAs, we formulate the DOA estimation problem as solving the penalized likelihood optimization problem

$$\arg \min_{\mathbf{r} \in \Omega} L_0(\mathbf{W}) + \lambda \|\mathbf{r}_{[1:k]}\|_1, \quad (15)$$

where

$$\Omega = \{\mathbf{t} \in \mathbb{R}^{k+m-1} \mid \mathbf{t} \geq 0, \mathbf{B} \text{Diag}(\mathbf{t}) \mathbf{B}^H + \mathbf{J}_m > 0\} \quad (16)$$

and $\lambda \geq 0$ is pre-specified. The l_1 -norm penalty term $\lambda \|\mathbf{r}_{[1:k]}\|_1$ would help deduce a sparse solution, and the penalty parameter λ controls the sparsity level [37].

Remark 1. We explain why the DOAs are not estimated by solving the plausible l_1 -penalized ML optimization problem

$$\arg \min_{\mathbf{p}, \sigma} L_0(\mathbf{R}) + \lambda \|\mathbf{p}\|_1. \quad (17)$$

Recalling Equation (6), we find for any $\lambda_1 > 0$ and $\lambda_2 > 0$,

$$L_0(\mathbf{R}) + \lambda_1 \|\mathbf{p}\|_1 = L_0 \left([\mathbf{A}, \mathbf{I}_m] \text{Diag} \left(\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \mathbf{p}; \frac{\lambda_1}{\lambda_2} \sigma \right) [\mathbf{A}, \mathbf{I}_m]^H \right) + \lambda_2 \left\| \frac{\lambda_1}{\lambda_2} \mathbf{p} \right\|_1. \quad (18)$$

Thus, $[\mathbf{p}, \sigma]$ is a locally optimal solution of Equation (17) with $\lambda = \lambda_1$ if and only if $\left[\frac{\lambda_1}{\lambda_2} \mathbf{p}; \frac{\lambda_1}{\lambda_2} \sigma \right]$ is a locally optimal solution of Equation (17) with $\lambda = \lambda_2$. This means if we estimate \mathbf{p} by Equation (17), the parameter λ cannot work theoretically in adjusting the sparsity level of the estimate of \mathbf{p} . Instead of considering Equation (17), we formulate the optimization (Equation (15)) for the DOA estimation. In Equation (15), noticing the constant matrix \mathbf{J}_m in Equation (12), we find that different values of λ would result in solutions with different sparsity levels.

3.2. Spatially Uniform White Noise

In the case of spatially uniform white sensor noise, $\sigma_1 = \dots = \sigma_m = \sigma$. Assuming $\sigma^2 > 0$, we denote

$$\mathbf{q} = \mathbf{p} / \sigma^2, \quad (19)$$

$$\mathbf{Q} = \mathbf{A} \text{Diag}(\mathbf{q}) \mathbf{A}^H + \mathbf{I}_m. \quad (20)$$

Estimating the DOAs means identifying the locations of nonzero elements of the vector \mathbf{q} . Considering $\mathbf{Q} = \mathbf{R}/\sigma^2$, we can estimate \mathbf{q} through solving the penalized likelihood optimization problem

$$\arg \min_{\mathbf{q} \geq 0} L_0(\mathbf{Q}) + \lambda \|\mathbf{q}\|_1 \quad (21)$$

with $\lambda \geq 0$, where the term $\lambda \|\mathbf{q}\|_1$ plays the same role as the penalty term in Equation (15).

Note that the number of unknown parameters to be estimated is k in the case of spatially uniform noise in contrast to $k + m - 1$ in the case of spatially non-uniform noise.

4. DOA Estimation Algorithms

In this section, we provide methods to solve the optimization problems in Equations (15) and (21) with λ fixed. As Equations (15) and (21) are non-convex, it is generally hard to give their globally optimal solutions. Based on the MM framework [39,40], we develop algorithms to find the locally optimal solutions of Equations (15) and (21).

A function $f(x)$ is said to be majorized by a function $g(x|x_0)$ at x_0 , if $f(x) \leq g(x|x_0)$ for all x and $f(x_0) = g(x_0|x_0)$. In the MM framework, the problem $\arg \min_x f(x)$ can be solved through iteratively solving $x_{u+1} = \arg \min_x g(x|x_u)$.

When solving the problems in Equations (15) and (21) in the MM framework, majorizing functions can be constructed based on the following two inequalities. For any positive definite matrices $\mathbf{U} \in \mathbb{C}^{m \times m}$ and $\mathbf{U}_u \in \mathbb{C}^{m \times m}$ [40],

$$\log \det(\mathbf{U}) \leq \log \det(\mathbf{U}_u) + \text{tr}(\mathbf{U}_u^{-1}\mathbf{U}) - m \quad (22)$$

and

$$\log(\mathbf{x}(t)^H \mathbf{U}^{-1} \mathbf{x}(t)) \leq \log(\mathbf{x}(t)^H \mathbf{U}_u^{-1} \mathbf{x}(t)) + \frac{\mathbf{x}(t)^H \mathbf{U}^{-1} \mathbf{x}(t)}{\mathbf{x}(t)^H \mathbf{U}_u^{-1} \mathbf{x}(t)} - 1, \quad (23)$$

where both equalities are achieved at $\mathbf{U} = \mathbf{U}_u$.

4.1. Algorithms for Spatially Non-Uniform White Noise

In this subsection, using two different majorizing functions of the objection function in Equation (15), we develop two different MM algorithms named MM1 and MM2 to solve Equation (15).

4.1.1. MM1 Algorithm

Denote

$$\mathbf{W}_u = \mathbf{B} \text{Diag}(\mathbf{r}_u) \mathbf{B}^H + \mathbf{J}_m, \quad (24)$$

where $\mathbf{r}_u \in \Omega$. Replacing the \mathbf{U} and \mathbf{U}_u in Equations (22) and (23) by \mathbf{W} and \mathbf{W}_u , respectively, we have for any $\mathbf{W} > 0, \mathbf{W}_u > 0$,

$$L_0(\mathbf{W}) \leq \text{tr}(\mathbf{W}_u^{-1}\mathbf{W}) + \frac{m}{n} \sum_{t=1}^n \frac{\mathbf{x}(t)^H \mathbf{W}^{-1} \mathbf{x}(t)}{\mathbf{x}(t)^H \mathbf{W}_u^{-1} \mathbf{x}(t)} + c_2 \quad (25)$$

$$= \text{tr}(\mathbf{W}_u^{-1}\mathbf{W}) + \text{tr}(\mathbf{M}_u \mathbf{W}^{-1}) + c_2 \quad (26)$$

where c_2 is a constant, the equality in Equation (25) is achieved at $\mathbf{W} = \mathbf{W}_u$, and

$$\mathbf{M}_u = \frac{m}{n} \sum_{t=1}^n \frac{\mathbf{x}(t)\mathbf{x}(t)^H}{\mathbf{x}(t)^H \mathbf{W}_u^{-1} \mathbf{x}(t)}. \quad (27)$$

Denote the sum of the first two terms on the right side of Equation (26) and the l_1 -norm penalty term in Equation (15) as $g_1(\mathbf{r} \mid \mathbf{r}_u, \lambda)$, and then due to Equation (12), $g_1(\mathbf{r} \mid \mathbf{r}_u, \lambda)$ can be rewritten as

$$g_1(\mathbf{r} \mid \mathbf{r}_u, \lambda) = \text{tr}(\mathbf{M}_u \mathbf{W}^{-1}) + \mathbf{w}_u^T \mathbf{r} \quad (28)$$

with

$$\mathbf{w}_u = \text{diag}(\mathbf{B}^H \mathbf{W}_u^{-1} \mathbf{B}) + [\lambda \mathbf{1}_k; \mathbf{0}_{m-1}]. \quad (29)$$

From Equation (25), $g_1(\mathbf{r} \mid \mathbf{r}_u, \lambda) + c_2$ is found to be a majorizing convex function of the objective function in Equation (15). Based on the MM framework, the problem in Equation (15) can be solved by iteratively solving the convex optimization problem

$$\mathbf{r}_{u+1} = \arg \min_{\mathbf{r} \in \Omega} g_1(\mathbf{r} \mid \mathbf{r}_u, \lambda). \quad (30)$$

Proposition 1 (The MM1 algorithm for Equation (15)). *The sequence $\{\mathbf{r}_u\}$ generated by*

$$\mathbf{r}_{u+1} = \arg \min_{\mathbf{r} \geq 0} g_1(\mathbf{r} \mid \mathbf{r}_u, \lambda) \quad (31)$$

with any initial value $\mathbf{r}_0 \in \Omega$ converges to a locally optimal solution of the optimization problem in Equation (15).

Proof. Since $g_1(\mathbf{r} \mid \mathbf{r}_u, \lambda)$ tends to $+\infty$ as \mathbf{W} goes to the boundary of the positive semidefinite cone, the optimization problems in Equations (30) and (31) are equivalent.

This proposition follows by the convergence property of the MM algorithm [40] and the fact that $L_0(\mathbf{W}) \rightarrow +\infty$ with probability 1 as \mathbf{W} tends to the boundary of the positive semidefinite cone [41]. \square

Due to Equation (11), $\mathbf{r}_0 = [(r_1)_0, \dots, (r_{k+m-1})_0]^T$ required in the iteration in Equation (31) can be the one obtained by inserting the initial estimate of $[\mathbf{p}; \boldsymbol{\sigma}]$ presented in [13]. Specifically,

$$(r_i)_0 = \frac{\mathbf{b}_i^H \hat{\mathbf{R}} \mathbf{b}_i}{\hat{\mathbf{R}}_{m,m} \|\mathbf{b}_i\|_2^4} > 0, \quad i = 1, \dots, k + m - 1, \quad (32)$$

where \mathbf{b}_i is the i th column of the matrix \mathbf{B} and

$$\hat{\mathbf{R}} = \frac{1}{n} \sum_{t=1}^n \mathbf{x}(t) \mathbf{x}(t)^H. \quad (33)$$

To solve the optimization problem in Equation (31) globally, we develop two available solvers: the coordinate descent (CD) and SPICE-like solvers. Denote $\mathbf{r} = [r_1, \dots, r_{k+m-1}]^T$ in the following.

