

Article

Sparse Method for Direction of Arrival Estimation Using Denoised Fourth-Order Cumulants Vector

Yangyu Fan, Jianshu Wang * , Rui Du and Guoyun Lv

School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China; fan_yangyu@nwpu.edu.cn (Y.F.); rdu@mail.nwpu.edu.cn (R.D.); lvguoyun101@nwpu.edu.cn (G.L.)

* Correspondence: wangjs123@mail.nwpu.edu.cn; Tel.: +86-151-2923-0640

Received: 11 April 2018; Accepted: 30 May 2018; Published: 4 June 2018



Abstract: Fourth-order cumulants (FOCs) vector-based direction of arrival (DOA) estimation methods of non-Gaussian sources may suffer from poor performance for limited snapshots or difficulty in setting parameters. In this paper, a novel FOCs vector-based sparse DOA estimation method is proposed. Firstly, by utilizing the concept of a fourth-order difference co-array (FODCA), an advanced FOCs vector denoising or dimension reduction procedure is presented for arbitrary array geometries. Then, a novel single measurement vector (SMV) model is established by the denoised FOCs vector, and efficiently solved by an off-grid sparse Bayesian inference (OGSBI) method. The estimation errors of FOCs are integrated in the SMV model, and are approximately estimated in a simple way. A necessary condition regarding the number of identifiable sources of our method is presented that, in order to uniquely identify all sources, the number of sources K must fulfill $K \leq (M^4 - 2M^3 + 7M^2 - 6M)/8$. The proposed method suits any geometry, does not need prior knowledge of the number of sources, is insensitive to associated parameters, and has maximum identifiability $O(M^4)$, where M is the number of sensors in the array. Numerical simulations illustrate the superior performance of the proposed method.

Keywords: direction of arrival estimation; fourth-order cumulants; non-Gaussian sources; fourth-order difference co-array; sparse Bayesian learning

1. Introduction

Direction of arrival (DOA) estimation is a topic that has received a great deal of attention in the last several decades. It arises in many practical scenarios, such as radar, sonar, and communications [1–4]. For example, in radar and sonar systems, objects including airplanes, birds, missiles, submarines, and fish torpedoes can be tracked by estimating their DOAs. In communication systems, the DOAs of the sources can be used to improve the communication quality.

Higher-order cumulants (HOCs)-based DOA estimation methods have been developed for non-Gaussian sources [1,5–8], mainly to overcome the limitations of second-order statistics-based DOA estimation methods such as MUltiple SIgnal Classification (MUSIC) [9] and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [10]. Their limitations include identifiable number of sources, colored Gaussian noise, and modeling errors. The famous 4-MUSIC method [5] uses the fourth-order cumulants (FOCs) to form a MUSIC-like method. Then, the $2q$ -MUSIC [6] and rectangular $2q$ -MUSIC [7] methods extend 4-MUSIC by utilizing $2q$ th-order cumulants. The rectangular $2q$ -MUSIC method achieves a trade-off between performance and maximal identifiable number of sources compared with the $2q$ -MUSIC method. To further increase the degrees of freedom (DOFs) of the FOCs-based methods, a kind of sparse linear array named multiple level nested array and the corresponding spatial smoothing MUSIC (SS-MUSIC) method were developed, which make the DOFs increase from $O(M^q)$ to $O(M^{2q})$ [8], where M is the number of sensors in the array. However,

HOCs-based methods always suffer from performance degradation when snapshots are limited, mainly because accurate estimations of HOCs require a large number of snapshots.

Recently, a series of covariance vector-based DOA methods were proposed [11–13] which built a single measurement vector (SMV) model by covariance vector and were solved by sparse methods such as L1-norm-based methods [14] and sparse Bayesian learning (SBL) [15]. Consequently, these methods always perform better than other covariance-based DOA methods such as MUSIC when snapshots are limited, mainly because estimation errors of covariance are utilized to build their sparse models. Moreover, the DOFs of these methods increase to $O(M^2)$, which is more than MUSIC. Analogously, the FOCs vector-based L1-norm method (4-L1) is applied in [16–18]. Compared with FOCs vector-based SS-MUSIC (4-SS-MUSIC) in [8], the 4-L1 method can utilize all the virtual sensors in the fourth-order difference co-array (FODCA) [8], which may give a better performance than 4-SS-MUSIC. Moreover, the 4-L1 method can be applied to any array, while 4-SS-MUSIC only suits for linear arrays. Nonetheless, the 4-L1 method is inconvenient to use in practice, since the parameter of allowable error bound is difficult to choose, and the solution of 4-L1 is sensitive to it. The allowable error bound usually needs to be chosen manually through trial-and-error for each scenario, as in [16–18]. Furthermore, it has been demonstrated both theoretically and empirically that the SBL technique induces less structural error (biased global minimum) and convergence error (failure in achieving the global minimum) than the L1 methods [11,19–21]. On the other hand, an advanced denoising procedure of the FOCs vector may further promote the performance of FOCs vector-based DOA estimation methods. An analogous procedure can be found in covariance vector-based DOA methods in [22].

In this paper, we establish a novel denoised SMV model by FOCs vector, and solve it by an off-grid sparse Bayesian inference (OGSBI) method [21]. By utilizing the concept of the fourth-order difference co-array, the advanced denoising or dimension reduction procedure of FOCs vector is presented for any geometry. The estimation errors of FOCs are integrated in the proposed SMV model, and are approximately estimated in a simple way. A necessary condition of our method regarding the number of identifiable sources is presented such that in order to uniquely identify all sources, the number of sources K must fulfill $K \leq (M^4 - 2M^3 + 7M^2 - 6M)/8$. The proposed method suits for any geometry, does not require prior knowledge of the number of sources, has maximum identifiability $O(M^4)$, and is insensitive to associated parameters. The off-grid parameter estimation in our method promotes superior performance in overdetermined cases when the number of snapshots is large and signal-noise-ratio (SNR) is high, and also ensures good performance when choosing a relatively coarse grid. Numerical simulations illustrate the superior performance of the proposed method.

2. Data Model

One-dimensional DOA estimation is discussed in this paper. Consider an array with any geometry which consists of M identical, omnidirectional, unpolarized and isotropic sensors. The sensor locations $\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_M$ are measured by $d = \lambda/2$, and collected by the vector set $\mathbb{S} \triangleq \{\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_M\} \subset \mathbb{R}^3$, where $\mathbf{n}_m = [n_{m,x}, n_{m,y}, n_{m,z}]^T$ is expressed in three dimensions for $m = 1, 2, \dots, M$, $(\cdot)^T$ is the transpose operator, λ is the wavelength, and \mathbb{R} denotes the set of real numbers. The cardinality of \mathbb{S} is denoted by $|\mathbb{S}|$, and $|\mathbb{S}| = M$. Each sensor receives the contribution of K zero-mean statistically independent stationary narrowband non-Gaussian sources. We denote by $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$, $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_K(t)]^T$ and $\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_M(t)]^T$ the observed signals of the sensors, the source signals, and the noises of the observed signals at the t th snapshot, respectively, where $t = 1, 2, \dots, +\infty$. The observed signals are modeled as:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t), \quad (1)$$

where $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{M \times K}$ denotes the array manifold matrix, $\mathbf{a}(\theta_k) = [e^{j2\pi\mathbf{u}_k^T \mathbf{n}_1}, e^{j2\pi\mathbf{u}_k^T \mathbf{n}_2}, \dots, e^{j2\pi\mathbf{u}_k^T \mathbf{n}_M}]^T$ is the steering vector for θ_k , j is an imaginary number, $\mathbf{u}_k = d/\lambda \cdot$

$[\cos(\theta_k), \sin(\theta_k), 0]^T$, θ_k denotes the azimuth angle of the k th source, and \mathbb{C} is the set of complex numbers. The noises are assumed to be statistically independent zero-mean white Gaussian and independent from each source.

Build the FOCs vector $\mathbf{c}_x \in \mathbb{C}^{M^4}$, where the i th entry is $[\mathbf{c}_x]_i = \text{cum}\{x_{i_1}(t), x_{i_2}(t), x_{i_3}^*(t), x_{i_4}^*(t)\}$ with $i = (i_1 - 1)M^3 + (i_2 - 1)M^2 + (i_3 - 1)M + i_4$, $\text{cum}\{\cdot\}$ denotes the cumulant operator, $(\cdot)^*$ denotes complex conjugate, and the indexes i_γ and γ are integers satisfying $1 \leq i_\gamma \leq M$ and $1 \leq \gamma \leq 4$, respectively. Due to model (1), the above assumptions of the sources and noises, and the properties of the cumulants [23,24], we can obtain

$$[\mathbf{c}_x]_i = \sum_{k=1}^K c_{s_k} a_{i_1}(\theta_k) a_{i_2}(\theta_k) a_{i_3}^*(\theta_k) a_{i_4}^*(\theta_k), \quad (2)$$

where $c_{s_k} = \text{cum}\{s_{k_1}(t), s_{k_2}(t), s_{k_3}^*(t), s_{k_4}^*(t)\}$ with $k_\gamma = k$, $a_{i_\gamma}(\theta_k)$ is the i_γ th element of $\mathbf{a}_\gamma(\theta_k)$, and $\mathbf{a}_\gamma(\theta_k) = \mathbf{a}(\theta_k)$. Then, we have

$$\mathbf{c}_x = \mathbf{B}\mathbf{c}_s \in \mathbb{C}^{M^4}, \quad (3)$$

where $\mathbf{c}_s = [c_{s_1}, c_{s_2}, \dots, c_{s_K}]^T$, $\mathbf{B} = [\mathbf{b}(\theta_1), \mathbf{b}(\theta_2), \dots, \mathbf{b}(\theta_K)]$, $\mathbf{b}(\theta_k) = \mathbf{a}_1(\theta_k) \otimes \mathbf{a}_2(\theta_k) \otimes \mathbf{a}_3^*(\theta_k) \otimes \mathbf{a}_4^*(\theta_k)$, and \otimes denotes the Kronecker product.

3. The Proposed Method

In this section, we present an FOCs vector-based method to perform DOA estimation. Firstly, we provide a dimension reduction method to build a novel data model by employing the concept of a fourth-order difference co-array [8]. This method can also be treated as a denoising procedure to reduce the estimation errors of FOCs. Next, according to the newly built data model, our sparse model is efficiently established and solved by the OGSBI method. Some information about the proposed method (e.g., identifiability, complexity, and comparison with the state-of-the-art methods) is also discussed.

3.1. Dimension Reduction or Denoising Procedure

In this subsection, the dimension of model (3) is reduced by utilizing a fourth-order difference co-array. Here we supplement several definitions and properties to facilitate dimension reduction.

Definition 1. *Fourth-order difference co-array (FODCA):* Given an array with arbitrary geometry \mathbb{S} , the fourth-order difference co-array \mathbb{D} is defined as the set of distinct vector elements from the set

$$\mathbb{D} = \{\mathbf{n}_{i_1} + \mathbf{n}_{i_2} - \mathbf{n}_{i_3} - \mathbf{n}_{i_4} : \forall \mathbf{n}_{i_1}, \mathbf{n}_{i_2}, \mathbf{n}_{i_3}, \mathbf{n}_{i_4} \in \mathbb{S}\}, \quad (4)$$

where $i_1, i_2, i_3, i_4 = 1, 2, \dots, M$, as described in Section 2.

Let $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{|\mathbb{D}|}$ denote the distinct elements in \mathbb{D} , where $|\mathbb{D}|$ is the cardinality of \mathbb{D} . The following definition gives the concept of a transformation matrix, which describes the relation between $\mathbf{b}(\theta)$ and the steering vector on \mathbb{D} (see Property 2 in this subsection).

