Signal Processing 128 (2016) 110–122

Contents lists available at ScienceDirect

Signal Processing

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

An efficient off-grid DOA estimation approach for nested array signal processing by using sparse Bayesian learning strategies

Jie Yang*, Guisheng Liao, Jun Li

National Laboratory of Radar Signal Processing, Xidian University, Xi'an, Shaanxi 710071, China

ARTICLE INFO

Article history:
Received 25 November 2015
Received in revised form
27 March 2016
Accepted 29 March 2016
Available online 31 March 2016

Keywords:
Direction-of-arrival (DOA) estimation
Nested array
Sparse Bayesian learning (SBL)
Off-grid effects
Polynomial rooting

ABSTRACT

In this paper, we address the problem of sparsity-inducing direction-of-arrival (DOA) estimation in the context of array covariance vectors when a recently proposed nested array structure is utilized. The proposed approach is based on a Sparse Bayesian learning (SBL) perspective while the off-grid effects inherent in traditional sparse recovery algorithms are taken into account simultaneously. Specifically, the notable advantages of the developed approach can be attributed to two aspects. First, through a linear transformation, the unknown noise variance can be eliminated effectively and the estimation error between the actual and sample covariance matrix can be normalized to an identity matrix, thereby facilitating a noise-free sparse representation. Second, a computationally tractable polynomial rooting procedure is introduced to calculate the discretized sampling grid error one by one based on the reconstructed spatial power spectrum, thus this novel algorithm can maintain high estimation accuracy even under a coarse sampling grid. Extensive simulation results finally demonstrate the superiority of the proposed approach in terms of DOA estimation precision and computational efficiency over off-the-shelves techniques.

1. Introduction

Direction-of-arrival (DOA) estimation, which determines the spatial spectra of the impinging electromagnetic/acoustic waves by using an array of sensors, has been an area of continued research interest due to its wide range of applications in source localization, radar imaging, mobile communication, etc. [1–6]. A considerable amount of attention in DOA estimation is the maximum number of sources that can be resolved, and the problem of detecting more sources than the number of sensors has received tremendous consideration in various research fields [7]. More specifically, the DOA estimation of speech, for instance, has many applications like teleconference systems and automatic conference minutes generators. Such applications usually face situations where more sources than the number of sensors are active simultaneously. However, most existing strategies for the DOA estimation are confined to the case of the uniform linear array (ULA) [8]. It is well known that conventional super-resolution methods, for example, the subspace-based ones [1], can resolve up to $M-1$ sources with an M-element ULA. To increase the number of resolvable sources beyond that offered by a ULA for a given number of physical sensors, numerous nonuniform array geometries, such as minimum redundancy arrays (MRAs) [9] and the recently proposed nested array [10], have been introduced for achieving a higher number of degrees-of-freedom (DOF). Toward this purpose, the difference co-array constructed by the original linear array is utilized to mimic a virtual array with wider aperture and more sensors. Unlike MRAs, which have no closed form expressions for the array configurations as well as the achievable DOF for an arbitrary number of sensor elements, the nested array is inherently capable of identifying $O(M^2)$ sources with only $O(M)$ physical sensors, and is obtained by systematically nesting two uniform linear subarrays with increasing spacing, thereby resulting an easy construction for the sensor locations and an exact expression for the available DOF corresponding to a given number of sensors.

In general, the idea behind nested array is to extend the concept of MRAs and span large aperture by using far fewer elements than prescribed by textbook antenna theory. This entails a strong motivation for studying DOA estimation methods that can handle more signals than the number of physical sensors, known also as underdetermined DOA estimation methods. The existing high-resolution DOA estimation approach for nested array can be accomplished based on the one proposed in [10], i.e., the so-called spatial smoothing-based MUSIC (SS-MUSIC) method. As its name suggests, this algorithm is a suitable extension of the traditional MUSIC technique [11], in which a spatial smoothing-style preprocessing procedure is incorporated to form a positive semi-

* Corresponding author.
E-mail address: yangjie_xidian@126.com (J. Yang).

http://dx.doi.org/10.1016/j.sigpro.2016.03.024
0165-1684/

definite augmented covariance matrix with enhanced DOF. Although this covariance matrix augmentation operation is specifically for MUSIC, the general ideas related to its development may be applicable mutatis mutandis to other classical subspace-based method such as ESPRIT [12].

Despite some positive and attractive features of these aforementioned approaches in super-resolution and DOA estimation precision under moderate settings, all of them share the same bottleneck. Indeed, they require adequate snapshots, moderately high signal-to-noise ratio (SNR) and the prior information of the incident signal number to obtain satisfying performance. Such constraint of these methods has blocked their applicability in various practical circumstances. Recently, the emerging technique of sparse reconstruction has attracted much attention from numerous areas, and it also provides a new perspective for DOA estimation. Much literatures [13–16] manifest that, the exploitation of the spatial sparsity of the incident signals helps to significantly enhance the adaptation of the DOA estimators to the above mentioned demanding scenarios, and they are also much immune to the lack of prior source number information. The existing sparsity-inducing DOA estimation methods mainly divide into two categories, one is the ℓ_p-norm ($0 \le p \le 1$)-based algorithms like joint ℓ_1-norm approximation and singular value decomposition (L1-SVD) [15], ℓ_1-norm-based sparse representation of array covariance vectors (L1-SRACV) [17], semi-parametric iterative covariance-based estimation (SPICE) [18,19], covariance matrix sparse representation (CMSR) [20], and so forth; The other is the Sparse Bayesian learning (SBL)-based algorithms [21–26]. Several preliminary literatures [27,28] have demonstrated both theoretically and empirically that, the SBL-based techniques induce less structural error (biased global minimum) and convergence error (failure in achieving the global minimum) than the ℓ_p-norm ($0 \le p \le 1$)-based ones, and they also perform well in capturing local signal properties to facilitate the refined DOA estimation process. Other merits of the SBL-based methods include improved resolution, flexibility in exploiting the possible structure of the signal to be recovered [29] and no requirement of practically unknown user-parameters, e.g., the noise statistics or regularization parameters, etc. Nonetheless, the main issue to cope with when applying the aforementioned sparsity-driven algorithms is the fact that the locations of all true targets must fall on the predefined discretization grid for guaranteeing reliable estimations [30]. Unfortunately, such a condition cannot easily verified since (1) a coarse grid leads to a high modeling error and (2) too dense a grid is computationally prohibitive and might result in computational instability (see, e.g., [31,32]). In addition, as has been mentioned the vectorized covariance matrix model can efficiently illustrate the array aperture extension, however, directly applying most sparsity-inducing algorithms (see, e.g., [15,17]) in this model is not rational as these methods are designed to work in the raw data domain. Consequently, by using nested array structure proposed in [10] in conjunction with the sparsity-based framework, as the array augmentation property is essential in the underdetermined DOA estimation scenario, new strategies for solving the correlation-aware support recovery problem is desired.