Proposition 2 (The CD solver for Equation (31)). *The sequence of \mathbf{r} generated by iterating*

$$r_i = \left(\left(\sqrt{\frac{\beta_i}{\alpha_i}} - 1 \right) \frac{1}{\gamma_i} \right) \delta(\beta_i - \alpha_i), \quad i = 1, \dots, k + m - 1 \quad (34)$$

with any initial value $\mathbf{r} \geq 0$ converges to the globally optimal solution \mathbf{r}_{u+1} of the problem in Equation (31), where $\delta(\cdot)$ is the indicator function of the interval $(0, \infty)$,

$$\alpha_i = (w_i)_u, \quad (35)$$

$$\beta_i = \mathbf{b}_i^H \mathbf{N}_i^{-1} \mathbf{M}_u \mathbf{N}_i^{-1} \mathbf{b}_i, \quad (36)$$

$$\gamma_i = \mathbf{b}_i^H \mathbf{N}_i^{-1} \mathbf{b}_i, \quad (37)$$

with $(w_i)_u$ being the i th element of the vector \mathbf{w}_u , and

$$\mathbf{N}_i = \sum_{l \neq i} r_l \mathbf{b}_l \mathbf{b}_l^H + \mathbf{J}_m. \quad (38)$$

Proof. By the general analysis of the CD method in [42], it is easy to find that the convex problem in Equation (31) can be globally solved by iterating

$$r_i = \arg \min_{r_i \geq 0} g_1(\mathbf{r} \mid \mathbf{r}_u, \lambda), \quad i = 1, \dots, k + m - 1 \quad (39)$$

until convergence.

Due to

$$\mathbf{W} = \mathbf{B} \text{Diag}(\mathbf{r}) \mathbf{B}^H + \mathbf{J}_m = r_i \mathbf{b}_i \mathbf{b}_i^H + \mathbf{N}_i \quad (40)$$

and

$$(r_i \mathbf{b}_i \mathbf{b}_i^H + \mathbf{N}_i)^{-1} = \mathbf{N}_i^{-1} - \frac{r_i \mathbf{N}_i^{-1} \mathbf{b}_i \mathbf{b}_i^H \mathbf{N}_i^{-1}}{1 + r_i \mathbf{b}_i^H \mathbf{N}_i^{-1} \mathbf{b}_i}, \quad (41)$$

the iteration in Equation (39) can be reformulated as

$$r_i = \arg \min_{r_i \geq 0} \alpha_i r_i - \frac{\beta_i r_i}{1 + \gamma_i r_i}, \quad i = 1, \dots, k + m - 1. \quad (42)$$

In addition, $\alpha_i > 0, \beta_i > 0, \gamma_i > 0$ for $i = 1, \dots, k + m - 1$. Solving the convex optimization problems in Equation (42) by the gradient method, we conclude that the equations in Equation (42) are equivalent to those in Equation (34). \square

As the solver in Proposition 2 is derived by the CD method, we call it the CD solver. The MM1 algorithm in Proposition 1 with the CD solver is specially named the MM1-CD algorithm.

For clarity, detailed steps of the MM1-CD algorithm for Equation (15) are presented in Algorithm 1.

Note that, in the CD solver, for $i = 1, \dots, k$, we have

$$\beta_i = \frac{m}{n} \sum_{t=1}^n \zeta_i \zeta_i^H \quad (43)$$

where

$$\zeta_i = \frac{\mathbf{a}_i^H \mathbf{N}_i^{-1} \mathbf{x}(t)}{\sqrt{\mathbf{x}(t)^H \mathbf{W}_u^{-1} \mathbf{x}(t)}} \quad (44)$$

with \mathbf{a}_i being the i th column of the matrix \mathbf{A} . The β_i can be interpreted as the correlation between the signal from a possible source at the grid θ_i and the array responses $\mathbf{x}(1), \dots, \mathbf{x}(n)$. From the indicator function $\delta(\beta_i - \alpha_i)$ in Equation (34), we find that it is more likely to force to zeros the powers of the assumed signals that are less correlated with the array responses.

Proposition 3 (The SPICE-like solver for Equation (31)). *The sequence of \mathbf{r} generated by iterating*

$$\mathbf{r} = \text{sqrt}(\text{diag}(\text{Diag}(\mathbf{r}) \mathbf{B}^H \mathbf{W}^{-1} \mathbf{M}_u \mathbf{W}^{-1} \mathbf{B} \text{Diag}(\mathbf{r})) \oslash \mathbf{w}_u) \quad (45)$$

with any initial value $\mathbf{r} > 0$ converges to the globally optimal solution \mathbf{r}_{u+1} of problem in Equation (31).

Algorithm 1 The MM1-CD algorithm for spatially non-uniform white noise.

```

1: Give  $\mathbf{r}_0$  and  $\epsilon$ ,
2:  $u = 0$ ,
3: repeat
4:    $\mathbf{W}_u = \mathbf{B} \text{Diag}(\mathbf{r}_u) \mathbf{B}^H + \mathbf{J}_m$ ,
5:   Calculate  $\mathbf{M}_u$  and  $\mathbf{w}_u$ ,
6:    $d = 0$ ,
7:    $\mathbf{r}^0 = \mathbf{r}_u, \mathbf{r} = \mathbf{r}^0$ ,
8:   repeat
9:     for  $i = 1 : k + m - 1$  do
10:       $\mathbf{N}_i = \sum_{l \neq i} r_l \mathbf{b}_l \mathbf{b}_l^H + \mathbf{J}_m$ ,
11:      Calculate  $\alpha_i, \beta_i$  and  $\gamma_i$ ,
12:      if  $\beta_i \leq \alpha_i$  then
13:         $r_i = 0$ ,
14:      else
15:         $r_i = (\sqrt{\beta_i / \alpha_i} - 1) / \gamma_i$ ,
16:      end if
17:    end for
18:     $d = d + 1$ ,
19:     $\mathbf{r}^d = [r_1, \dots, r_{k+m-1}]^T$ ,
20:    until  $\|\mathbf{r}^d - \mathbf{r}^{d-1}\|_2 / \|\mathbf{r}^{d-1}\|_2 < \epsilon$ ,
21:     $u = u + 1$ ,
22:     $\mathbf{r}_u = \mathbf{r}^d$ ,
23: until  $\|\mathbf{r}_u - \mathbf{r}_{u-1}\|_2 / \|\mathbf{r}_{u-1}\|_2 < \epsilon$ .

```

Proof. It is obvious that Equation (31) and the SPICE criterion in [15] are with similar forms. By the same way as the SPICE criterion is analyzed in [15], we find that the globally optimal solution of the problem in Equation (31) is identical to the minimizer \mathbf{r} of the optimization problem

$$\begin{aligned} \min_{\mathbf{r} \geq 0, \mathbf{E} \in \mathbb{C}^{(k+m) \times n}} \quad & \text{tr}(\mathbf{E}^H (\text{Diag}([\mathbf{r}; 1]))^{-1} \mathbf{E}) + \mathbf{w}_u^T \mathbf{r} \\ \text{s.t.} \quad & [\mathbf{A}, \mathbf{I}_m] \mathbf{E} = \mathbf{X}_u, \end{aligned} \quad (46)$$

where $\mathbf{X}_u = [\tilde{\mathbf{x}}(1), \dots, \tilde{\mathbf{x}}(n)]$ with

$$\tilde{\mathbf{x}}(t) = \sqrt{\frac{m}{n}} \frac{\mathbf{x}(t)}{\sqrt{\mathbf{x}(t)^T \mathbf{W}_u^{-1} \mathbf{x}(t)}}, \quad t = 1, \dots, n. \quad (47)$$

For a fixed \mathbf{r} , the matrix \mathbf{E} minimizing Equation (46) can be verified to be [15]

$$\mathbf{E} = \text{Diag}([\mathbf{r}; 1]) [\mathbf{A}, \mathbf{I}_m]^H \mathbf{W}_u^{-1} \mathbf{X}_u, \quad (48)$$

and for a fixed \mathbf{E} , the vector \mathbf{r} minimizing Equation (46) can be readily given by

$$\mathbf{r} = \text{sqrt}((\text{diag}(\mathbf{E} \mathbf{E}^H))_{[1:k+m-1]} \oslash \mathbf{w}_u). \quad (49)$$

The sequences of \mathbf{E} and \mathbf{r} , generated by alternately iterating Equations (48) and (49) from $\mathbf{r} > 0$, converge to the globally optimal solution of the convex problem in Equation (46) [14,15].

Due to $\mathbf{X}_u \mathbf{X}_u^H = \mathbf{M}_u$, iterating Equation (49) is just iterating Equation (45). Thus, the sequence of \mathbf{r} generated by Equation (45) converges to the minimizer \mathbf{r} of Equation (46). \square

The solver in Proposition 3 is named the SPICE-like solver, and the MM1 algorithm in Proposition 1 with it is called the MM1-SPICE algorithm. Algorithm 2 summarily illustrates the MM1-SPICE algorithm for Equation (15).