Definition 2. *The transformation matrix \mathbf{J} :* Given array \mathbb{S} , the associated FODCA \mathbb{D} can be obtained by Definition 1. The transformation matrix \mathbf{J} is an $M^4 \times |\mathbb{D}|$ matrix, where the ξ th column corresponds to the vector $\mathbf{m}_\xi \in \mathbb{D}$, and the (i, ξ) th entry of \mathbf{J} satisfies

$$\mathbf{J}_{i,\xi} = \begin{cases} 1, & \text{if } \mathbf{n}_{i_1} + \mathbf{n}_{i_2} - \mathbf{n}_{i_3} - \mathbf{n}_{i_4} = \mathbf{m}_\xi, \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where the index $\xi = 1, 2, \dots, |\mathbb{D}|$, and recall that $i = (i_1 - 1)M^3 + (i_2 - 1)M^2 + (i_3 - 1)M + i_4$ in Section 2.

Property 1. The transformation matrix \mathbf{J} has full column rank.

Property 2. Let $\mathbf{a}_{\mathbb{D}}(\theta)$ be the steering vector of \mathbb{D} (i.e., $\mathbf{a}_{\mathbb{D}}(\theta) = [e^{j2\pi\mathbf{u}^T\mathbf{m}_1}, e^{j2\pi\mathbf{u}_k^T\mathbf{m}_2}, \dots, e^{j2\pi\mathbf{u}_k^T\mathbf{m}_{|\mathbb{D}|}}]^T$), we have

$$\mathbf{b}(\theta) = \mathbf{J}\mathbf{a}_{\mathbb{D}}(\theta), \quad (6)$$

where $\mathbf{b}(\theta) = \mathbf{a}_1(\theta) \otimes \mathbf{a}_2(\theta) \otimes \mathbf{a}_3^*(\theta) \otimes \mathbf{a}_4^*(\theta)$, $\mathbf{a}_\gamma(\theta)$ is the steering vector for θ , and $\gamma = 1, 2, 3, 4$.

Property 3. Define $\mathbb{H} = \{|\mathbb{D}| : \mathbb{D} = f(\mathbb{S}), \mathbb{S} = \{\mathbf{n}_1, \dots, \mathbf{n}_M\}, \forall \mathbf{n}_1, \dots, \mathbf{n}_M \in \mathbb{R}^3\}$, where $f(\cdot)$ is the function that \mathbb{S} maps to \mathbb{D} by Definition 1. Then the supremum of \mathbb{H} is

$$\sup\{\mathbb{H}\} = (M^4 - 2M^3 + 7M^2 - 6M + 4)/4. \quad (7)$$

Properties 1, 2, and 3 are proved in Appendix A, B, and C, respectively.

From Property 2, we have $\mathbf{B} = \mathbf{J}\mathbf{A}_{\mathbb{D}}$, where $\mathbf{A}_{\mathbb{D}} = [\mathbf{a}_{\mathbb{D}}(\theta_1), \mathbf{a}_{\mathbb{D}}(\theta_2), \dots, \mathbf{a}_{\mathbb{D}}(\theta_K)]$. Substituting it to (3), we obtain $\mathbf{c}_x = \mathbf{J}\mathbf{A}_{\mathbb{D}}\mathbf{c}_s$. Due to Property 1, we have

$$\mathbf{J}^+\mathbf{c}_x = \mathbf{A}_{\mathbb{D}}\mathbf{c}_s \in \mathbb{C}^{|\mathbb{D}|}, \quad (8)$$

where $\mathbf{J}^+ = (\mathbf{J}^H\mathbf{J})^{-1}\mathbf{J}^H$ is the Moore–Penrose inverse of \mathbf{J} . Comparing the model (8) with (3), it is clear that the row dimension decreases from M^4 to $|\mathbb{D}|$.

Remark 1: An analogous procedure can be found in covariance vector-based DOA estimation methods [22]. Among HOCs-based DOA estimation methods, rectangular $2q$ -MUSIC provides two strategies to reduce the dimension of a rectangle FOCs matrix (i.e., selection strategy and averaging strategy), by eliminating redundancies of associated left and right virtual steering vectors [7]. From this perspective, the proposed dimension reduction procedure is the essential realization of the averaging strategy applying to FODCA, with an explicit formulation (8). It is clear that the averaging strategy is better than the selection strategy due to the averaging of the noise in the associated observation components [7], if the computation burden is not taken into consideration. Thus, the averaging strategy can also be treated as a denoising procedure. In addition, it can be seen that FODCA has the same geometry as a fourth-order virtual array with parameters $(q, l) = (4, 2)$ [25], which is deduced from an eighth-order cumulants-based array processing problem. In TABLE IX from [25], the max number of identical virtual sensors in the fourth-order virtual array, with parameters $(q, l) = (4, 2)$ and space diversity (without angular and polarization diversity), is illustrated, and it is equal to (7) after simplification due to the same geometry. The detailed derivation of (7) is provided in Appendix C, since it is not provided in [25].

Remark 2: It is clear that when $2 \leq K < M$, the limiting root mean square error (RMSE) of covariance vector-based methods is not necessarily zero [22,26], while the RMSE of the traditional MUSIC method converges to zero as signal–noise-ratio (SNR) approaches infinity. As the FODCA can be regarded as two levels of second-order difference co-array [8,18], applied to construct covariance vector-based methods, it can be deduced that the FOCs vector-based DOA estimation methods, including [8,16–18], will be outperformed by traditional MUSIC in high-SNR regions when $2 \leq K < M$. From this point, we only suggest to use these methods (including the proposed method) in underdetermined cases ($K \geq M$).

3.2. Sparse Model

In this subsection we develop the sparse model from (8). Let $\tilde{\boldsymbol{\theta}} = [\tilde{\theta}_1, \dots, \tilde{\theta}_I]$ be a fixed sampling grid in the DOA range, where $I \gg |\mathbb{D}|$. Let $\tilde{\boldsymbol{\theta}}$ be a uniform grid with a grid resolution $r = \tilde{\theta}_2 - \tilde{\theta}_1 \propto I^{-1}$.

Suppose $\tilde{\theta}_{w_k}$ is the nearest grid point to θ_k , where $w_k \in \{1, \dots, I\}$. Then, the steering vector $\mathbf{a}_{\mathbb{D}}(\theta_k)$ can be approximated by its first-order Taylor series

$$\mathbf{a}_{\mathbb{D}}(\theta_k) \approx \mathbf{a}_{\mathbb{D}}(\tilde{\theta}_{w_k}) + \mathbf{a}'_{\mathbb{D}}(\tilde{\theta}_{w_k})(\theta_k - \tilde{\theta}_{w_k}). \quad (9)$$

Denote

$$\tilde{\mathbf{A}}_{\mathbb{D}} = [\mathbf{a}_{\mathbb{D}}(\tilde{\theta}_1), \dots, \mathbf{a}_{\mathbb{D}}(\tilde{\theta}_I)], \quad (10)$$

$$\tilde{\mathbf{A}}'_{\mathbb{D}} = [\mathbf{a}'_{\mathbb{D}}(\tilde{\theta}_1), \dots, \mathbf{a}'_{\mathbb{D}}(\tilde{\theta}_I)]. \quad (11)$$

Denote $\boldsymbol{\beta} = [\beta_1, \dots, \beta_I]^T \in [-r/2, r/2]^I$ and $\boldsymbol{\Phi} = \tilde{\mathbf{A}}_{\mathbb{D}} + \tilde{\mathbf{A}}'_{\mathbb{D}} \text{diag}\{\boldsymbol{\beta}\}$, where for $w = 1, \dots, I$, $\beta_w = \theta_k - \tilde{\theta}_{w_k}$ if $w = w_k$ for any $k = 1, \dots, K$ and $\beta_w = 0$ otherwise, and $\text{diag}\{\cdot\}$ denotes the operator to form a diagonal matrix by a vector or to form a vector by diagonal entries of a matrix. Neglecting approximation errors of (9), model (8) can be expressed as:

$$\mathbf{J}^+ \mathbf{c}_x = \boldsymbol{\Phi} \mathbf{z}, \quad (12)$$

where $\mathbf{z} = [z_1, z_2, \dots, z_I]^T$ is the zero-padded extension of \mathbf{c}_s from $[\theta_1, \dots, \theta_K]$ to $[\tilde{\theta}_1, \dots, \tilde{\theta}_I]$.

Consider the case of existing estimation errors of FOCs. Let $\hat{\mathbf{c}}_x = \mathbf{c}_x + \mathbf{e}$, where $\hat{\mathbf{c}}_x$ is the estimate of \mathbf{c}_x , and \mathbf{e} is the estimation errors vector. Then, model (12) becomes

$$\mathbf{J}^+ \hat{\mathbf{c}}_x = \boldsymbol{\Phi} \mathbf{z} + \mathbf{J}^+ \mathbf{e}. \quad (13)$$

Assume $\mathbf{e} \sim \text{AsCN}(\mathbf{0}_{M^4}, \mathbf{Q})$ [1], where AsCN denotes the asymptotic normal distribution, $\mathbf{0}_{M^4}$ denotes zero vector with M^4 elements, and $\mathbf{Q} \in \mathbb{C}^{M^4 \times M^4}$ is a positive definite (Hermitian) matrix. The estimation of FOCs and the matrix \mathbf{Q} are discussed in Section 3.3. Following this assumption, we have $\mathbf{J}^+ \mathbf{e} \sim \text{AsCN}(\mathbf{0}_{|\mathbb{D}|}, \tilde{\mathbf{Q}})$, where $\tilde{\mathbf{Q}} = \mathbf{J}^+ \mathbf{Q} (\mathbf{J}^+)^T$. Furthermore, we can obtain $\tilde{\mathbf{Q}}^{-\frac{1}{2}} \mathbf{J}^+ \mathbf{e} \sim \text{AsCN}(\mathbf{0}_{|\mathbb{D}|}, \mathbf{I}_{|\mathbb{D}| \times |\mathbb{D}|})$, where $\mathbf{I}_{|\mathbb{D}| \times |\mathbb{D}|}$ is $|\mathbb{D}|$ dimensional identity matrix. We denote by

$$\tilde{\mathbf{c}}_x = \tilde{\mathbf{Q}}^{-\frac{1}{2}} \mathbf{J}^+ \hat{\mathbf{c}}_x \in \mathbb{C}^{|\mathbb{D}|}, \quad (14)$$

$$\tilde{\boldsymbol{\Phi}} = \tilde{\mathbf{Q}}^{-\frac{1}{2}} \boldsymbol{\Phi} = \tilde{\mathbf{Q}}^{-\frac{1}{2}} (\tilde{\mathbf{A}}_{\mathbb{D}} + \tilde{\mathbf{A}}'_{\mathbb{D}} \text{diag}\{\boldsymbol{\beta}\}) \in \mathbb{C}^{|\mathbb{D}| \times I}, \quad (15)$$

$$\tilde{\mathbf{e}} = \tilde{\mathbf{Q}}^{-\frac{1}{2}} \mathbf{J}^+ \mathbf{e} \in \mathbb{C}^{|\mathbb{D}|}. \quad (16)$$

Then, we can obtain the following SMV off-grid sparse model

$$\tilde{\mathbf{c}}_x = \tilde{\boldsymbol{\Phi}} \mathbf{z} + \tilde{\mathbf{e}}, \quad (17)$$

with $\tilde{\mathbf{e}} \sim \text{AsCN}(\mathbf{0}_{|\mathbb{D}|}, \mathbf{I}_{|\mathbb{D}| \times |\mathbb{D}|})$. Note that if setting $\beta_w = 0$ for $w = 1, \dots, I$, the model (17) is simplified to an on-grid DOA model.