As it becomes clear from the above discussion, the successes of SBL-based methods in past contexts motivate us to consider their extension to the problem of target localization for nested array. More specifically, through vectorizing the estimated covariance matrix and reducing the redundancy elements, we can circumvent the procedure of evaluating the noise variance by resorting to a linear transformation. This is due to the fact that the Bayesian inference method suffered from the underestimation of the noise variance, which is even not attainable under the underdetermined scenarios. Additionally, the estimation error of the covariance matrix is also normalized to an identity matrix as a by-product of the denoising scheme. Then a novel but noise-free sparse representation, with measurement matrix from extended steering vectors which provide an increased DOF of original array, is obtained, and the DOA estimation can be cast as a problem of recovering a nonnegative sparse vector. To overcome grid mismatch of traditional sparsity-based methods, the true dictionary is approximated as a summation of a presumed dictionary and a structured parameterized matrix via the second-order Taylor expansion. As a result, the off-grid model has a much smaller modeling error than the on-grid one. Besides these efforts, by exploiting the underlying joint sparsity structure between the original signal and the grid mismatch, a SBL-based algorithm is proposed to iteratively refine the DOA estimates, in which the discretized sampling grid error is calculated one by one via a computationally cheap polynomial rooting procedure. As we shall show in the subsequent Sections, the proposed method has a smaller root-mean-square-error (RMSE) and a superior detection probability in comparison with the existing techniques. Moreover, the proposed method can work without the knowledge of the number of signals, and is capable of handling more sources than the number of actual physical sensors. Analytical and numerical results will be both given to support each other in the ensuing development.

The rest of the paper is organized as follows. The model of DOA estimation in nested array and its relation with sparse recovery are presented in Section 2. The proposed SBL-based algorithm is discussed in detail in Section 3, where some mathematically analyses with the maximal source number that it is able to separate are also performed. Section 4 provides extensive numerical simulations demonstrating the advantages of our method in terms of estimation accuracy, DOF, and detection probability. Such results reaffirm and demonstrate the usefulness of the results described in Section 3. Finally, some conclusions are drawn in Section 5.

Notations: We use lower-case (upper-case) bold characters to denote vectors (matrices). In particular, $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ imply conjugate, transpose, conjugate transpose and inverse, respectively. $\mathbf{z}_p$ represents the pth element of $\mathbf{z}$, $\mathbf{Z}_{p,:}$, $\mathbf{Z}_{:,p}$ and $\mathbf{Z}_{p_1,p_2}$ represent the pth row, pth column and (p_1, p_2)th element of $\mathbf{Z}$, respectively. $\mathrm{vec}(\cdot)$ denotes the vectorization operator that turns a matrix into a vector by stacking all columns on top of the another, and $\mathrm{diag}(\cdot)$ means forming a diagonal matrix by taking the given vector as the diagonal elements. The symbol $\mathbb{R}$ and $\mathbb{C}$ represent the sets of real numbers and complex numbers, respectively. $|\cdot|$ denotes the absolute value of a scalar or determinant of a matrix. $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ mean taking the real and imaginary parts of a complex value, respectively. $\|\cdot\|_0$ and $\|\cdot\|_2$ respectively denote the ℓ_0 and Euclidean (ℓ_2) norms. $\mathrm{Tr}\{\cdot\}$ is the trace operator, and $\mathrm{E}\{\cdot\}$ is the expectation operator of stochastic variables. $\circ$, $\oplus$ and $\otimes$ stand for the KR product, Hadamard product and Kronecker product, respectively. $CN(\mathbf{a}, \mathbf{B})$ denotes joint complex Gaussian distribution with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{B}$. j is reserved for the imaginary unit $\sqrt{-1}$. $\boldsymbol{\chi}'(\theta)$ and $\boldsymbol{\chi}''(\theta)$ are the first-order and second-order derivative of $\boldsymbol{\chi}(\theta)$ with respect to θ, respectively.

2. Signal model and preliminaries

A prototype nested array [10] with M scalar sensors, as described in the previous section, consists of two concatenated ULAs: the inner ULA has M_1 sensors with inter-sensor spacing d and the

outer ULA has M_2 sensors with inter-sensor spacing $(M_1+1)d$. The unit inter-element spacing, d, is usually set as half wavelength, or $\lambda/2$. More precisely the array sensors are positioned at $\mathbf{D}=\mathbf{r}d$, where $\mathbf{r}=\{r_i, i=1, 2, \ldots, M\}$

$=\{0, 1, \ldots, M_1-1, M_1, 2(M_1+1)-1, \ldots, M_2(M_1+1)-1\}$ is an integer vector denoting all the sensors' position information. Note that the first sensor is assumed as the reference, i.e., $r_1=0$.