Algorithm 2 The MM1-SPICE algorithm for spatially non-uniform white noise.

```

1: Give  $\mathbf{r}_0$  and  $\epsilon$ ,
2:  $u = 0$ ,
3: repeat
4:    $\mathbf{W}_u = \mathbf{B} \text{Diag}(\mathbf{r}_u) \mathbf{B}^H + \mathbf{J}_m$ ,
5:   Calculate  $\mathbf{M}_u$  and  $\mathbf{w}_u$ ,
6:    $d = 0$ ,
7:    $\mathbf{r}^0 = \mathbf{r}_u$ ,
8:   repeat
9:      $\mathbf{W}^d = \mathbf{B} \text{Diag}(\mathbf{r}^d) \mathbf{B}^H + \mathbf{J}_m$ ,
10:     $\mathbf{r}^{d+1} = \text{sqrt}(\text{diag}(\text{Diag}(\mathbf{r}^d) \mathbf{B}^H (\mathbf{W}^d)^{-1} \mathbf{M}_u (\mathbf{W}^d)^{-1} \mathbf{B} \text{Diag}(\mathbf{r}^d)) \oslash \mathbf{w}_u)$ ,
11:     $d = d + 1$ ,
12:    until  $\|\mathbf{r}^d - \mathbf{r}^{d-1}\|_2 / \|\mathbf{r}^{d-1}\|_2 < \epsilon$ ,
13:     $u = u + 1$ ,
14:     $\mathbf{r}_u = \mathbf{r}^d$ ,
15: until  $\|\mathbf{r}_u - \mathbf{r}_{u-1}\|_2 / \|\mathbf{r}_{u-1}\|_2 < \epsilon$ .

```

From Propositions 1–3, it is found that the proposed MM1 algorithm is by an inner–outer iteration loop. The MM1-CD and the MM1-SPICE algorithms are with the identical outer loop but different nested inner loops. In addition, relationship and difference between the MM1-CD and the MM1-SPICE are discussed in Section 4.3.

4.1.2. MM2 Algorithm

When $\mathbf{r}_u > 0$, it is found from [36] that, for any $\mathbf{r} > 0$,

$$\begin{bmatrix} \mathbf{C}_u (\text{Diag}([\mathbf{r}; 1]))^{-1} \mathbf{C}_u^H & \mathbf{I}_m \\ \mathbf{I}_m & \mathbf{W} \end{bmatrix} = \mathbf{F}^H \mathbf{F} \geq 0, \quad (50)$$

where

$$\mathbf{C}_u = \mathbf{W}_u^{-1} [\mathbf{A}, \mathbf{I}_m] \text{Diag}([\mathbf{r}_u; 1]), \quad (51)$$

$$\mathbf{F} = [(\text{Diag}([\mathbf{r}; 1]))^{-\frac{1}{2}} \mathbf{C}_u^H, (\text{Diag}([\mathbf{r}; 1]))^{\frac{1}{2}} [\mathbf{A}, \mathbf{I}_m]^H]. \quad (52)$$

From Equation (50), we have

$$\mathbf{W}^{-1} \leq \mathbf{C}_u (\text{Diag}([\mathbf{r}; 1]))^{-1} \mathbf{C}_u^H \quad (53)$$

with the equality achieved at $\mathbf{r} = \mathbf{r}_u$. The inequality in Equation (53) is also valid for any $\mathbf{r} \in \Omega$ due to $\mathbf{W} > 0$.

Denote

$$g_2(\mathbf{r} | \mathbf{r}_u, \lambda) = \text{tr}(\mathbf{C}_u^H \mathbf{M}_u \mathbf{C}_u (\text{Diag}([\mathbf{r}; 1]))^{-1}) + \mathbf{w}_u^T \mathbf{r}. \quad (54)$$

It is clear from Equation (53) that, when $\mathbf{r}_u > 0$, for any $\mathbf{r} \in \Omega$,

$$g_1(\mathbf{r} | \mathbf{r}_u, \lambda) \leq g_2(\mathbf{r} | \mathbf{r}_u, \lambda), \quad (55)$$

where the equality is achieved at $\mathbf{r} = \mathbf{r}_u$. Therefore, at any $\mathbf{r}_u > 0$, $g_2(\mathbf{r} | \mathbf{r}_u, \lambda) + c_2$ majorizes the objective function in Equation (15) for $\mathbf{r} \in \Omega$.

Proposition 4 (The MM2 algorithm for Equation (15)). *The sequence $\{\mathbf{r}_u\}$ generated by*

$$\mathbf{r}_{u+1} = \text{sqrt}(\text{diag}(\text{Diag}(\mathbf{r}_u)\mathbf{B}^H\mathbf{W}_u^{-1}\mathbf{M}_u\mathbf{W}_u^{-1}\mathbf{B}\text{Diag}(\mathbf{r}_u)) \oslash \mathbf{w}_u) \quad (56)$$

with any initial value $\mathbf{r}_0 > 0$ converges to a locally optimal solution of the problem in Equation (15).

Proof. Through the convergence analysis in [36], we have that, although Equation (55) is valid only when $\mathbf{r}_u > 0$, the sequence $\{\mathbf{r}_u\}$ generated

$$\mathbf{r}_{u+1} = \arg \min_{\mathbf{r} \in \Omega} g_2(\mathbf{r} | \mathbf{r}_u, \lambda) \quad (57)$$

with any initial value $\mathbf{r}_0 > 0$ converges to a locally optimal solution of the problem in Equation (15). It is worth mentioning that the elements of the coefficient vector $\text{diag}(\mathbf{C}_u^H\mathbf{M}_u\mathbf{C}_u)$ in Equation (54) are positive. By solving the optimization problem in Equation (57) using the gradient method, the iteration procedure in Equation (57) is found to be exactly Equation (56) \square

The \mathbf{r}_0 involved in the iteration procedure in Equation (56) can be the one given in Equation (32). Summarily, Algorithm 3 gives the detailed steps of the MM2 algorithm for the problem in Equation (15).

Algorithm 3 The MM2 algorithm for spatially non-uniform white noise.

- 1: Give \mathbf{r}_0 and ϵ ,
 - 2: $u = 0$,
 - 3: **repeat**
 - 4: $\mathbf{W}_u = \mathbf{B}\text{Diag}(\mathbf{r}_u)\mathbf{B}^H + \mathbf{J}_m$,
 - 5: Calculate \mathbf{M}_u and \mathbf{w}_u ,
 - 6: $\mathbf{r}_{u+1} = \text{sqrt}(\text{diag}(\text{Diag}(\mathbf{r}_u)\mathbf{B}^H\mathbf{W}_u^{-1}\mathbf{M}_u\mathbf{W}_u^{-1}\mathbf{B}\text{Diag}(\mathbf{r}_u)) \oslash \mathbf{w}_u)$,
 - 7: $u = u + 1$,
 - 8: **until** $\|\mathbf{r}_u - \mathbf{r}_{u-1}\|_2 / \|\mathbf{r}_{u-1}\|_2 < \epsilon$.
-

4.2. DOA Estimation for Spatially Uniform White Noise

The DOA estimation in the case of spatially uniform white noise amounts to solving the optimization problem in Equation (21). Problems in Equations (21) and (15) are with similar forms, but Equation (21) involves a smaller number of unknown parameters. By the same way as the problem in Equation (15) is analyzed in Section 4.1, we can naturally derive the algorithms to solve Equation (21). The algorithms for Equation (21) are still named MM1 (including MM1-CD and MM1-SPICE) and MM2.

4.2.1. MM1 Algorithm

Proposition 5 (The MM1 algorithm for Equation (21)). *Denote*

$$h_1(\mathbf{q} | \mathbf{q}_u, \lambda) = \text{tr}(\mathbf{Z}_u\mathbf{Q}^{-1}) + \mathbf{e}_u^T\mathbf{q} \quad (58)$$

with

$$\mathbf{Z}_u = \frac{m}{n} \sum_{t=1}^n \frac{\mathbf{x}(t)\mathbf{x}(t)^H}{\mathbf{x}(t)^H \mathbf{Q}_u^{-1} \mathbf{x}(t)}, \quad (59)$$

$$\mathbf{e}_u = \text{diag}(\mathbf{A}^H \mathbf{Q}_u^{-1} \mathbf{A}) + \lambda \mathbf{1}_k, \quad (60)$$

$$\mathbf{Q}_u = \mathbf{A} \text{Diag}(\mathbf{q}_u) \mathbf{A}^H + \mathbf{I}_m, \quad (61)$$

and then the sequence $\{\mathbf{q}_u\}$ generated by

$$\mathbf{q}_{u+1} = \arg \min_{\mathbf{q} \geq 0} h_1(\mathbf{q} \mid \mathbf{q}_u, \lambda) \quad (62)$$

with any initial value $\mathbf{q} \geq 0$ converges to a locally optimal solution of the problem in Equation (21).

To solve the optimization problem in Equation (62), using the same method as Equation (31) is solved, we offer two different iterative solvers for Equation (62).

Proposition 6 (The CD solver for Equation (62)). *The sequence of \mathbf{q} generated by iterating*

$$q_j = \left(\frac{\tilde{\beta}_j}{\tilde{\alpha}_j} - 1 \right) \frac{1}{\tilde{\gamma}_j} \delta(\tilde{\beta}_j - \tilde{\alpha}_j), \quad j = 1, \dots, k \quad (63)$$

with any initial value $\mathbf{q} \geq 0$ converges to the globally optimal solution of the problem in Equation (62), where

$$\tilde{\alpha}_j = (e_j)_u, \quad (64)$$

$$\tilde{\beta}_j = \mathbf{a}_j^H (\mathbf{H}_j)^{-1} \mathbf{Z}_u (\mathbf{H}_j)^{-1} \mathbf{a}_j, \quad (65)$$

$$\tilde{\gamma}_j = \mathbf{a}_j^H (\mathbf{H}_j)^{-1} \mathbf{a}_j, \quad (66)$$

with $(e_j)_u$ being the j th element of the vector \mathbf{e}_u , \mathbf{a}_j being the j th column of the matrix \mathbf{A} and

$$\mathbf{H}_j = \sum_{l \neq j} q_l \mathbf{a}_l \mathbf{a}_l^H + \mathbf{I}_m. \quad (67)$$

Proposition 6 introduces the nested CD inner loop of the MM1 algorithm for the problem in Equation (21), and can be proven similar to Proposition 2.

