Signal Processing 153 (2018) 47–57

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

DOA estimation with planar array via spatial finite rate of innovation reconstruction

Yujian Pan, Guo Qing Luo*, Huayan Jin, Xiao Hong Zhang, Chuan Yin

Key Laboratory of RF Circuits & System of Ministry of Education, Institute of Antennas and Microwave Technology, Hangzhou Dianzi University, Xiasha High Education Park, Hangzhou 310018, China

ARTICLE INFO

Article history:
Received 1 December 2017
Revised 26 May 2018
Accepted 2 July 2018
Available online 4 July 2018

Keywords:
Direction-of-arrival estimation
Finite rate of innovation
Correlated sources
Direct data domain
Pseudo-inverse
Gridless estimation

ABSTRACT

Recently, a new DOA estimation algorithm called FRIDA was proposed. It is based on spatial finite rate of innovation (FRI) reconstruction and exhibits some attractive features. In this paper, we propose a variation of FRIDA which is called FRIDA-V (suffix "-V" means variation) . FRIDA-V improves FRIDA in three aspects. Firstly, a multiple measurement vectors (MMV) spatial FRI model is built in the direct data domain rather than the covariance domain, which avoids the model residual error in FRIDA and makes the new method capable of handling both incoherent and coherent sources. Secondly, the full column rank constraint on the mapping matrix is relaxed by adopting the pseudo-inverse technique, which makes it possible to reduce the Bessel function approximation error to a negligible level. Thirdly, the multiple random initializations in the iterative calculations are replaced by a single straightforward initialization, which will speed up the convergence to the optimal solution and save the computational resource. Theoretical derivations and numerical simulation results are given to demonstrate the effectiveness of the proposed method. Compared with the representative methods, the new method is gridless and possesses higher performance with closely-spaced sources and under low signal-to-noise ratio (SNR), small number of snapshots scenarios.

1. Introduction

The problem of DOA estimation is an attractive research subject with decades of history [1]. It belongs to the field of array signal processing and finds applications in military and civilian fields, e.g., radar, sonar, wireless communication and radio astronomy. For most DOA estimation systems, multi-target resolution is an important functional requirement. So, in addition to the amplitude or phase comparison method, the conventional beamforming method was developed by introducing the time-domain Fourier transformation to the spatial domain. However, the resolution is limited by the Rayleigh threshold which is relevant to the array aperture. Resolution that breaks through the Rayleigh threshold is called the super-resolution. By exploiting the orthogonality between subspaces, the subspace-based algorithms represented by the multiple signal classification (MUSIC) [2] and the estimation of signal parameter via rotational invariance technique (ESPRIT) [3] are able to realize super-resolution. But, the high performance of the subspace-based algorithms is based on sufficient SNR or snapshot number. Once the conditions are not satisfied, the subspaces can not be accurately estimated and the sources are unable to be accurately resolved. In addition, ESPRIT is difficult to apply to the irregular array and MUSIC needs to scan over all spatial grids.

For certain applications confronted with weak targets or real-time requirement, sufficient SNR or snapshot number is hard to obtain. To improve the DOA estimation performance under these demanding scenarios, the sparse representation based methods are imported. Among them, the most well-known algorithms are the $\ell 1$-SVD [4] which is based on $\ell 1$-norm constraint and the SBL-DOA [5] which is based on sparse Bayesian Learning (SBL). By the reconstruction of sparse signals from the over-complete dictionary which is built over the uniform spatial grids, both of the two algorithms can achieve better resolution performance under low SNR or small number of snapshots scenarios. However, the $\ell 1$-norm constraint has structural error [6], which makes $\ell 1$-SVD an biased estimator especially for the closely-spaced sources. The SBL-DOA algorithm has no structural error, but it usually converges slowly since it is based on the expectation-maximization (EM) algorithm [7]. Besides, for the sparse representation based methods, the over-complete dictionary is built on the preset spatial grids and only the DOAs on grids are treated as candidates. Off-grid DOAs will bring the model mismatch, which will degrade the estimation performance [8]. Denser grids scheme seems to be a rational solu-

* Corresponding author.
E-mail addresses: pyj@hdu.edu.cn (Y. Pan), luoguoqing@hdu.edu.cn (G.Q. Luo).

https://doi.org/10.1016/j.sigpro.2018.07.001
0165-1684/

tion. However, it will increase the computational burden and may contradict the restricted isometry property (RIP) constraint on the over-complete dictionary for reliable sparse recovery [9].

So, realizing super-resolution DOA estimation under low SNR or small number of snapshots scenarios with a gridless algorithm is the next research objective. A recent proposed algorithm named as FRI-based DOA estimation (FRIDA) shows its potential [10,11]. The implementation details of this algorithm can be found in [12]. In this algorithm, a single measurement vector (SMV) spatial FRI model are built in the covariance domain via the first kind Bessel functions approximation. The DOAs are estimated by exploiting the time-domain FRI reconstruction principle [13,14] and the optimization theory. This algorithm is gridless and exhibits better resolution performance for closely-spaced sources. Moreover, it works for arbitrary array layouts[1] and can resolve more sources than sensors because of the covariance domain application. However, it has some shortcomings. Firstly, because of the covariance domain application, there will exists model residual error because of finite snapshots (which will be explained in this paper) and the algorithm is unable to handle the correlated sources. Secondly, the mapping matrix in the FRI model which contains Bessel functions is constrained to be column full rank. This limits the order of the Bessel function, leading to an unneglectable approximation error. Thirdly, to enhance the probability of obtaining the optimal solution, the algorithm needs to perform iterative calculations with multiple random initializations. This will increase the computational burden. Furthermore, the model residual error and the Bessel approximation error will make the FRIDA a biased estimator. The spatial FRI model which built in the direct data domain has been proposed in [11]. However, it is a SMV model and in-depth studies are needed.

In this paper, aiming at realizing the gridless super-resolution DOA estimation under low SNR or small number of snapshots scenarios , we propose a variation of FRIDA which is called FRIDA-V. FRIDA-V improves FRIDA in three aspects. Firstly, the spatial FRI model is built in the direct data domain, which avoids the model residual error and makes the new algorithm capable of handling both incoherent and coherent sources. To make use of multiple snapshots, the original SMV FRI model is extended to the MMV model. Secondly, in the FRI model, the full column rank constraint on the mapping matrix is relaxed owing to the combination of iterative calculations with the pseudo-inverse technique when solving the FRI model. This means the mapping matrix can be a flat matrix with higher orders of Bessel functions in it and the Bessel approximation error can be reduced to a negligible level. At last, the iteration is initialized with a single straightforward initialization obtained from a low resolution coarse DOA pre-estimation instead of with the multiple random initializations. This will speed up the convergence to the optimal solution and save the computational resource. To reduce the computational load under large number of snapshots, a computational complexity reduction method which is based on singular value decomposition (SVD) pre-processing is also presented. Moreover, the proposed algorithm can be applied to the planar array with arbitrary layouts as long as it is unambiguous in $[0, 2\pi]$ azimuth range. Through numerical simulations, we demonstrate that the new proposed algorithm have higher performance with closely-spaced sources and under low SNR, small number of snapshots scenarios compared with other representative methods.

Notations used in the paper are introduced as follows. $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ are denoted as the transpose, conjugate and conjugate transpose operator, respectively. $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the $\ell 2$ norm and the Frobenius norm, respectively. tr$(\cdot)$, $(*)$ and $\otimes$ are the trace, convolution, and Kronecker product operator respectively. diag$(\cdot)$ is to form a diagonal matrix with entries of a vector and E$(\cdot)$ denotes the expectation of a variable. a_j denotes the jth element of $\boldsymbol{a}$. $\boldsymbol{A}_{\cdot j}$ denotes the jth column of $\boldsymbol{A}$. $\boldsymbol{A}_{i,j}$ denotes the element at ith row and jth column of $\boldsymbol{A}$.

2. MMV spatial FRI model building

In this section, we will show how to build the MMV spatial FRI model in the direct data domain using the array outputs. In order to show that the proposed method is applicable to the planar array with arbitrary layouts, we take a general array configuration for model building.

Consider an array located in a two dimensional Cartesian coordinate system with M identical omnidirectional sensors and the Cartesian coordinate vector of the mth sensor is $\boldsymbol{p}_m = [r_m cos(\theta_m), r_m sin(\theta_m)]$ where $[r_m, \theta_m]$ is its polar coordinate vector. K far-field narrowband sources on the $x-y$ plane from DOAs $\boldsymbol{\varphi} = [\varphi_1, \varphi_2, \ldots, \varphi_K]$ which are within $[0, 2\pi]$ impinge on the array. Set the coordinate origin as the phase reference point, and then the output of the sensors at snapshot n can be represented by a vector $\boldsymbol{x}(n) = [x_1(n), x_2(n), \ldots, x_M(n)]^T$ and its mth element is

$$\begin{aligned} x_m(n) &= \sum_{k=1}^{K} \exp\left[-\mathrm{j}\boldsymbol{\upsilon}_k \boldsymbol{p}_m^T\right] s(\varphi_k, n) + \varepsilon_m(n) \\ &= \sum_{k=1}^{K} \exp\left[\mathrm{j}\beta r_m \cos(\varphi_k - \theta_m)\right] s(\varphi_k, n) + \varepsilon_m(n) \end{aligned} \tag{1}$$

where $\boldsymbol{\upsilon}_k = -2\pi[cos(\varphi_k), sin(\varphi_k)]/\lambda$ is the wave vector of the kth signal, $\beta = 2\pi/\lambda$, λ is the signal wavelength, $s(\varphi_k, n)$ is the signal value at DOA φ_k and $\varepsilon_m(n)$ is the Gaussian noise component which is temporally and spatially white with variance σ^2.