Consider a scenario in which a number of K narrowband far-field sources impinging on the array from directions $\boldsymbol{\theta}=\{\theta_k, k=1, 2, \ldots, K\}$, and their discretized baseband waveforms are expressed as $s_k(t)$, $t=1, 2, \ldots, N$, for $k=1, 2, \ldots, K$. Then, the data vector received at the nested array is expressed as

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

where

$$\mathbf{a}(\theta_k)=\left[1, e^{j2\pi r_2 d\sin\theta_k/\lambda}, \ldots, e^{j2\pi r_M d\sin\theta_k/\lambda}\right]^T \tag{2}$$

is the steering vector of the array corresponding to θ_k, $\mathbf{A}=[\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_K)]$ is the array manifold matrix, and $\mathbf{s}(t)=[s_1(t), s_2(t), \ldots, s_K(t)]^T$. We suppose the source signals follow circular complex Gaussian distributions, i.e., $s_k(t)\sim CN(0, \sigma_k^2)$, and they are all independent of each other. The noise vector $\mathbf{n}(t)$ is assumed to be white Gaussian with power σ_n^2, and uncorrelated with the sources.

Based on our assumptions, the source autocorrelation matrix $\mathbf{R_S}$ is diagonal: $\mathbf{R_S}=\mathrm{E}[\mathbf{s}(t)\mathbf{s}^H(t)]=\mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2)$. Then the covariance matrix of the received signal $\mathbf{x}(t)$ for nested array is obtained as

$$\mathbf{R_x}=\mathrm{E}\left[\mathbf{x}(t)\mathbf{x}^H(t)\right]=\mathbf{A}\mathbf{R_S}\mathbf{A}^H+\sigma_n^2\mathbf{I}_M$$
$$=\sum_{k=1}^{K}\sigma_k^2\mathbf{a}(\theta_k)\mathbf{a}^H(\theta_k)+\sigma_n^2\mathbf{I}_M \tag{3}$$

In practice, we estimate $\mathbf{R_x}$ from a finite number (say, N) snapshots as

$$\hat{\mathbf{R}}_\mathbf{x}=\frac{1}{N}\sum_{t=1}^{N}\mathbf{x}(t)\mathbf{x}^H(t) \tag{4}$$

A central assumption in the existing DOA estimation methods which utilize the covariance matrix of the received signal is that the number of sources is less than the number of sensors, i.e., $K<M$. Under this condition the matrix product $\mathbf{A}\mathbf{R_S}\mathbf{A}^H\in\mathbb{C}^{M\times M}$ is low rank (rank K), and subspace-based methods estimate the DOAs by exploiting this low rank. When $K<M$, the matrix $\mathbf{A}$ is tall, and we call such a signal model as "overdetermined". In contrast, this paper considers the situation when $K>M$, possibly $K=O(M^2)$. In this case $\mathbf{A}$ is fat, and such a model is "underdetermined". In this regime, we consider utilizing the structure of the nested array to build a suitable bigger matrix, starting from $\mathbf{R_x}$, that has a low rank component even when $K>M$.

Please note that the entries of the covariance matrix $\mathbf{R_x}$ correspond to different lags. From antennas located at the mth and nth positions in $\mathbf{D}$, the correlation $E[\mathbf{x}_m(t)\mathbf{x}_n^*(t)]$ yields an entry in $\mathbf{R_x}$ with lag $(r_m-r_n)d$. As such, all the available values of m and n, where $1\le m\le M$ and $1\le n\le M$, yields the following set of all pairwise differences

$$\mathbf{C_D}=\left\{\mathbf{D}_i-\mathbf{D}_j, 1\le i, j\le M\right\} \tag{5}$$

This is known as the difference co-array and for the nested geometry, it can be shown that $\mathbf{C_D}$ contains $M^2/2+M-1$ distinct consecutive integers. The significance of the difference co-array is that the correlation of the received signal can be calculated at all differences in set $\mathbf{C_D}$. Any application which depends only on such correlation (e.g., DOA estimation) can exploit all the DOF offered by the resulting co-array structure. After vectorizing the correlation matrix $\mathbf{R_x}$, we have

$$\mathbf{y}=\mathrm{vec}(\mathbf{R_x})=\mathrm{vec}\left[\sum_{k=1}^{K}\sigma_k^2\mathbf{a}(\theta_k)\mathbf{a}^H(\theta_k)\right]+\mathrm{vec}\left(\sigma_n^2\mathbf{I}_M\right)$$
$$=(\mathbf{A}^*\circ\mathbf{A})\mathbf{p}+\sigma_n^2\mathbf{1}_n \tag{6}$$

where $(\mathbf{A}^*\circ\mathbf{A})=[\mathbf{a}^*(\theta_1)\otimes\mathbf{a}(\theta_1), \mathbf{a}^*(\theta_2)\otimes\mathbf{a}(\theta_2), \ldots, \mathbf{a}^*(\theta_K)\otimes\mathbf{a}(\theta_K)]$, $\mathbf{p}=[\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2]^T$, $\mathbf{1}_n=[\mathbf{e}_1^T, \mathbf{e}_2^T, \ldots, \mathbf{e}_M^T]^T$ with $\mathbf{e}_i$ being a vector of all zeros except a 1 at the ith position, and the last equation follows the property for KR product [33]. A straightforward assertion from (6) is that the aperture of the virtual array is determined by the number of distinct entries in $\mathbf{a}^*(\theta)\otimes\mathbf{a}(\theta)$. Considering $\mathbf{A}^*\circ\mathbf{A}$ in model (6), we construct a new matrix $\tilde{\mathbf{A}}$ of size $(M^2/2+M-1)\times K$ by removing the repeated rows from $\mathbf{A}^*\circ\mathbf{A}$ and also sorting them so that the ith row corresponds to the sensor location $(-M^2/4-M/2+i)d$ in the difference co-array of the nested array. This is equivalent to removing the corresponding rows from the observation vector $\mathbf{y}$ and sorting them to get a new vector