Proposition 7 (The SPICE-like solver for Equation (62)). *The sequence of \mathbf{q} generated by iterating*

$$\mathbf{q} = \text{sqrt}(\text{diag}(\text{Diag}(\mathbf{q}) \mathbf{A}^H \mathbf{Q}^{-1} \mathbf{Z}_u \mathbf{Q}^{-1} \mathbf{A} \text{Diag}(\mathbf{q})) \oslash \mathbf{e}_u) \quad (68)$$

with any initial value $\mathbf{q} > 0$ converges to the globally optimal solution of the problem in Equation (62).

The iterative procedure in Proposition 7 above is the nested SPICE-like inner loop of the MM1 algorithm for Equation (21), and can be proven similar to Proposition 3.

The MM procedure in Proposition 5 with the CD nested loop in Proposition 6 is the MM1-CD algorithm for Equation (21). The MM1-SPICE algorithm for Equation (21) is the MM procedure in Proposition 5 with the SPICE-like nested loop in Proposition 7.

4.2.2. MM2 Algorithm

Proposition 8 (The MM2 algorithm for Equation (21)). *The sequence $\{\mathbf{q}_u\}$ generated by*

$$\mathbf{q}_{u+1} = \text{sqrt}(\text{diag}(\text{Diag}(\mathbf{q}_u) \mathbf{A}^H \mathbf{Q}_u^{-1} \mathbf{Z}_u \mathbf{Q}_u^{-1} \mathbf{A} \text{Diag}(\mathbf{q}_u)) \oslash \mathbf{e}_u) \quad (69)$$

with any initial value $\mathbf{q} > 0$ converges to a locally optimal solution of the problem in Equation (21).

Due to Equation (19), by means of inserting in \mathbf{q} the initial estimates of \mathbf{p} and σ^2 given in [13], $\mathbf{q}_0 = [(q_1)_0, \dots, (q_k)_0]^T$ required by the iterations in Equations (62) and (69) can be

$$(q_j)_0 = \frac{\mathbf{a}_i^H \hat{\mathbf{R}} \mathbf{a}_i}{\hat{\sigma} \|\mathbf{a}_i\|_2^4} > 0, j = 1, \dots, k, \quad (70)$$

where $\hat{\mathbf{R}}$ is given by Equation (33) and $\hat{\sigma}$ is the mean of the first m smallest values of the set

$$\left\{ \frac{\mathbf{a}_1^H \hat{\mathbf{R}} \mathbf{a}_1}{\|\mathbf{a}_1\|_2^2}, \dots, \frac{\mathbf{a}_k^H \hat{\mathbf{R}} \mathbf{a}_k}{\|\mathbf{a}_k\|_2^2}, \hat{\mathbf{R}}_{1,1}, \dots, \hat{\mathbf{R}}_{m,m} \right\}. \quad (71)$$

4.3. Discussions on the MM1 and MM2 Algorithms

The following arguments are focused on the case of spatially non-uniform noise, which are also applicable to the case of spatially uniform noise.

Both the MM1 and MM2 algorithms decrease the objective function in Equation (15) at each MM iteration and converge locally, in which the different majorizing functions are adopted. Compared to the MM2 algorithm, the MM1 algorithm has a majorizing function closer to the objective function in Equation (15) due to Equation (55). The computational burdens of the two algorithms are mainly caused by the matrix inversion operations.

Although the MM1-CD and MM1-SPICE algorithms have different nested inner iteration procedures, they converge to the same local solution theoretically because their outer MM iteration procedures are both Equation (31). Each nested inner iteration of the MM1-CD algorithm, detailed by Steps 9–17 in Algorithm 1, requires $k + m - 1$ matrix inverse operations. In each nested inner iteration of the MM1-SPICE algorithm, presented by Steps 9–10 in Algorithm 2, only one matrix inverse operation is entailed.

It is somewhat interesting to find that the iteration of the MM2 algorithm (see Equation (56)) and the nested inner iteration of the MM1-SPICE algorithm (see Equation (45)) have similar forms. In each iteration of the MM2 algorithm, only one matrix inverse operation is needed.

4.4. Selection of Penalty Parameter

The penalty parameter λ in Equations (15) and (21) affects the sparsity levels of the estimates of \mathbf{r} and \mathbf{q} . A modified Bayesian information criterion (BIC) method [37], which is common and statistically sound, is provided here to choose an appropriate λ . Let $\hat{\mathbf{r}}_\lambda$ and $\hat{\mathbf{q}}_\lambda$ be the solutions of the problems in Equations (15) and (21) with a fixed λ , respectively, and denote $\hat{\mathbf{W}}_\lambda = \mathbf{B} \text{Diag}(\hat{\mathbf{r}}_\lambda) \mathbf{B}^H + \mathbf{J}_m$ and $\hat{\mathbf{Q}}_\lambda = \mathbf{A} \text{Diag}(\hat{\mathbf{q}}_\lambda) \mathbf{A}^H + \mathbf{I}_m$. The appropriate λ for spatially non-uniform noise and uniform noise are

$$\arg \min_{\lambda} L_0(\hat{\mathbf{W}}_\lambda) + \hat{z}_{\lambda,1} \log(n) / (2n) \quad (72)$$

and

$$\arg \min_{\lambda} L_0(\hat{\mathbf{Q}}_\lambda) + \hat{z}_{\lambda,2} \log(n) / (2n), \quad (73)$$

respectively, where $\hat{z}_{\lambda,1}$ and $\hat{z}_{\lambda,2}$ are the numbers of nonzero elements of the vectors $(\hat{\mathbf{r}}_\lambda)_{[1:k]}$ and $\hat{\mathbf{q}}_\lambda$, respectively.

Note that the elements of $(\hat{\mathbf{r}}_\lambda)_{[1:k]}$ and $\hat{\mathbf{q}}_\lambda$ can be treated as 0 if they are smaller than some certain values respectively, e.g., $\varepsilon \max((\hat{\mathbf{r}}_\lambda)_{[1:k]})$ and $\varepsilon \max(\hat{\mathbf{q}}_\lambda)$ with a very small $\varepsilon > 0$.

Notice that the $L_0(\hat{\mathbf{W}}_\lambda)$ and $L_0(\hat{\mathbf{Q}}_\lambda)$ in Equations (72) and (73) are substitutes for the marginal likelihood also called as Bayesian evidence required in the BIC criterion [43]. By the way,

when modeling in the Bayesian framework, the marginal likelihood usually cannot be analytically calculated, but can be approximated by several computational methods in [44–46].

5. Numerical Experiments

In this section, we numerically show the performance of the proposed methods. Consider a uniform linear array (ULA), and, for each $j = 1, \dots, k_0$, the steering vector corresponding to the DOA ζ_j is

$$\mathbf{a}(\zeta_j) = [1, \exp(i\pi \cos(\zeta_j)), \dots, \exp(i\pi(m-1) \cos(\zeta_j))]^T. \quad (74)$$

The array output data $\mathbf{x}(t)$ are generated from the model

$$\mathbf{x}(t) = \sum_{j=1}^{k_0} \mathbf{a}(\zeta_j) \bar{s}_j(t) + \mathbf{v}(t), \quad t = 1, \dots, n, \quad (75)$$

and both the source signals $\bar{s}_1(t), \dots, \bar{s}_{k_0}(t)$ and the observation noise $\mathbf{v}(t)$ are temporally independent. The SNR is defined as

$$\text{SNR}(\text{dB}) = 10 \log_{10} \left(\frac{\text{tr}(\bar{\mathbf{A}} \bar{\mathbf{P}} \bar{\mathbf{A}}^H)}{\text{tr}(\mathbf{V})} \right), \quad (76)$$

where $\bar{\mathbf{A}} = [\mathbf{a}(\zeta_1), \dots, \mathbf{a}(\zeta_{k_0})]$, $\bar{\mathbf{P}}$ is the covariance matrix of $\bar{\mathbf{s}}(t) = [\bar{s}_1(t), \dots, \bar{s}_{k_0}(t)]^T$, and \mathbf{V} is the covariance matrix of $\mathbf{v}(t)$. The root mean-square error (RMSE) of the DOA estimate is employed to evaluate the estimation performance, which is approximated by $R = 500$ Monte Carlo runs as

$$\text{RMSE} = \frac{1}{R} \sum_{i=1}^R \sqrt{\frac{1}{k_0} \sum_{j=1}^{k_0} (\hat{\zeta}_{j,i} - \zeta_j)^2}, \quad (77)$$

where $\hat{\zeta}_{j,i}$ is the estimate of ζ_j in the i th Monte Carlo run. In the following experiments, we applied the proposed methods and the SPICE [13,15] and LIKES methods [15] to estimate the DOAs. Set the grid points $\theta = \{0^\circ, 1^\circ, \dots, 180^\circ\}$, and the tolerance $\epsilon = 10^{-8}$ for convergence. All methods were coded in MATLAB and executed on a workstation with two 2.10 GHz Intel Xeon CPUs.