3.3. Estimation of FOCs and the Matrix \mathbf{Q}

To facilitate the implementation of the proposed method, in this subsection we present the estimation methods of FOCs and the matrix \mathbf{Q} .

Let us introduce the following general notations: $\mathbf{m}_{4,x}(i_1, i_2, i_3^*, i_4^*) = \mathbb{E}\{x_{i_1}(t)x_{i_2}(t)x_{i_3}^*(t)x_{i_4}^*(t)\}$, $\mathbf{m}_{2,x}(i_1, i_2^*) = \mathbb{E}\{x_{i_1}(t)x_{i_2}^*(t)\}$, and $\mathbf{m}_{2,x}(i_1, i_2) = \mathbb{E}\{x_{i_1}(t)x_{i_2}(t)\}$. Their empirical estimators are $\hat{\mathbf{m}}_{4,x}(i_1, i_2, i_3^*, i_4^*) = \frac{1}{L} \sum_{t=1}^L \{x_{i_1}(t)x_{i_2}(t)x_{i_3}^*(t)x_{i_4}^*(t)\}$, $\hat{\mathbf{m}}_{2,x}(i_1, i_2^*) = \frac{1}{L} \sum_{t=1}^L \{x_{i_1}(t)x_{i_2}^*(t)\}$, and

$\hat{\mathbf{m}}_{2,x}(i_1, i_2) = \frac{1}{L} \sum_{t=1}^L \{x_{i_1}(t)x_{i_2}(t)\}$, respectively, where L is the number of snapshots. Due to the assumptions of sources and noises, $[\mathbf{c}_x]_i$ can be expressed as [25]:

$$\begin{aligned} [\mathbf{c}_x]_i &= \text{cum}\{x_{i_1}(t), x_{i_2}(t), x_{i_3}^*(t), x_{i_4}^*(t)\} \\ &= \mathbf{m}_{4,x}(i_1, i_2, i_3^*, i_4^*) - \mathbf{m}_{2,x}(i_1, i_2)\mathbf{m}_{2,x}^*(i_3, i_4) \\ &\quad - \mathbf{m}_{2,x}(i_1, i_3^*)\mathbf{m}_{2,x}(i_2, i_4^*) - \mathbf{m}_{2,x}(i_1, i_4^*)\mathbf{m}_{2,x}(i_2, i_3^*). \end{aligned} \quad (18)$$

Utilizing the corresponding empirical estimators of terms in (18), $[\mathbf{c}_x]_i$ can be estimated by

$$\begin{aligned} [\hat{\mathbf{c}}_x]_i &= \hat{\mathbf{m}}_{4,x}(i_1, i_2, i_3^*, i_4^*) - \hat{\mathbf{m}}_{2,x}(i_1, i_2)\mathbf{m}_{2,x}^*(i_3, i_4) \\ &\quad - \hat{\mathbf{m}}_{2,x}(i_1, i_3^*)\mathbf{m}_{2,x}(i_2, i_4^*) - \hat{\mathbf{m}}_{2,x}(i_1, i_4^*)\mathbf{m}_{2,x}(i_2, i_3^*). \end{aligned} \quad (19)$$

Consider complex variables $X(t) = x_{i_1}(t)x_{i_2}(t)x_{i_3}^*(t)x_{i_4}^*(t) - x_{i_1}(t)x_{i_2}(t)\sum_{t_0=1}^L x_{i_3}^*(t_0)x_{i_4}^*(t_0)/L - x_{i_1}(t)x_{i_3}^*(t)\sum_{t_1=1}^L x_{i_2}(t_1)x_{i_4}^*(t_1)/L - x_{i_1}(t)x_{i_4}^*(t)\sum_{t_2=1}^L x_{i_2}(t_2)x_{i_3}^*(t_2)/L$ for $t = 1, 2, \dots, L$. It is obvious that $\lim_{L \rightarrow +\infty} \mathbb{E}\{X(t)\} = [\mathbf{c}_x]_i$. According to central limit theorem, as $L \rightarrow +\infty$, the sample average $\bar{X} = \sum_{t=1}^L X(t) = [\hat{\mathbf{c}}_x]_i$ converges in distribution to a complex normal. This solution theoretically supports the assumption that $\mathbf{e} = \hat{\mathbf{c}}_x - \mathbf{c}_x \sim \text{AsCN}(\mathbf{0}_{M^4}, \mathbf{Q})$ in Section 3.2, and we aim for a robust estimation of \mathbf{Q} in the following derivation.

Consider signals given by

$$\mathbf{y}(t) = \mathbf{x}(t) \otimes \mathbf{x}(t) - \mathbb{E}\{\mathbf{x}(t) \otimes \mathbf{x}(t)\}. \quad (20)$$

It is obvious that $\mathbb{E}\{\mathbf{y}(t)\} = \mathbf{0}_{M^2}$. Let $y_g(t)$ denote the g th element of $\mathbf{y}(t)$, $g_1 = (i_1 - 1)M + i_2$, and $g_2 = (i_3 - 1)M + i_4$, then we have

$$\mathbf{m}_{2,y}(g_1, g_2^*) = \mathbf{m}_{4,x}(i_1, i_2, i_3^*, i_4^*) - \mathbf{m}_{2,x}(i_1, i_2)\mathbf{m}_{2,x}^*(i_3, i_4). \quad (21)$$

Rewrite (18) as:

$$\begin{aligned} [\mathbf{c}_x]_i &= \mathbf{m}_{2,y}(g_1, g_2^*) - \mathbf{m}_{2,x}(i_1, i_3^*)\mathbf{m}_{2,x}(i_2, i_4^*) \\ &\quad - \mathbf{m}_{2,x}(i_1, i_4^*)\mathbf{m}_{2,x}(i_2, i_3^*), \end{aligned} \quad (22)$$

where $\mathbf{y}(t)$ is given by (20).

Lemma 1. *Given two joint distributed complex-valued random variables Y_t and Z_t with snapshots index $t = 1, 2, \dots, L$, Y_t and Z_t are independent from snapshot to snapshot, respectively. As L approaches infinity, it holds true that*

$$\text{var}(\overline{Y_t Z_t}) / \text{var}(\overline{Y_t} \overline{Z_t}) = O(L), \quad (23)$$

where $\text{var}(\cdot)$ is variance, and $\overline{Y_t Z_t}$, $\overline{Y_t}$, and $\overline{Z_t}$ are sample averages of $Y_t Z_t$, Y_t , and Z_t with L snapshots, respectively.

Proof. See Appendix D.

Due to Lemma 1, from (19) and (22), it is easy to find that when the snapshots number L is large, the term $\hat{\mathbf{m}}_{2,y}(g_1, g_2^*)$ has much larger variance than the other two terms in (22). This implies that the estimation error of $[\hat{\mathbf{c}}_x]_i$ is mainly caused by the term $\hat{\mathbf{m}}_{2,y}(g_1, g_2^*)$, giving rise to the following approximate expression of e_i :

$$e_i = [\hat{\mathbf{c}}_x]_i - [\mathbf{c}_x]_i \approx \hat{\mathbf{m}}_{2,y}(g_1, g_2^*) - \mathbf{m}_{2,y}(g_1, g_2^*). \quad (24)$$

Since the signals are independent from snapshot to snapshot, and $E\{\mathbf{y}(t)\} = \mathbf{0}_{M^2}$, we can obtain

$$\begin{aligned} E\{e_i e_j^*\} &\approx \frac{1}{L^2} \sum_{t_1=1}^L \sum_{t_2=1}^L E\{[y_{g_1}(t_1)y_{g_2}^*(t_1)][y_{h_1}(t_2)y_{h_2}^*(t_2)]^*\} \\ &\quad - m_{2,y}(g_1, g_2^*)m_{2,y}^*(h_1, h_2^*) \\ &= \text{cum}\{y_{g_1}(t), y_{g_2}^*(t), y_{h_1}^*(t), y_{h_2}(t)\}/L \\ &\quad + m_{2,y}(g_1, h_1^*)m_{2,y}(g_2^*, h_2)/L \\ &\quad + m_{2,y}(g_1, h_2)m_{2,y}(g_2^*, h_1^*)/L, \end{aligned} \quad (25)$$

where $h_1 = (j_1 - 1)M + j_2$, $h_2 = (j_3 - 1)M + j_4$, $j = (h_1 - 1)M^2 + h_2$, and $j_1, j_2, j_3, j_4 = 1, 2, \dots, M$.

In (25), the term $\text{cum}\{y_{g_1}(t), y_{g_2}^*(t), y_{h_1}^*(t), y_{h_2}(t)\}/L$ contains the eighth-order statistics of $\mathbf{x}(t)$, which significantly aggravates the inaccuracy of estimation if the number of snapshots is not large enough. In addition, as the number of snapshots number approaches infinity, we get $E\{e_i e_j^*\} \approx 0$. Therefore, we neglect this term to obtain a robust estimation of \mathbf{Q} , given by

$$\begin{aligned} \hat{\mathbf{Q}}_{i,j} &= \hat{m}_{2,y}(g_1, h_1^*)\hat{m}_{2,y}(g_2^*, h_2)/L \\ &\quad + \hat{m}_{2,y}(g_1, h_2)\hat{m}_{2,y}(g_2^*, h_1^*)/L. \end{aligned} \quad (26)$$

Rewrite (26) in matrix form

$$\hat{\mathbf{Q}} = \hat{\mathbf{R}}_y \otimes \hat{\mathbf{R}}_y^T/L + [\hat{\mathbf{T}}_y \otimes \hat{\mathbf{T}}_{y,1}^*, \hat{\mathbf{T}}_y \otimes \hat{\mathbf{T}}_{y,2}^*, \dots, \hat{\mathbf{T}}_y \otimes \hat{\mathbf{T}}_{y,M^2}^*]/L, \quad (27)$$

where $\hat{\mathbf{R}}_y = \sum_{t=1}^L \mathbf{y}(t)\mathbf{y}^H(t)/L$, $\hat{\mathbf{T}}_y = \sum_{t=1}^L \mathbf{y}(t)\mathbf{y}^T(t)/L$, and $\hat{\mathbf{T}}_{y,h_1}$ is the h_1 th column of $\hat{\mathbf{T}}_y$.

Note that the estimator $\hat{\mathbf{Q}}$ in (30) need to be revised in certain cases. It is clear that, for the noise-free case, $\hat{\mathbf{R}}_y = (\mathbf{A} \otimes \mathbf{A}) \sum_{t=1}^L \{[\mathbf{s}(t)\mathbf{s}^H(t)] \otimes [\mathbf{s}(t)\mathbf{s}^H(t)]\}(\mathbf{A} \otimes \mathbf{A})^H/L$, and $\hat{\mathbf{T}}_y = (\mathbf{A} \otimes \mathbf{A}) \sum_{t=1}^L \{[\mathbf{s}(t)\mathbf{s}^T(t)] \otimes [\mathbf{s}(t)\mathbf{s}^T(t)]\}(\mathbf{A} \otimes \mathbf{A})^T/L$. When $K < M$, the ranks of both matrices $\hat{\mathbf{R}}_y$ and $\hat{\mathbf{T}}_y$ are K^2 . As a result, the estimator (27) may be singular, which may lead to a singular matrix $\hat{\mathbf{Q}}$. This is unallowable for model (17). Therefore, when $K < M$, and SNR is high, to avoid above problems, \mathbf{Q} is estimated by

$$\hat{\mathbf{Q}} = \hat{\mathbf{R}}_y \otimes \hat{\mathbf{R}}_y^T/L + [\hat{\mathbf{T}}_y \otimes \hat{\mathbf{T}}_{y,1}^*, \hat{\mathbf{T}}_y \otimes \hat{\mathbf{T}}_{y,2}^*, \dots, \hat{\mathbf{T}}_y \otimes \hat{\mathbf{T}}_{y,M^2}^*]/L + \epsilon_1 \mathbf{I}_{|\mathbb{D}| \times |\mathbb{D}|}, \quad (28)$$

where ϵ_1 is a small positive number.