The signal component $s(\varphi_k, n)$ in Eq. (1) can be regarded as a value of a space-time function $s(\varphi, n)$. In the spatial domain, $s(\varphi, n)$ has a period of 2π and only has nonzero value at DOA $\varphi_k, k = 1, 2, \ldots, K$, within $[0, 2\pi]$ at snapshot n. This means $s(\varphi, n)$ can be described as the sum of Dirac delta functions weighted by the signal values in the spatial domain and expressed as

$$\begin{aligned} s(\varphi, n) &= \sum_{k=1}^{K} s(\varphi_k, n) \sum_{l=-\infty}^{+\infty} \delta(\varphi - \varphi_k - 2\pi l) \\ &= \sum_{q=-\infty}^{+\infty} \frac{1}{2\pi} \eta(q, \boldsymbol{\varphi}, n) \exp(\mathrm{j}q\varphi) \end{aligned} \tag{2}$$

$s(\varphi, n)$ is the periodic stream of Diracs in spatial domain. $s(\varphi_k, n)$ and φ_k is the innovation of $s(\varphi, t)$ and the rate of innovation is K/π. φ_k is the DOA and here we call it the spatial FRI. The second equality in Eq. (2) is the Fourier series expansion of $s(\varphi, t)$. $\eta(q, \boldsymbol{\varphi}, n)/2\pi$ is the Fourier series coefficients and

$$\eta(q, \boldsymbol{\varphi}, n) = \sum_{k=1}^{K} s(\varphi_k, n) \exp(-\mathrm{j}q\varphi_k) \tag{3}$$

From Eq. (2), we have

$$\begin{aligned} s(\varphi_k, n) &= \int_0^{2\pi} s(\varphi, n) \sum_{l=-\infty}^{+\infty} \delta(\varphi - \varphi_k - 2\pi l) d\varphi \\ &= \int_0^{2\pi} s(\varphi, n) \delta(\varphi - \varphi_k) d\varphi \end{aligned} \tag{4}$$

Substituting Eq. (4) into Eq. (1) and utilizing the Jacobi–Anger expansion, we will obtain (see Appendix A for a detailed derivation)

$$x_m(n) = \sum_{q=-\infty}^{+\infty} \mathrm{j}^q J_q(\beta r_m) \exp(\mathrm{j}q\theta_m) \eta(q, \boldsymbol{\varphi}, n) + \varepsilon_m(n) \tag{5}$$

[1] Theoretically speaking, this algorithm can not be applied to the linear array. See Section 3.5 for the reason.

where $J_q(\cdot)$ is the qth order Bessel function of the first kind. As we see, now $x_m(n)$ is represented as a summation of infinite items. However, this can not be realized in the practical computation. So, we limit q in the range of $[-Q, Q]$ and use finite orders of Bessel function to approximate $x_m(n)$. According to the property of the Bessel function, the larger the value of Q is, the smaller the approximation error will be. Construct $\boldsymbol{B} \in \mathbb{C}^{M\times(2Q+1)}$ where $\boldsymbol{B}_{m,q+Q+1} = \mathrm{j}^q J_q(\beta r_m)\exp(\mathrm{j}q\theta_m)$, $\boldsymbol{\eta}(\boldsymbol{\varphi}, n) = [\eta(-Q, \boldsymbol{\varphi}, n), \eta(-Q+1, \boldsymbol{\varphi}, n), \ldots, \eta(Q, \boldsymbol{\varphi}, n)]^T$ and $\boldsymbol{\varepsilon}(n) = [\varepsilon_1(n), \varepsilon_2(n), \ldots, \varepsilon_M(n)]^T$, and then the vector expression of approximate version of Eq. (5) is

$$\boldsymbol{x}(n) \approx \boldsymbol{B}\boldsymbol{\eta}(\boldsymbol{\varphi}, n) + \boldsymbol{\varepsilon}(n) \tag{6}$$

we can find that $\boldsymbol{\eta}(\boldsymbol{\varphi}, n)$ is a collection of sums of K complex exponential functions. So it can be annihilated by a $K+1$ elementary filter with coefficients $c_k, k = 1, 2, \ldots, K+1$ whose one-side z-transform is [13]

$$\sum_{k=1}^{K+1} c_k z^{-(k-1)} = \prod_{k=1}^{K}\left[1 - \frac{\exp(-\mathrm{j}\varphi_k)}{z}\right] \tag{7}$$

Construct $\boldsymbol{c} = [c_1, c_2, \ldots, c_{K+1}]^T$, and then the annihilation between $\boldsymbol{\eta}(\boldsymbol{\varphi}, n)$ and $\boldsymbol{c}$ can be expressed as (see Appendix B for a detailed derivation)

$$\boldsymbol{\eta}(\boldsymbol{\varphi}, n) * \boldsymbol{c} = \boldsymbol{0} \tag{8}$$

So far, we only consider the SMV model. To achieve high performance, MMV model should be introduced. Consider N snapshots, construct $\boldsymbol{X} = [\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(N)]$, $\boldsymbol{H} = [\boldsymbol{\eta}(\boldsymbol{\varphi}, 1), \boldsymbol{\eta}(\boldsymbol{\varphi}, 2), \ldots, \boldsymbol{\eta}(\boldsymbol{\varphi}, N)]$ and $\boldsymbol{E} = [\boldsymbol{\varepsilon}(1), \boldsymbol{\varepsilon}(2), \ldots, \boldsymbol{\varepsilon}(N)]$. Then, by absorbing the approximation error into the noise component, Eqs. (6) and (8) are transformed into

$$\begin{cases} \boldsymbol{X} = \boldsymbol{B}\boldsymbol{H} + \boldsymbol{E} \\ \boldsymbol{H}_{.n} * \boldsymbol{c} = \boldsymbol{0}, n = 1, 2, \ldots, N \end{cases} \tag{9}$$

This is the MMV spatial FRI model for DOA estimation. In next section, we will show how to solve this model.

3. DOA estimation via spatial FRI reconstruction

3.1. Spatial FRI reconstruction

To solve Eq. (9), the conventional FRI reconstruction methods which all belong to the SMV model no longer work. Inspired by the method in [12], we convert it to a MMV optimization problem which is

$$\begin{aligned} &\min_{\boldsymbol{H},\boldsymbol{c}\in\mathbb{C}} \|\boldsymbol{X} - \boldsymbol{B}\boldsymbol{H}\|_{\mathrm{F}}^2 \\ &\text{s. t.} \quad \boldsymbol{R}(\boldsymbol{c})\boldsymbol{H} = \boldsymbol{0} \\ &\boldsymbol{\omega}^H\boldsymbol{c} = 1 \end{aligned} \tag{10}$$

This is a constrained minimization problem. The first constrain in Eq. (10) is based on the commutativity of the convolution and that the convolution of two vectors can be constructed as a matrix multiplication, where one of the inputs is converted into a Toeplitz matrix. Here $\boldsymbol{R}(\boldsymbol{c}) \in \mathbb{C}^{(2Q+1-K)\times(2Q+1)}$ is a Toeplitz matrix constructed from $\boldsymbol{c}$ [12,13]. The second constrain is to make $\boldsymbol{c}$ unique and $\boldsymbol{\omega}$ is a constant vector. Generally, there are two choice for $\boldsymbol{\omega}$. The one is the one-hot vector in which all elements are 0 except one 1 and the other is the initialization value of $\boldsymbol{c}$ which is $\boldsymbol{c}^{(0)}$ [15]. The latter choice builds a hyperplane constraint related to the initialization of $\boldsymbol{c}$, which offers more robust performance. So our choice is setting $\boldsymbol{\omega} = \boldsymbol{c}^{(0)}$.

In Eq. (10), the variables need to be resolved are $\boldsymbol{H}$, $\boldsymbol{c}$ and the DOAs can be extracted from $\boldsymbol{c}$ according to Eq. (8). The classical method to handle the constrained optimization problem is the method of Lagrange multipliers. By introducing a matrix variable $\boldsymbol{\rho} \in \mathbb{C}^{(2Q+1-K)\times(2Q+1)}$ and a scalar variable μ, a real Lagrange function [16] can be built as

$$\begin{aligned} \mathcal{L}(\boldsymbol{H}, \boldsymbol{c}, \boldsymbol{\rho}, \mu) = &\|\boldsymbol{X} - \boldsymbol{B}\boldsymbol{H}\|_{\mathrm{F}}^2 + \mathrm{tr}\left[\boldsymbol{\rho}^{\mathrm{H}}\boldsymbol{R}(\boldsymbol{c})\boldsymbol{H}\right] \\ &+ \mathrm{tr}\left[\boldsymbol{\rho}^H\boldsymbol{R}(\boldsymbol{c})\boldsymbol{H}\right]^* + \mu^*\left(\boldsymbol{\omega}^H\boldsymbol{c} - 1\right) + \mu\left(\boldsymbol{\omega}^H\boldsymbol{c} - 1\right)^* \end{aligned} \tag{11}$$

Our goal is to find the optimal solution $\boldsymbol{H}$ and $\boldsymbol{c}$ that minimize $\mathcal{L}(\boldsymbol{H}, \boldsymbol{c}, \boldsymbol{\rho}, \mu)$. Nevertheless, we find minimizing over the both variables simultaneously is not straightforward. So, we first fix $\boldsymbol{c}$, and then the optimal $\boldsymbol{H}$ is

$$\boldsymbol{H} = \arg\min_{\boldsymbol{H}} \mathcal{L}(\boldsymbol{H}, \boldsymbol{c}, \boldsymbol{\rho}, \mu) \tag{12}$$

By setting $\partial\mathcal{L}(\boldsymbol{H}, \boldsymbol{c}, \boldsymbol{\rho}, \mu)/\partial\boldsymbol{H}^* = \boldsymbol{0}$ and $\partial\mathcal{L}(\boldsymbol{H}, \boldsymbol{c}, \boldsymbol{\rho}, \mu)/\partial\boldsymbol{\rho}^* = \boldsymbol{0}$ [17], we can obtain (see Appendix C for a detailed derivation)

$$\boldsymbol{H} = \boldsymbol{Y} - \left(\boldsymbol{B}^H\boldsymbol{B}\right)^+\boldsymbol{R}^H(\boldsymbol{c})\left[\boldsymbol{R}(\boldsymbol{c})\left(\boldsymbol{B}^H\boldsymbol{B}\right)^+\boldsymbol{R}^H(\boldsymbol{c})\right]^+\boldsymbol{R}(\boldsymbol{c})\boldsymbol{Y} \tag{13}$$

where

$$\boldsymbol{Y} = \left(\boldsymbol{B}^H\boldsymbol{B}\right)^+\boldsymbol{B}^H\boldsymbol{X} \tag{14}$$

It can be found we adopt the pseudo-inverse instead of the normal inverse. This is because smaller approximation error in Eq. (6) requires larger number of orders of the Bessel function in $\boldsymbol{B}$, which can make $\boldsymbol{B}$ a flat matrix and lead to the non-existence of the inverse of $\boldsymbol{B}^H\boldsymbol{B}$. The pseudo-inverse on $\boldsymbol{R}(\boldsymbol{c})(\boldsymbol{B}^H\boldsymbol{B})^+\boldsymbol{R}^H(\boldsymbol{c})$ is also to avoid the problem of singularity of matrix. In fact, the solution for an underdetermined equation is not unique. But, here, taking the minimum norm solution in Eq. (13) is reasonable, which is proved in Appendix C. Substituting Eq. (13) back into Eq. (11), we obtain the cost function for solving $\boldsymbol{c}$ which is

$$\begin{aligned} \mathcal{L}(\boldsymbol{c}, \mu) = &\mathrm{tr}\left\{\boldsymbol{Y}^H\boldsymbol{R}^H(\boldsymbol{c})\left[\boldsymbol{R}(\boldsymbol{c})(\boldsymbol{B}^H\boldsymbol{B})^+\boldsymbol{R}^H(\boldsymbol{c})\right]^+\boldsymbol{R}(\boldsymbol{c})\boldsymbol{Y}\right\} \\ &+ \|\boldsymbol{X} - \boldsymbol{B}\boldsymbol{Y}\|_F^2 + \mu^*\left(\boldsymbol{\omega}^H\boldsymbol{c} - 1\right) + \mu\left(\boldsymbol{\omega}^H\boldsymbol{c} - 1\right)^* \end{aligned} \tag{15}$$

