Digital Signal Processing 82 (2018) 187–193

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Sparse Bayesian learning for off-grid DOA estimation with nested arrays

Fangfang Chen [a], Jisheng Dai [a,*], Nan Hu [b], Zhongfu Ye [c]

[a] Department of Electronic Engineering, Jiangsu University, Zhenjiang 212013, China
[b] School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China
[c] Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China

ARTICLE INFO

Article history:
Available online 10 August 2018

Keywords:
Sparse Bayesian learning (SBL)
DOA estimation
Nested array
Off-grid

ABSTRACT

The existing off-grid sparse Bayesian learning (SBL) DOA estimation method for nested arrays suffers from two major drawbacks: reduced array aperture and high modeling error. To solve these issues, a new data model formulation is first presented in this paper, in which we take the noise variance as a part of the unknown signal of interest, so as to learn its value by the Bayesian inference inherently. Then, we provide a novel grid refining procedure to eliminate the modeling error caused by off-grid gap, where we consider the locations of grid points as adjustable parameters and proceed to refine the grid point iteratively. Simulation results demonstrate that our method significantly improves the DOA estimation performance especially using a coarse grid.

1. Introduction

Direction-of-arrival (DOA) estimation of multiple narrow-band sources using antenna arrays has attracted considerable interest in radar, sonar and mobile communication etc. [1]. Many high-resolution algorithms have been proposed to estimate DOAs during the last decades. Usually, only up to $M-1$ sources can be resolved by an M-element uniform linear array (ULA). To achieve a dramatic increase in the degrees of freedom (DOF), many non-uniform arrays and co-arrays have been introduced [2–4], among which the nested array is a popular non-uniform array geometry because it has closed-form expressions for array configurations and provides more DOF than a ULA with a given number of sensors [5].

A number of efficient approaches that can be applied to nested arrays have been put forward [5–9]. The so-called spatial smoothing based MUSIC (SS-MUSIC) approach [5] constructed an augmented covariance matrix with the spatially smoothing technique, and then applied the traditional subspace based algorithms, such as MUSIC [10] and ESPRIT [11], to identify up to $\mathcal{O}(M^2)$ sources with an M-element two level nested array. However, it is well known that the traditional subspace methods require a large number of snapshots to obtain the signal or noise subspace accurately, and their performance may deteriorate significantly when the signal-to-noise ratio (SNR) is low.

Recently, the rapid development of sparse representation technique has greatly promoted the rise of sparse representation methods in DOA estimation [12,13]. These methods, as well as the sparse Bayesian learning (SBL) based methods [14–16], have shown many excellent characteristics, e.g., improved robustness to noise, limited number of snapshots and blind number of sources. In CMSR [17], the lower left triangular covariance elements were aligned to form a vector, and the signal components were recovered from this vector via sparse representation for DOA estimation. The grid-less method ANM [18] used a convex relaxation approach in which the atomic norm was adopted to recover the frequency and they solved the atomic norm via semi-definite programming. The ICSPICE [19] combined the Capon [20] and SPICE [21] methods to estimate the value of angle deviation from the grid and then execute the DOA estimation algorithm iteratively under the stochastic maximum likelihood (SML) criterion. The SBL-based methods [22–26] exploit the sparsity information with a sparse prior assumption for the signal of interest from a Bayesian perspective, and include the l_1-norm minimization method as a special case when a maximum *a posteriori* (MAP) optimal estimate is adopted with a Laplace signal prior. Theoretical and empirical results show that the SBL-based methods can achieve enhanced performance over l_1 regularized optimization [27].

To the best of our knowledge, the SBL method for nested arrays was first addressed in [28], which has several notable contributions, e.g., 1) through eliminating the effect of noise and normalizing the approximation error of covariance matrix with an identity variance matrix, it facilitated a noise-free SBL problem;

* Corresponding author.
E-mail addresses: ffchenujs@gmail.com (F. Chen), jsdai@ujs.edu.cn (J. Dai), hunan@suda.edu.cn (N. Hu), yezf@ustc.edu.cn (Z. Ye).

https://doi.org/10.1016/j.dsp.2018.08.004
1051-2004/

2) a second-order Taylor expansion instead of first-order Taylor expansion was used to alleviate the modeling error caused by off-grid gap. Nevertheless, the main drawbacks of this technique are as follows:

- **Reduced Working Array Aperture**. In order to eliminate the effect of noise, only a triming virtual signal vector was used for the DOA estimation. In this case, the working array aperture is reduced from M^2 to $M^2/2+M-2$, which in return will degrade the overall DOA estimation performance significantly.
- **High Modeling Error with Coarse Grids**. Adopting a Taylor expansion can alleviate the modeling error caused by off-grid gap, but it cannot bring a significant improvement, because the distance between the true DOA and the nearest grid point remains unchanged in each iteration. When a coarse grid is used, it may lead to a high modeling error.

In this paper, we propose a novel off-grid SBL method for DOA estimation with nested arrays. To overcome the first aforementioned drawback, we adopt the whole virtual signal vector to estimate DOAs, rather than a triming one used in [28]. The main difficulty in handling the whole virtual signal vector is that the formulation is not in a standard form of SBL. To solve this problem, we take the noise variance as a part of the unknown signal of interest, and then its value can be learned by the Bayesian inference inherently. In order to overcome the second drawback, we further propose a new grid refining procedure to eliminate the modeling error caused by off-grid gap, where we consider the locations of grid points as adjustable parameters and proceed to update the grid points in each iteration. Simulation results demonstrate that the DOA estimation performance of our method is greatly enhanced.

2. Data model and the existing algorithm review

Consider K narrow-band far-field sources impinging on an M-element nested array that consists of two concatenated ULAs, where the inner and outer ULAs have M_1 and M_2 sensors with inter-sensor spacing d and $(M_1+1)d$, respectively. Specifically, if the first sensor of inner ULA is assumed as the reference, all the sensors' position can be denoted as:

$$d\cdot\mathbf{r}=d\cdot\{0,\ 1,\ldots,\ (M_1-1),\ M_1,\ 2(M_1+1)-1,\ldots,\ M_2(M_1+1)-1\}. \tag{1}$$