Specially, if the observed signals are circular, we have $m_{2,x}(l_1, l_2) = 0$ and $m_{4,x}(l_1, l_2, l_3, l_4) = 0$ for $\forall l_1, l_2, l_3, l_4 = 1, \dots, M$. Then, we obtain $\mathbf{y}(t) = \mathbf{x}(t) \otimes \mathbf{x}(t)$, and \mathbf{c}_i can be estimated by

$$\begin{aligned} [\hat{\mathbf{c}}_x]_i &= \hat{m}_{4,x}(i_1, i_2, i_3^*, i_4^*) - \hat{m}_{2,x}(i_1, i_3^*)\hat{m}_{2,x}(i_2, i_4^*) \\ &\quad - \hat{m}_{2,x}(i_1, i_4^*)\hat{m}_{2,x}(i_2, i_3^*). \end{aligned} \quad (29)$$

Furthermore, it is obvious that $m_{2,y}(g_1, h_2)$ and $m_{2,y}(g_2^*, h_1^*)$ become zero in (25). Then, the estimator of \mathbf{Q} in (27) becomes:

$$\hat{\mathbf{Q}} = \hat{\mathbf{R}}_y \otimes \hat{\mathbf{R}}_y^T/L. \quad (30)$$

Accordingly, when $K < M$ and SNR is high, for circular sources, \mathbf{Q} is estimated by

$$\hat{\mathbf{Q}} = \hat{\mathbf{R}}_y \otimes \hat{\mathbf{R}}_y^T/L + \epsilon_2 \mathbf{I}_{|\mathbb{D}| \times |\mathbb{D}|}, \quad (31)$$

where ϵ_2 is a small positive number.

Remark 3: Approximate expressions for the covariances of FOCs which are more precise than (25) are derived in [5]. However, it is not easy to use in our method in practice, because several higher-order (greater than four) moments need to be estimated, which may cause the estimation

of \mathbf{Q} to be not as robust as FOCs. For example, if the number of snapshots is large enough to generate precise estimates of FOCs, but not large enough to generate ones of higher-order (greater than four) moments, the DOA estimation results may suffer from a mass of outliers in independent experiments. In comparison, (26) contains at most fourth-order statistics, which means that it is more robust and has lower computational cost. We will show that our estimators work well for the proposed method in numerical simulations.

3.4. DOA Estimation by OGSBI Method

In this subsection, the sparse model (17) is solved to obtain the DOAs by the off-grid sparse Bayesian inference (OGSBI) method. Recall the sparse model (17):

$$\tilde{\mathbf{c}}_x = \tilde{\mathbf{\Phi}}\mathbf{z} + \tilde{\mathbf{e}}. \quad (32)$$

In Section 3.3, we showed that the estimation errors of FOCs obey $\mathbf{e} = \hat{\mathbf{c}}_x - \mathbf{c}_x \sim \text{AsCN}(\mathbf{0}_{M^4}, \mathbf{Q})$, and the robust estimators of \mathbf{Q} are also given. As $\tilde{\mathbf{e}} = \tilde{\mathbf{Q}}^{-\frac{1}{2}}\mathbf{J}^+\mathbf{e} \in \mathbb{C}^{|\mathbb{D}|}$, it is obvious that $\tilde{\mathbf{e}} \sim \text{AsCN}(\mathbf{0}_{|\mathbb{D}|}, \mathbf{I}_{|\mathbb{D}| \times |\mathbb{D}|})$. Following the sparse Bayesian formulation in [21], further assume $\tilde{\mathbf{e}} \sim \text{CN}(\mathbf{0}_{|\mathbb{D}|}, \mathbf{I}_{|\mathbb{D}| \times |\mathbb{D}|})$ for simplicity. Then, we have

$$p(\tilde{\mathbf{c}}_x | \mathbf{z}, \boldsymbol{\beta}) = \text{CN}(\tilde{\mathbf{\Phi}}\mathbf{z}, \mathbf{I}_{|\mathbb{D}| \times |\mathbb{D}|}). \quad (33)$$

Adopt the two-stage hierarchical prior for \mathbf{z} that $p(\mathbf{z}; \rho) = \int p(\mathbf{z} | \boldsymbol{\alpha}) p(\boldsymbol{\alpha}; \rho) d\boldsymbol{\alpha}$, in which

$$p(\mathbf{z} | \boldsymbol{\alpha}) = \text{CN}(\mathbf{0}_I, \boldsymbol{\Lambda}), \quad (34)$$

$$p(\boldsymbol{\alpha}; \rho) = \prod_{w=1}^I \Gamma(\alpha_w; 1, \rho), \quad (35)$$

where $\rho > 0$, $\boldsymbol{\alpha} \in \mathbb{R}^I$, $\boldsymbol{\Lambda} = \text{diag}\{\boldsymbol{\alpha}\}$, $\Gamma(\alpha_w; 1, \rho) = \rho e^{-\rho\alpha_w}$ is the probability density function (PDF) of the Gamma distribution with parameters $(1, \rho)$, and α_w is the w th entry of $\boldsymbol{\alpha}$. Here the Gamma hyperprior is assumed for α_w , since it is a conjugate prior of the Gaussian distribution, and it is widely used in SBL techniques and demonstrated with good performance and robustness [13,21,27,28].

Assume a uniform prior for $\boldsymbol{\beta}$: $\boldsymbol{\beta} \sim \mathcal{U}([-r/2, r/2]^I)$. By combining the stages of the hierarchical Bayesian model, the joint PDF is

$$p(\mathbf{z}, \tilde{\mathbf{c}}_x, \boldsymbol{\alpha}, \boldsymbol{\beta}) = p(\tilde{\mathbf{c}}_x | \mathbf{z}, \boldsymbol{\beta}) p(\mathbf{z} | \boldsymbol{\alpha}) p(\boldsymbol{\alpha}) p(\boldsymbol{\beta}). \quad (36)$$

The posterior distribution of \mathbf{z} can be obtained

$$p(\mathbf{z} | \tilde{\mathbf{c}}_x, \boldsymbol{\Lambda}, \boldsymbol{\beta}) = \text{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (37)$$

with

$$\boldsymbol{\mu} = \boldsymbol{\Sigma} \tilde{\mathbf{\Phi}}^H \tilde{\mathbf{c}}_x \quad (38)$$

$$\boldsymbol{\Sigma} = (\tilde{\mathbf{\Phi}}^H \tilde{\mathbf{\Phi}} + \boldsymbol{\Lambda}^{-1})^{-1} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda} \tilde{\mathbf{\Phi}}^H (\mathbf{I} + \tilde{\mathbf{\Phi}} \boldsymbol{\Lambda} \tilde{\mathbf{\Phi}}^H)^{-1} \tilde{\mathbf{\Phi}} \boldsymbol{\Lambda}. \quad (39)$$

As a single snapshot case of the OGSBI method, we can obtain the following updates of $\boldsymbol{\alpha}$ from the deduction in [21]:

$$\alpha_w^{\text{new}} = (\sqrt{1 + 4\rho E\{|z_w|^2\}} - 1)/(2\rho), \quad (40)$$

where $E\{|z_w|^2\} = |\mu_w|^2 + \Sigma_{ww}$, μ_w is the w th entry of $\boldsymbol{\mu}$, and Σ_{ww} is (w, w) th entry of $\boldsymbol{\Sigma}$. For $\boldsymbol{\beta}$, by maximizing $E\{\log p(\tilde{\mathbf{c}}_x | \mathbf{z}, \boldsymbol{\beta}) p(\boldsymbol{\beta})\}$, we can obtain the updates

$$\boldsymbol{\beta}^{\text{new}} = \arg \min_{\boldsymbol{\beta} \in [-r/2, r/2]^L} \{\boldsymbol{\beta}^T \mathbf{P} \boldsymbol{\beta} - 2\mathbf{q}^T \boldsymbol{\beta}\}, \quad (41)$$

where $\mathbf{P} = \Re\{[(\tilde{\mathbf{A}}_{\mathbb{D}}')^H \tilde{\mathbf{A}}_{\mathbb{D}}']^* \odot (\boldsymbol{\mu} \boldsymbol{\mu}^H + \boldsymbol{\Sigma})\}$, $\mathbf{q} = \Re\{\text{diag}\{\boldsymbol{\mu}^*\} (\tilde{\mathbf{A}}_{\mathbb{D}}')^H (\tilde{\mathbf{c}}_x - \tilde{\mathbf{A}}_{\mathbb{D}} \boldsymbol{\mu})\} - \Re\{\text{diag}\{(\tilde{\mathbf{A}}_{\mathbb{D}}')^H \tilde{\mathbf{A}}_{\mathbb{D}} \boldsymbol{\Sigma}\}\}$, $\Re(\cdot)$ takes the real part of a complex variable and \odot is the Hadamard product. An easier way to estimate $\boldsymbol{\beta}$ is provided in [21]. We use $\boldsymbol{\beta}$, \mathbf{P} , and \mathbf{q} hereafter to denote their truncated versions for simplicity. By (41) and $\frac{\partial}{\partial \boldsymbol{\beta}} \{\boldsymbol{\beta}^T \mathbf{P} \boldsymbol{\beta} - 2\mathbf{q}^T \boldsymbol{\beta}\} = 2(\mathbf{P} \boldsymbol{\beta} - \mathbf{q})$, we have $\boldsymbol{\beta}^{\text{new}} = \check{\boldsymbol{\beta}}$ if \mathbf{P} is invertible and $\check{\boldsymbol{\beta}} = \mathbf{P}^{-1} \mathbf{q} \in [-r/2, r/2]^K$. Otherwise, we update $\boldsymbol{\beta}$ elementwise. That is, at each step we update one β_w by fixing up the other entries of $\boldsymbol{\beta}$. For $w = 1, \dots, K$, we first let $\check{\boldsymbol{\beta}} = \frac{q_k - (\mathbf{P}_k)_{-k}^T \boldsymbol{\beta}_{-k}}{P_{kk}}$, where $\boldsymbol{\beta}_{-k}$ denotes $\boldsymbol{\beta}$ without the k th entry. Then, by constraining $\beta_k \in [-r/2, r/2]$, we have

$$\beta_k^{\text{new}} = \begin{cases} \check{\beta}_k, & \text{if } \check{\beta}_k \in [-r/2, r/2]; \\ -r/2, & \text{if } \check{\beta}_k < -r/2; \\ r/2, & \text{otherwise.} \end{cases} \quad (42)$$

The OGSBI is terminated if $\|\boldsymbol{\alpha}^{\kappa+1} - \boldsymbol{\alpha}^{\kappa}\|_2 / \|\boldsymbol{\alpha}^{\kappa}\|_2 < \tau$ or the maximum number of iterations is reached, where κ is the iteration and τ is the predefined tolerance parameter. Then, $\boldsymbol{\alpha}$ can be treated as the pseudo-spectrum

$$\mathcal{P}_w = \alpha_w. \quad (43)$$

By finding the grid indices of highest K peaks of \mathcal{P} , denoted by $\hat{p}_k, k = 1, \dots, K$, we can obtain K estimated DOAs by

$$\hat{\theta}_k = \hat{\theta}_{w_k} + \hat{\beta}_{w_k}. \quad (44)$$

Note that the parameter ρ in the method should be given in advance. According to associated information in [21] and our testing results, the OGSBI method is insensitive to ρ if ρ is not too large. Furthermore, we have observed that by setting $\rho = 0.01$, the sparse Bayesian learning methods in [13,21,28] always achieve a good performance in a wide range of scenarios. By contrast, the 4-L1 method [16–18] is sensitive to the parameter of allowable error bound, which is not easy to estimate (it was chosen to give the best results through trial-and-error for each scenario in the corresponding simulations in [16–18]). Therefore, compared with the 4-L1 method, the OGSBI method has the advantage of convenience in setting parameters.