$$\mathbf{z}=\tilde{\mathbf{A}}\mathbf{p}+\sigma_n^2\bar{\mathbf{e}} \tag{7}$$

where $\bar{\mathbf{e}}\in\mathbb{R}^{(M^2/2+M-1)\times 1}$ is a vector of all zeros except for a 1 at the center position. It is interesting to note from (7) that $\mathbf{z}$ is reminiscent of an array signal model where $\tilde{\mathbf{A}}=[\tilde{\mathbf{a}}(\theta_1), \tilde{\mathbf{a}}(\theta_2), \ldots, \tilde{\mathbf{a}}(\theta_K)]\in\mathbb{C}^{(M^2/2+M-1)\times K}$ is virtually the steering matrix of a ULA with a consecutive range of virtual sensors in the set $[(-M^2/4-M/2+1)d, (M^2/4+M/2-1)d]$, and $\mathbf{p}$ is the source signal vector. Therefore, (7) can be regarded as a ULA detecting a coherent source $\mathbf{p}$ with deterministic noise term $\sigma_n^2\bar{\mathbf{e}}$. By applying MUSIC with spatial smoothing, Piya Pal et al. [10] showed that up to $O(M^2)$ sources can be detected, using this approach.

Alternatively, (7) can be solved using the sparsity-inducing approach. One of the main motivations behind using this technique is that the aforementioned SS-MUSIC method works worse when the estimated covariance matrix dismatches the true one severely (i.e., for small N or low SNR). Starting from the key observation that the signals impinging on the antenna array are intrinsically sparse in the spatial domain, we discretize the potential DOA range $[0, \pi]$ of the incident sources to yield a direction set $\Theta=[\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_I]$ (Fig. 1) and forms the corresponding manifold dictionary $\bar{\mathbf{A}}=[\bar{\mathbf{a}}(\theta_1), \bar{\mathbf{a}}(\theta_2), \ldots, \bar{\mathbf{a}}(\theta_I)]$, where I denotes the sampling grid number and satisfies $I>>(M^2/2+M-1)>K$. Thus the overcomplete basis representation for (7) can be formulated as

$$\mathbf{z}=\bar{\mathbf{A}}\bar{\mathbf{p}}+\sigma_n^2\bar{\mathbf{e}} \tag{8}$$

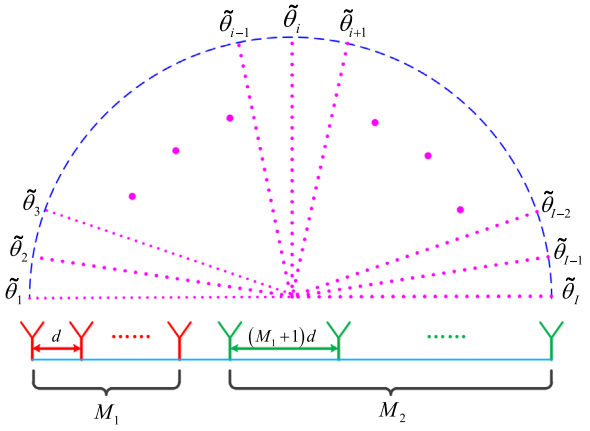

Fig. 1. Spatial scope discretization.

where $\bar{\mathbf{p}}$ is a zero-padded extension of $\mathbf{p}$ in which only the positions of entries corresponding to the actual DOAs are occupied by non-zero entries. That is, the signal directions can be estimated according to the locations of the nonzero magnitudes in $\bar{\mathbf{p}}$ if it can be estimated. Typically, the direction set Θ is often selected empirically by taking super-resolution into consideration, and an equal spatial sampling strategy is always used to simplify notation with a dense enough interval to enhance super-resolution and decrease quantization error.

3. Sparse recovery with off-grid targets

In this Section, we first review the overcomplete model for the sparsity-inducing DOA estimation methods with the off-grid effects in mind. Then in Section 3.2, we introduce the basic idea of the proposed method to be used to resolve the independent signal components. After that, we discuss in detail how the original on-grid DOA and the off-grid distance (the distance from the true DOA to the nearest grid point) are updated iteratively via a SBL inference. Finally, we outline the proposed algorithm by joining the main procedures together and complementing them with some reasonable remarks (see, for instance, the maximum separable signal number and the stochastic Cramer–Rao Lower Bound) in Section 3.3.

3.1. Modeling grid mismatch

Actually, the discrete set Θ is never dense enough to include all the possibilities of the true source directions, and the grid mismatch will deteriorate the validity of the overcomplete model (8) significantly. In order to describe the actual array output model more precisely, we take the dictionary mismatches into account in this section.