The K signals, $s_1(t), s_2(t), \ldots, s_K(t)$, arrive at the array from distinct directions $\boldsymbol{\theta}=\{\theta_k, k=1,2,\ldots,K\}$, with respect to the array axis. Then, the output vector received by the nested array is given by

$$\mathbf{x}(t)=\sum_{k=1}^{K}\mathbf{a}(\theta_k)s_k(t)+\mathbf{n}(t)=\mathbf{A}\mathbf{s}(t)+\mathbf{n}(t),\quad t=1,2,\ldots,T \tag{2}$$

where T is the number of snapshots, $\mathbf{x}(t)=[x_1(t),x_2(t),\ldots,x_M(t)]^T$, $\mathbf{s}(t)=[s_1(t),s_2(t),\ldots,s_K(t)]^T$, the narrowband signal $s_n(t)=\rho_n\exp(j2\pi f_{d_n}t)$ with f_{d_n} being the Doppler frequency and ρ_n being the amplitude which is influenced mainly by the radar cross section (RCS), the antenna gain in the target direction, the propagation loss and so on, $\mathbf{A}=[\mathbf{a}(\theta_1),\mathbf{a}(\theta_2),\ldots,\mathbf{a}(\theta_K)]$, $\mathbf{a}(\theta_k)=[1,\nu_{\theta_k}^{r_2},\ldots,\nu_{\theta_k}^{r_M}]^T$, $\nu_{\theta_k}=\exp(-j2\pi d\sin(\theta_k)/\lambda)$, λ is the wavelength of the source, r_m denotes the mth element of $\mathbf{r}$, and $\mathbf{n}(t)=[n_1(t),n_2(t),\ldots,n_M(t)]^T$ is an unknown noise vector. All the source signals $s_k(t)$s are independent and each has a complex Gaussian distribution with zero mean and variance σ_k^2, $k\in\{1,2,\ldots,K\}$, and elements in the noise vector $\mathbf{n}$ are also independent and have zero mean complex Gaussian distributions with a common variance σ_n^2. Assume that both the source signals and the noise are considered stationary, then the covariance matrix of the output vector $\mathbf{x}(t)$ can be expressed as

$$\mathbf{R}_x=E\left\{\mathbf{x}(t)\mathbf{x}^H(t)\right\}=\mathbf{A}\mathbf{R}_s\mathbf{A}^H+\sigma_n^2\mathbf{I}_M \tag{3}$$

where $E\{\cdot\}$ is the expectation operator, $\mathbf{R}_s=E\left\{\mathbf{s}(t)\mathbf{s}^H(t)\right\}=\mathrm{diag}\{\sigma_1^2,\sigma_2^2,\ldots,\sigma_K^2\}$, $\mathrm{diag}\{\cdot\}$ represents the diagonal matrix operator, and $\mathbf{I}_M$ denotes the $M\times M$ identity matrix. After vectoring the correlation matrix $\mathbf{R}_x$, we are able to obtain a column data vector

$$\mathbf{y}\triangleq\mathrm{vec}(\mathbf{R}_x)=\left(\mathbf{A}^*\circ\mathbf{A}\right)\mathbf{p}+\sigma_n^2\mathbf{1}_n \tag{4}$$

where $\mathbf{A}^*\circ\mathbf{A}=[\dot{\boldsymbol{a}}(\theta_1),\dot{\boldsymbol{a}}(\theta_2),\ldots,\dot{\boldsymbol{a}}(\theta_K)]$ with $\dot{\boldsymbol{a}}(\theta_k)=\mathbf{a}^*(\theta_k)\otimes\mathbf{a}(\theta_k)$, $\mathbf{p}=[\sigma_1^2,\sigma_2^2,\ldots,\sigma_K^2]^T$, $\mathbf{1}_n=[\mathbf{e}_1^T,\mathbf{e}_2^T,\ldots,\mathbf{e}_M^T]^T$, $\mathbf{e}_m$ denotes a vector of all zeros except for the mth element being one, and $\circ$ and $\otimes$ denote KR product and Kronecker product, respectively.

In practice, the covariance matrix $\mathbf{R}_x$ is usually estimated from a finite number of snapshots, i.e.,

$$\hat{\mathbf{R}}_x=\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}^H(t) \tag{5}$$

which will bring about approximation error. It was revealed in [29] that the approximation error of the vectorized covariance matrix follows an asymptotic complex Gaussian distribution:

$$\boldsymbol{\varepsilon}\triangleq\mathrm{vec}(\hat{\mathbf{R}}_x-\mathbf{R}_x)\sim\mathcal{CN}(\mathbf{0},\frac{1}{T}\mathbf{R}_x^T\otimes\mathbf{R}_x). \tag{6}$$

Therefore, we have

$$\hat{\mathbf{y}}\triangleq\mathrm{vec}(\hat{\mathbf{R}}_x)=\left(\mathbf{A}^*\circ\mathbf{A}\right)\mathbf{p}+\sigma_n^2\mathbf{1}_n+\boldsymbol{\varepsilon}. \tag{7}$$

In the traditional SS-MUSIC method [5], a virtual ULA steering matrix $\tilde{\mathbf{A}}\in\mathbb{C}^{(M^2/2+M-1)\times K}$ was obtained by removing the repeat rows from $\mathbf{A}^*\circ\mathbf{A}$ and sorting the remaining rows accordingly, where the ith row of $\tilde{\mathbf{A}}$ corresponds to the sensor location at $(-M^2/4-M/2+i)d$. If we define the selecting matrix $\mathbf{F}\in\mathbb{C}^{(M^2/2+M-1)\times M^2}$ such that $\tilde{\mathbf{A}}=\mathbf{F}(\mathbf{A}^*\circ\mathbf{A})$, the corresponding virtual output vector can be written as

$$\tilde{\mathbf{y}}=\tilde{\mathbf{A}}\mathbf{p}+\sigma_n^2\tilde{\mathbf{e}}+\tilde{\boldsymbol{\varepsilon}} \tag{8}$$

where $\tilde{\mathbf{y}}=\mathbf{F}\hat{\mathbf{y}}$, $\tilde{\boldsymbol{\varepsilon}}=\mathbf{F}\boldsymbol{\varepsilon}$, $\tilde{\mathbf{e}}$ is a vector of all zeros except for the center element being one. With (8), the classical smoothing technique can be applied to estimate DOAs. On the other hand, Yang *et al.* [28] noticed that the term $\sigma_n^2\tilde{\mathbf{e}}$ in (8) has only one nonzero center element, and then they removed the center row from $\tilde{\mathbf{y}}$ directly so as to obtain a noise-free received signal. Although such operation can avoid the estimation of noise variance, the working array aperture is reduced from M^2 to $M^2/2+M-2$, which in return may significantly degrade the overall performance. Besides, it is somewhat difficult to compute the perturbation variance matrix of $\tilde{\boldsymbol{\varepsilon}}$, because it is no longer $\frac{1}{T}\mathbf{R}_x^T\otimes\mathbf{R}_x$, but depends on the selecting matrix $\mathbf{F}$.