Finally, we summarize our method in Algorithm 1.

Algorithm 1 The proposed sparse method using denoised FOCs vector

- 1) Input: \mathbf{x} , \mathbb{S} , and d .
 - 2) Initialization: r , $\boldsymbol{\alpha}$, ρ , ϵ , and τ .
 - 3) According to the associated scenarios, calculate $\hat{\mathbf{c}}_x$ by (19) or (29).
 - 4) According to the associated scenarios, calculate $\hat{\mathbf{Q}}$ by (27), (28), (30), or (31).
 - 5) Calculate \mathbf{J} , $\tilde{\mathbf{A}}_{\mathbb{D}}$, $\tilde{\mathbf{A}}_{\mathbb{D}}'$, and $\tilde{\mathbf{c}}_x$ by Definition 2, (10), (11), and (14), respectively.
 - 6) Repeat the following:
 - a) Calculate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by (38) and (39), respectively.
 - b) Update $\boldsymbol{\alpha}$ by (40).
 - c) Update $\boldsymbol{\beta}$ by (42).
 - d) If $\|\boldsymbol{\alpha}^{\kappa+1} - \boldsymbol{\alpha}^{\kappa}\|_2 / \|\boldsymbol{\alpha}^{\kappa}\|_2 < \tau$ or the maximum number of iterations is reached, iteration terminates.
 - 7) Find the grid indices of the highest peaks of (43), and output the DOAs by (44).
-

3.5. Identifiability of the Number of Sources

According to the theorems about parameter identifiability in [29–31] and the model (8), we have the following proposition.

Proposition 1. Any K sources can be uniquely identified from model (17) if and only if

$$K < \text{spark}\{\mathbb{B}_\theta\}/2, \quad (45)$$

where $\text{spark}\{\mathbb{B}_\theta\}$ is defined as the smallest number of elements in \mathbb{B}_θ that are linearly dependent [32], $\mathbb{B}_\theta = \{\mathbf{b}(\theta) : \theta \in \Theta\} \subset \mathbb{C}^{|\mathbb{D}|}$, and Θ is $[-\pi/2, \pi/2)$ for linear array and $[-\pi, \pi)$ for planar array.

It is generally difficult to compute $\text{spark}\{\mathbb{B}_\theta\}$, except for when \mathbb{D} is a uniform linear array (ULA). In this circumstance, \mathbb{D} is hole-free, $\text{spark}\{\mathbb{B}_\theta\} = |\mathbb{D}| + 1$, and then $K < (|\mathbb{D}| + 1)/2$ sources can be identified. However, hole-free FODCA with $O(M^4)$ DOFs has not been studied so far to the best of our knowledge. Recently, [16–18] provided several methods to construct FODCA with larger ULA segments, with which $K < (|\mathbb{D}_{\text{ULA}}| + 1)/2$ sources can be identified, where \mathbb{D}_{ULA} is defined as the largest ULA segment in \mathbb{D} .

For general arrays, because $\mathbb{B}_\theta \subset \mathbb{C}^{|\mathbb{D}|}$, it is obvious that

$$\text{spark}\{\mathbb{B}_\theta\} \leq |\mathbb{D}| + 1. \quad (46)$$

According to Property 3, (46), and Proposition 1, we can obtain a necessary condition regarding the identifiability of the number of sources.

Theorem 1. Consider model (17), a necessary condition to uniquely identify all sources is that the number of sources K fulfills

$$K \leq (M^4 - 2M^3 + 7M^2 - 6M)/8. \quad (47)$$

Proof. See Appendix E.

In other words, any more than $(M^4 - 2M^3 + 7M^2 - 6M)/8$ sources cannot be uniquely identified by model (17).

In addition, there are $O(M^4)$ distinct elements in \mathbb{D}_{ULA} for certain sparse linear arrays (SLAs), such as four-level nested arrays [8] and arrays constructed from an expanding and shift scheme which consists of coprime arrays or nested arrays [18]. Therefore, our proposed method can identify $O(M^4)$ sources at most, if these arrays (not limited to) are used.

3.6. Complexity

Consider the worst case of complexity. For non-circular sources, the estimation of \mathbf{c}_x using (19) needs $O(LM^4)$ multiplications. The estimation of \mathbf{Q} using (27) or (28) needs $O(LM^8)$ multiplications. Note that all the symmetries of \mathbf{c}_x and \mathbf{Q} are not taken into account here, while it does not affect the order of complexity. Each iteration of (38) and (39) needs $O(I^2|\mathbb{D}|)$ multiplications, which are much more than each iteration of (40) and (42). Let T denote the iteration number, and consider the worst case that $O(|\mathbb{D}|) = O(M^4)$. Combining all of the above processes, we can obtain that the complexity of the proposed method is $O(LM^8 + I^2M^4T)$.

3.7. Comparison with State-of-the-Art Methods

In this subsection, we compare the proposed method with state-of-the-art FOCs-based ones, in terms of required geometry, identifiability, prior knowledge of the number of sources, ability to

handle correlated sources, numerical complexity, and sensitivity to parameters. Associated methods are listed as follows:

- 4-MUSIC [5,6]: non-redundant version with averaging strategy [7], denoted by A-4-MUSIC;
- 4-SS-MUSIC [8]: with averaging strategy, denoted by A-4-SS-MUSIC;
- 4-L1 [16–18].

Associated comparison results are illustrated in Table 1.

Table 1. Comparison with state-of-the-art methods. 4-L1: fourth-order cumulants (FOCs) vector-based L1-norm method; A-4-MUSIC: non-redundant version of 4-MUSIC with averaging strategy; A-4-SS-MUSIC: FOCs vector-based spatial smoothing (SS)-MUSIC with averaging strategy; FODCA: fourth-order difference co-array; ULA: uniform linear array.

	A-4-MUSIC	A-4-SS-MUSIC	4-L1	Proposed
Required geometry	Any geometry	LA, with FODCA-containing ULA segment	Any geometry	Any geometry
Maximum identifiability	$O(M^2)$	$O(M^4)$	$O(M^4)$	$O(M^4)$
Prior knowledge of the number of sources	Need	Need	No need	No need
Handles correlated sources	Yes	No	No	No
Parameters and sensitivity	With no parameters	With no parameters	With sensitive parameters	With insensitive parameters
Complexity	$O(LM^4 + IM^4 + M^6)$	$O(LM^4 + IM^8 + M^{12})$	$O(LM^4 + I^3T_2)$	$O(LM^8 + I^2M^4T_1)$

In Table 1, T_1 and T_2 denote the iteration number of the proposed method and 4-L1, respectively. All comparison results are obvious or were discussed in the previous subsections, except for the comparison of complexity. Recall that M is the number of sensors, L is the number of snapshots, and I is the number of grid points. It should be noted that the complexities of four methods are derived under the assumption that $O(|\mathbb{D}|) = O(M^4)$, which gives the best identifiability and highest complexities for all methods. Due to the condition $I \gg |\mathbb{D}|$ required to perform the sparsity in the proposed method and 4-L1, it is clear that $O(I) \geq O(M^4)$, which results in $O(I^3T_2) > O(IM^8)$ and $O(I^2M^4T_1) > O(IM^8)$. Consequently, the proposed method and 4-L1 always have higher complexity than A-4-MUSIC and A-4-SS-MUSIC.

Next, we make a comparison of complexity between the proposed method and 4-L1. Firstly, if the number of snapshots L is large enough that the terms $O(I^3T_2)$ and $O(I^2M^4T_1)$ can be negligible, the proposed method has higher complexity than 4-L1 due to the term $O(LM^8)$ which is caused by the estimation of \mathbf{Q} . Secondly, if L is small, we limit the comparison between $O(I^3T_2)$ and $O(I^2M^4T_1)$. It is clear that $O(I^3) \geq O(I^2M^4)$, implying that the proposed method has lower complexity than 4-L1 in one iteration. However, it is a fact that 4-L1 always has less iterations than the proposed method (i.e., $T_1 > T_2$). Thus, it is not easy to draw a conclusion. In fact, to achieve the super resolution, 4-L1 needs a dense grid, while the proposed method only needs a coarser grid due to the off-grid parameter estimation, which gives our method an advantage for complexity comparison. It should be noted that off-grid parameter estimation technique has been successfully applied to L1-norm minimization methods in [33,34]. However, to the best of our knowledge, the off-grid version of the 4-L1 method has not been published, and it is beyond the scope of this paper.

4. Numerical Simulations

Numerical simulations were carried out to demonstrate the superior performance of the proposed method. In the following simulations, the sampled analytic signals of statistically independent

quaternary phase-shift keying (QPSK) signals were used as circular sources. The RMSE was chosen as the performance criterion, and is computed by

$$RMSE = \sqrt{\frac{1}{FK} \sum_{f=1}^F \sum_{k=1}^K |\hat{\theta}_{k,f} - \theta_k|^2}, \quad (48)$$

where F is the number of independent experiments, and $\hat{\theta}_{k,f}$ is the estimate of θ_k in the f th independent experiment. The RMSE performances of the four methods in Table 1 with respect to number of snapshots, grid, adjacent sources, SNR, modelling errors, geometry, and circularity of sources were further investigated. In the proposed method, we set that $\rho = 0.01$ and $\tau = 10^{-4}$ for all scenarios, and set $\epsilon_1 = 10^{-5}$ or $\epsilon_2 = 10^{-5}$ for the case that $K < M$ and $SNR \geq 8$ dB. 4-L1 was solved by the SeDuMi toolbox [35], and the associated allowable error bound was chosen to give the best results through trial-and-error for each scenario as done in [16–18]. Each simulation executed $F = 200$ independent experiments for all methods.

4.1. Identifiability

Consider an SLA with three sensors located at $\mathbb{S} = \{1, 2, 6\} \times d$. It can be deduced that the associated FODCA has 19 distinct elements. That is, $|\mathbb{D}| = 19$, which is equal to $\sup\{\mathbb{H}\}$ when $M = 3$. According to Theorem 1, any greater than nine sources cannot be uniquely identified for the proposed method. Let us consider nine sources with DOAs $\{-60^\circ, -45^\circ, -30^\circ, -15^\circ, 0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ\}$. The number of snapshots and SNR were set as $L = 10^4$ and $SNR = 20$ dB, respectively. Grid resolution was set to $r = 0.5^\circ$. Figure 1a,b shows the normalized spectra of the proposed method and 4-L1, respectively. As can be seen, all DOAs of the sources were identified successfully. Note that this was unachievable for A-4-MUSIC and A-4-SS-MUSIC. The virtual array of A-4-MUSIC contained seven distinct virtual sensors, which could identify six sources at most. The largest ULA segment of the FODCA contained 13 distinct elements, giving rise to a maximum identifiable number of six for A-4-SS-MUSIC.