The sparsity-driven model developed so far are based on the exact knowledge of the covariance matrix of the observations. In practice, we can only estimate the covariance matrix from multiple measurements which share a common sparse support. The formulation presented in (8) might not always hold in a strict sense. We now derive a practical off-grid model to correlation-aware sparse support recovery by explicitly accounting for such deviations from the ideal model. Without loss of generality, we assume that the minimum distance between two sampling points of the grid set Θ is r. Suppose $\theta_k \notin \{\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_I\}$ for source $k \in \{1, 2, \ldots, K\}$ and that $\tilde{\theta}_{l_k}$, $l_k \in \{1, 2, \ldots I\}$, is the nearest grid point to θ_k. We approximate the steering vector $\tilde{\mathbf{a}}(\theta_k)$ using linearization: $\tilde{\mathbf{a}}(\theta_k) \approx \tilde{\mathbf{a}}(\tilde{\theta}_{l_k}) + \mathbf{b}(\tilde{\theta}_{l_k})(\theta_k - \tilde{\theta}_{l_k}) + \mathbf{c}(\tilde{\theta}_{l_k})(\theta_k - \tilde{\theta}_{l_k})^2$ with $\mathbf{b}(\tilde{\theta}_{l_k}) = \tilde{\mathbf{a}}'(\tilde{\theta}_{l_k})$ and $\mathbf{c}(\tilde{\theta}_{l_k}) = \tilde{\mathbf{a}}''(\tilde{\theta}_{l_k})/2$. Denote $\mathbf{B} = [\mathbf{b}(\tilde{\theta}_1), \mathbf{b}(\tilde{\theta}_2), \ldots, \mathbf{b}(\tilde{\theta}_I)]$, $\mathbf{C} = [\mathbf{c}(\tilde{\theta}_1), \mathbf{c}(\tilde{\theta}_2), \ldots, \mathbf{c}(\tilde{\theta}_I)]$, $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_I]^T \in [-r/2, r/2]^I$ and $\boldsymbol{\Phi}(\boldsymbol{\beta}) = \bar{\mathbf{A}} + \mathbf{B}\mathrm{diag}(\boldsymbol{\beta}) + \mathbf{C}\mathrm{diag}(\boldsymbol{\beta}^2)$, where we define (using the MATLAB notations) $\boldsymbol{\beta}^2 = \boldsymbol{\beta} \oplus \boldsymbol{\beta}$, and for $i = 1, 2, \ldots, I$

$$\begin{cases} \beta_i = \theta_k - \tilde{\theta}_{l_k}, \ \ \bar{\mathbf{p}}_i = \mathbf{p}_{l_k}, \\ \quad \text{if} \ \ i = I_k \ \text{ for any } \ k \in \{1, 2, \ldots, K\} \\ \beta_i = 0, \ \ \bar{\mathbf{p}}_i = 0, \qquad\qquad \text{otherwise} \end{cases} \tag{9}$$

In addition, as just described previously, an immediate consequence of substituting $\hat{\mathbf{R}}_{\mathbf{x}}$ for $\mathbf{R}_{\mathbf{x}}$ is that the equality constraint in (8) collapses. To address the challenge of recovering support under such non-ideal conditions, the unknown $\mathbf{y}$ in (6) is estimated from N snapshots, viz., $\hat{\mathbf{y}} = \mathrm{vec}(\hat{\mathbf{R}}_{\mathbf{x}}) = \mathbf{y} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon} = \mathrm{vec}(\hat{\mathbf{R}}_{\mathbf{x}} - \mathbf{R}_{\mathbf{x}})$ is the vectorized estimation error of the covariance matrix. Similar to the operation in (8), extracting all the distinct lag samples of $\hat{\mathbf{y}}$ and approximating the ideal measurement matrix by using the second-order Taylor expansion around the predefined grid Θ lead to

$$\hat{\mathbf{z}} = \boldsymbol{\Phi}(\boldsymbol{\beta})\bar{\mathbf{p}} + \sigma_n^2\bar{\mathbf{e}} + \bar{\boldsymbol{\varepsilon}} \tag{10}$$

It should be noted that (10) is the off-grid model to be used in the remainder of this paper, and in this case, the formulation in (10) captures the effect of a finite number of multiple measurement vectors. Another noteworthy point is that this model is closely related to the on-grid one that can be obtained by setting $\boldsymbol{\beta} = \mathbf{0}$ in (10) ($\boldsymbol{\Phi}(\mathbf{0}) = \bar{\mathbf{A}}$).

3.2. Sparse Bayesian learning for DOA estimation

In this Subsection, we realize DOA estimation based on the "off-grid and imperfect correlation-awareness" framework in (10) by exploiting the SBL inference. The proposed algorithm is developed via an iterative procedure to jointly recover the sparse support and estimate the unknown grid mismatches. As we pointed out earlier, the SBL strategies often lead to extremely inaccurate estimation of noise variance. But unfortunately, the performance of these SBL-based techniques is apt to deteriorate when somewhat significant estimation biases within the noise variance exist, therefore we turn to a linear transformation operation to remove the unknown noise component from $\hat{\mathbf{z}}$ to promote the estimation accuracy of the SBL algorithm. Furthermore, according to the statistical property of the covariance matrix estimation error (viz., $\bar{\boldsymbol{\varepsilon}}$), an additional normalization operation is also incorporated in the above mentioned linear transformation procedure to simplify the hyperparameter learning rule of our method. Followed by this preprocessing management, a gradual and interweaved process is implemented to refine the sparse support and the unknown sampling grid error.

In Ref. [34], it was revealed that the perturbation of the vectorized covariance matrix satisfies an asymptotic Gaussian distribution

$$\boldsymbol{\varepsilon} = \mathrm{vec}\left(\hat{\mathbf{R}}_{\mathbf{x}} - \mathbf{R}_{\mathbf{x}}\right) \sim CN\left(\mathbf{0}, \mathbf{R}_{\mathbf{x}}^T \otimes \mathbf{R}_{\mathbf{x}}/N\right) \tag{11}$$

Nonetheless, by observing the structures of $\boldsymbol{\varepsilon}$ and $\bar{\boldsymbol{\varepsilon}}$, we can verify that the perturbation variance given in (10) is different from that in $\hat{\mathbf{y}}$, it depends not only on the entries of $\mathbf{R}_{\mathbf{x}}$, but also on the transformation involves removing the duplicated information in the vectorized covariance matrix, which makes it difficult to compute. Therefore, by recalling the construction of $\hat{\mathbf{z}}$, which extracts the entries of $\hat{\mathbf{y}}$ corresponding to the distinct components in $\mathbf{C}_{\mathbf{D}}$, we establish the relationship between $\boldsymbol{\varepsilon}$ and $\bar{\boldsymbol{\varepsilon}}$ as follows.