To minimize the cost function, we use the iterative method since it is a nonlinear optimization problem. In the ith iteration, $\left[\boldsymbol{R}(\boldsymbol{c})(\boldsymbol{B}^H\boldsymbol{B})^+\boldsymbol{R}^H(\boldsymbol{c})\right]^+$ is regarded as a constant with $\boldsymbol{c} = \boldsymbol{c}^{(i-1)}$ [12,18]. $\boldsymbol{c}^{(i-1)}$ is the solution of $\boldsymbol{c}$ in the $(i-1)$th iteration. According to the commutativity of convolution,

$$\boldsymbol{R}(\boldsymbol{c})\boldsymbol{Y} = [\boldsymbol{T}(\boldsymbol{Y}_{.1})\boldsymbol{c}, \boldsymbol{T}(\boldsymbol{Y}_{.2})\boldsymbol{c}, \ldots, \boldsymbol{T}(\boldsymbol{Y}_{.N})\boldsymbol{c}] \tag{16}$$

where $\boldsymbol{T}(\cdot)$ is a Toeplitz operation like $\boldsymbol{R}(\cdot)$. Denoting $\boldsymbol{T}(\boldsymbol{Y}) = [\boldsymbol{T}^T(\boldsymbol{Y}_{.1}), \boldsymbol{T}^T(\boldsymbol{Y}_{.2}), \ldots, \boldsymbol{T}^T(\boldsymbol{Y}_{.N})]^T \in \mathbb{C}^{N(2Q+1-K)\times(K+1)}$, we can rewrite Eq. (15) as

$$\begin{aligned} \mathcal{L}\left(\boldsymbol{c}^{(i)}, \mu\right) = &\boldsymbol{c}^{(i)H}\boldsymbol{T}^H(\boldsymbol{Y})\left\{\boldsymbol{I}_N \otimes \left[\boldsymbol{R}\left(\boldsymbol{c}^{(i-1)}\right)\left(\boldsymbol{B}^H\boldsymbol{B}\right)^+\boldsymbol{R}^H\left(\boldsymbol{c}^{(i-1)}\right)\right]^+\right\} \\ &\boldsymbol{T}(\boldsymbol{Y})\boldsymbol{c}^{(i)} \\ &+\|\boldsymbol{X} - \boldsymbol{B}\boldsymbol{Y}\|_F^2 + \mu^*\left(\boldsymbol{\omega}^H\boldsymbol{c}^{(i)} - 1\right) + \mu\left(\boldsymbol{\omega}^H\boldsymbol{c}^{(i)} - 1\right)^* \end{aligned} \tag{17}$$

and

$$\boldsymbol{c}^{(i)} = \arg\min_{\boldsymbol{c}^{(i)}} \mathcal{L}\left(\boldsymbol{c}^{(i)}, \mu\right) \tag{18}$$

Eq. (17) is the classical Lagrange multipliers for a constrained quadratic programming problem and it has a well-known solution. However, to enhance the numerical stability, we refer to the method in [12]. By setting $\partial\mathcal{L}(\boldsymbol{c}^{(i)}, \mu)/\partial\boldsymbol{c}^{(i)*} = \boldsymbol{0}$ and $\partial\mathcal{L}(\boldsymbol{c}^{(i)}, \mu)/\partial\mu^* = 0$ and applying some math tricks (detailed derivations are shown in Appendix C), the solution for Eq. (17) can be achieved by solving the following system of equations

$$\begin{bmatrix} \mathbf{0} & \boldsymbol{T}^H(\boldsymbol{Y}) & \mathbf{0} & \boldsymbol{\omega} \\ \boldsymbol{T}(\boldsymbol{Y}) & \mathbf{0} & -\boldsymbol{I}_N \otimes \boldsymbol{R}(\boldsymbol{c}^{(i-1)}) & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{I}_N \otimes \boldsymbol{R}^H(\boldsymbol{c}^{(i-1)}) & \boldsymbol{I}_N \otimes (\boldsymbol{B}^H\boldsymbol{B}) & \mathbf{0} \\ \boldsymbol{\omega}^H & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{c}^{(i)} \\ \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ 1 \end{bmatrix} \tag{19}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are the new introduced auxiliary variables. $\boldsymbol{c}^{(i)}$ can be obtained by resolving Eq. (19). To avoid the problem of singularity of the coefficient matrix on the left side of Eq. (19), we calculate the pseudo-inverse instead of the normal inverse of it when solving the system of equations. If the coefficient matrix is singular, Eq. (19) will have infinite solution. But, since each solution can minimize the cost function in Eq. (17), i.e., satisfying Eq. (18), we will simply choose the minimum norm solution.

So, in order to retrieve $\boldsymbol{c}$, we need to firstly set an initial value $\boldsymbol{c}^{(0)}$. Then perform calculations iteratively according to Eq. (19) until the termination of iteration. The termination of iteration happens when the iteration converges or the maximum number of iterations I is reached. The iteration is considered to have converged when

$$\xi(i) = \frac{\left\|\boldsymbol{c}^{(i)} - \boldsymbol{c}^{(i-1)}\right\|_2}{\left\|\boldsymbol{c}^{(i-1)}\right\|_2} \le \zeta \tag{20}$$

Here we call $\xi(i)$ the relative update ratio in the ith iteration and ζ is the iteration termination threshold. Since the cost function here is non-convex, convergence to the global minimum is hard to ensure. So, initialization is important. We will discuss it in the next subsection.

3.2. A Straightforward Initialization

For the iteration above, the initialization decides whether we can get the optimal solution and we hope the initialization is as close as possible to the optimal solution. The author in [12] use multiple random initializations to increase the probability of obtaining the optimal solution, which increase the computational burden. Here we present a new initialization method which we call the straightforward initialization. From Eq. (7), we notice the relation between $\boldsymbol{c}$ and $\boldsymbol{\varphi}$. So, our method is first to perform a coarse DOA pre-estimation via the conventional beamforming (CBF). Then substitute the pre-estimated DOAs into Eq. (7). The initialization $\boldsymbol{c}^{(0)}$ can be obtained via the inverse one-sided Z-transform on Eq. (7). Assuming $\varphi_k^{CBF}, k = 1, 2, \ldots, \hat{K}$ is the pre-estimated DOAs and $\hat{K}$ is the number of peaks of the CBF spectrum, the initialization $\boldsymbol{c}^{(0)}$ can be written as

$$\boldsymbol{c}^{(0)} = \mathcal{Z}^{-1}\left\{\prod_{k=1}^{\hat{K}}\left[1 - \frac{\exp(-\mathrm{j}\varphi_k^{CBF})}{z}\right]\right\} \tag{21}$$

The true number of sources K is assumed to be known or estimated by the source number estimators such as the Akaike information criterion (AIC) [19] and minimum description length (MDL) [20] methods. If there are more than one source locate in the same spatial beam, the peak number of the CBF spectrum $\hat{K}$ will not coincide with the true number of sources K, i.e., $\hat{K} < K$. As $\boldsymbol{c}^{(0)} \in \mathbb{Z}^{(K+1)\times 1}$, now there will be $K - \hat{K}$ zero elements in $\boldsymbol{c}^{(0)}$ according to Eq. (21). However, through numerical simulation, we will find this will not influence the resolution of the closely-spaced sources in the following fine estimation. Moreover, we will show that the error in the DOA pre-estimation has little impact on the accuracy of the DOA fine estimation in the simulation part.

Compared to the multiple random initializations, the straightforward initialization can provide a good initialization in a single implementation. If the sources distribute in different spatial beams and there is no noise, then the initialization obtained via Eq. (21) will be exactly the optimal solution. In other conditions, the straightforward initialization can provide an initialization which is close to the optimal solution of the problem.

3.3. DOA estimation

After the iterative FRI reconstruction procedure converges, we get the estimation of $\boldsymbol{c}$ denoted as $\hat{\boldsymbol{c}} = [\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_{K+1}]^T$. Then, according to Eq. (7), the fine estimated DOAs $\hat{\boldsymbol{\varphi}} = [\hat{\varphi}_1, \hat{\varphi}_2, \ldots, \hat{\varphi}_K]$ can be written as

$$\hat{\boldsymbol{\varphi}} = -\mathrm{angle}\left\{\mathrm{root}\left[\sum_{k=1}^{K+1}\hat{c}_k z^{K-k-1} = 0\right]\right\} \tag{22}$$

So far, we have showed the derivation of the proposed DOA estimation algorithm. Since it is a variation of FRIDA, we give it the name FRIDA-V. The steps of the algorithm is concluded in Algorithm 1 .

Algorithm 1: A variation of FRIDA (FRIDA-V).

Input: The MMV of the array output $\boldsymbol{X}$, the number of sources K,the mapping matrix $\boldsymbol{B}$, the maximum number of iterations I, the iteration termination threshold ζ.

Output: The estimated DOA vector $\hat{\boldsymbol{\varphi}}$.

1 Calculate $\boldsymbol{Y}$ via Eq. (14), initialize $\boldsymbol{c}$ with $\boldsymbol{c}^{(0)}$ according to Eq. (21) and assign $\boldsymbol{\omega} = \boldsymbol{c}^{(0)}$.

2 **for** $i = 1, 2, \ldots, I$ **do**

  Build the coefficient matrix with $\boldsymbol{Y}$, $\boldsymbol{c}^{(i-1)}$ and update $\boldsymbol{c}^{(i)}$ by solving Eq. (19)

  if $\left\|\boldsymbol{c}^{(i)} - \boldsymbol{c}^{(i-1)}\right\|_2 / \left\|\boldsymbol{c}^{(i-1)}\right\|_2 \le \zeta$ **then**

    Terminate the loop

  end

end

3 $\hat{\boldsymbol{c}} \leftarrow \boldsymbol{c}^{(i)}$ and calculate $\hat{\boldsymbol{\varphi}}$ according to Eq. (22).

3.4. Computational complexity reducing

It can be seen the dimension of the coefficient matrix in Eq. (19) is $[N(4Q-K+2)+K+2] \times [N(4Q-K+2)+K+2]$. Calculating the pseudo-inverse of the coefficient matrix requires the SVD operation whose computational load is proportional to the third power of the size of the matrix. As the dimension of the coefficient matrix increases linearly with the number of snapshots N, the computational load will be very heavy when the snapshot number is large.