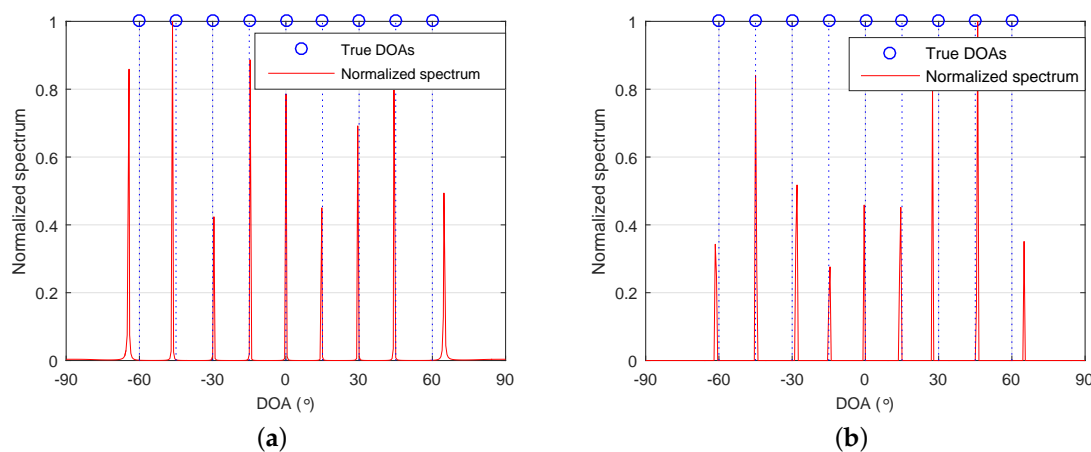


Figure 1. Normalized spectra of (a) the proposed method and (b) 4-L1. $\mathbb{S} = \{1, 2, 6\} \times d$, $K = 9$, $L = 10^4$, signal–noise ratio (SNR) = 20, and $r = 0.5^\circ$. DOA: direction of arrival.

4.2. Impact of the Grid Resolution

Consider an SLA $\{1, 2, 4, 11, 25\} \times d$, which is constructed from an expanding and shift scheme that consists of two nested arrays $\{1, 2, 4\} \times d$ and $\{1, 2, 4\} \times d$ [18]. It can be obtained that the number of distinct elements of the corresponding FODCA and associated largest ULA segment is $|\mathbb{D}| = 85$ and

$|\mathbb{D}|_{\text{ULA}} = 55$, respectively. Consider two independent sources with DOAs $\{0^\circ, 20^\circ\} + \zeta_1$, where ζ_1 is a random variable chosen uniformly within $[-1^\circ, 1^\circ]$. SNR was set at $\text{SNR} = 3$ dB.

Figure 2a,b shows the RMSE performance of the four methods with respect to number of snapshots, with grid resolution $r = 0.5^\circ$ and $r = 0.1^\circ$, respectively. In Figure 2a, when snapshots number was large, the RMSEs of A-4-MUSIC, A-4-SS-MUSIC, and 4-L1 did not reduce as snapshots increased, mainly because of the mismatch errors between grid and true DOAs. Due to the off-grid parameters estimation, the proposed method outperformed the other three methods for most of simulated snapshot numbers. In Figure 2b, the grid resolution is $r = 0.1^\circ$, and it is clear that the RMSEs of A-4-MUSIC, A-4-SS-MUSIC, and 4-L1 increased to the same level of the proposed method when snapshots were greater than 50 and less than 2000. Nonetheless, the proposed method still outperformed the other three methods when snapshots were more than or equal to 2000. It can also be observed that the proposed method had smaller RMSE than A-4-MUSIC and 4-L1 when snapshots were less than 100. The RMSEs of A-4-SS-MUSIC were slightly larger than the other three methods when snapshots were greater than 50 and less than 5000, which may be caused by the fact that 4-SS-MUSIC does not utilize all the virtual sensors in associated FODCA. In addition, the dashed line in Figure 2b gives the RMSEs of the proposed method for $r = 0.5^\circ$, and it is clear that the grid resolution seldom affected the RMSEs of the proposed method when snapshots were less than 2000, which implies that when snapshots are not too large, the proposed method can reduce the computational complexity by selecting a relatively coarser grid without heavy loss of RMSE performance.

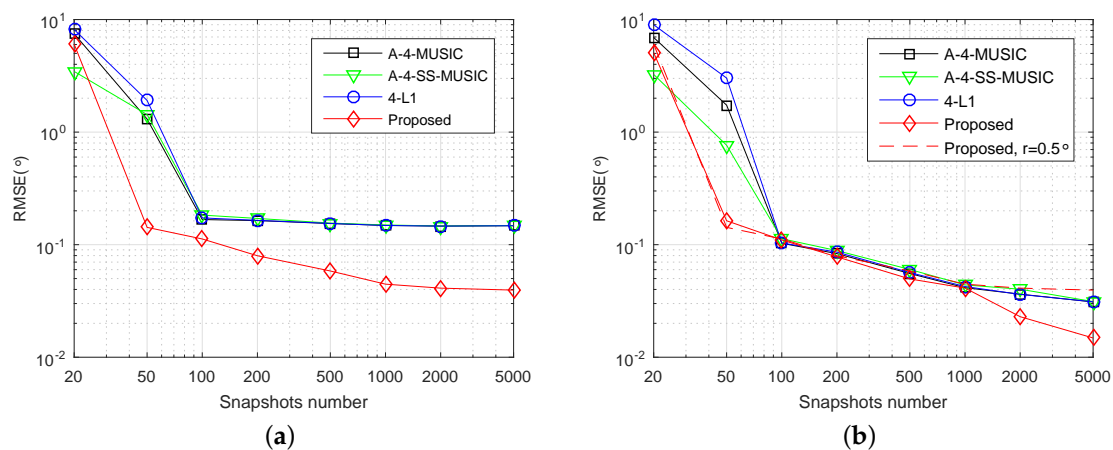


Figure 2. Root mean square error (RMSE) performance of four methods versus number of snapshots. $\mathbb{S} = \{1, 2, 4, 11, 25\} \times d$, $K = 2$, $\text{SNR} = 3$ dB, (a) $r = 0.5^\circ$ and (b) $r = 0.1^\circ$.

4.3. Identifiability for Adjacent Sources

Next, we investigated the identifiability of adjacent sources using the proposed method. The array and parameters were set as in Section 4.2, except the grid resolution was set to $r = 0.1^\circ$, and the DOAs of sources were set to $\{\zeta_2, \zeta_2 + \Delta\theta\}$, where ζ_2 is a random variable chosen uniformly within $[-1^\circ, 1^\circ]$, and $\Delta\theta = 0.5^\circ, 1^\circ, \dots, 5^\circ$ are the different DOA intervals. Figure 3a,b shows the RMSE performance of four methods with different DOA intervals, when the number of snapshots is 300 and SNR is 3 dB, and when the number of snapshots is 1000 and SNR is 10 dB, respectively. It is clear that the proposed method had smallest RMSEs for most simulated DOA intervals in both simulations, demonstrating the superior identifiability of adjacent sources with the proposed method.

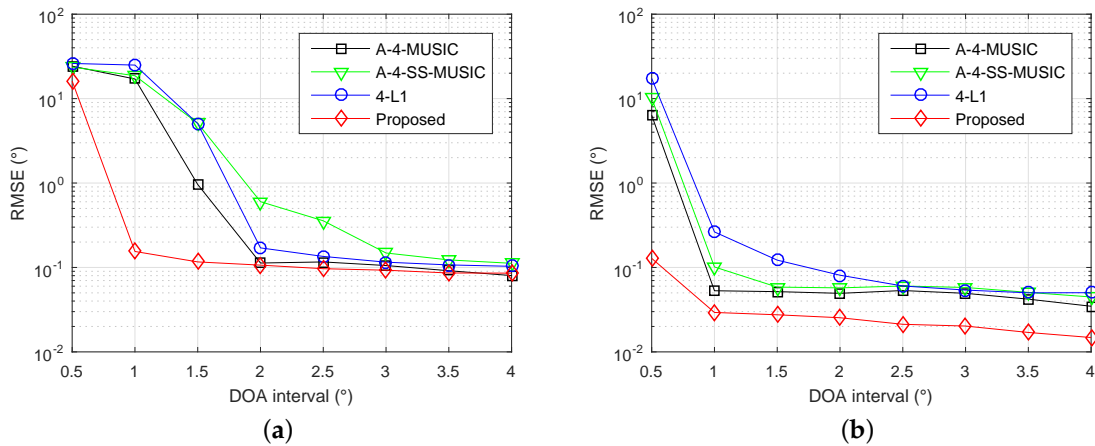


Figure 3. RMSE performance of four methods with different DOA intervals, $\mathcal{S} = \{1, 2, 4, 11, 25\} \times d$, $K = 2$, $r = 0.1^\circ$, (a) $L = 300$ and $SNR = 3$ dB, (b) $L = 1000$ and $SNR = 10$ dB.

4.4. Impact of SNR

The array and parameters were set as in Section 4.2, except the grid resolution was set to $r = 0.1^\circ$. Figure 4 plots the RMSEs of four methods as a function of SNR, where the number of snapshots was set to $L = 500$. It is clear that the four methods had almost equal RMSEs when SNR was smaller than 4 dB. In the high SNR region, a relatively better performance of the proposed method appeared, benefiting from the off-grid parameter estimation as discussed in the large snapshot number region in Section 4.2.

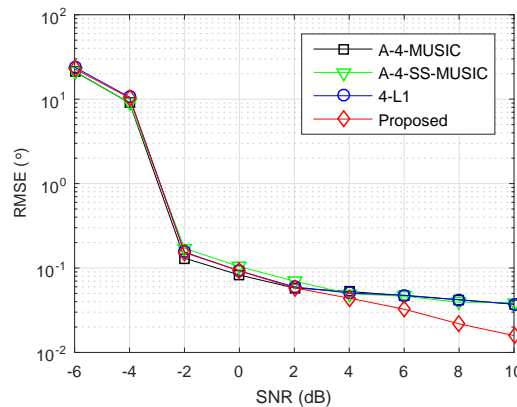


Figure 4. RMSE performance of four methods versus SNR, $\mathcal{S} = \{1, 2, 4, 11, 25\} \times d$, $K = 2$, $r = 0.1^\circ$, $L = 500$.

4.5. Impact of Modelling Errors

We also studied the effect of modelling errors on the proposed method. We used the same array geometry and parameters, and the same way of generating $K = 2$ DOAs as in Section 4.4. We perturbed the steering vector by adding a modelling error vector $\mathbf{e}_{\mathbf{a}(\theta_k)}$ to get the perturbed steering vector as $\tilde{\mathbf{a}}(\theta_k) = \mathbf{a}(\theta_k) + \mathbf{e}_{\mathbf{a}(\theta_k)}$, where $k = 1, \dots, K$, and $\mathbf{e}_{\mathbf{a}(\theta_k)}$ are assumed to be zero mean statistically independent Gaussian distributed with $E\{\mathbf{e}_{\mathbf{a}(\theta_k)} \mathbf{e}_{\mathbf{a}(\theta_k)}^H\} = \sigma_{\mathbf{a}(\theta_k)}^2 \mathbf{I}$, where $\sigma_{\mathbf{a}(\theta_k)}^2 = 0.0025$. Figure 5 shows the RMSE performance of four methods with such modelling errors. In order to facilitate comparison, the RMSEs of the proposed methods with no modelling errors are also plotted in a dashed line. It can be observed that the RMSE performance of the four methods did not degrade heavily. However, when snapshots were more than 1000, the superior performance of the proposed method disappeared,

suggesting that modelling errors did have a large impact on the off-grid parameters estimation of the proposed method.

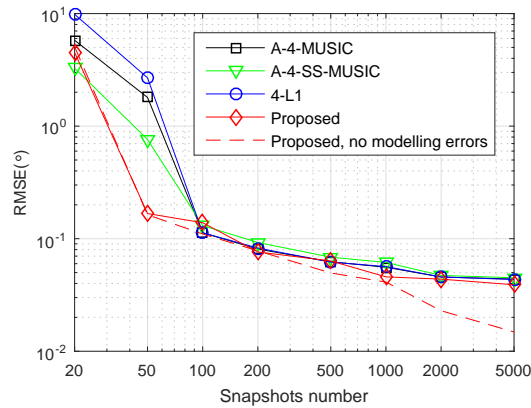


Figure 5. RMSE performance of four methods with modelling errors, $\mathbb{S} = \{1, 2, 4, 11, 25\} \times d$, $K = 2$, $r = 0.1^\circ$, $SNR = 3$ dB.