In essence, a nested array can be considered as a ULA with "missing" sensors, i.e., a nested array takes only a subset, say $\mathbf{r}$ as defined in Section 2, of the sensors of a ULA. Thus a nested array can be represented by its sensor index set $\mathbf{r} \subset \{0, 1, \ldots, M_2(M_1 + 1) - 1\}$. Without loss of generality, we assume that $\mathbf{r}$ is sorted ascendingly with $r_1 = 0$ and $r_M = M_2(M_1 + 1) - 1$, where M denotes the nested array size. Then the steering vector of the aforementioned $(M_2M_1 + M_2)$-element ULA for source k, denoted by $\mathbf{d}(\theta_k)$, is

$$\mathbf{d}(\theta_k) = \left[1, e^{j2\pi d \sin\theta_k/\lambda}, \ldots, e^{j2\pi(M_2M_1+M_2-1)d\sin\theta_k/\lambda}\right] \tag{12}$$

Denote $\boldsymbol{\Psi}_{\mathbf{r}} \in \{0, 1\}^{M\times[M_2(M_1+1)]}$ a selection matrix such that the ith row of $\boldsymbol{\Psi}_{\mathbf{r}}$ contains all 0 s but a single 1 at the $(r_i + 1)$th position. It is clear that

$$\mathbf{a}(\theta_k) = \boldsymbol{\Psi}_{\mathbf{r}}\mathbf{d}(\theta_k) \tag{13}$$

As a consequence, we have

$$\mathbf{a}^*(\theta_k)\otimes\mathbf{a}(\theta_k)=\left[\boldsymbol{\Psi}_r\mathbf{d}(\theta_k)\right]^*\otimes\left[\boldsymbol{\Psi}_r\mathbf{d}(\theta_k)\right]=\left[\boldsymbol{\Psi}_r\mathbf{d}^*(\theta_k)\right]\otimes\left[\boldsymbol{\Psi}_r\mathbf{d}(\theta_k)\right]$$
$$=\left(\boldsymbol{\Psi}_r\otimes\boldsymbol{\Psi}_r\right)\left[\mathbf{d}^*(\theta_k)\otimes\mathbf{d}(\theta_k)\right] \tag{14}$$

where the last equality is due to the property of Kronecker product as defined by [35]. We also notice that for a vector $\mathbf{d}(\theta_k)$,

$$\begin{aligned}\mathbf{d}^*(\theta_k)\otimes\mathbf{d}(\theta_k)&=\left[\mathbf{d}(\theta_k),\, e^{-j2\pi d\sin\theta_k/\lambda}\mathbf{d}(\theta_k),\,\ldots,\, e^{-j2\pi(M_2M_1+M_2-1)d\sin\theta_k/\lambda}\mathbf{d}(\theta_k)\right]^T\\&=\Big[1,\, e^{j2\pi d\sin\theta_k/\lambda},\,\ldots,\, e^{j2\pi(M_2M_1+M_2-1)d\sin\theta_k/\lambda},\\&\quad e^{-j2\pi d\sin\theta_k/\lambda},\,1,\,\ldots,\, e^{j2\pi(M_2M_1+M_2-2)d\sin\theta_k/\lambda},\\&\quad\ldots\ldots,\, e^{-j2\pi(M_2M_1+M_2-1)d\sin\theta_k/\lambda},\, e^{-j2\pi(M_2M_1+M_2-2)d\sin\theta_k/\lambda},\,\ldots,\,1\Big]^T\end{aligned} \tag{15}$$

This means that $\mathbf{d}^*(\theta_k)\otimes\mathbf{d}(\theta_k)$ can be alternatively represented by

$$\mathbf{d}^*(\theta_k)\otimes\mathbf{d}(\theta_k)=\mathbf{G}\tilde{\mathbf{a}}(\theta_k) \tag{16}$$

where

$$\mathbf{G}=\begin{bmatrix}0&\cdots&0&1&0&\cdots&0\\0&\cdots&0&0&1&\cdots&0\\\vdots&\ddots&\vdots&\vdots&\vdots&\ddots&\vdots\\0&\cdots&0&0&0&\cdots&1\\0&\cdots&1&0&0&\cdots&0\\0&\cdots&0&1&0&\cdots&0\\\vdots&\ddots&\vdots&\vdots&\ddots&\vdots&\vdots\\0&\cdots&0&0&\cdots&1&0\\\vdots&\vdots&\vdots&\vdots&\vdots&\vdots&\vdots\\1&0&\cdots&0&0&\cdots&0\\0&1&\cdots&0&0&\cdots&0\\\vdots&\vdots&\ddots&\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&1&0&\cdots&0\end{bmatrix}\in\mathbb{R}^{\left[M_2(M_1+1)\right]^2\times\left[2M_2(M_1+1)-1\right]} \tag{17}$$