So, Algorithm 1 with large snapshot number maybe not practical under the current processing power. Fortunately, inspired by the work of Malioutov [4], we can present a computational complexity reduction method. First, we perform the SVD operation on the MMV of the array output $\boldsymbol{X}$ and $\boldsymbol{X} = \boldsymbol{USV}^H$. Supposing $N \ge K$, as we know, the signal subspace is only located in the space which is spanned by K singular vectors of $\boldsymbol{X}$ corresponding to the largest K singular values [4]. So, we can replace $\boldsymbol{X}$ with $\boldsymbol{X}_s$ which consists of the largest K left singular vectors of $\boldsymbol{X}$ weighted by the corresponding K singular values. Assuming the singular value in $\boldsymbol{S}$ is in descending order, we can write $\boldsymbol{X}_s \in \mathbb{Z}^{M\times K}$ as

$$\begin{aligned} \boldsymbol{X}_s &= [\boldsymbol{U}_{\cdot 1}, \boldsymbol{U}_{\cdot 2}, \ldots, \boldsymbol{U}_{\cdot K}]\mathrm{diag}\{[\boldsymbol{S}_{1,1}, \boldsymbol{S}_{2,2}, \ldots, \boldsymbol{S}_{K,K}]\} \\ &= \boldsymbol{XVD}_k \end{aligned} \tag{23}$$

where $\boldsymbol{D}_k = [\boldsymbol{I}_K, \boldsymbol{0}]^T \in \mathbb{R}^{N\times K}$. Denoting $\boldsymbol{H}_s = \boldsymbol{H}\boldsymbol{V}\boldsymbol{D}_K$ and $\boldsymbol{E}_s = \boldsymbol{E}\boldsymbol{V}\boldsymbol{D}_K$ and utilizing the linearity of convolution, we can transform Eq. (9) into

$$\begin{cases} \boldsymbol{X}_s = \boldsymbol{B}\boldsymbol{H}_s + \boldsymbol{E}_s \\ (\boldsymbol{H}_s)_{.k} * \boldsymbol{c} = \boldsymbol{0}, k = 1, 2, \ldots, K \end{cases} \tag{24}$$

The form of Eq. (24) is similar to Eq. (9). So the steps of computational complexity reduction version are the same with FRIDA-V except the input $\boldsymbol{X}$ is replaced with $\boldsymbol{X}_s$ and the snapshot number N is replaced with K. As the dimension of the coefficient matrix in Eq. (19) now is $[K(4Q-K+3)+2]\times[K(4Q-K+3)+2]$, the computational complexity is reduced.

For situations mixed with coherent sources, the dimension of the signal subspace diminishes. Then the K in Eq. (23) should be changed to the number of incoherent sources.

It can be seen computational complexity reduction version is only valid when $N \geq K$. So, our strategy is to adopt FRIDA-V when $N \leq K$ and adopt computational complexity reduction version when $N > K$.

3.5. Some discussions

Advantage of direct data domain application. In order to show the advantage of direct data domain application, we study the disadvantage of building the FRI model in the covariance domain which is the case of FRIDA. According to (1) and assuming two signals existing in a noiseless scenario, omitting the temporal index, we can write the covariance between the mth and the nth array element output as

$$\begin{aligned} \boldsymbol{R}_{m,n} &= \mathrm{E}[x_m x_n^*] \\ &= \mathrm{E}\Big[\sum_{k=1}^{2} \exp(\mathrm{j}\beta r_m \cos(\varphi_k - \theta_m)) s(\varphi_k) \\ &\quad \sum_{k=1}^{2} \exp(-\mathrm{j}\beta r_n \cos(\varphi_k - \theta_n)) s^*(\varphi_k)\Big] \\ &= \sum_{k=1}^{2} \exp\left[\mathrm{j}2\beta r \sin\left(\varphi_k - \frac{\theta_m+\theta_n}{2}\right)\sin\frac{\theta_m-\theta_n}{2}\right] p(\varphi_k) \\ &+ \exp\left[\mathrm{j}2\beta r \sin\left(\frac{\Sigma_\varphi - \theta_m - \theta_n}{2}\right)\sin\frac{\Delta_\varphi + \theta_m - \theta_n}{2}\right] \\ &\quad \mathrm{E}[s(\varphi_2)s^*(\varphi_1)] \\ &+ \exp\left[\mathrm{j}2\beta r \sin\left(\frac{\Sigma_\varphi - \theta_m - \theta_n}{2}\right)\sin\frac{-\Delta_\varphi + \theta_m - \theta_n}{2}\right] \\ &\quad \mathrm{E}[s(\varphi_1)s^*(\varphi_2)] \end{aligned} \tag{25}$$

where we made a simplification as $r = r_m = r_n$ which is convenient for analysis and $p(\varphi_k) = \mathrm{E}[s(\varphi_k)s^*(\varphi_k)]$, $\Sigma_\varphi = \varphi_1 + \varphi_2$, $\Delta_\varphi = \varphi_1 - \varphi_2$. If signals are uncorrelated, $\mathrm{E}[s(\varphi_2)s^*(\varphi_1)] = \mathrm{E}[s(\varphi_1)s^*(\varphi_2)] = 0$ and (25) becomes the basis of FRIDA. However, in practice, estimated covariance between general uncorrelated signals (eg., random signals) will not be exactly zero due to finite snapshots. So the last two terms in (25) will be the model residual error and we can find the residual error is related to the angle separation. This residual error will bring model mismatch and make the FRIDA a biased estimator. Only with orthogonal signals (eg., sin and cos) is FRIDA unbiased since the estimated covariance between orthogonal signals will be zero even with finite snapshots. However, it should be noted that spatially resolving orthogonal signals is lack of practical significance since orthogonal signals can be resolved by orthogonal decomposition. Direct data domain application avoids the model residual error and can work normally with both incoherent and coherent signals.

Computational complexity. For algorithm FRIDA-V, the most time consuming step is the iterative calculation for $\hat{\boldsymbol{c}}$. In each iteration, for solving Eq. (19), the computational complexity is $O([N(4Q-K+2)+K+2]^3)$ which mainly arises from the computation of the pseudo-inverse of the coefficient matrix. After performing the computational complexity reduction method, each iteration has a lower computational complexity which is $O([K(4Q-K+3)+2]^3)$. However, there is an extra step which is expressed in Eq. (23) in the computational complexity reduction method. This requires a SVD operation on $\boldsymbol{X}$. Since we only need the largest K singular values and the corresponding singular vectors, we can calculate the economy version of the SVD which only outputs the largest M singular values and corresponding singular vectors. The computational complexity of the economy-sized SVD for $\boldsymbol{X}$ is only $O(M^2N)$ [21].

Applicable array layout. The DOA estimation method proposed above is a parametric method and we have not imposed special restrictions on the array layout except it should be unambiguous in $[0, 2\pi]$ azimuth range since we assume $s(\varphi, n)$ has a period of 2π in the spatial domain in Eq. (2). So, in theory, this method can be applied on the planar array with arbitrary layouts. Different array layout corresponds to different mapping matrix $\boldsymbol{B}$. But, for the linear array, the method will be invalid as linear array is only unambiguous in $[0, \pi]$ azimuth range. For successful application on linear array, some modifications should be done, which we put in the future works.

The convergence. The iterative calculating procedure for Eq. (15) is similar to the method of iterative quadratic maximum likelihood (IQML) [18]. The convergence properties of IQML have not been fully studied so far. The dominant point of view is that IQML will convergence in most cases [22–24]. So, for the proposed algorithm, its convergence properties are still open questions and require further theoretical investigations. However, in the numerical simulation part, we have not come across any non-convergence cases and its performance can reach the Cramer-Rao lower bound (CRLB) under appropriate conditions.

4. Numerical simulations

In this section, some experiments are performed to evaluate the performance of the algorithm. The root-mean-square-error (RMSE) of DOA estimation is used as the performance indicator. A uniform circular array (UCA) with $M = 15$ elements is exposed to two narrowband far-field equal-power sources. Unless otherwise specified, the radius of the UCA is set as $r = \lambda$ and the signals are set as Gaussian random processes. For the general array model in Eq. (1) being compatible with UCA, we set $r_m = \lambda$ and $\theta_m = 2\pi(m-1)/M$ where $m = 1, 2, \ldots, M$. For the iterative FRI reconstruction, the maximum number of iterations is set as $I = 50$ and iteration termination threshold is set as $\varsigma = 10^{-8}$. The simulation results are obtained by averaging 500 Monte Carlo trials. The algorithms we choose for performance comparison are the MUSIC [2], ℓ1-SVD [4] and FRIDA [10]. The CRLB is also included as the performance reference. For the ℓ1-SVD, its overcomplete dictionary is built on the spatial grids which locate within $[-180^\circ, 180^\circ]$ with the step of 1°. Grid refinement technique is utilized to improve the DOA estimation accuracy of ℓ1-SVD [4]. FRIDA is originally for the wideband application. Here we regard the narrowband signal as the wideband signal with only one subband. For the FRIDA with UCA, to make the corresponding mapping matrix whose dimension is $M(M-1)\times(2Q+1)$ [10] full column rank, the maximum order of the first kind Bessel function can only be set as $Q = M-1$ (verified by numerical calculation). For FRIDA iteration, we follow the strategy of 15 random initializations with a limit of 50 inner iterations.

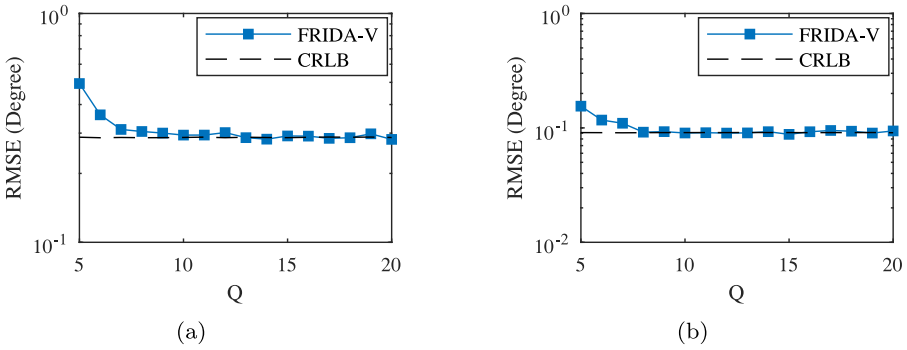

Fig. 1. The impact of Bessel function approximation error on the DOA estimation performance. (a) SNR = 10 dB, (b) SNR = 20 dB.

4.1. Simulation analyses of FRIDA-V

In this subsection, we will test the effect of the pseudo-inverse technique and the straightforward initialization method in FRIDA-V.

In the first experiment, we show the effect of pseudo-inverse technique on FRIDA-V. With pseudo-inverse technique, the maximum order of the first kind Bessel function Q can be increased to show the impact of Bessel function approximation error in the mapping matrix $\boldsymbol{B}$ on the DOA estimation performance. The two sources are located at 30.2° and 42.7°. The number of snapshots is set as $N = 50$. The Q is varied from 5 to 20. Simulations are conducted under two different SNR which are 10 dB and 20 dB, respectively. The results are shown in Fig. 1. From this figure, we can observe that if the full column rank requirement of mapping matrix remains, i.e., $Q \leq 7$ for our algorithm, the Bessel function approximation error is visible. However, if we increase Q with the pseudo-inverse, the DOA estimation error approaches the CRLB gradually. When $Q \geq 13$ for 10 dB SNR or $Q \geq 8$ for 20 dB SNR , the DOA estimation error reaches the CRLB and the Bessel function approximation error can be neglected. Large Q helps reduce the approximation error, but it can also lead to the non-existence of $(\boldsymbol{B}^H\boldsymbol{B})^{-1}$. Fortunately, with the pseudo-inverse technique used in the proposed algorithm, this problem can be solved perfectly.