3. The proposed method

In this section, we try to propose a novel SBL-based method for the DOA estimation with nested arrays. To avoid reducing the working array aperture, we adopt the original virtual signal model (7), rather than the triming one (8). The main difficulty in di-

rectly handling (7) via SBL is that σ_n^2 is unknown and $\boldsymbol{\varepsilon}$ follows an asymptotic complex Gaussian distribution, which is not in a standard SBL form. To solve this problem, we take σ_n^2 as a part of the unknown signal of interest, i.e.,

$$\hat{\mathbf{y}} = \begin{bmatrix} \mathbf{A}^* \circ \mathbf{A} & \mathbf{1}_n \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \sigma_n^2 \end{bmatrix} + \boldsymbol{\varepsilon} \tag{9}$$

where $\begin{bmatrix} \mathbf{A}^* \circ \mathbf{A} & \mathbf{1}_n \end{bmatrix}$ can be seen as a reconstructed measurement matrix. Following the convention in sparse representation methods [12], we let $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_i\}_{i=1}^{\hat{K}}$ be a fixed sampling grid that uniformly covers the DOA range $[-\pi/2, \pi/2]$, where $\hat{K}$ denotes the grid number. If the grid is fine enough such that the true DOAs lie on (or, practically, close to) the grid, we can use the following model for $\hat{\mathbf{y}}$:

$$\hat{\mathbf{y}} = \begin{bmatrix} \bar{\mathbf{A}} & \mathbf{1}_n \end{bmatrix} \begin{bmatrix} \bar{\mathbf{p}} \\ \sigma_n^2 \end{bmatrix} + \boldsymbol{\varepsilon} \tag{10}$$

where $\bar{\mathbf{A}} \triangleq [\dot{\boldsymbol{a}}(\hat{\theta}_1), \dot{\boldsymbol{a}}(\hat{\theta}_2), \ldots, \dot{\boldsymbol{a}}(\hat{\theta}_{\hat{K}})]$ and $\bar{\mathbf{p}}$ is a zero-padded extension of $\mathbf{p}$ whose non-zero elements correspond to the true DOAs at $\{\theta_k, k = 1, 2, \ldots, K\}$. Note that the assumption that the true DOAs are located on the predefined spatial grids is not always valid in practical implementations. To handle the off-grid gap, a linear approximation method with the first-order Taylor expansion is used as in [23]. Specifically if $\theta_k \notin \{\hat{\theta}_i\}_{i=1}^{\hat{K}}$ and $\hat{\theta}_{n_k}, n_k \in \{1, 2, \ldots, \hat{K}\}$, is the nearest grid point to θ_k, the steering vector $\dot{\mathbf{a}}(\theta_k)$ is approximated by the linearization:

$$\dot{\mathbf{a}}(\theta_k) \approx \dot{\mathbf{a}}(\hat{\theta}_{n_k}) + \mathbf{b}(\hat{\theta}_{n_k}) \left(\theta_k - \hat{\theta}_{n_k}\right) \tag{11}$$

where $\mathbf{b}(\hat{\theta}_{n_k}) = \dot{\mathbf{a}}'(\hat{\theta}_{n_k})$, $(\cdot)'$ represents the first-order derivative product. Then, the observation model (10) can be rewritten as

$$\hat{\mathbf{y}} = \begin{bmatrix} \bar{\mathbf{A}}(\boldsymbol{\beta}) & \mathbf{1}_n \end{bmatrix} \begin{bmatrix} \bar{\mathbf{p}} \\ \sigma_n^2 \end{bmatrix} + \boldsymbol{\varepsilon} \tag{12}$$

where $\bar{\mathbf{A}}(\boldsymbol{\beta}) = \bar{\mathbf{A}} + \mathbf{B}\text{diag}\{\boldsymbol{\beta}\}$, $\mathbf{B} = [\mathbf{b}(\hat{\theta}_1), \ldots, \mathbf{b}(\hat{\theta}_{\hat{K}})]$, and $\boldsymbol{\beta}$ is a zero vector except that the n_k-th element $[\boldsymbol{\beta}]_{n_k} = \theta_k - \hat{\theta}_{n_k}, k = 1, 2, \ldots, K$.

It is true that the modeling error caused by off-grid gap can be alleviated by the new model (12), but it can not be fully eliminated. [28] proposed to use a second-order Taylor expansion instead of a first-order Taylor expansion. This idea is sound, as second-order Taylor expansion can give a better approximation than first-order one; however, it cannot bring a significant improvement, because the distance between the true DOA θ_k and the nearest grid point to θ_k (denoted by $\hat{\theta}_{n_k}$) remains unchanged in each iteration. When a coarse grid is used, it may lead to a high modeling error. Our previous work [16] provided an alternative to fix the off-grid gap with a very high accuracy, but applying the rooting method for nested array might be very time-consuming, because the order of polynomial is rather high. Hence, we try to propose a new grid refining method in this paper (please refer to Section 3.3).

3.1. Sparse Bayesian formulation

For ease of presentation, we define $\boldsymbol{\Phi}(\boldsymbol{\beta})$ and $\mathbf{d}$ as follows:

$$\hat{\mathbf{y}} = \underbrace{\begin{bmatrix} \bar{\mathbf{A}}(\boldsymbol{\beta}) & \mathbf{1}_n \end{bmatrix}}_{\triangleq \boldsymbol{\Phi}(\boldsymbol{\beta})} \underbrace{\begin{bmatrix} \bar{\mathbf{p}} \\ \sigma_n^2 \end{bmatrix}}_{\triangleq \mathbf{d}} + \boldsymbol{\varepsilon}. \tag{13}$$

From (6), we have

$$p(\hat{\mathbf{y}}|\mathbf{d}; \boldsymbol{\beta}) = \mathcal{CN}(\hat{\mathbf{y}}|\boldsymbol{\Phi}(\boldsymbol{\beta})\mathbf{d}, \mathbf{W}) \tag{14}$$

where $\mathbf{W} \triangleq \frac{1}{T}\mathbf{R}_x^T \otimes \mathbf{R}_x$. A typical SBL treatment of $\mathbf{d}$ begins by assigning a non-stationary Gaussian prior distribution with a variance δ_i for each element of $\mathbf{d}$. Letting $\boldsymbol{\delta} = [\delta_1, \delta_2, \ldots, \delta_{\hat{K}+1}]^T$, we obtain

$$p(\mathbf{d}|\boldsymbol{\delta}) = \mathcal{CN}(\mathbf{d}|\mathbf{0}, \boldsymbol{\Delta}) \tag{15}$$

where $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta})$. The hyper-parameter δ_is are further modeled as independent Gamma distributions [14,15], i.e.,

$$p(\boldsymbol{\delta}) = \prod_{i=1}^{\hat{K}+1} \Gamma(\delta_i;\ 1, \rho) \tag{16}$$

where ρ is a small positive constraint (e.g., $\rho = 0.01$ [23,15]).