5.1. Experimental Settings

5.1.1. Experiment 1

Consider a scenario with Gaussian source signals and noise:

- The number of sensors in the array is $m = 15$, and the number of snapshots is $n = 100$.
- There are $k_0 = 4$ sources locating at $\zeta_1 = 60^\circ, \zeta_2 = 80^\circ, \zeta_3 = 100^\circ, \zeta_4 = 120^\circ$, respectively.
- The signal $\bar{\mathbf{s}}(t)$ is zero-mean circular complex Gaussian with covariance matrix $\bar{\mathbf{P}} = \text{Diag}([100, 100, 100, 100])$.
- The noise $\mathbf{v}(t)$ is zero-mean circular complex Gaussian with covariance matrix $\mathbf{V} = \sigma^2 \mathbf{I}_m$.

5.1.2. Experiment 2

Consider the scenario where both the source signals and the observation noise are non-Gaussian. Let $m = 15$. We used the experimental settings:

- The $k_0 = 4$ source signals from $\zeta_1 = 10^\circ, \zeta_2 = 40^\circ, \zeta_3 = 40^\circ + \Delta\zeta, \zeta_4 = 70^\circ$ are, respectively, as $\bar{s}_1(t) = 5 \exp(i\varphi_1(t)), \bar{s}_2(t) = 10 \exp(i\varphi_2(t)), \bar{s}_3(t) = 10 \exp(i\varphi_3(t)), \bar{s}_4(t) = 5 \exp(i\varphi_4(t))$, where $\Delta\zeta$ is the angle separation between ζ_2 and ζ_3 , and $\varphi_1(t), \dots, \varphi_4(t)$ are independent and respectively, distributed as

$$\begin{aligned}
 \varphi_1(t) &\sim 2\pi\text{Gamma}(1, 1), \\
 \varphi_2(t) &\sim 2\pi\text{Gamma}(1, 1), \\
 \varphi_3(t) &\sim \text{U}[0, 2\pi], \\
 \varphi_4(t) &\sim \text{U}[0, 2\pi],
 \end{aligned}
 \tag{78}$$

where $\text{Gamma}(a, b)$ denotes the Gamma distribution with the shape parameter a and scale parameter b , and $\text{U}[a, b]$ denotes the uniform distribution on the interval $[a, b]$. Note that the source signals $\bar{s}_1(t)$, $\bar{s}_2(t)$, $\bar{s}_3(t)$ and $\bar{s}_4(t)$ have constant modulus, which is common in communication applications [13].

- The observation noise $\mathbf{v}(t)$ is distributed as

$$\mathbf{v}(t) \sim \varrho \mathbf{n}_1, \tag{79}$$

where $\varrho \sim \text{Gamma}(1, 6)$ and \mathbf{n}_1 is zero-mean Gaussian with covariance matrix $\mathbf{V} = \sigma^2 \mathbf{I}_m$.

5.1.3. Experiment 3

Consider a scenario with non-Gaussian source signals and non-Gaussian spatially non-uniform white noise. Let $m = 8$ and $n = 100$. The $k_0 = 3$ source signals locating at $\xi_1 = 40^\circ$, $\xi_2 = 50^\circ$, $\xi_3 = 80^\circ$ are $\bar{s}_1(t) = 10 \exp(i\varphi_1(t))$, $\bar{s}_2(t) = 30 \exp(i\varphi_2(t))$ and $\bar{s}_3(t) = 10 \exp(i\varphi_3(t))$ with $\varphi_1(t)$, $\varphi_2(t)$, $\varphi_3(t)$ being given by Equation (78). The observation noise is distributed as

$$\mathbf{v}(t) \sim \varrho \mathbf{n}_2, \tag{80}$$

where $\varrho \sim \text{Gamma}(2, 2)$ and \mathbf{n}_2 is zero-mean Gaussian with covariance matrix $\mathbf{V} = \sigma^2 \text{Diag}([m, m - 1, \dots, 1])$.

5.2. Experimental Results

In Experiment 1, the MM1-CD, MM1-SPICE and MM2 algorithms with $\lambda = 0$ were firstly applied. Table 1 reports the iteration numbers n_1, n_2, n_3 , the computational time τ (in seconds), and the DOA estimation RMSEs (in degrees). Specifically, n_1 is the number of total iterations, n_2 is the number of iterations in the outer loop, and n_3 is the average of the iterations in a nested inner loop. Besides the RMSEs, each value in Table 1 is the average of the results of 500 Monte Carlo runs. Note that, even though $n_1 = n_2 \times n_3$ in a Monte Carlo run, n_1 is not exactly equal to $n_2 \times n_3$ in Table 1 because they are the average of 500 Monte Carlo runs.

Table 1. Comparison of computational complexities and RMSEs.

SNR	Method	n_1	n_2	n_3	τ	RMSE
−1 dB	MM1-CD	525.5600	16.0440	32.7550	27.9710	0.0877
	MM1-SPICE	317.6500	16.9280	18.7370	0.3805	0.0926
	MM2	129.5000	–	–	0.3266	0.1059
−3 dB	MM1-CD	539.2700	16.0000	33.7040	28.7700	0.2345
	MM1-SPICE	310.0300	16.6260	18.6070	0.3714	0.2434
	MM2	129.3800	–	–	0.3260	0.2568
−5 dB	MM1-CD	561.0900	16.4380	34.1500	29.3910	0.4292
	MM1-SPICE	303.1400	16.3340	18.5130	0.3590	0.4681
	MM2	129.1400	–	–	0.3195	0.4840

As shown in Table 1, the MM1-CD and MM1-SPICE algorithms had similar RMSEs and the MM1-CD algorithm took much more time than the MM1-SPICE algorithm. Considering that the MM1-CD and MM1-SPICE algorithms theoretically converge to the identical local solution, we believe

that they will always have similar RMSEs. Moreover, the MM2 algorithm had the smallest iteration numbers, the least computational time and satisfactory RMSEs. In the following, we present the DOA estimation performance only of the MM1-SPICE and MM2 algorithms. Note that the MM1-SPICE and MM2 algorithms may converge to different local solutions, and then they are referred to as two different estimation methods hereinafter. Specifically, we compared the following six methods:

1. LIKES: The ML method proposed in [15] under the assumption that the array output $\mathbf{x}(t)$ is Gaussian distributed.
2. SPICE: A sparse covariance-based estimation method proposed in [13] with no distribution assumption.
3. MM1-SPICE-0: The MM1-SPICE method with the penalty parameter $\lambda = 0$;
4. MM1-SPICE-P: The MM1-SPICE method with the penalty parameter selected by the criterion in Section 4.4.
5. MM2-0: The MM2 method with the penalty parameter $\lambda = 0$.
6. MM2-P: The MM2 method with the penalty parameter selected by the criterion in Section 4.4.

It is worth presenting that in, Experiments 1–3, the standard MUSIC algorithm was found to be almost ineffective, thus we do not illustrate it.

The RMSE comparisons of the above six methods are illustrated in Figures 1–3 for Experiments 1–3 with different SNRs, respectively. Figures 4 and 5 for Experiment 2 show the DOA estimation RMSEs of the above six methods versus the number of snapshots and the angle separation, respectively.

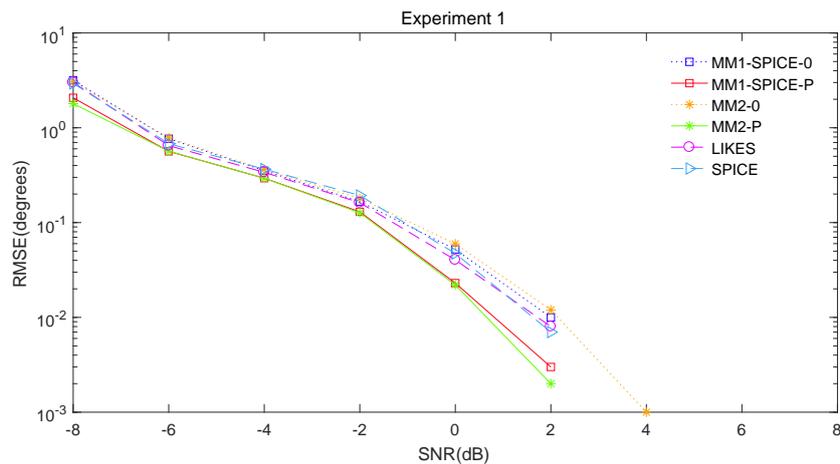


Figure 1. DOA estimation RMSE versus SNR for Experiment 1.

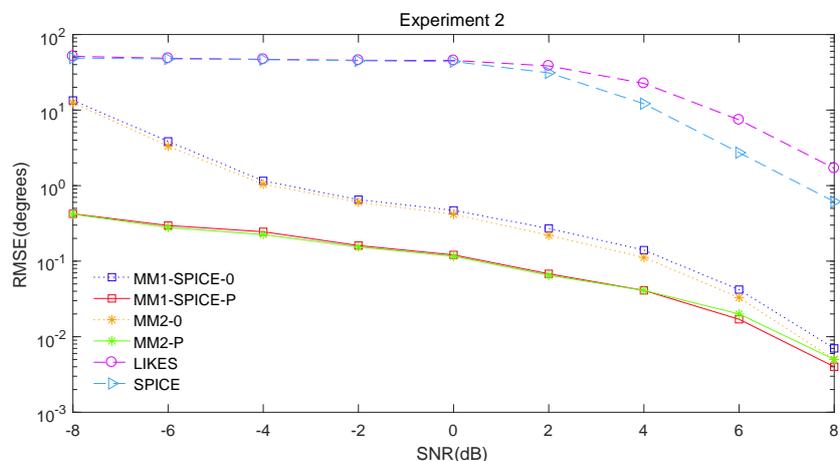


Figure 2. DOA estimation RMSE versus SNR for Experiment 2 with $n = 100$ and $\Delta\xi = 15^\circ$.