Since the original FRIDA is mainly proposed for wideband application, we will also test the performance of FRIDA-V in different subbands. Here, the array radius is fixed to r and the signal wavelength λ is varied from $0.5r$ to $2r$ to represent the center wavelength of different subbands. The SNR is set as 20 dB and other conditions are the same as previous simulation. We simulate two scenarios which are $Q = 8$ and $Q = 10$. The results are shown Fig. 2. From Fig. 1(b), we know $Q = 8$ is a reasonable setting for $\lambda = r$ under 20 dB SNR. From Fig. 2, we can find this setting is still reasonable for lower frequencies but will bring performance degradation for higher frequencies. This phenomenon can be explained by Eq. (5). That is, for fixed Q, higher frequency will increase the Bessel approximation error. The higher frequency adaptation problem can be alleviated by higher Q, which is confirmed by Fig. 2. However, this will increase the computational burden. So, for wideband DOA estimation using FRIDA-V, we suggest making the array radius close to the smallest wavelength within the bandwidth or adopting the coherent focussing technique [25] which focuses all subband frequencies to a suitable frequency.

In the second experiment, we test the effect of the proposed straightforward initialization. we evaluate the impact of DOA pre-estimation error in initialization on the performance of the final DOA estimation. For this purpose, we consider two situations. One is two sources locate in the same CBF beam called *DOAs in beam*. Another is two sources locate in the different CBF beams called *DOAs out beam*. For *DOAs in beam*, the sources are located

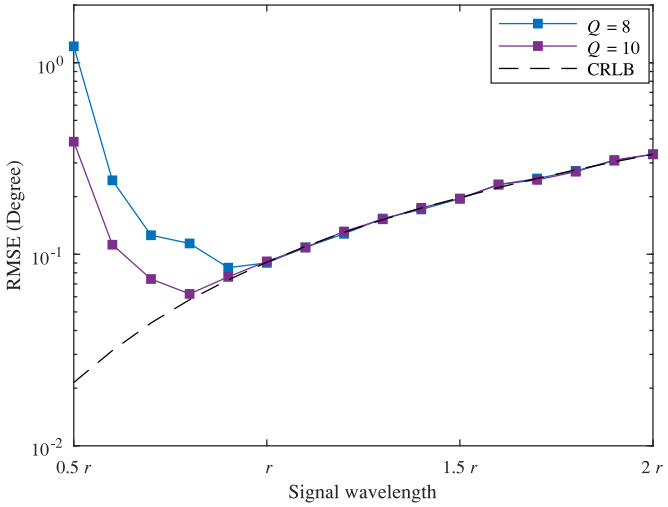

Fig. 2. RMSE of DOA estimation under different signal wavelengths with fixed Q.

at 30.2° and 42.7° and the CBF estimator outputs one DOA which is $(30.2° + 42.7°)/2 + \Delta\varphi_1$. For *DOAs out beam*, the sources are located at 30.2° and 60° and the CBF estimator outputs two DOAs which are $30.2° + \Delta\varphi_2$ and $60° + \Delta\varphi_3$. $\Delta\varphi_1$, $\Delta\varphi_2$ and$\Delta\varphi_3$ follow the Gaussian distribution with zero mean and $\sigma_{\Delta\varphi}$ standard deviation. The other conditions are set as $N = 50$, SNR = 10 dB and $Q = 14$. We vary $\sigma_{\Delta\varphi}$ from 0° to 60° and the simulation results are exhibited in Fig. 3. It can been seen that the multiple initializations on $\boldsymbol{c}$ is successfully substituted by a single straightforward initialization in the angle domain realized by low resolution DOA pre-estimation. Whether the DOAs reside in different beams or not, the DOA estimation error almost remains unchanged even when RMSE of DOA pre-estimation is as large as 60°. This means the DOA estimation performance is insensitive to the DOA pre-estimation error.

In this experiment, we also plot the convergence curve of new algorithm. Four situations are considered. They are the *DOAs out beam* with $\sigma_{\Delta\varphi} = 0°$, *DOAs out beam* with $\sigma_{\Delta\varphi} = 60°$, *DOAs in beam* with $\sigma_{\Delta\varphi} = 0°$ and *DOAs out beam* with $\sigma_{\Delta\varphi} = 60°$. The averaged relative update ratio is used to represent the convergence curve. In each trial, the iteration is terminated by the maximum number of iterations I instead of the termination threshold ζ. So, the averaged relative update ratio $\bar{\xi}(i), i = 1, 2, \ldots, I$ is the average of the relative update ratio $\xi(i), i = 1, 2, \ldots, I$ in (20) in each Monte Carlo trial. The result is shown in Fig. 4. As we see, the algorithm converges well even when the RMSE of DOA pre-estimation is as large as 60°. The convergence speed increases as the error in DOA pre-estimation decreases. For *DOAs out beam* with $\sigma_{\Delta\varphi} = 0°$, it only takes about 6 iterations to reach our termination threshold $\zeta = 10^{-8}$. If we continue the iteration, the relative update ratio

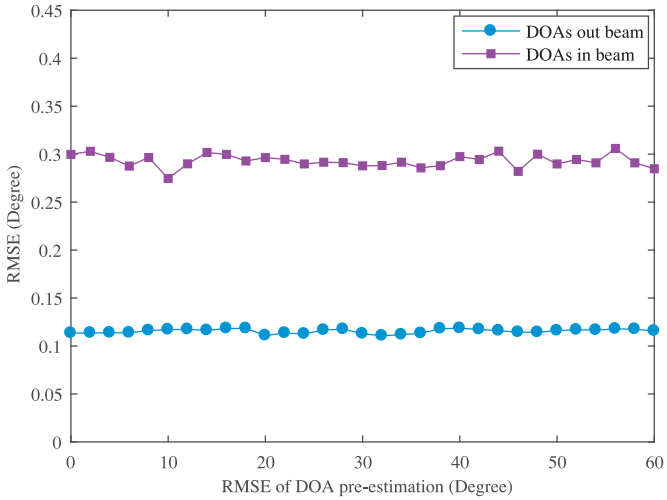

Fig. 3. The impact of DOA pre-estimation error on the DOA estimation performance.

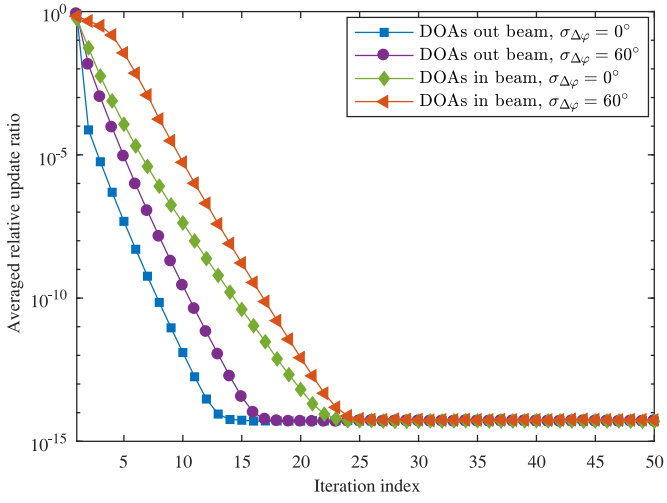

Fig. 4. The convergence curve under different situations.

Table 1
RMSE of DOA estimation for four different initialization methods .

	DOAs out beam	DOAs in beam
Straightforward initialization	0.1168°	0.2955°
Random DOAs initialization	15.9038°	16.4617°
Single random $\boldsymbol{c}$ initialization	9.1459°	11.6695°
Multiple random $\boldsymbol{c}$ initialization	0.1165°	0.2976°

can be further reduced. However, this is of little help to improve the DOA estimation performance.

Since the DOA estimation performance is insensitive to the DOA pre-estimation error for the proposed algorithm, we compare the straightforward initialization with the random DOAs initialization where random DOAs which are independently and uniformly distributed within $[0, 2\pi]$ are used to obtain $\boldsymbol{c}^{(0)}$ according to (21). Also the single random $\boldsymbol{c}$ initialization where $\boldsymbol{c}^{(0)}$ is randomly generated only once and multiple random $\boldsymbol{c}$ initializations which is the initialization method of FRIDA (15 random initializations with a limit of 50 inner iterations) are included for performance comparison. The simulation conditions are the same as previous and the RMSE of DOA estimation for the four different initialization methods is shown in Table 1. From this table, we can find that both the random DOAs initialization and the single random $\boldsymbol{c}$ initialization fail to resolve the two signals under two scenarios. The multiple random $\boldsymbol{c}$ initialization has similar performance with the straightforward initialization . However, multiple random $\boldsymbol{c}$ initialization leads to higher computational burden because of multiple runs.

Table 2
Other variations of FRIDA .

	Algorithm description
FRIDA-p	FRIDA with pseudo-inverse ($Q = 23$)
FRIDA-s	FRIDA with straightforward initialization
FRIDA-o	FRIDA with orthogonal signals (sin, cos)
FRIDA-op	FRIDA-p with orthogonal signals (sin, cos)
FRIDA-ops	FRIDA-op with straightforward initialization

4.2. FRIDA-V versus FRIDA

In this subsection, we compare the performance of FRIDA-V and FRIDA and try to figure out what factors contribute to these performance differences. So , in the third experiment, a noise free condition is considered and two sources are located at 30.2° and $30.2° + \Delta\varphi$. $\Delta\varphi$ is varied from 3° to 60°. The other conditions are set as $N = 50$, and $Q = 14$ for FRIDA-V. Apart from FRIDA-V and FRIDA, we build other variations of FRIDA which is listed in Table 2. For example, FRIDA-p is used to reduce the Bessel approximation error and FRIDA-o is used to remove the model residual error. The bias and RMSE of the DOA estimation of these methods are plotted in Figs. 5 and 6.

From Fig. 5, we can find that FRIDA is a biased estimator and FRIDA-p and FRIDA-o reduce the bias a little. Only the bias of FRIDA-op and FRIDA-ops are as small as FRIDA-V. This indicates that the bias of FRIDA comes from both the model residual error explained in (25) and the Bessel approximation error. From Fig. 6, we can find that when bias exists, the straightforward initialization can reduce the RMSE of DOA estimation of closely-spaced sources. When the bias is removed, the RMSE of FRIDA can be as small as FRIDA-V and straightforward initialization and multiple random initializations has the same effect except different computational burden. Another interesting phenomenon is that both the bias and RMSE are correlated with the angle separation. This is because the model residual error is correlated with the angle separation (which is explained in (25)) and we can also infer the Bessel approximation error leads to the correlation too (which can be observed from the results of FRIDA-o).

4.3. FRIDA-V versus some representative algorithms

In this section, FRIDA-V is compared with some representative algorithms under different SNRs, snapshot numbers, angle separations and correlation coefficients . For FRIDA, the original algorithm is used. In the fourth experiment, we compare the performance of the proposed algorithm with other representative algorithms under different SNRs. The two sources are located at 30.2° and 42.7°. The SNR is varied from -4 dB to 20 dB. The other conditions are set as $N = 50$ and $Q = 14$. The simulation results are presented in Fig. 7. From this simulation, we can find that the performance of MUSIC is the lowest under low SNR. The RMSE of $\ell 1$-SVD deviates CRLB under high SNR since it is biased with closely-spaced sources. The RMSE of FRIDA deviates CRLB under high SNR because of the bias caused by the model residual error and Bessel function approximation error. The RMSE of proposed algorithm FRIDA-V reaches the CRLB when SNR = 4 dB and is consistent with the tendency of CRLB as SNR increases. Moreover, for the proposed algorithm, we have not seen any signs of the Bessel function approximation error.