4.6. Underdetermined Case

We investigated the performance of the proposed method in underdetermined cases. The same array $\{1, 2, 4, 11, 25\} \times d$ was used, and the grid resolution $r = 0.1^\circ$, while six independent sources with DOAs $\{-45^\circ, -30^\circ, -15^\circ, 0^\circ, 25^\circ, 40^\circ\}$ were set. Figure 6 shows the RMSE performance of the four methods. In Figure 6a, the number of snapshots varying from 100 to 10,000 and the SNR was fixed at 20 dB. It is clear that the proposed method had the smallest RMSEs for most of the simulated snapshot numbers. In Figure 6b, SNR varied from -4 dB to 28 dB and the snapshots number was fixed at 1000. It can be observed that the proposed method had smaller RMSEs when $SNR > 4$ dB, but had slightly larger values when $SNR \leq 4$ dB, when compared with the other three methods. The RMSE performance of A-4-SS-MUSIC was obviously outperformed by the other three methods in the large snapshot number and high SNR regions, because 4-SS-MUSIC does not utilize all the virtual sensors in the associated FODCA. Moreover, the RMSEs of each method seemed to converge to a positive constant as SNR increased. A similar phenomenon can be observed and explained in covariance vector-based methods in the underdetermined case [22,26].

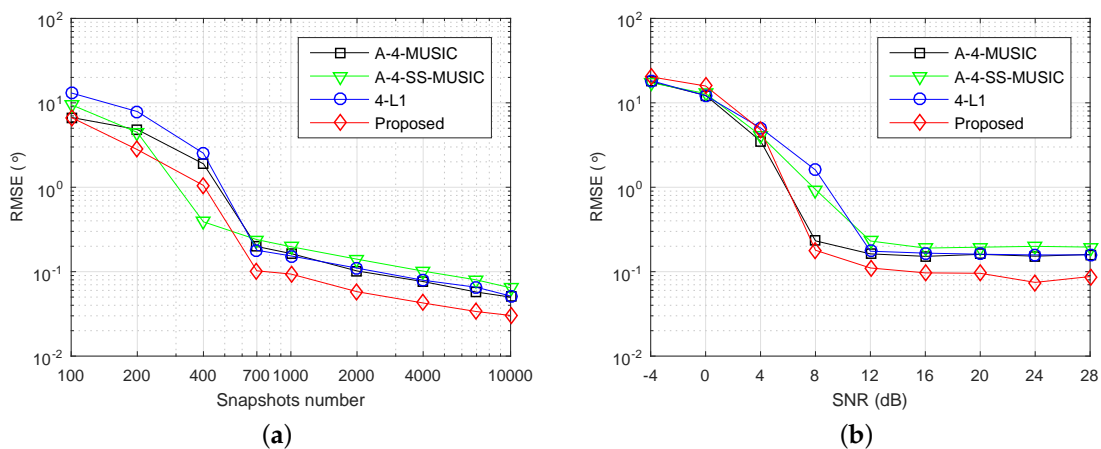


Figure 6. RMSE performance of four methods in the underdetermined case, $\mathbb{S} = \{1, 2, 4, 11, 25\} \times d$, $K = 6$, $r = 0.1^\circ$, (a) $SNR = 20$ dB, (b) snapshots number is $L = 1000$.

4.7. Non-Linear Array Case

We now give an example to show the effectiveness of the proposed method in the non-linear array case. We randomly generated a planar array as shown in Figure 7a, where the minimum distance between sensors was half of the wavelength. The corresponding FODCA geometry is also illustrated, which had $|\mathbb{D}| = 131$ distinct elements. Note that the A-4-SS-MUSIC cannot be applied for the non-linear array case. We set three sources with DOAs $\{-60^\circ, 0^\circ, 40^\circ\} + \zeta_3$, where ζ_3 is a random variable chosen uniformly within $[-1^\circ, 1^\circ]$. SNR was set at $SNR = 6$ dB, and the RMSEs of A-4-MUSIC, 4-L1, and the proposed method were plotted as a function of the number of snapshots (Figure 7b). It can also be seen that the proposed method had better RMSE performance when snapshots were greater than 200.

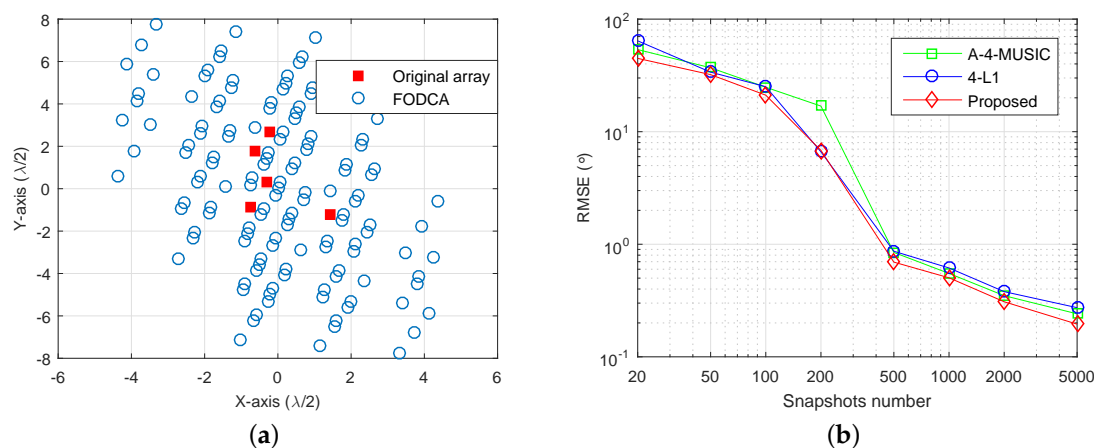


Figure 7. RMSE performance of three methods in the non-linear array case. $M = 5$, $K = 3$, $r = 0.1^\circ$, (a) the array geometry and the corresponding FODCA geometry, (b) RMSE versus the number of snapshots, $SNR = 6$ dB.

4.8. Effectiveness for Non-Circular Sources

Finally, we consider the case of non-circular sources, using the same array $\{1, 2, 4, 11, 25\} \times d$, and six independent sources with DOAs $\{-45^\circ, -30^\circ, -15^\circ, 0^\circ, 25^\circ, 40^\circ\}$ as in Section 4.6. In contrast, we replaced the imaginary parts of sources by their real parts. That is, $\mathbf{s}_{NC}(t) = \Re\{\mathbf{s}(t)\} + j\Re\{\mathbf{s}(t)\}$, where $\mathbf{s}_{NC}(t)$ is the non-circular sources to be simulated. We set $SNR = 20$ dB and $r = 0.1^\circ$. We estimated the FOCs by (19) for all the methods, and estimated $\hat{\mathbf{Q}}$ by (27) for the proposed method. Figure 8 shows the RMSEs of the four methods versus the number of snapshots. Similar to Figure 6a, it is clear that the proposed method had the smallest RMSEs for all simulated snapshot numbers. Beyond doubt, all the methods were effective for non-circular sources with the estimation of FOCs by (19).

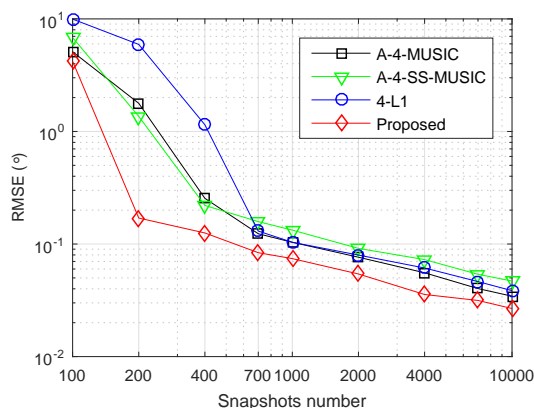


Figure 8. RMSE performance of four methods with non-circular sources, $M = 5$, $K = 6$, $r = 0.1^\circ$, $SNR = 6$ dB.

5. Conclusion

In this paper, a novel SMV sparse model for non-Gaussian sources using denoised FOCs vector is established to perform DOA estimation, and solved efficiently by the OGSBI method. An advanced denoising and dimension reduction procedure of FOCs vector is provided. The estimation errors of FOCs are integrated in the proposed SMV model, and are approximately estimated in a simple way. The proposed method suits any geometry, does not need prior knowledge of the number of sources, has maximum identifiability $O(M^4)$, and is insensitive to associated parameters. The off-grid parameter estimation in our method promotes superior performance when the number of snapshots is large and SNR is high, and also ensures good performance when choosing a relatively coarse grid. Numerical simulations illustrate that the proposed method has superior identifiability or RMSE performance in different scenarios, namely grid resolution, adjacent sources, SNR, modelling errors, geometry, and non-circular sources, when compared with other state-of-the-art methods.

Author Contributions: J.W. designed the method and wrote the paper; J.W. and Y.F. conceived and designed the simulations; R.D. performed the experiments; J.W., Y.F., and G.L. analyzed the simulation results; R.D. contributed SeDuMi toolbox in MATLAB.

Funding: This research was funded by State Commission of Science Technology of China (grant no. 17H863-05-2T-002-028-02) and Foundation of Key Laboratory of Underwater Acoustic Countermeasure (grant no. kmb5494).

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Appendix A. The Proof of Property 1

Proof. Consider the linear equation $\sum_{\zeta=1}^{|\mathbb{D}|} v_{\zeta} \mathbf{J}_{\zeta} = \mathbf{0}$, where \mathbf{J}_{ζ} is the ζ th column vector of \mathbf{J} . For any $\mathbf{n}_{i_1}, \mathbf{n}_{i_2}, \mathbf{n}_{i_3}, \mathbf{n}_{i_4} \in \mathbb{S}$, it can be deduced that $0 = [\sum_{\zeta=1}^{|\mathbb{D}|} v_{\zeta} \mathbf{J}_{\zeta}]_i = \sum_{\zeta=1}^{|\mathbb{D}|} v_{\zeta} \mathbf{J}_{i,\zeta} = v_{\zeta_0}$, where the index ζ_0 satisfies that associated vector $\mathbf{m}_{\zeta_0} = \mathbf{n}_{i_1} + \mathbf{n}_{i_2} - \mathbf{n}_{i_3} - \mathbf{n}_{i_4}$. Due to the arbitrariness of $\mathbf{n}_{i_1}, \mathbf{n}_{i_2}, \mathbf{n}_{i_3}, \mathbf{n}_{i_4} \in \mathbb{S}$, the coefficients $v_{\zeta} = 0$ for all $v_{\zeta} = 1, 2, \dots, |\mathbb{D}|$, implying that \mathbf{J} has full column rank. The proof is completed. \square

Appendix B. The Proof of Property 2

Proof. It is obvious that $\mathbf{J}_{\mathbb{D}}(\theta) = \sum_{\zeta=1}^{|\mathbb{D}|} \mathbf{J}_{\zeta} e^{j2\pi \mathbf{u}^T \mathbf{m}_{\zeta}}$. According to Definition 2, we can obtain $[\sum_{\zeta=1}^{|\mathbb{D}|} \mathbf{J}_{\zeta} e^{j2\pi \mathbf{u}^T \mathbf{m}_{\zeta}}]_i = e^{j2\pi \mathbf{u}^T (\mathbf{n}_{i_1} + \mathbf{n}_{i_2} - \mathbf{n}_{i_3} - \mathbf{n}_{i_4})}$. On the other hand, the i th row of $\mathbf{b}(\theta)$ is $[\mathbf{b}(\theta)]_i =$

$a_{i_1}(\theta)a_{i_2}(\theta)a_{i_3}^*(\theta)a_{i_4}^*(\theta) = e^{j2\pi\mathbf{u}^T(\mathbf{n}_{i_1}+\mathbf{n}_{i_2}-\mathbf{n}_{i_3}-\mathbf{n}_{i_4})}$. Thus, we have $\mathbf{b}(\theta) = \mathbf{J}_{\mathbb{D}}(\theta)$. The proof is completed. \square

Appendix C. The Proof of Property 3

Proof. Consider Definition 1. For groups $\{\mathbf{n}_{i_1}, \mathbf{n}_{i_2}\}$ and $\{\mathbf{n}_{i_3}, \mathbf{n}_{i_4}\}$, we have two cases as follows.