3.2. Bayesian inference

As $p(\mathbf{d}, \boldsymbol{\delta}|\hat{\mathbf{y}}; \boldsymbol{\beta})$ cannot be explicitly calculated, we will exploit an EM algorithm to perform the Bayesian inference. We first treat $\mathbf{d}$ as a hidden variable, whose posterior distribution is [23,14]:

$$p(\mathbf{d}|\hat{\mathbf{y}}, \boldsymbol{\delta}; \boldsymbol{\beta}) = \mathcal{CN}(\mathbf{d}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{17}$$

where

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}\boldsymbol{\Phi}^H(\boldsymbol{\beta})\mathbf{W}^{-1}\hat{\mathbf{y}} \tag{18}$$

$$\boldsymbol{\Sigma} = (\boldsymbol{\Phi}^H(\boldsymbol{\beta})\mathbf{W}^{-1}\boldsymbol{\Phi}(\boldsymbol{\beta}) + \boldsymbol{\Delta}^{-1})^{-1}. \tag{19}$$

Then, E-step and M-step are repeatedly proceeded until a prescribed accuracy is achieved. Note that since the following results about E-step and M-step can be derived by following the similar procedure as in [14] (also refer to [23]), the detailed derivations are omitted for brevity.

- In E-step, we try to construct a lower-bound on the evidence function $p(\boldsymbol{\delta}|\hat{\mathbf{y}}; \boldsymbol{\beta})$, or equivalently, $\ln p(\hat{\mathbf{y}}, \boldsymbol{\delta}; \boldsymbol{\beta})$:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\delta}; \boldsymbol{\beta}) &= E\left\{\ln p(\mathbf{d}, \hat{\mathbf{y}}, \boldsymbol{\delta}; \boldsymbol{\beta})\right\}_{p(\mathbf{d}|\hat{\mathbf{y}}, \boldsymbol{\delta}; \boldsymbol{\beta})} \\ &= E\left\{\ln p(\hat{\mathbf{y}}|\mathbf{d}; \boldsymbol{\beta}) p(\mathbf{d}|\boldsymbol{\delta}) p(\boldsymbol{\delta})\right\}_{p(\mathbf{d}|\hat{\mathbf{y}}, \boldsymbol{\delta}; \boldsymbol{\beta})} \end{aligned} \tag{20}$$

  where $E\{\cdot\}_{p(x)}$ denotes the expectation with respect to $p(x)$.
- In M-step, the hyperparameter updates that maximize the lower bound are determined by solving the following optimization problem

$$(\boldsymbol{\delta}^\star, \boldsymbol{\beta}^\star) = \arg\max_{\boldsymbol{\delta}, \boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\delta}; \boldsymbol{\beta}). \tag{21}$$

  Then, the hyperparameter updates for δ_is can be obtained as:

$$\delta_i^{new} = \frac{\sqrt{1 + 4\rho[\boldsymbol{\Xi}]_{ii}} - 1}{2\rho}, \quad \forall i \tag{22}$$

  where $\boldsymbol{\Xi} \triangleq \boldsymbol{\mu}\boldsymbol{\mu}^H + \boldsymbol{\Sigma}$, $[\cdot]_{ii}$ denotes the (i, i)th element of a matrix. However, as the structure of measurement matrix $\boldsymbol{\Phi}(\boldsymbol{\beta})$ is rather different from the one in [23], the existing updating derivation for $\boldsymbol{\beta}$ is not applicable to our problem. We will address the update for $\boldsymbol{\beta}$ in the next subsection, as well as a new grid refining procedure.

3.3. Grid refining

In this subsection, we try to provide a new grid refining method,[1] which can be regarded as an improved version of the off-grid method proposed in [23].

Ignoring the independent terms, (20) becomes

$$E\left\{\ln p(\hat{\mathbf{y}}|\mathbf{d};\boldsymbol{\beta})\right\}_{p(\mathbf{d}|\hat{\mathbf{y}},\boldsymbol{\delta};\boldsymbol{\beta})}$$

$$= -E\left\{(\hat{\mathbf{y}}-\boldsymbol{\Phi}(\boldsymbol{\beta})\mathbf{d})^H\mathbf{W}^{-1}(\hat{\mathbf{y}}-\boldsymbol{\Phi}(\boldsymbol{\beta})\mathbf{d})\right\}_{p(\mathbf{d}|\hat{\mathbf{y}},\boldsymbol{\delta};\boldsymbol{\beta})} \tag{23}$$

$$= -E\left\{\|\mathbf{W}^{-1/2}(\hat{\mathbf{y}}-\boldsymbol{\Phi}(\boldsymbol{\beta})\mathbf{d})\|_2^2\right\}_{p(\mathbf{d}|\hat{\mathbf{y}},\boldsymbol{\delta};\boldsymbol{\beta})} \tag{24}$$

$$= -\|\mathbf{z}-\mathbf{W}^{-1/2}\boldsymbol{\Phi}(\boldsymbol{\beta})\boldsymbol{\mu}\|_2^2 - \mathrm{tr}\left(\mathbf{W}^{-1}\boldsymbol{\Phi}(\boldsymbol{\beta})\boldsymbol{\Sigma}\boldsymbol{\Phi}^H(\boldsymbol{\beta})\right) + \mathrm{const} \tag{25}$$

where $\mathbf{z}=\mathbf{W}^{-1/2}\hat{\mathbf{y}}$ and $\mathbf{W}^{-1/2}$ denotes the Hermitian square root of $\mathbf{W}^{-1}$. The first term in the right side of (25) can be rewritten by

$$\|\mathbf{z}-\mathbf{W}^{-1/2}\boldsymbol{\Phi}(\boldsymbol{\beta})\boldsymbol{\mu}\|_2^2$$

$$= \left\|\mathbf{z}-\left(\underbrace{\mathbf{W}^{-1/2}\bar{\mathbf{A}}}_{\triangleq\mathbf{A}_W}+\underbrace{\mathbf{W}^{-1/2}\mathbf{B}}_{\triangleq\mathbf{B}_W}\mathrm{diag}\{\boldsymbol{\beta}\}\right)\boldsymbol{\mu}_- - \mu_0\underbrace{\mathbf{W}^{-1/2}\mathbf{1}_n}_{\triangleq\mathbf{1}_W}\right\|_2^2 \tag{26}$$