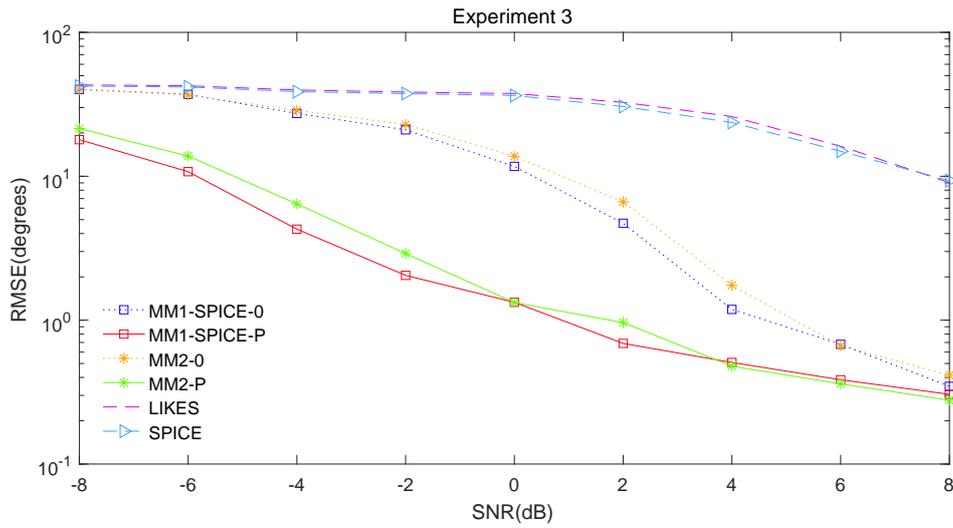


Figure 3. DOA estimation RMSE versus SNR for Experiment 3.

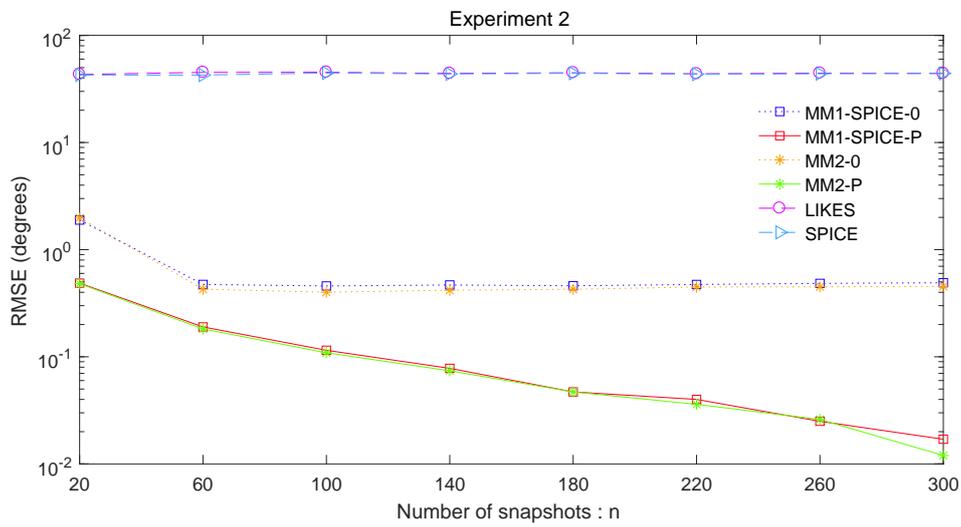


Figure 4. DOA estimation RMSE versus the number of snapshots n for Experiment 2 with SNR = 0 dB and $\Delta\zeta = 15^\circ$.

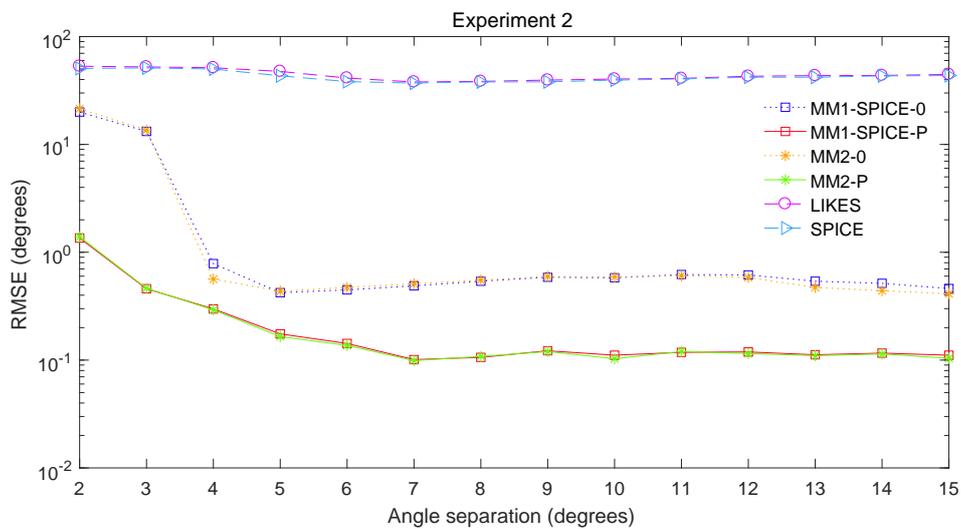


Figure 5. DOA estimation RMSE versus the angle separation $\Delta\zeta$ for Experiment 2 with $n = 100$ and SNR = 0 dB.

In Figure 1, we can see that the MM1-SPICE-0 and MM2-0 methods were comparable to the LIKES and SPICE methods in the Gaussian cases. As shown in Figures 2–5, the scenarios of non-Gaussian random source signals and heavy-tailed random noise, the MM1-SPICE-0 and MM2-0 methods performed much better than the LIKES and SPICE methods. In other words, the ML methods designed for the CES outputs were effective in the simulation scenarios of Gaussian and non-Gaussian distributions.

As shown in Figures 1–5, we found that the penalized ML methods, i.e., the MM1-SPICE-P and MM2-P methods, had smaller RMSEs than the other four methods. As shown in Figure 4, with the increase of the number of snapshots, the performance of the MM1-SPICE-P and MM2-P methods improved but the performance of the MM1-SPICE-0 and MM2-0 methods remained virtually unchanged. Figure 5 shows that, as $\Delta\zeta$ increased from 2° to 6° , the performance of the MM1-SPICE-P and MM2-P methods became better, while, when $\Delta\zeta$ was larger than 4° , the performance of the MM1-SPICE-0 and MM2-0 methods no longer improved. As shown in Figures 1–3, unsurprisingly, as the SNR increased, the performance of all the six methods became better.

For illustrating the difference between the penalized and un-penalized ML methods, we evaluated the normalized spectrum (NS):

$$NS = \frac{\hat{\mathbf{r}}_{[1:k]}}{\max(\hat{\mathbf{r}}_{[1:k]})} \text{ or } \frac{\hat{\mathbf{q}}}{\max(\hat{\mathbf{q}})}, \quad (81)$$

where $\hat{\mathbf{r}}$ and $\hat{\mathbf{q}}$ are, respectively, the estimates of \mathbf{r} and \mathbf{q} .

Figures 6–10 report the NSs of Experiments 1–3 by the MM2-0 and MM2-P methods (the curves of the NSs of the MM-SPICE-0 and MM2-SPICE-P methods are similar to those of the MM2-0 and MM2-P methods, respectively). Figures 6 and 10 are for Experiments 1 and 3 with SNR = 0 dB, respectively. Figures 7 and 8 are both for Experiment 2 with $n = 100$ and SNR = 0 dB, but they involve two scenarios with different angle separations. Particularly, Figure 8 is for the very small $\Delta\zeta = 3^\circ$, as opposed to Figure 7 for $\Delta\zeta = 15^\circ$. Figure 9 is also for Experiment 2, but it is about the scenario with small snapshot size $n = 20$, small SNR = -10 dB and moderate angle separation $\Delta\zeta = 10^\circ$. Note that the red dashed lines mark the true DOAs and the NS curves in each figure are the results of a randomly selected realization. We can clearly see in Figures 6–10 that the MM2-P method having the l_1 -norm penalty added yielded a higher angular resolution.

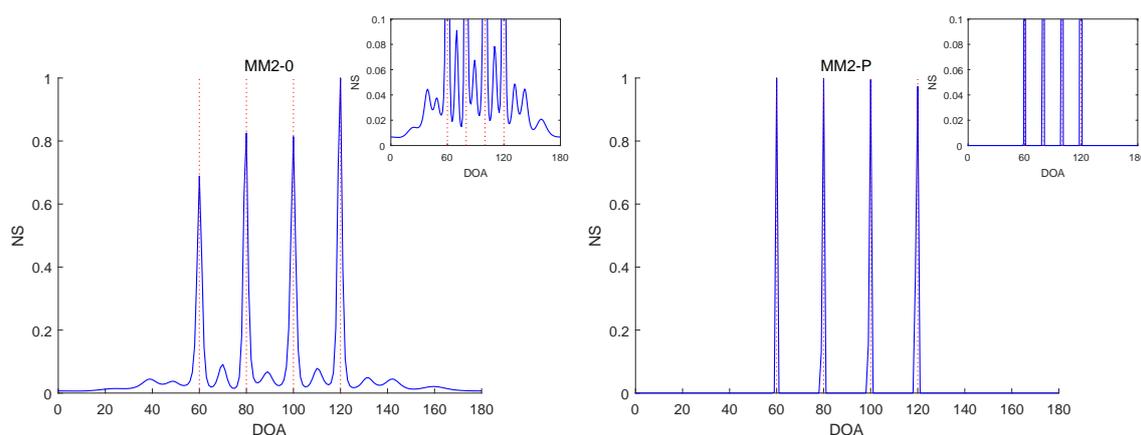


Figure 6. NSs of Experiment 1 with SNR = 0 dB.