Using these representations, we can re-express $\mathbf{A}^*\circ\mathbf{A}$ as

$$\mathbf{A}^*\circ\mathbf{A}=\left(\boldsymbol{\Psi}_r\otimes\boldsymbol{\Psi}_r\right)\mathbf{G}\tilde{\mathbf{A}} \tag{18}$$

where $\tilde{\mathbf{A}}=[\tilde{\mathbf{a}}(\theta_1),\,\tilde{\mathbf{a}}(\theta_2),\,\ldots,\,\tilde{\mathbf{a}}(\theta_K)]\in\mathbb{C}^{(2M_2M_1+2M_2-1)\times K}$ is a dimension-reduced virtual array response matrix. Actually, it is interesting to note that the replicas of those repeated rows in $\mathbf{A}^*\circ\mathbf{A}$ or $\boldsymbol{\varepsilon}$ can be excluded in a same way. That is, the dimension reducing transformation $[(\boldsymbol{\Psi}_r\otimes\boldsymbol{\Psi}_r)\mathbf{G}]^{-1}$ made previously on the matrix $\mathbf{A}^*\circ\mathbf{A}$, apply verbatim to $\boldsymbol{\varepsilon}$ as well

$$\boldsymbol{\varepsilon}=\left(\boldsymbol{\Psi}_r\otimes\boldsymbol{\Psi}_r\right)\mathbf{G}\bar{\boldsymbol{\varepsilon}} \tag{19}$$

As stated before, it is, in general, difficult to previously determine the unknown noise variance in a heuristic manner, especially in the underdetermined DOA estimation circumstances while the data covariance matrix no longer contains a low rank component spanning the noise subspace. Below, we derive a reliable formula for eliminating the unknown noise variance effectively and easily. It is interesting to observe that the vector $\sigma_n^2\bar{\mathbf{e}}$ in (10) has only one nonzero element which equals to σ_n^2. Based on this observation, we can remove the element of $\hat{\mathbf{z}}$ corresponding to the position of σ_n^2 in $\sigma_n^2\bar{\mathbf{e}}$; this also means that the rest $(M^2/2+M-2)$ entries of $\hat{\mathbf{z}}$ corresponding to those positions of zero elements in $\sigma_n^2\bar{\mathbf{e}}$ are preserved. Mathematically, this operation can be formulated as performing a projection

$$\bar{\mathbf{z}}=\mathbf{J}\hat{\mathbf{z}}=\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})\bar{\mathbf{p}}+\mathbf{J}\bar{\boldsymbol{\varepsilon}} \tag{20}$$

Here, $\mathbf{J}$ is a $[2M_2(M_1+1)-2]\times[2M_2(M_1+1)-1]$ selecting matrix and can be represented as

$$\mathbf{J}=\begin{bmatrix}\mathbf{I}_{\left[M_2(M_1+1)-1\right]}&\mathbf{0}_{\left[M_2(M_1+1)-1\right]\times\left[M_2(M_1+1)\right]}\\\mathbf{0}_{\left[M_2(M_1+1)-1\right]\times\left[M_2(M_1+1)\right]}&\mathbf{I}_{\left[M_2(M_1+1)-1\right]}\end{bmatrix} \tag{21}$$

This elimination operation avoids the estimation of noise variance, therefore facilitating a noise-free sparse representation. Now, the inherent perturbation variance of (20) is different from that in (10). More concretely, by revisiting the formulation (11), the perturbation $\boldsymbol{\varepsilon}$ is asymptotically Gaussian distribution, which along with (19) and (20) yields

$$\mathbf{J}\bar{\boldsymbol{\varepsilon}}\sim CN\left(0,\,\mathbf{J}\mathbf{F}^{-1}\left(\mathbf{R}_x^T\otimes\mathbf{R}_x\right)\left(\mathbf{J}\mathbf{F}^{-1}\right)^T/N\right) \tag{22}$$

Here, for notational simplicity, we use the symbol $\mathbf{F}$ to represent the matrix product $(\boldsymbol{\Psi}_r\otimes\boldsymbol{\Psi}_r)\mathbf{G}\in\mathbb{R}^{M^2\times[2M_2(M_1+1)-1]}$. Combining (20) and (22), it can be deduced that

$$\mathbf{W}^{-1/2}\left[\bar{\mathbf{z}}-\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})\bar{\mathbf{p}}\right]\sim CN\left(0,\,\mathbf{I}_{\left[2M_2(M_1+1)-2\right]}\right) \tag{23}$$

where the weighted matrix $\mathbf{W}^{-1/2}$ is the Hermitian square root of $\mathbf{W}^{-1}$, $\mathbf{W}=\mathbf{J}\mathbf{F}^{-1}(\mathbf{R}_x^T\otimes\mathbf{R}_x)(\mathbf{J}\mathbf{F}^{-1})^T/N$. That is to say, such a normalization conversion makes the estimation error of the weighted covariance vector, $\mathbf{W}^{-1/2}\bar{\mathbf{z}}$, satisfies the asymptotic standard normal distribution. Then, the novel noise-variance-eliminated data model used to estimate off-grid DOAs can be written as

$$\tilde{\mathbf{z}}=\mathbf{W}^{-1/2}\bar{\mathbf{z}}=\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\bar{\mathbf{p}}+\tilde{\boldsymbol{\varepsilon}} \tag{24}$$

where $\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})=\mathbf{W}^{-1/2}\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})\in\mathbb{C}^{[2M_2(M_1+1)-2]\times I}$ is the new manifold matrix, $\tilde{\boldsymbol{\varepsilon}}=\mathbf{W}^{-1/2}\mathbf{J}\bar{\boldsymbol{\varepsilon}}\in\mathbb{C}^{[2M_2(M_1+1)-2]\times 1}$.