$$= \left\|(\mathbf{z}-\mathbf{A}_W\boldsymbol{\mu}_- - \mu_0\mathbf{1}_W) - \mathbf{B}_W\mathrm{diag}\{\boldsymbol{\beta}\}\boldsymbol{\mu}_-\right\|_2^2 \tag{27}$$

$$= \boldsymbol{\beta}^T\left(\mathbf{B}_W^T\mathbf{B}_W^*\odot\boldsymbol{\mu}_-\boldsymbol{\mu}_-^H\right)\boldsymbol{\beta}$$

$$-2\mathrm{Re}\left\{\mathrm{diag}\{\boldsymbol{\mu}_-^*\}\mathbf{B}_W^H(\mathbf{z}-\mathbf{A}_W\boldsymbol{\mu}_- - \mu_0\mathbf{1}_W)\right\}^T\boldsymbol{\beta} + \mathrm{const} \tag{28}$$

where $\boldsymbol{\mu}_-$ denotes a sub-vector of $\boldsymbol{\mu}$ whose elements are indexed from 1 to $\hat{K}$, μ_0 denotes the end element of $\boldsymbol{\mu}$, $\odot$ denotes Hadamard product, and $\mathrm{Re}(\cdot)$ represent the real part of a complex variable. If we partition the matrix $\boldsymbol{\Sigma}$ into $\begin{bmatrix}\boldsymbol{\Upsilon} & \boldsymbol{\gamma}\\ \boldsymbol{\gamma}^H & \gamma_0\end{bmatrix}$, the second term in the right side of (25) can be rewritten as

$$\mathrm{tr}\left(\mathbf{W}^{-1}\boldsymbol{\Phi}(\boldsymbol{\beta})\boldsymbol{\Sigma}\boldsymbol{\Phi}^H(\boldsymbol{\beta})\right)$$

$$= \mathrm{tr}\left(\boldsymbol{\Phi}^H(\boldsymbol{\beta})(\mathbf{W}^{-1/2})^H\mathbf{W}^{-1/2}\boldsymbol{\Phi}(\boldsymbol{\beta})\boldsymbol{\Sigma}\right)$$

$$= \mathrm{tr}\left((\mathbf{A}_W+\mathbf{B}_W\mathrm{diag}\{\boldsymbol{\beta}\})^H(\mathbf{A}_W\boldsymbol{\Upsilon}+\mathbf{B}_W\mathrm{diag}\{\boldsymbol{\beta}\}\boldsymbol{\Upsilon}+\mathbf{1}_W\boldsymbol{\gamma}^H)\right)$$

$$+\mathbf{1}_W^H\mathbf{B}_W\mathrm{diag}\{\boldsymbol{\beta}\}\boldsymbol{\gamma} + \mathrm{const} \tag{29}$$

$$= 2\mathrm{Re}\left\{\mathrm{tr}\left(\mathbf{B}_W^H\mathbf{A}_W\boldsymbol{\Upsilon}\mathrm{diag}\{\boldsymbol{\beta}\}\right)\right\} + \mathrm{tr}\left(\mathrm{diag}\{\boldsymbol{\beta}\}\boldsymbol{\Upsilon}\mathrm{diag}\{\boldsymbol{\beta}\}\mathbf{B}_W^H\mathbf{B}_W\right)$$

$$+\mathrm{tr}\left(\mathrm{diag}\{\boldsymbol{\beta}\}\mathbf{B}_W^H\mathbf{1}_W\boldsymbol{\gamma}^H\right) + \mathbf{1}_W^H\mathbf{B}_W\mathrm{diag}\{\boldsymbol{\beta}\}\boldsymbol{\gamma} + \mathrm{const} \tag{30}$$

$$= 2\mathrm{Re}\{(\mathrm{diag}\{\mathbf{B}_W^H\mathbf{A}_W\boldsymbol{\Upsilon}\})^T + \mathbf{1}_W^H\mathbf{B}_W\mathrm{diag}\{\boldsymbol{\gamma}\}\}\boldsymbol{\beta}$$

$$+\boldsymbol{\beta}^T(\boldsymbol{\Upsilon}\odot(\mathbf{B}_W^T\mathbf{B}_W^*))\boldsymbol{\beta} + \mathrm{const}. \tag{31}$$

[1] We only consider a first-order Taylor expansion in the proposed method. The DOA estimation performance can be further improved, if a second-order expansion is adopted instead. Since the extension to the second-order expansion is straightforward, it is omitted for brevity. Moreover, the extension to higher order statistics with the technique proposed in [8] is also straightforward.

Substituting (28) and (31) into (25), we have

$$E\left\{\ln p(\hat{\mathbf{y}}|\mathbf{d};\boldsymbol{\beta})\right\}_{p(\mathbf{d}|\hat{\mathbf{y}},\boldsymbol{\delta};\boldsymbol{\beta})} = -\boldsymbol{\beta}^T\mathbf{P}\boldsymbol{\beta} + 2\mathbf{v}^T\boldsymbol{\beta} + \mathrm{const} \tag{32}$$

where $\mathbf{v}=\mathrm{Re}\{\mathrm{diag}\{\boldsymbol{\mu}_-^*\}\mathbf{B}_W^H(\mathbf{z}-\mathbf{A}_W\boldsymbol{\mu}_- - \mu_0\mathbf{1}_W)\} - \mathrm{Re}\{\mathrm{diag}\{\mathbf{B}_W^H\mathbf{A}_W\boldsymbol{\Upsilon}\} + \mathrm{diag}\{\boldsymbol{\gamma}\}\mathbf{B}_W^T\mathbf{1}_W^*\}$ and $\mathbf{P}=\mathrm{Re}\{\mathbf{B}_W^T\mathbf{B}_W^*\odot(\boldsymbol{\mu}_-\boldsymbol{\mu}_-^H+\boldsymbol{\Upsilon})\}$. Finally, setting the derivative of (32), with respect to $\boldsymbol{\beta}$, to zero and solving for $\boldsymbol{\beta}$ gives:

$$\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{v} \tag{33}$$

if $\mathbf{P}$ is invertible; otherwise,

$$\beta_i = \frac{[\mathbf{v}]_i}{[\mathbf{P}]_{i,i}}, \quad \forall i. \tag{34}$$