In the fifth experiment, we compare the performance of the proposed algorithm with other representative algorithms under

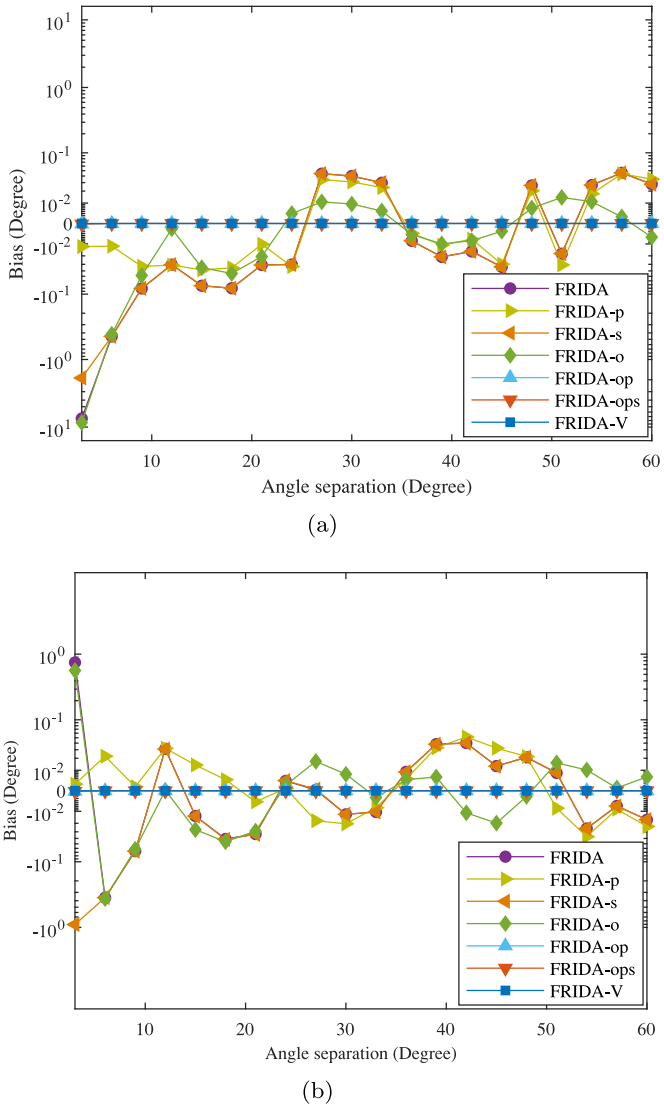

Fig. 5. The bias the DOA estimation versus different angle separations under noise free condition. (a) the first source, (b) the second source.

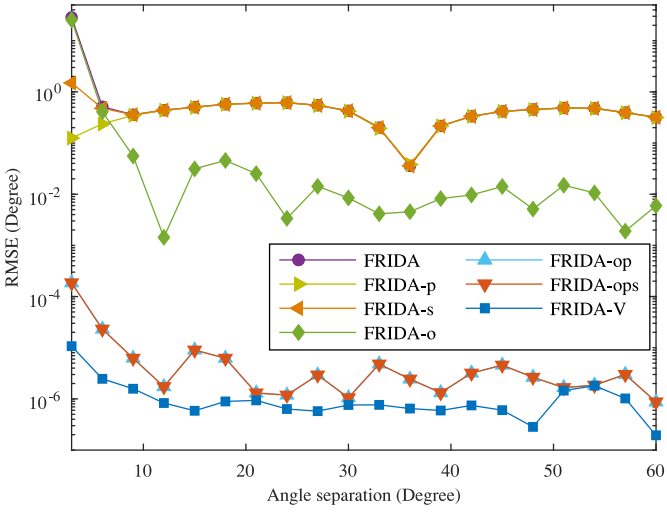

Fig. 6. RMSE of DOA estimation versus different angle separations under noise free condition.

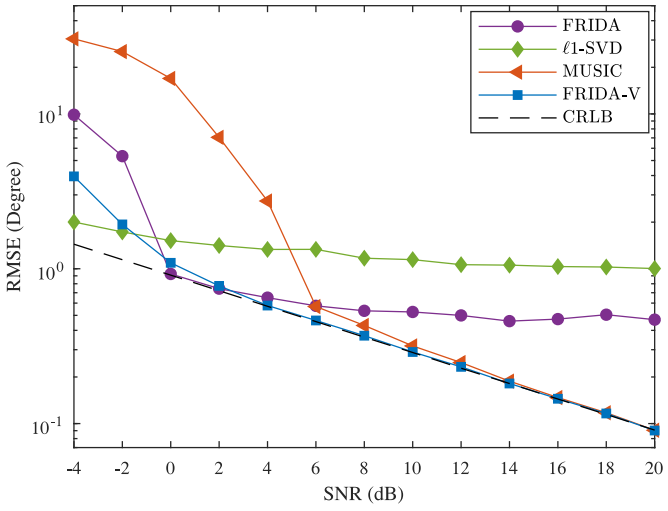

Fig. 7. RMSE of DOA estimation versus different SNRs.

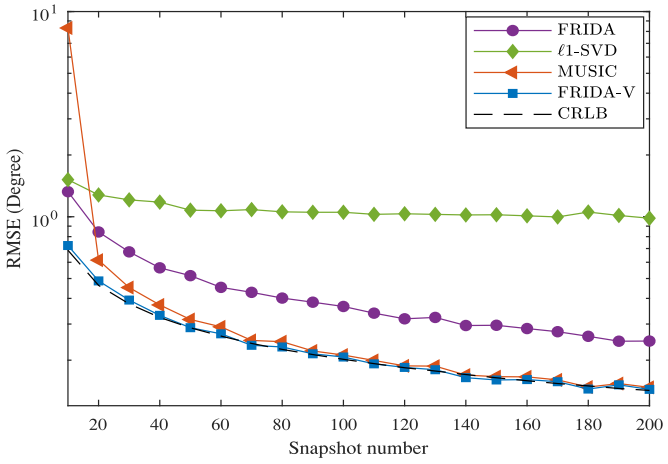

Fig. 8. RMSE of DOA estimation versus different snapshot numbers.

different snapshot numbers. The two sources are located at 30.2° and 42.7°. The snapshot number is varied from 10 to 200. The other conditions are set as SNR = 10 dB and $Q = 14$. The simulation results are presented in Fig. 8. From this simulation, we can find that the performance of MUSIC is the lowest under small number of snapshots. The RMSE of $\ell 1$-SVD deviates CRLB under large number of snapshots. The RMSE of FRIDA can not reach the CRLB because of estimation bias. FRIDA-V has the highest performance. It reaches the CRLB with 50 snapshots and is consistent with the tendency of CRLB as the number of snapshots increases.

In the sixth experiment, we compare the performance of the proposed algorithm with other representative algorithms under different angle separations between sources. The two sources are located at 30.2° and $30.2° + \Delta\varphi$. $\Delta\varphi$ is varied from 3° to 23°. The other conditions are set as $N = 50$, SNR = 10 dB and $Q = 14$. The simulation results are shown in Fig. 9. We can find that the performance of MUSIC is the lowest for closely-spaced sources. The RMSE of $\ell 1$-SVD reaches the CRLB with the sources angle separation of 23°. This means $\ell 1$-SVD gets less biased with sources of large angle separation. The RMSE of FRIDA deviates the CRLB because of estimation bias. And its RMSE increases as angle separation increases results from that the bias is correlated with angle separation, which is confirmed in the third experiment. The RMSE of proposed algorithm FRIDA-V reaches the CRLB when angle sep-

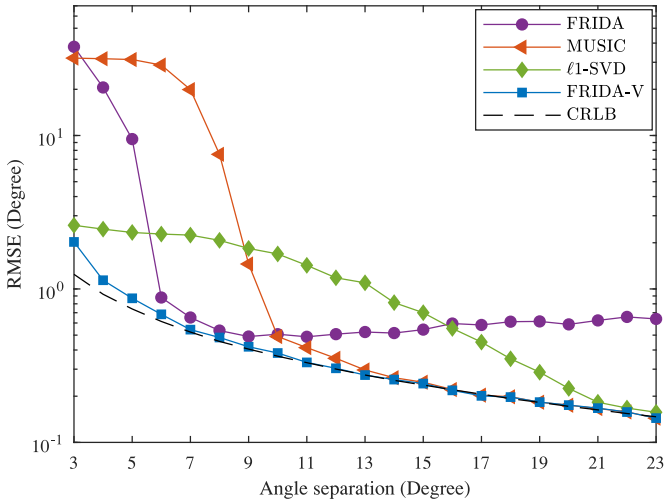

Fig. 9. RMSE of DOA estimation versus different angle separations.

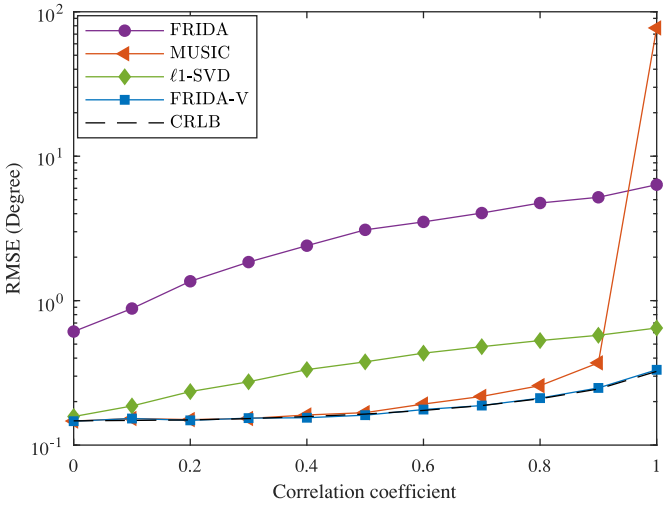

Fig. 10. RMSE of DOA estimation versus different correlation coefficients between signals.

aration is 11° and is consistent with the tendency of CRLB as angle separation increases.

In the last experiment, we compare the performance of the proposed algorithm with other representative algorithms under signals with different correlation coefficients. To remove the effect of angle separation, we expand the angle interval and the two sources are located at 30.2° and 53.2° according to Fig. 9. The correlation coefficient is varied from 0 to 1. The other conditions are set as $N = 50$, SNR = 10 dB and $Q = 14$. The simulation results are placed in Fig. 10. The results indicate that the MUSIC has the lowest performance when signals are coherent since it is based on the estimation of signal or noise subspace. The FRIDA deviates from the CRLB since it is based on the assumption of non-correlation between signals. The $\ell 1$-SVD deviates from the CRLB more slowly and has a acceptable DOA estimation error when signals are coherent. The proposed algorithm FRIDA-V possesses the highest performance and is consistent with the CRLB under any correlation coefficient. This is normal because no restriction about correlation between signals is imposed in the proposed method.