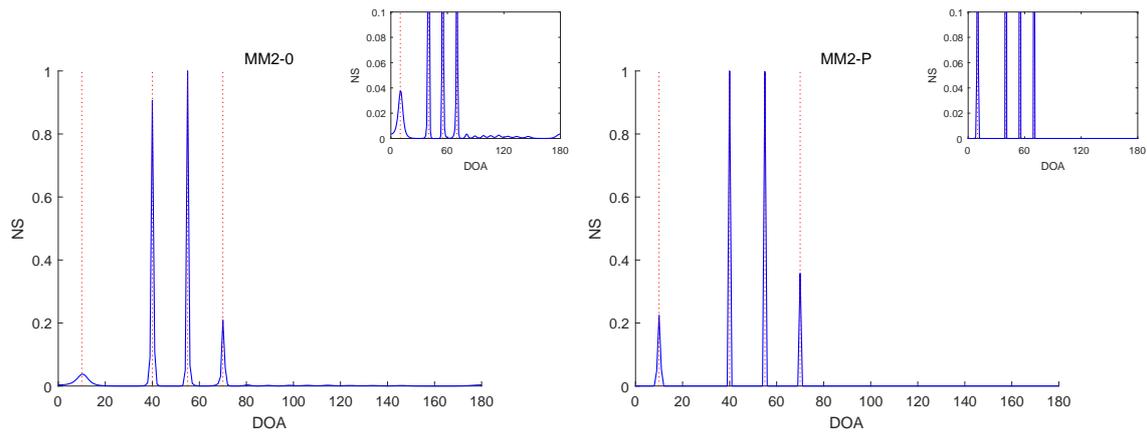


Figure 7. NSs of Experiment 2 with $n = 100$, $\text{SNR} = 0$ dB and $\Delta\xi = 15^\circ$.

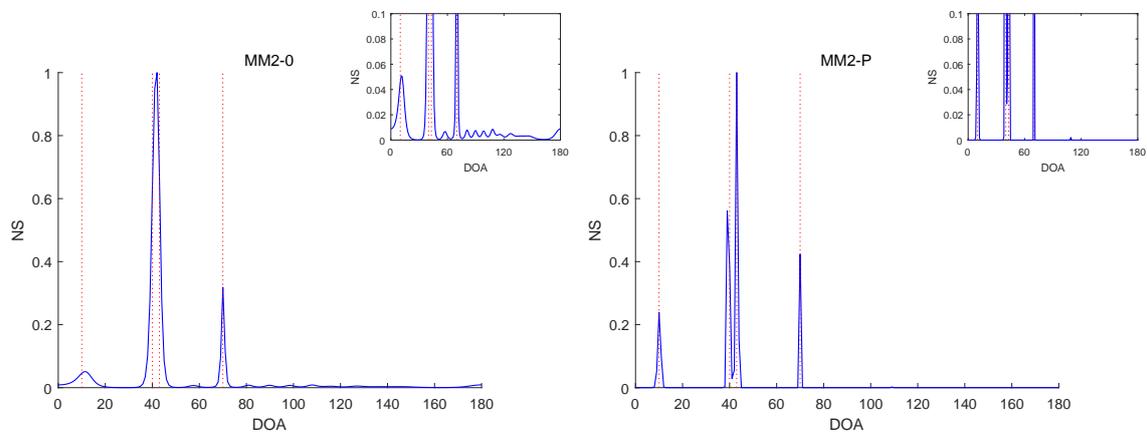


Figure 8. NSs of Experiment 2 with $n = 100$, $\text{SNR} = 0$ dB and $\Delta\xi = 3^\circ$.

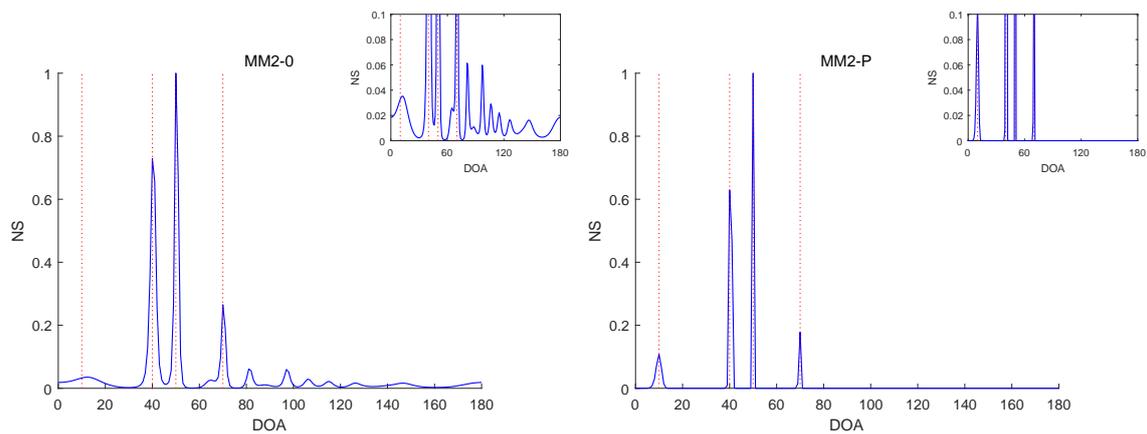


Figure 9. NSs of Experiment 2 with $n = 20$, $\text{SNR} = -10$ dB and $\Delta\xi = 10^\circ$.

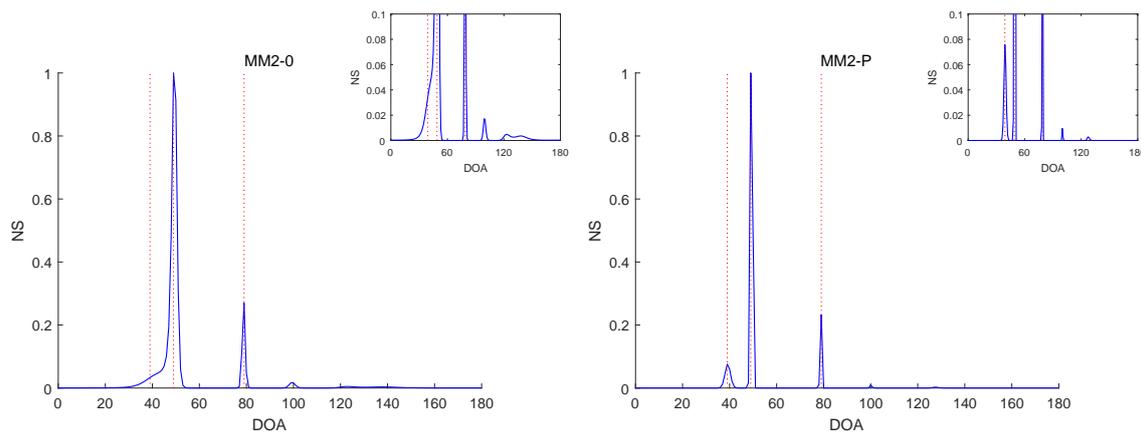


Figure 10. NSs of Experiment 3 with SNR = 0 dB.

6. Conclusions

This paper provides a penalized ML method for DOA estimation under the assumption that the array output has a CES distribution. Two MM algorithms, named MM1 and MM2, are developed for solving the non-convex penalized ML optimization problem for spatially uniform or non-uniform noise. They converge locally with different rates. The numerical experiments showed that the MM2 ran faster and performed as well as the MM1 when estimating the DOAs. Their rates of convergence will be further explored theoretically in future.

It is worth mentioning that, in numerical simulations, the proposed l_1 -norm penalized likelihood method effectively estimated the DOAs, although the values of nonzero elements of \mathbf{r} or \mathbf{p} were not estimated very accurately. By replacing the l_1 -norm penalty by other proper penalties (e.g., the smoothly clipped absolute deviation (SCAD) [47], adaptive LASSO [48], or l_p -norm penalty ($0 \leq p < 1$)), more accurate estimates of the unknown parameter may be derived in the future.

If the l_1 -norm penalty is replaced by the adaptive LASSO penalty, then the algorithms proposed in this paper can be applied almost without modification because the adaptive LASSO penalty is a weighted l_1 -norm penalty. When the non-convex SCAD or the l_p -norm penalty ($0 \leq p < 1$) is employed, the convex majorizing functions given in [40,49] can be exploited based on the MM framework, and then the algorithms in this paper with minor modifications are also applicable.

When we have some informative prior knowledge on the directions of source signals, the problem of estimating the DOAs can be formulated as maximizing a sparse posterior likelihood from the Bayesian perspective. The sparse Bayesian method of estimating DOAs from the CES array outputs is interesting and is worth studying, while how to formulate and solve a sparse posterior likelihood optimization and how to do model selection are challenges. The BIC criterion can be used for the model selection, in which the Bayesian evidence difficult to be analytically derived can be approximated by the methods in [44–46]. More research along this line will be done in the future.

Author Contributions: Conceptualization, C.C. and J.Z.; Funding acquisition, J.Z.; Methodology, C.C., J.Z. and M.T.; Software, C.C.; Supervision, J.Z.; Validation, C.C.; Writing—original draft, C.C.; and Writing—review and editing, J.Z. and M.T.