Case 1: the same element does not exist between $\{\mathbf{n}_{i_1}, \mathbf{n}_{i_2}\}$ and $\{\mathbf{n}_{i_3}, \mathbf{n}_{i_4}\}$. If $\mathbf{n}_{i_1} = \mathbf{n}_{i_2}$ and $\mathbf{n}_{i_3} = \mathbf{n}_{i_4}$, there are $\binom{M}{1}\binom{M-1}{1}$ distinct elements at most; if $\mathbf{n}_{i_1} = \mathbf{n}_{i_2}$ and $\mathbf{n}_{i_3} \neq \mathbf{n}_{i_4}$, there are $\binom{M}{1}\binom{M-1}{2}$ distinct elements at most; if $\mathbf{n}_{i_1} \neq \mathbf{n}_{i_2}$ and $\mathbf{n}_{i_3} = \mathbf{n}_{i_4}$, there are $\binom{M}{2}\binom{M-2}{1}$ distinct elements at most; if $\mathbf{n}_{i_1} \neq \mathbf{n}_{i_2}$ and $\mathbf{n}_{i_3} \neq \mathbf{n}_{i_4}$, there are $\binom{M}{2}\binom{M-2}{2}$ distinct elements at most.

Case 2: the same element exists between $\{\mathbf{n}_{i_1}, \mathbf{n}_{i_2}\}$ and $\{\mathbf{n}_{i_3}, \mathbf{n}_{i_4}\}$. Without loss of generality, assuming $\mathbf{n}_{i_2} = \mathbf{n}_{i_4}$, if $\mathbf{n}_{i_1} = \mathbf{n}_{i_3}$, there is one element 0; if $\mathbf{n}_{i_1} \neq \mathbf{n}_{i_3}$, there are $M(M-1)$ distinct elements at most.

If there is no other redundancy, the max $|\mathbb{D}|$ is obtained by summing over all the cases, which is

$$|\mathbb{D}|_{\max}(M) = (M^4 - 2M^3 + 7M^2 - 6M + 4)/4. \quad (\text{A1})$$

Next, we provide an example to show that the maximum $|\mathbb{D}|_{\max}(M)$ can be achieved for any M . Consider the linear array $\mathbb{S} = \{10^0, 10^1, \dots, 10^{M-1}\}$. It is not hard to find that in both of the above cases all the maxima can be achieved due to the different orders of magnitude of the elements in \mathbb{S} . Thus, $|\mathbb{D}|_{\max}(M)$ can be achieved for any M . Therefore, combining the definition of \mathbb{H} , it is clear that the supremum of \mathbb{H} is $\sup\{\mathbb{H}\} = (M^4 - 2M^3 + 7M^2 - 6M + 4)/4$ for given M . The proof is completed. \square

Appendix D. The Proof of Lemma 1

Proof. It is clear that $\text{var}(\overline{Y_t Z_t}) = \text{var}(\sum_{t=1}^L Y_t Z_t / L) = \text{var}(Y_t Z_t) / L$, and $\text{var}(\overline{Y_t} \overline{Z_t}) = \text{var}(\sum_{t_1=1}^L Y_{t_1} \cdot \sum_{t_2=1}^L Z_{t_2} / L^2) = \text{var}(Y_t Z_t) / L^3 + \text{var}(Y_t) \text{var}(Z_t) (L-1) / L^3$. Since $\text{var}(Y_t Z_t)$ and $\text{var}(Y_t) \text{var}(Z_t)$ are independent from L , it can be deduced that as L approaches to infinity, $\text{var}(\overline{Y_t Z_t}) / \text{var}(\overline{Y_t} \overline{Z_t}) = O(L)$. The proof is completed. \square

Appendix E. The Proof of Theorem 1

Proof. Due to Proposition 1, if any K sources can be uniquely identified from model (8), we have $K < \text{spark}\{\mathbb{B}_\theta\} / 2$. From (46), we obtain $K < (|\mathbb{D}| + 1) / 2$. Applying Property 3, we can obtain $K < (M^4 - 2M^3 + 7M^2 - 6M) / 8 + 1$. As K is an integer, it is obvious $K \leq (M^4 - 2M^3 + 7M^2 - 6M) / 8$. The proof is completed. \square

References

1. Liu, J.; Zhou, W.; Wang, X. Fourth-order cumulants-based sparse representation approach for DOA estimation in MIMO radar with unknown mutual coupling. *Signal Process.* **2016**, *128*, 123–130. [CrossRef]
2. Saucan, A.; Chonavel, T.; Sintès, C.; Le, J. CPHD-DOA tracking of multiple extended sonar targets in impulsive environments. *IEEE Trans. Signal Process.* **2016**, *64*, 1147–1160. [CrossRef]
3. Wan, L.; Han, G.; Jiang, J.; Zhu, C.; Shu, L. A DOA estimation approach for transmission performance guarantee in D2D communication. *Mob. Netw. Appl.* **2017**, *22*, 998–1009. [CrossRef]
4. Chang, J.; Chang, A. DOA estimation in the time-space CDMA system using modified particle swarm optimization. *AEU Int. J. Electron. Commun.* **2013**, *67*, 340–347. [CrossRef]
5. Porat, B.; Friedlander, B. Direction finding algorithms based on high-order statistics. *IEEE Trans. Signal Process.* **1991**, *39*, 2016–2024. [CrossRef]
6. Chevalier, P.; Ferreol, A.; Albera, L. High-resolution direction finding from higher order statistics: The 2q-MUSIC algorithm. *IEEE Trans. Signal Process.* **2006**, *54*, 2986–2997. [CrossRef]
7. Becker, H.; Chevalier, P.; Haardt, M. Higher order direction finding from rectangular cumulant matrices: The rectangular 2q-MUSIC algorithms. *Signal Process.* **2017**, *133*, 240–249.

8. Pal, P.; Vaidyanathan, P. Multiple level nested array: An efficient geometry for $2q$ th order cumulant based array processing. *IEEE Trans. Signal Process.* **2012**, *60*, 1253–1269. [[CrossRef](#)]
9. Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [[CrossRef](#)]
10. Roy, R.; Kailath, T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984–995. [[CrossRef](#)]
11. Liu, Z.; Huang, Z.; Zhou, Y. Sparsity-inducing direction finding for narrowband and wideband signals based on array covariance vectors. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 1–12. [[CrossRef](#)]
12. He, Z.; Shi, Z.; Huang, L. Covariance sparsity-aware DOA estimation for nonuniform noise. *Digit. Signal Process.* **2014**, *28*, 75–81. [[CrossRef](#)]
13. Zhao, Y.; Zhang, L.; Gu, Y. Array covariance matrix-based sparse Bayesian learning for off-grid direction-of-arrival estimation. *Electron. Lett.* **2016**, *52*, 401–402. [[CrossRef](#)]
14. Malioutov, D.; Cetin, M.; Willsky, A. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.* **2005**, *53*, 3010–3022. [[CrossRef](#)]
15. Tipping, M. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
16. Shen, Q.; Liu, W.; Cui, W.; Wu, S. Extension of co-prime arrays based on the fourth-order difference co-array concept. *IEEE Signal Process. Lett.* **2016**, *23*, 615–619. [[CrossRef](#)]
17. Shen, Q.; Liu, W.; Cui, W.; Wu, S. Extension of nested arrays with the fourth-order difference co-array enhancement. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, China, 20–25 March 2016; pp. 2991–2995.
18. Cai, J.; Liu, W.; Zong, R.; Shen, Q. An expanding and shift scheme for constructing fourth-order difference co-Arrays. *IEEE Signal Process. Lett.* **2017**, *24*, 480–484. [[CrossRef](#)]
19. Wipf, D.; Rao, B. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Trans. Signal Process.* **2007**, *55*, 3704–3716. [[CrossRef](#)]
20. Stoica, P.; Selen, Y. Model-order selection: A review of information criterion rules. *IEEE Signal Process. Mag.* **2004**, *21*, 36–47. [[CrossRef](#)]
21. Yang, Z.; Xie, L.; Zhang, C. Off-grid direction of arrival estimation using sparse Bayesian inference. *IEEE Trans. Signal Process.* **2011**, *61*, 38–43. [[CrossRef](#)]
22. Liu, C.; Vaidyanathan, P. Cramér–Rao bounds for coprime and other sparse arrays, which find more sources than sensors. *Digit. Signal Process.* **2016**, *61*, 43–61. [[CrossRef](#)]
23. Mendel, J.M. Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications. *Proc. IEEE* **1991**, *79*, 278–305. [[CrossRef](#)]
24. Dogan, M.; Mendel, J. Applications of cumulants to array processing. I. Aperture extension and array calibration. *IEEE Trans. Signal Process.* **1995**, *43*, 1200–1216. [[CrossRef](#)]
25. Chevalier, P.; Albera, L.; Ferreol, A.; Comon, P. On the virtual array concept for higher order array processing. *IEEE Trans. Signal Process.* **2005**, *53*, 1254–1271. [[CrossRef](#)]
26. Wang, M.; Nehorai, A. Coarrays, MUSIC, and the Cramér–Rao Bound. *IEEE Trans. Signal Process.* **2017**, *65*, 933–946. [[CrossRef](#)]
27. Babacan, S.; Molina, R.; Katsaggelos, A. Bayesian compressive sensing using Laplace priors. *IEEE Trans. Image Process.* **2010**, *19*, 53–63. [[CrossRef](#)] [[PubMed](#)]
28. Dai, J.; So, C. Sparse Bayesian learning approach for outlier-resistant Direction-of-arrival estimation. *IEEE Trans. Signal Process.* **2018**, *66*, 744–756. [[CrossRef](#)]
29. Yang, Z.; Li, J.; Stoica, P.; Xie, L. Sparse methods for direction-of-arrival estimation. *arXiv* **2016**, arXiv:1609.09596.
30. Chen, J.; Huo, X. Theoretical results on sparse representations of multiple-measurement vectors. *IEEE Trans. Signal Process.* **2006**, *54*, 4634–4643. [[CrossRef](#)]
31. Davies, M.; Eldar, Y. Rank awareness in joint sparse recovery. *IEEE Trans. Inf. Theory* **2010**, *58*, 1135–1146. [[CrossRef](#)]
32. Kruskal, J. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* **1977**, *18*, 95–138. [[CrossRef](#)]
33. Shen, Q.; Cui, W.; Liu, W.; Wu, S.; Zhang, D.; Amin, M. Underdetermined wideband DOA estimation of off-grid sources employing the difference co-array concept. *Signal Process.* **2017**, *130*, 299–304. [[CrossRef](#)]

34. Liu, Q.; So, H.; Gu, Y. Off-grid DOA estimation with nonconvex regularization via joint sparse representation. *Signal Process.* **2018**, *142*, 87–95. [[CrossRef](#)]
35. Sturm, J. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.* **1999**, *11*, 625–653. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).