5. Conclusion

In this paper, we proposed a DOA estimation algorithm called FRIDA-V which is variation of a recent proposed algorithm named FRIDA . This algorithm can be applied on the planar array with arbitrary layouts. Combined with the SVD pre-processing, its computational complexity can be reduced. Compared with FRIDA, FRIDA-V has four advantages. Firstly, it avoids the model residual error. Secondly, it is able to deal with both incoherent and coherent sources. Thirdly, it can reduce the Bessel function approximation error to a negligible level. Fourthly, it needs only a single coarse initialization for the iteration. Compared with the representative methods, the new method is gridless and possesses higher performance with closely-spaced sources and under low SNR, small number of snapshots scenarios.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (no. 61701145 and no. 61722107) and the Natural Science Foundation of Zhejiang Province (no. LQ18F010003).

The authors would like to thank the anonymous reviewers for the improvements of this paper.

Appendix A. The derivation of Eq. (5)

According to the Jacobi–Anger expansion, we have

$$\exp\left[\mathrm{j}\beta r_m \cos(\varphi_k - \theta_m)\right] = \sum_{p=-\infty}^{+\infty} \mathrm{j}^p J_p(\beta r_m) \exp\left[\mathrm{j}p(\varphi_k - \theta_m)\right] \quad \text{(A.1)}$$

Substitute Eqs. (A.1), (4) into Eq. (1) and refer to Eq. (2), and then we have

$$\begin{aligned} x_m(n) &= \sum_{k=1}^{K} \sum_{p=-\infty}^{+\infty} \mathrm{j}^p J_p(\beta r_m) \exp\left[\mathrm{j}p(\varphi_k - \theta_m)\right] \\ &\quad \int_0^{2\pi} s(\varphi, n)\delta(\varphi - \varphi_k)d\varphi + \varepsilon_m(n) \\ &= \int_0^{2\pi} \sum_{p=-\infty}^{+\infty} \mathrm{j}^p J_p(\beta r_m) \exp\left[\mathrm{j}p(\varphi - \theta_m)\right] \\ &\quad s(\varphi, n) \sum_{k=1}^{K} \delta(\varphi - \varphi_k)d\varphi + \varepsilon_m(n) \\ &= \int_0^{2\pi} \sum_{p=-\infty}^{+\infty} \mathrm{j}^p J_p(\beta r_m) \exp\left[\mathrm{j}p(\varphi - \theta_m)\right] s(\varphi, n)d\varphi + \varepsilon_m(n) \\ &= \int_0^{2\pi} \sum_{p=-\infty}^{+\infty} \mathrm{j}^p J_p(\beta r_m) \exp\left[\mathrm{j}p(\varphi - \theta_m)\right] \\ &\quad \sum_{q=-\infty}^{+\infty} \frac{1}{2\pi} \eta(q, \boldsymbol{\varphi}, n) \exp(\mathrm{j}q\varphi)d\varphi \\ &\quad + \varepsilon_m(n) \end{aligned} \quad \text{(A.2)}$$

When $p \neq -q$, the above integral is equal to zero. So, with the relationship $J_{-q}(\cdot) = (-1)^q J_q(\cdot)$, $x_m(n)$ can be written as

$$x_m(n) = \sum_{q=-\infty}^{+\infty} \mathrm{j}^q J_q(\beta r_m) \exp(\mathrm{j}q\theta_m) \eta(q, \boldsymbol{\varphi}, n) + \varepsilon_m(n) \quad \text{(A.3)}$$

Appendix B. Proof of the annihilation between $\boldsymbol{\eta}(\boldsymbol{\varphi}, n)$ and $\boldsymbol{c}$

As $\eta_{q+Q+1}(\boldsymbol{\varphi}, n) = \sum_{k=1}^{K} s(\varphi_k, n) \exp(-\mathrm{j}q\varphi_k)$, $q = -Q, -Q+1, \ldots, Q$, the convolution of $\boldsymbol{c}$ and $\boldsymbol{\eta}(\boldsymbol{\varphi}, n)$ is calculated as fol-

lows

$$\begin{aligned}[\boldsymbol{c} * \boldsymbol{\eta}(\boldsymbol{\varphi}, n)]_q &= \sum_{k'=1}^{K+1} \boldsymbol{c}_{k'} \eta_{q-k'}(\boldsymbol{\varphi}, n) \\ &= \sum_{k'=1}^{K+1} \boldsymbol{c}_{k'} \sum_{k=1}^{K} s(\varphi_k, n) \exp\left[-\mathrm{j}(q-k'-Q-1)\varphi_k\right] \\ &= \sum_{k=1}^{K} s(\varphi_k, n) \sum_{k'=1}^{K+1} \boldsymbol{c}_{k'} \exp\left[\mathrm{j}(k'-1)\varphi_k\right] \\ &\quad \exp[\mathrm{j}(Q-q+2)\varphi_k] \end{aligned} \tag{B.1}$$

According to Eq. (7), we have $\sum_{k'=1}^{K+1} \boldsymbol{c}_{k'} \exp[\mathrm{j}(k'-1)\varphi_k] = 0$. So $\boldsymbol{c} * \boldsymbol{\eta}(\boldsymbol{\varphi}, n) = \boldsymbol{0}$ is proved.

Appendix C. The derivations for the iterative FRI reconstruction

Proposition C.1. *For any matrix $\boldsymbol{A}$, if it resides in column space of matrix $\boldsymbol{G}$ and $\boldsymbol{G}$ is a hermitian matrix, then we have*

$$\begin{cases} (\boldsymbol{I} - \boldsymbol{G}\boldsymbol{G}^{+})\boldsymbol{A} = \boldsymbol{0} \\ \boldsymbol{A}^{H}(\boldsymbol{I} - \boldsymbol{G}^{+}\boldsymbol{G}) = \boldsymbol{0} \end{cases} \tag{C.1}$$

Proof. Since $\boldsymbol{A}$ resides in column space $\boldsymbol{G}$, we have $\boldsymbol{G}\boldsymbol{W} = \boldsymbol{A}$ and $\boldsymbol{W}$ is an arbitrary nonzero matrix. Hence,

$$\boldsymbol{G}\boldsymbol{G}^{+}\boldsymbol{A} = \boldsymbol{G}\boldsymbol{G}^{+}\boldsymbol{G}\boldsymbol{W} = \boldsymbol{G}\boldsymbol{W} = \boldsymbol{A} \tag{C.2}$$

So, $(\boldsymbol{I} - \boldsymbol{G}\boldsymbol{G}^{+})\boldsymbol{A} = \boldsymbol{0}$. As $\boldsymbol{G}$ is a hermitian matrix, we obtain $\boldsymbol{A}^{H}(\boldsymbol{I} - \boldsymbol{G}^{+}\boldsymbol{G}) = \boldsymbol{0}$. □

For minimizing Eq. (12) with fixed $\boldsymbol{c}$, setting $\partial\mathcal{L}(\boldsymbol{H}, \boldsymbol{c}, \boldsymbol{\rho}, \mu)/\partial\boldsymbol{H}^{*} = \boldsymbol{0}$, $\partial\mathcal{L}(\boldsymbol{H}, \boldsymbol{c}, \boldsymbol{\rho}, \mu)/\partial\boldsymbol{\rho}^{*} = \boldsymbol{0}$, we can obtain

$$\begin{cases} \boldsymbol{B}^{H}(\boldsymbol{B}\boldsymbol{H} - \boldsymbol{X}) + \boldsymbol{R}^{H}(\boldsymbol{c})\boldsymbol{\rho} = \boldsymbol{0} \\ \boldsymbol{R}(\boldsymbol{c})\boldsymbol{H} = \boldsymbol{0} \end{cases} \tag{C.3}$$

For the first expression in Eq. (C.3), we have

$$\boldsymbol{B}^{H}\boldsymbol{B}\boldsymbol{H} = \boldsymbol{B}^{H}\boldsymbol{X} - \boldsymbol{R}^{H}(\boldsymbol{c})\boldsymbol{\rho} \tag{C.4}$$

Since $\boldsymbol{B}^{H}\boldsymbol{B}$ is singular in most cases, we adopt the pseudo-inverse technique and denote $\boldsymbol{G} = \boldsymbol{B}^{H}\boldsymbol{B}$, and then the solution $\boldsymbol{H}$ is

$$\boldsymbol{H} = \boldsymbol{G}^{+}\left[\boldsymbol{B}^{H}\boldsymbol{X} - \boldsymbol{R}^{H}(\boldsymbol{c})\boldsymbol{\rho}\right] + (\boldsymbol{I} - \boldsymbol{G}^{+}\boldsymbol{G})\boldsymbol{W}_1 \tag{C.5}$$

Here, H contains the all solutions of Eq. (C.4) and $\boldsymbol{W}_1$ is an arbitrary matrix. Denote $\boldsymbol{F} = \boldsymbol{R}(\boldsymbol{c})\boldsymbol{G}^{+}\boldsymbol{R}^{H}(\boldsymbol{c})$, substitute $\boldsymbol{H}$ into the second expression of Eq. (C.3) and notice $\boldsymbol{R}^{H}(\boldsymbol{c})$ resides in the column space of $\boldsymbol{G}$ from Eq. (C.4), and then we have

$$\begin{aligned} \boldsymbol{F}\boldsymbol{\rho} =& \boldsymbol{R}(\boldsymbol{c})\left[\boldsymbol{G}^{+}\boldsymbol{B}^{H}\boldsymbol{X} + (\boldsymbol{I} - \boldsymbol{G}^{+}\boldsymbol{G})\boldsymbol{W}_1\right] \\ =& \boldsymbol{R}(\boldsymbol{c})\boldsymbol{G}^{+}\boldsymbol{B}^{H}\boldsymbol{X} \end{aligned} \tag{C.6}$$

As $\boldsymbol{F}$ is also singular in most cases, adopting the pseudo-inverse technique, we can write the solution $\boldsymbol{\rho}$ as

$$\boldsymbol{\rho} = \boldsymbol{F}^{+}\boldsymbol{R}(\boldsymbol{c})\boldsymbol{G}^{+}\boldsymbol{B}^{H}\boldsymbol{X} + (\boldsymbol{I} - \boldsymbol{F}^{+}\boldsymbol{F})\boldsymbol{W}_2 \tag{C.7}$$

where $\boldsymbol{W}_2$ is an arbitrary matrix. Substituting Eqs. (C.7) and (14) into Eq. (C.5), and noticing $\boldsymbol{R}(\boldsymbol{c})$ resides in the column space of $\boldsymbol{F}$ from Eq. (C.6) we have

$$\boldsymbol{H} = \boldsymbol{Y} - \boldsymbol{G}^{+}\boldsymbol{R}^{H}(\boldsymbol{c})\boldsymbol{F}^{+}\boldsymbol{R}(\boldsymbol{c})\boldsymbol{Y} \tag{C.8}$$

where

$$\boldsymbol{Y} = \left(\boldsymbol{B}^{H}\boldsymbol{B}\right)^{+}\boldsymbol{B}^{H}\boldsymbol{X} \tag{C.9}$$