Note that the original off-grid method [23] uses the obtained value of $\boldsymbol{\beta}$ to update the measurement matrix $\boldsymbol{\Phi}(\boldsymbol{\beta})$ and then goes back to the next iteration. Clearly, the updated $\boldsymbol{\Phi}(\boldsymbol{\beta})$ can provide a better approximation for the true steering matrix. However, the distance between the true DOA θ_k and the nearest grid point to θ_k (denoted by $\hat{\theta}_{n_k}$) remains unchanged. In other words, the modeling error in (11) caused by off-grid gap always exists [16,30]. To fix this problem, we alternatively consider the locations of grid points as adjustable parameters, and proceed to update the grid point directly, i.e.,

$$\hat{\theta}_i^{new} = \hat{\theta}_i + \beta_i, \quad \forall i, \tag{35}$$

rather than updating the $\boldsymbol{\Phi}(\boldsymbol{\beta})$. The motivation behind (35) comes from the fact that $\hat{\theta}_{n_k}^{new}$ is much closer to θ_k than $\hat{\theta}_{n_k}$. After several iterations, the updated grid points will tend to approach the true DOAs. In the practical implementation, we may further screen the candidate from where the refined grid point falls into. The refined grid point is finally accepted if $\hat{\theta}_{n_k}+\beta_{n_k}$ falls into the set of $[\hat{\theta}_{n_k-1},\hat{\theta}_{n_k+1}]$; otherwise, it is rejected and the corresponding grid point retains unchanged.

4. Numerical simulations

In order to evaluate the performance of our proposed method, several simulations will be conducted in this section. We will compare the proposed method with the SS-MUSIC method [5], the second-order SBL method [28], the CMSR method [17], the gridless method ANM [18] and the ICSPICE [19].

We consider the inner ULA of the nested arrays with intersensor spacing of $d=\lambda/2$, and all sources have identical powers. We measure the effectiveness of these algorithms with root mean square error (RMSE), which is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{LK}\sum_{t=1}^{L}\sum_{k=1}^{K}(\hat{\theta}_{k,l}-\theta_k)^2} \tag{36}$$

where $L=200$ is the number of Monte-Carlo runs, $\hat{\theta}_{k,l}$ is the estimate of the kth signal in the lth Mont-Carlo trial.

Simulation 1 verifies the performance improvement of the proposed method in terms of the root mean square error (RMSE) of DOA estimation, where two uncorrelated sources uniformly come from intervals $[-30°,-20°]$ and $[0°,10°]$, respectively. That is to say the DOAs of the two sources are random values coming from these two intervals in each Monte-Carlo run. Fig. 1 shows the RMSE of DOA estimates versus SNR based on 200 Monte Carlo runs. A nested array composed of $M_1=M_2=2$ sensors is adopted in Figs. 1-a and 1-b, and the number of snapshots is $T=200$; while a nested array composed of $M_1=M_2=3$ sensors is adopted in Figs. 1-c and 1-d, and the number of snapshots is $T=100$. It is

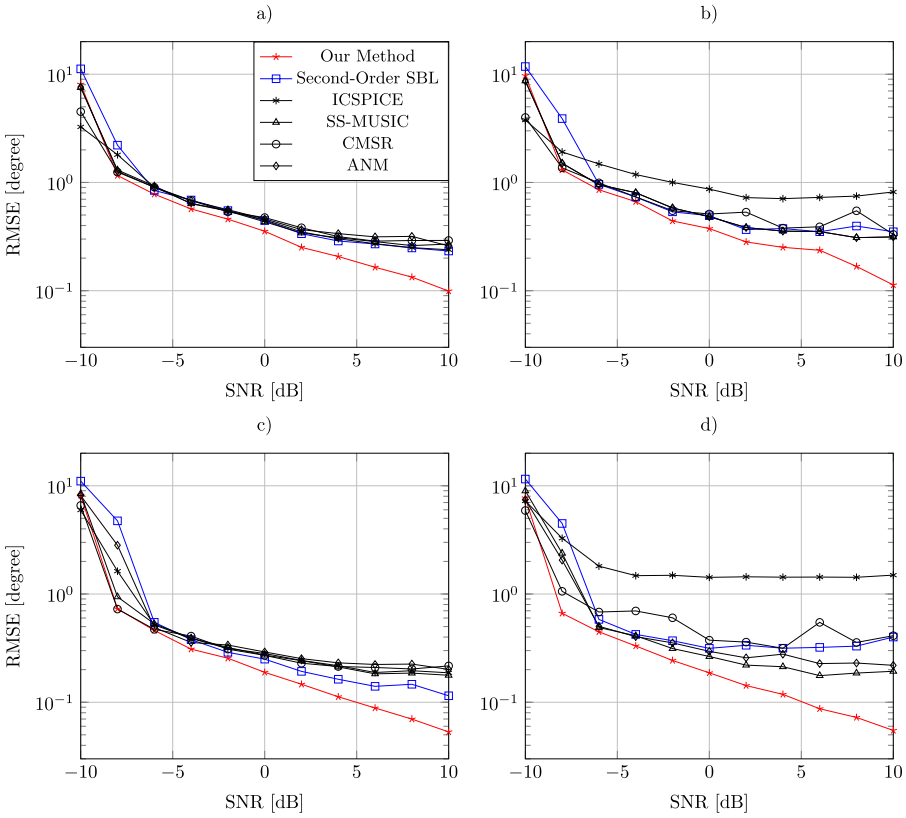

Fig. 1. RMSE of DOA estimates versus SNR. a) $M_1 = M_2 = 2$, $T = 200$ and $r = 1°$; b) $M_1 = M_2 = 2$, $T = 200$ and $r = 5°$; c) $M_1 = M_2 = 3$, $T = 100$ and $r = 1°$; d) $M_1 = M_2 = 3$, $T = 100$ and $r = 5°$.