Funding: This work was supported in part by the National Natural Science Foundation of China under grant 61374027, the Application Basic Research Project of Sichuan Province under grant 2019YJ0122, and the Program for Changjiang Scholars and Innovative Research Team in University under grant IRT_16R53 from the Chinese Education Ministry.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Schmidt, R.O. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [[CrossRef](#)]
2. Roy, R.; Kailath, T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984–995. [[CrossRef](#)]
3. Stoica, P.; Sharman, K.C. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Trans. Acoust. Speech Signal Process.* **1990**, *38*, 1132–1143. [[CrossRef](#)]
4. Friedlander, B. The root-MUSIC algorithm for direction finding with interpolated arrays. *Signal Process.* **1993**, *30*, 15–29. [[CrossRef](#)]
5. Yang, Z.; Li, J.; Stoica, P.; Xie, L. Sparse methods for direction-of-arrival estimation. In *Academic Press Library in Signal Processing, Volume 7: Array, Radar and Communications Engineering*; Chellappa, R., Theodoridis, S., Eds.; Academic Press: London, UK, 2018; Chapter 11, pp. 509–581.
6. Gorodnitsky, I.F.; Rao, B.D. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Trans. Signal Process.* **1997**, *45*, 600–616. [[CrossRef](#)]
7. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
8. Wipf, D.P.; Rao, B.D. Sparse Bayesian learning for basis selection. *IEEE Trans. Signal Process.* **2004**, *52*, 2153–2163. [[CrossRef](#)]
9. Malioutov, D.; Çetin, M.; Willsky, A.S. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.* **2005**, *53*, 3010–3022. [[CrossRef](#)]
10. Babacan, S.D.; Molina, R.; Katsaggelos, A.K. Bayesian compressive sensing using Laplace priors. *IEEE Trans. Image Process.* **2010**, *19*, 53–63. [[CrossRef](#)] [[PubMed](#)]
11. Hyder, M.M.; Mahata, K. Direction-of-arrival estimation using a mixed $l_{2,0}$ norm approximation. *IEEE Trans. Signal Process.* **2010**, *58*, 4646–4655. [[CrossRef](#)]
12. Xu, X.; Wei, X.; Ye, Z. DOA estimation based on sparse signal recovery utilizing weighted l_1 -norm penalty. *IEEE Signal Process. Lett.* **2012**, *19*, 155–158. [[CrossRef](#)]
13. Stoica, P.; Babu, P.; Li, J. SPICE: A sparse covariance-based estimation method for array processing. *IEEE Trans. Signal Process.* **2011**, *59*, 629–638. [[CrossRef](#)]
14. Stoica, P.; Babu, P.; Li, J. New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data. *IEEE Trans. Signal Process.* **2011**, *59*, 35–46. [[CrossRef](#)]
15. Stoica, P.; Babu, P. SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation. *Signal Process.* **2012**, *92*, 1580–1590. [[CrossRef](#)]
16. Stoica, P.; Zachariah, D.; Li, J. Weighted SPICE: A unifying approach for hyperparameter-free sparse estimation. *Digit. Signal Process.* **2014**, *33*, 1–12. [[CrossRef](#)]
17. Yin, J.; Chen, T. Direction-of-arrival estimation using a sparse representation of array covariance vectors. *IEEE Trans. Signal Process.* **2011**, *59*, 4489–4493. [[CrossRef](#)]
18. Zheng, J.; Kaveh, M. Sparse spatial spectral estimation: A covariance fitting algorithm, performance and regularization. *IEEE Trans. Signal Process.* **2013**, *61*, 2767–2777. [[CrossRef](#)]
19. Liu, Z.M.; Huang, Z.T.; Zhou, Y.Y. Array signal processing via sparsity-inducing representation of the array covariance matrix. *IEEE Trans. Aerosp. Electron. Syst.* **2013**, *49*, 1710–1724. [[CrossRef](#)]
20. He, Z.Q.; Shi, Z.P.; Huang, L. Covariance sparsity-aware DOA estimation for nonuniform noise. *Digit. Signal Process.* **2014**, *28*, 75–81. [[CrossRef](#)]
21. Chen, T.; Wu, H.; Zhao, Z. The real-valued sparse direction of arrival (DOA) estimation based on the Khatri-Rao product. *Sensors* **2016**, *16*, 693. [[CrossRef](#)]
22. Wang, Y.; Hashemi-Sakhtsari, A.; Trinkle, M.; Ng, B.W.H. Sparsity-aware DOA estimation of quasi-stationary signals using nested arrays. *Signal Process.* **2018**, *144*, 87–98. [[CrossRef](#)]
23. Shutin, D.; Fleur, B.H. Sparse variational Bayesian SAGE algorithm with application to the estimation of multipath wireless channels. *IEEE Trans. Signal Process.* **2011**, *59*, 3609–3623. [[CrossRef](#)]
24. Hu, L.; Shi, Z.; Zhou, J.; Fu, Q. Compressed sensing of complex sinusoids: An approach based on dictionary refinement. *IEEE Trans. Signal Process.* **2012**, *60*, 3809–3819.
25. Yang, Z.; Zhang, C.; Xie, L. Robustly stable signal recovery in compressed sensing with structured matrix perturbation. *IEEE Trans. Signal Process.* **2012**, *60*, 4658–4671. [[CrossRef](#)]

26. Yang, Z.; Xie, L.; Zhang, C. Off-grid direction of arrival estimation using sparse Bayesian inference. *IEEE Trans. Signal Process.* **2013**, *61*, 38–43. [[CrossRef](#)]
27. Fan, Y.; Wang, J.; Du, R.; Lv, G. Sparse method for direction of arrival estimation using denoised fourth-order cumulants vector. *Sensors* **2018**, *18*, 1815. [[CrossRef](#)] [[PubMed](#)]
28. Wu, X.; Zhu, W.P.; Yan, J.; Zhang, Z. Two sparse-based methods for off-grid direction-of-arrival estimation. *Signal Process.* **2018**, *142*, 87–95. [[CrossRef](#)]
29. Bhaskar, B.N.; Tang, G.; Recht, B. Atomic norm denoising with applications to line spectral estimation. *IEEE Trans. Signal Process.* **2013**, *61*, 5987–5999. [[CrossRef](#)]
30. Yang, Z.; Xie, L.; Zhang, C. A discretization-free sparse and parametric approach for linear array signal processing. *IEEE Trans. Signal Process.* **2014**, *62*, 4959–4973. [[CrossRef](#)]
31. Yang, Z.; Xie, L. Enhancing sparsity and resolution via reweighted atomic norm minimization. *IEEE Trans. Signal Process.* **2016**, *64*, 995–1006. [[CrossRef](#)]
32. Wang, X.; Wang, W.; Li, X.; Liu, Q.; Liu, J. Sparsity-aware DOA estimation scheme for noncircular source in MIMO Radar. *Sensors* **2016**, *16*, 539. [[CrossRef](#)] [[PubMed](#)]
33. Jing, X.; Liu, X.; Liu, H. A sparse recovery method for DOA estimation based on the sample covariance vectors. *Circuits Syst. Signal Process.* **2017**, *36*, 1066–1084. [[CrossRef](#)]
34. Stoica, P.; Babu, P. Sparse estimation of spectral lines: Grid selection problems and their solutions. *IEEE Trans. Signal Process.* **2012**, *60*, 962–967. [[CrossRef](#)]
35. Ollila, E.; Tyler, D.E.; Koivunen, V.; Vincent, H. Complex elliptically symmetric distributions: Survey, new results and applications. *IEEE Trans. Signal Process.* **2012**, *60*, 5597–5625. [[CrossRef](#)]
36. Sun, Y.; Babu, P.; Palomar, D.P. Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions. *IEEE Trans. Signal Process.* **2016**, *64*, 3576–3590. [[CrossRef](#)]
37. Pourahmadi, M. *High-Dimensional Covariance Estimation*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
38. Sangston, K.J.; Gini, F.; Greco, M.S. Coherent radar target detection in heavy-tailed compound-Gaussian clutter. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 64–77. [[CrossRef](#)]
39. Hunter, D.R.; Lange, K. A tutorial on MM algorithms. *Am. Stat.* **2004**, *58*, 30–37. [[CrossRef](#)]
40. Sun, Y.; Babu, P.; Palomar, D.P. Majorization-Minimization algorithms in signal processing, communications, and machine learning. *IEEE Trans. Signal Process.* **2017**, *65*, 794–816. [[CrossRef](#)]
41. Kent, J.T.; Tyler, D.E. Maximum likelihood estimation for the wrapped Cauchy distribution. *J. Appl. Stat.* **1988**, *15*, 247–254. [[CrossRef](#)]
42. Luo, Z.Q.; Tseng, P. On the convergence of the coordinate descent method for convex differentiable minimization. *J. Optim. Theory Appl.* **1992**, *72*, 7–35. [[CrossRef](#)]
43. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
44. Friel, N.; Wyse, J. Estimating the evidence—A review. *Stat. Neerl.* **2012**, *66*, 288–308. [[CrossRef](#)]
45. Martino, L.; Elvira, V.; Luengo, D.; Corander, J. An adaptive population importance sampler: Learning from uncertainty. *IEEE Trans. Signal Process.* **2015**, *63*, 4422–4437. [[CrossRef](#)]
46. Martino, L.; Elvira, V.; Luengo, D.; Corander, J. Layered adaptive importance sampling. *Stat. Comput.* **2017**, *27*, 599–623. [[CrossRef](#)]
47. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [[CrossRef](#)]
48. Zou, H. The adaptive Lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [[CrossRef](#)]
49. Fan, J.; Feng, Y.; Wu, Y. Network exploration via the adaptive LASSO and SCAD penalties. *Ann. Appl. Stat.* **2009**, *3*, 521–541. [[CrossRef](#)] [[PubMed](#)]