Substituting Eq. (C.8) into Eq. (11), we can obtain Eq. (15). For minimizing Eq. (18), set $\partial\mathcal{L}(\boldsymbol{c}^{(i)}, \mu)/\partial\boldsymbol{c}^{(i)*} = \boldsymbol{0}$ and $\partial\mathcal{L}(\boldsymbol{c}^{(i)}, \mu)/\partial\mu^{*} = 0$, and then we can obtain

$$\begin{cases} \boldsymbol{T}^{H}(\boldsymbol{Y})\left\{\boldsymbol{I}_N \otimes \left[\boldsymbol{R}(\boldsymbol{c}^{(i-1)})(\boldsymbol{B}^{H}\boldsymbol{B})^{+}\boldsymbol{R}^{H}(\boldsymbol{c}^{(i-1)})\right]^{+}\right\}\boldsymbol{T}(\boldsymbol{Y})\boldsymbol{c}^{(i)} + \mu\boldsymbol{\omega} = \boldsymbol{0} \\ \boldsymbol{\omega}^{H}\boldsymbol{c} = 0 \end{cases} \tag{C.10}$$

Introduce a new auxiliary variable $\boldsymbol{\alpha}$

$$\boldsymbol{\alpha} = \left\{\boldsymbol{I}_N \otimes \left[\boldsymbol{R}\left(\boldsymbol{c}^{(i-1)}\right)\left(\boldsymbol{B}^{H}\boldsymbol{B}\right)^{+}\boldsymbol{R}^{H}\left(\boldsymbol{c}^{(i-1)}\right)\right]^{+}\right\}\boldsymbol{T}(\boldsymbol{Y})\boldsymbol{c}^{(i)} \tag{C.11}$$

So

$$\boldsymbol{T}^{H}(\boldsymbol{Y})\boldsymbol{\alpha} + \mu\boldsymbol{\omega} = \boldsymbol{0} \tag{C.12}$$

As we have $\boldsymbol{T}(\boldsymbol{Y}) = [\boldsymbol{T}^{T}(\boldsymbol{Y}_{.1}), \boldsymbol{T}^{T}(\boldsymbol{Y}_{.2}), \ldots, \boldsymbol{T}^{T}(\boldsymbol{Y}_{.N})]^{T}$ and $\boldsymbol{T}(\boldsymbol{Y}_{.n}), n = 1, 2, \ldots, N$ resides in the column space of $\boldsymbol{R}(\boldsymbol{c})$ which contains the column space of $\left[\boldsymbol{R}(\boldsymbol{c}^{(i-1)})(\boldsymbol{B}^{H}\boldsymbol{B})^{+}\boldsymbol{R}^{H}(\boldsymbol{c}^{(i-1)})\right]^{+}$. According to Proposition C.1, Eq. (C.11) can be transformed into

$$\left\{\boldsymbol{I}_N \otimes \left[\boldsymbol{R}\left(\boldsymbol{c}^{(i-1)}\right)\left(\boldsymbol{B}^{H}\boldsymbol{B}\right)^{+}\boldsymbol{R}^{H}\left(\boldsymbol{c}^{(i-1)}\right)\right]\right\}\boldsymbol{\alpha} = \boldsymbol{T}(\boldsymbol{Y})\boldsymbol{c}^{(i)} \tag{C.13}$$

Introduce another new auxiliary variable $\boldsymbol{\beta}$

$$\boldsymbol{\beta} = \left\{\boldsymbol{I}_N \otimes \left[\left(\boldsymbol{B}^{H}\boldsymbol{B}\right)^{+}\boldsymbol{R}^{H}\left(\boldsymbol{c}^{(i-1)}\right)\right]\right\}\boldsymbol{\alpha} \tag{C.14}$$

So, Eq. (C.13) can be transformed into

$$\left[\boldsymbol{I}_N \otimes \boldsymbol{R}\left(\boldsymbol{c}^{(i-1)}\right)\right]\boldsymbol{\beta} = \boldsymbol{T}(\boldsymbol{Y})\boldsymbol{c}^{(i)} \tag{C.15}$$

For Eq. (C.14), multiply both sides by $\boldsymbol{I}_N \otimes (\boldsymbol{B}^{H}\boldsymbol{B})$. Since $\boldsymbol{R}(\boldsymbol{c})$ resides in the column space of $\boldsymbol{B}^{H}\boldsymbol{B}$ according to Eqs. (C.4), Eq. (C.15) can be transformed into

$$\left[\boldsymbol{I}_N \otimes \left(\boldsymbol{B}^{H}\boldsymbol{B}\right)\right]\boldsymbol{\beta} = \left[\boldsymbol{I}_N \otimes \boldsymbol{R}^{H}\left(\boldsymbol{c}^{(i-1)}\right)\right]\boldsymbol{\alpha} \tag{C.16}$$

Put Eqs. (C.10), (C.12), (C.15) and (C.16) together, we have

$$\begin{cases} \boldsymbol{T}^{H}(\boldsymbol{Y})\boldsymbol{\alpha} + \mu\boldsymbol{\omega} = \boldsymbol{0} \\ \left[\boldsymbol{I}_N \otimes \boldsymbol{R}\left(\boldsymbol{c}^{(i-1)}\right)\right]\boldsymbol{\beta} = \boldsymbol{T}(\boldsymbol{Y})\boldsymbol{c}^{(i)} \\ \left[\boldsymbol{I}_N \otimes \left(\boldsymbol{B}^{H}\boldsymbol{B}\right)\right]\boldsymbol{\beta} = \left[\boldsymbol{I}_N \otimes \boldsymbol{R}^{H}\left(\boldsymbol{c}^{(i-1)}\right)\right]\boldsymbol{\alpha} \\ \boldsymbol{\omega}^{H}\boldsymbol{c} = 0 \end{cases} \tag{C.17}$$

Rearrange Eq. (C.17) in a matrix form, we have Eq. (19).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.sigpro.2018.07.001

References

[1] H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach, IEEE Signal Process. Mag. 13 (4) (1996) 67–94, doi:10.1109/79.526899.

[2] R.O. Schmidt, Multiple emitter location and signal parameter estimation, in: RADC Spectrum Estimation Workshop, 1979, pp. 243–258.

[3] R. Roy, A. Paulraj, T. Kailath, Direction-of-arrival estimation by subspace rotation methods - ESPRIT, in: Proc. and Signal Processing ICASSP '86. IEEE Int. Conf. Acoustics, Speech, 11, 1986, pp. 2495–2498, doi:10.1109/ICASSP.1986.1168673.

[4] D. Malioutov, M. Cetin, A.S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (2005) 3010–3022, doi:10.1109/TSP.2005.850882.

[5] Z.M. Liu, Z.T. Huang, Y.Y. Zhou, An efficient maximum likelihood method for direction-of-arrival estimation via sparse bayesian learning, IEEE Trans. Wireless Commun. 11 (10) (2012) 1–11, doi:10.1109/TWC.2012.090312.111912.

[6] D.P. Wipf, B.D. Rao, Sparse bayesian learning for basis selection, IEEE Trans. Signal Process. 52 (8) (2004) 2153–2164, doi:10.1109/TSP.2004.831016.

[7] J.A. Fessler, A.O. Hero, Space-alternating generalized expectation-maximization algorithm, IEEE Trans. Signal Process. 42 (10) (1994) 2664–2677, doi:10.1109/78.324732.
[8] Z. Yang, L. Xie, C. Zhang, Off-grid direction of arrival estimation using sparse bayesian inference, IEEE Trans. Signal Process. 61 (1) (2013) 38–43, doi:10.1109/TSP.2012.2222378.
[9] E.J. Candes, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (12) (2005) 4203–4215, doi:10.1109/TIT.2005.858979.
[10] H. Pan, R. Scheibler, E. Bezzam, I. Dokmanic, M. Vetterli, Frida: Fri-based doa estimation for arbitrary array layouts, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 3186–3190, doi:10.1109/ICASSP.2017.7952744.
[11] R. Scheibler, Rake, Peel, Sketch: The Signal Processing Pipeline Revisited, EPFL, 2017 Ph.D. thesis. doi: 10.5075/epfl-thesis-7651.
[12] H. Pan, T. Blu, M. Vetterli, Towards generalized fri sampling with an application to source resolution in radioastronomy, IEEE Trans. Signal Process. 65 (4) (2017) 821–835, doi:10.1109/TSP.2016.2625274.
[13] M. Vetterli, P. Marziliano, T. Blu, Sampling signals with finite rate of innovation, IEEE Trans. Signal Process. 50 (6) (2002) 1417–1428, doi:10.1109/TSP.2002.1003065.
[14] T. Blu, P.L. Dragotti, M. Vetterli, P. Marziliano, L. Coulot, Sparse sampling of signal innovations, IEEE Signal Process. Mag. 25 (2) (2008) 31–40, doi:10.1109/MSP.2007.914998.
[15] Z. Doan, C. Gilliam, T. Blu, D.V.D. Ville, Reconstruction of finite rate of innovation signals with model-fitting approach, IEEE Trans. Signal Process. 63 (22) (2015) 6024–6036, doi:10.1109/TSP.2015.2461513.
[16] D.H. Brandwood, A complex gradient operator and its application in adaptive array theory, IEE Proc. F Commun. Radar Signal Process. 130 (1) (1983) 11–16, doi:10.1049/ip-f-1.1983.0003.
[17] A. Hjrungnes, Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications, first ed., Cambridge University Press, New York, NY, USA, 2011.
[18] Y. Bresler, A. Macovski, Exact maximum likelihood parameter estimation of superimposed exponential signals in noise, Signal Process. IEEE Trans. Acoust. Speech 34 (5) (1986) 1081–1089, doi:10.1109/TASSP.1986.1164949.
[19] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Contr. 19 (6) (1974) 716–723, doi:10.1109/TAC.1974.1100705.
[20] M. Wax, T. Kailath, Detection of signals by information theoretic criteria, Signal Process. IEEE Trans. Acoust. Speech 33 (2) (1985) 387–392, doi:10.1109/TASSP.1985.1164557.
[21] G.H. Golub, C.F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, MD, USA, 1996.
[22] J. Li, P. Stoica, Z.-S. Liu, Comparative study of iqml and mode direction-of-arrival estimators, IEEE Trans. Signal Process. 46 (1) (1998) 149–160, doi:10.1109/78.651203.
[23] P. Stoica, J. Li, T. Sderstrm, On the inconsistency of iqml, Signal Process. 56 (2) (1997) 185–190, doi:10.1016/S0165-1684(96)00167-3.
[24] P. Babu, P. Stoica, T.L. Marzetta, An iqml type algorithm for ar parameter estimation from noisy covariance sequences, in: Proc. 17th European Signal Processing Conf, 2009, pp. 1022–1026, doi:10.5281/zenodo.41403.
[25] H. Hung, M. Kaveh, Focussing matrices for coherent signal-subspace processing, IEEE Trans. Acoust. 36 (8) (1988) 1272–1281, doi:10.1109/29.1655.