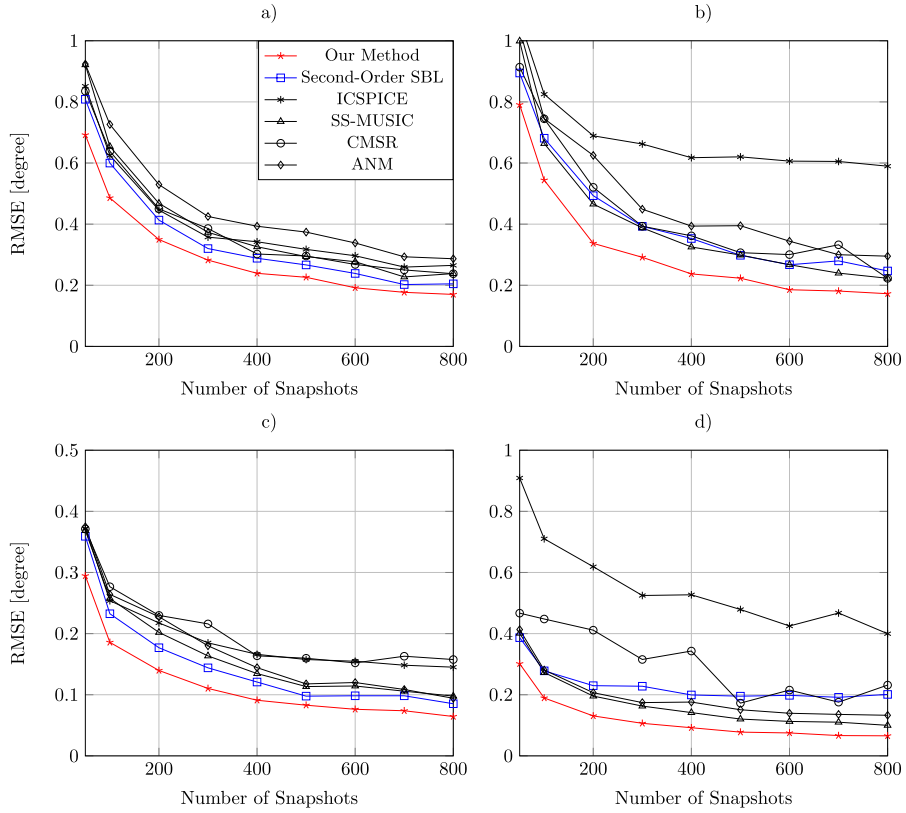

Fig. 2. RMSE of DOA estimates versus number of snapshots with SNR = 0 dB. a) $M_1 = M_2 = 2$ and $r = 1°$; b) $M_1 = M_2 = 2$ and $r = 4°$; c) $M_1 = M_2 = 3$ and $r = 1°$; d) $M_1 = M_2 = 3$ and $r = 4°$.

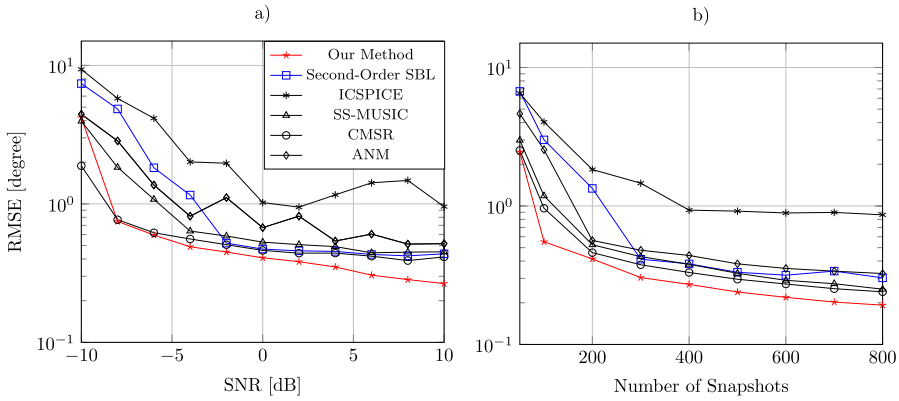

Fig. 3. RMSE of DOA estimates versus SNR and number of snapshots, respectively, where $M_1 = M_2 = 3$ and $r = 4°$. a) $T = 200$; b) SNR $= 0$ dB.

shown that RMSEs of all the methods decrease as SNR gets higher. The second-order SBL method outperforms SS-MUSIC, CMSR, ICSPICE and ANM at the grid interval of 1° (see Figs. 1-a and 1-c); however, when the grid interval is broadened from 1° to 5°, it brings a serious performance loss (see Figs. 1-b and 1-d). The performance loss of the second-order SBL method is due to the fact that the adopted Taylor expansion will cause a high modeling error if a coarse grid is used. The ICSPICE is the most sensitive to grid interval, because there are many simplified expressions in calculating the off-grid gaps, which may bring about large estimation errors when using a coarse grid. Our method always achieves the best performance regardless of what grid interval is used, because it can avoid reducing the working array aperture and properly handle the modeling error through viewing the sampled locations as the adjustable parameters.

Fig. 2 shows the RMSE of DOA estimates versus the number of snapshots with the SNR fixed to 0 dB. A nested array composed of $M_1 = M_2 = 2$ sensors is adopted in Figs. 2-a and 2-b; while a nested array composed of $M_1 = M_2 = 3$ sensors is adopted in Figs. 2-c and 2-d. From Fig. 2, it confirms that our method always outperforms the state-of-the-art methods, and the choice of the grid interval does not affect the performance of our method much. On the contrary, the performances of ICSPICE, the second-order SBL method and CMSR degrade significantly when the grid interval is large (see Figs. 2-b and 2-d).

The last simulation studies the performance of the proposed method in estimating more sources than the number of physical sensors. Assume the nested array composed of $M_1 = M_2 = 3$ sensors is used to receive $K = 7$ uncorrelated sources coming from $[-55.5°, -37.2°, -22.6°, 3.5°, 20.3°, 38.9°, 52.6°]$. The grid interval is assigned as $r = 4°$. Fig. 3 shows the RMSE of DOA estimates versus SNR and number of snapshots, respectively, where the number of snapshots is fixed to $T = 200$ in Fig. 3-a and the SNR is fixed to 0 dB in Fig. 3-b. It can be seen that all the displayed methods are capable of resolving more sources than the number of sensor elements but the RMSEs of all the methods increase compared to simulation 1. The ICSPICE performs the worst all the time. The second-order SBL method works worse than others except for ICSPICE when the SNR is low, but it gets better than CMSR, with the increasing of SNR (≥ -2 dB). Our method has a good performance in under-determined DOA estimation problem and outperforms the others significantly, especially at a small number of snapshots.

5. Conclusion

We have proposed an efficient SBL method that approaches the problem of off-grid DOA estimation with nested arrays. To avoid reducing the working array aperture and also simplify the derivation of the perturbation variance, we adopt the whole virtual signal vector to estimate DOAs, rather than a triming one used in the second-order SBL method. The main contribution of this paper is to give a new data model formulation and a novel grid refining procedure. In fact, since the proposed method has no specific requirements for the array structure, it can be applied to more higher level nested arrays. Simulation results demonstrate that our method can improve the DOA estimation performance significantly especially at a coarse grid.

Acknowledgments

Fangfang Chen, Jisheng Dai and Zhongfu Ye provided the idea of this work. Nan Hu conceived of and designed the experiments. Nan Hu and Fangfang Chen performed the experiments and provided all of the figures and data for the paper. Zhongfu Ye and Jisheng Dai prepared the literature and analyzed the data. Fangfang Chen and Jisheng Dai wrote the paper. Correspondence and requests for the paper should be addressed to Jisheng Dai. This work is supported by the National Natural Science Foundation of China (NSFC) under Project 61571211.

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.
[2] A. Moffet, Minimum-redundancy linear arrays, IEEE Trans. Antennas Propag. 16 (2) (1968) 172–175.
[3] M. Rubsamen, A.B. Gershman, Direction-of-arrival estimation for nonuniform sensor arrays: from manifold separation to Fourier domain music methods, IEEE Trans. Signal Process. 57 (2) (2009) 588–599.
[4] R.T. Hoctor, S.A. Kassam, The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging, Proc. IEEE 78 (4) (1990) 735–752.
[5] P. Pal, P. Vaidyanathan, Nested arrays: a novel approach to array processing with enhanced degrees of freedom, IEEE Trans. Signal Process. 58 (8) (2010) 4167–4181.
[6] K. Han, A. Nehorai, Improved source number detection and direction estimation with nested arrays and ULAs using jackknifing, IEEE Trans. Signal Process. 61 (23) (2013) 6118–6128.
[7] P. Pal, P. Vaidyanathan, Multiple level nested array: an efficient geometry for $2q$ th order cumulant based array processing, IEEE Trans. Signal Process. 60 (3) (2012) 1253–1269.
[8] Q. Shen, W. Liu, W. Cui, S. Wu, Extension of nested arrays with the fourth-order difference co-array enhancement, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2016, pp. 2991–2995.
[9] A. Ahmed, Y.D. Zhang, B. Himed, Effective nested array design for fourth-order cumulant-based DOA estimation, in: Radar Conference, RadarConf, IEEE, 2017, pp. 0998–1002.
[10] R. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (3) (1986) 276–280.
[11] R. Roy, T. Kailath, Esprit-estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust. Speech Signal Process. 37 (7) (1989) 984–995.
[12] 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.

[13] J. Dai, X. Xu, D. Zhao, Direction-of-arrival estimation via real-valued sparse representation, IEEE Antennas Wirel. Propag. Lett. 12 (2013) 376–379.
[14] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (Jun) (2001) 211–244.
[15] S.D. Babacan, R. Molina, A.K. Katsaggelos, Bayesian compressive sensing using Laplace priors, IEEE Trans. Image Process. 19 (1) (2010) 53–63.
[16] J. Dai, X. Bao, W. Xu, C. Chang, Root sparse Bayesian learning for off-grid DOA estimation, IEEE Signal Process. Lett. 24 (1) (2017) 46–50.
[17] Z.-M. Liu, Z.-T. Huang, Y.-Y. Zhou, Array signal processing via sparsity-inducing representation of the array covariance matrix, IEEE Trans. Aerosp. Electron. Syst. 49 (3) (2013) 1710–1724.
[18] Z. Yang, L. Xie, Exact joint sparse frequency recovery via optimization methods, IEEE Trans. Signal Process. 64 (19) (2014) 5145–5157.
[19] S. Cai, G. Wang, J. Zhang, K.-K. Wong, H. Zhu, Efficient direction of arrival estimation based on sparse covariance fitting criterion with modeling mismatch, Signal Process. 137 (2017) 264–273.
[20] J. Capon, High-resolution frequency-wavenumber spectrum analysis, Proc. IEEE 57 (8) (1969) 1408–1418.
[21] P. Stoica, P. Babu, J. Li, SPICE: a sparse covariance-based estimation method for array processing, IEEE Trans. Signal Process. 59 (2) (2011) 629–638.
[22] Z.-M. Liu, Y.-Y. Zhou, A unified framework and sparse Bayesian perspective for direction-of-arrival estimation in the presence of array imperfections, IEEE Trans. Signal Process. 61 (15) (2013) 3786–3798.
[23] 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.
[24] N. Hu, B. Sun, J. Wang, J. Dai, C. Chang, Source localization for sparse array using nonnegative sparse Bayesian learning, Signal Process. 127 (2016) 37–43.
[25] Q. Huang, G. Zhang, Y. Fang, Real-valued DOA estimation for spherical arrays using sparse Bayesian learning, Signal Process. 125 (2016) 79–86.
[26] J. Dai, H.C. So, Sparse Bayesian learning approach for outlier-resistant direction-of-arrival estimation, IEEE Trans. Signal Process. 66 (3) (2018) 744–756.
[27] D.P. Wipf, B.D. Rao, Sparse Bayesian learning for basis selection, IEEE Trans. Signal Process. 52 (8) (2004) 2153–2164.
[28] J. Yang, G. Liao, J. Li, An efficient off-grid DOA estimation approach for nested array signal processing by using sparse Bayesian learning strategies, Signal Process. 128 (2016) 110–122.
[29] B. Ottersten, P. Stoica, R. Roy, Covariance matching estimation techniques for array signal processing applications, Digit. Signal Process. 8 (3) (1998) 185–210.
[30] J. Dai, A. Liu, V.K. Lau, FDD massive MIMO channel estimation with arbitrary 2D-array geometry, IEEE Trans. Signal Process. 66 (10) (2018) 2584–2599.

Fangfang Chen received B.Eng. degree in electronic information engineering from Jiangsu University, Zhenjiang, China, in 2015 and now she is a Master student at the School of Electrical and Information Engineering, Jiangsu University. Her main research interest is target angle detection in MIMO radar.

Jisheng Dai received the B.Eng. degree in electronic engineering from Nanjing University of Technology, Nanjing, China, in 2005 and the Ph.D. degree in information and communication engineering from the University of Science and Technology of China, Hefei, China, in 2010. He is currently a Research Associate with the School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, China.

Nan Hu received the B.S. degree in electronic and information engineering and the Ph.D. degree in signal and information processing from the University of Science and Technology of China, Hefei, China, in 2008 and 2013, respectively. He is currently a Research Associate with the School of Electronic and Information Engineering, Soochow University, Suzhou, China. His research interests include sensor array signal processing and electroencephalography signal processing.

Zhongfu Ye received his B.E. and M.S. degrees in electronics and information engineering from the Hefei University of Technology, Hefei, China in 1982 and 1986, respectively and his Ph.D. degree from the University of Science and Technology of China, Hefei, China in 1995. He is currently a professor at the University of Science and Technology of China. His research interests are in statistical and array signal processing and image processing.