It should also be noted that, in practical applications, the covariance matrix $\mathbf{R}_x$ can only be estimated using finite length of snapshots collected at time instants, hence the estimate of $\mathbf{W}^{-1/2}$ can be denoted as

$$\mathbf{W}^{-1/2}=\left[\mathbf{J}\mathbf{F}^{-1}\left(\hat{\mathbf{R}}_x^T\otimes\hat{\mathbf{R}}_x\right)\left(\mathbf{J}\mathbf{F}^{-1}\right)^T/N\right]^{-1/2} \tag{25}$$

After obtaining the above overcomplete formulation (24) of the covariance vector, the following task is to extract the basis set of true sources from $\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})$ to approximate $\tilde{\mathbf{z}}$ by invoking the SBL criterion. The joint off-grid distance compensation and sparse support recovery method that is developed follows an iteration routine. To do this, we assume that the rows of $\bar{\mathbf{p}}$ are circular complex Gaussian distributed, and a hyperparameter vector $\boldsymbol{\gamma}=[\gamma_1,\,\gamma_2,\,\ldots,\,\gamma_I]^T$ can be introduced to denote its row variance, i.e., $\bar{\mathbf{p}}\sim CN(\mathbf{0},\,\boldsymbol{\Gamma})$ and $\boldsymbol{\Gamma}=\mathrm{diag}(\boldsymbol{\gamma})$. Then the likelihood function of $\tilde{\mathbf{z}}$ with respect to (w.r.t.) $\boldsymbol{\gamma}$ by integrating out the same powers is given as follows:

$$\begin{aligned}p(\tilde{\mathbf{z}}|\boldsymbol{\gamma})&=\int p(\tilde{\mathbf{z}}|\bar{\mathbf{p}})p(\bar{\mathbf{p}}|\boldsymbol{\gamma})d\bar{\mathbf{p}}\\&=\left|\pi\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}\right|^{-1}\exp\left\{-\tilde{\mathbf{z}}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}^{-1}\tilde{\mathbf{z}}\right\}\times\int\left|\pi\boldsymbol{\Sigma}_{\bar{\mathbf{p}}}\right|^{-1}\exp\left\{-\left[\left(\bar{\mathbf{p}}-\boldsymbol{\mu}\right)^H\boldsymbol{\Sigma}_{\bar{\mathbf{p}}}^{-1}\left(\bar{\mathbf{p}}-\boldsymbol{\mu}\right)\right]\right\}d\bar{\mathbf{p}}\\&=\left|\pi\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}\right|^{-1}\exp\left\{-\tilde{\mathbf{z}}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}^{-1}\tilde{\mathbf{z}}\right\}\end{aligned} \tag{26}$$

where

$$\boldsymbol{\mu}=\boldsymbol{\Gamma}\tilde{\boldsymbol{\Phi}}^H(\boldsymbol{\beta})\left[\mathbf{I}_{\left[2M_2(M_1+1)-2\right]}+\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\boldsymbol{\Gamma}\tilde{\boldsymbol{\Phi}}^H(\boldsymbol{\beta})\right]^{-1}\tilde{\mathbf{z}} \tag{27}$$

$$\boldsymbol{\Sigma}_{\bar{\mathbf{p}}}=\boldsymbol{\Gamma}-\boldsymbol{\Gamma}\tilde{\boldsymbol{\Phi}}^H(\boldsymbol{\beta})\left[\mathbf{I}_{\left[2M_2(M_1+1)-2\right]}+\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\boldsymbol{\Gamma}\tilde{\boldsymbol{\Phi}}^H(\boldsymbol{\beta})\right]^{-1}\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\boldsymbol{\Gamma} \tag{28}$$

$$\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}=\mathbf{I}_{\left[2M_2(M_1+1)-2\right]}+\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\boldsymbol{\Gamma}\tilde{\boldsymbol{\Phi}}^H(\boldsymbol{\beta}) \tag{29}$$

Eq. (26) is a type-II maximum likelihood problem [36], and it reveals the relationship between $\tilde{\mathbf{z}}$ and $\boldsymbol{\gamma}$. $\boldsymbol{\gamma}$ can be optimized by maximizing the likelihood function $p(\tilde{\mathbf{z}}|\boldsymbol{\gamma})$, and such problems are generally solved using the Expectation-Maximization (EM) algorithm [37]. After that, the first order posterior moment of $\bar{\mathbf{p}}$ is

calculated according to (27) and the indexes of its nonzero elements indicate the signal directions. We should also emphasize that the coefficient elements of $\bar{\mathbf{p}}$ in (24) should be coherent as they represent the spatial power distribution of the incident signals. However, since non-Gaussian assumptions generally greatly block Bayesian parameter estimation processes, we abandon this prior information and follow the guideline of SBL to append a Gaussian distribution to $\bar{\mathbf{p}}$.

Taking the logarithm of (26) and neglecting the constants yields the following objective function for optimizing $\boldsymbol{\gamma}$,

$$L(\boldsymbol{\gamma}) = \ln\left|\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}\right| + \tilde{\mathbf{z}}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{z}}}^{-1}\tilde{\mathbf{z}} \tag{30}$$

The EM algorithm, which consists of iterative E-steps and M-steps, can then be used to estimate $\boldsymbol{\gamma}$ by minimizing this objective function. The E-step calculates the first- and second-order posterior moments of $\bar{\mathbf{p}}$ with (27) and (28), and the M-step updates $\boldsymbol{\gamma}$ by minimizing $L(\boldsymbol{\gamma})$ according to $\partial L(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma} = \mathbf{0}$, which results in the following update strategy of $\boldsymbol{\gamma}$

$$\gamma_i^{(q)} = \left|\boldsymbol{\mu}_i^{(q)}\right|^2 + \left(\boldsymbol{\Sigma}_{\bar{\mathbf{p}}}^{(q)}\right)_{i,i} \tag{31}$$

where the superscript $(\cdot)^q$ represents the qth iteration, $\boldsymbol{\mu}^{(q)}$ and $\boldsymbol{\Sigma}_{\bar{\mathbf{p}}}^{(q)}$ are calculated with (27) and (28) in the qth iteration. The update strategy given in (31) can be substituted with a fixed-point iteration as $\gamma_i^{(q)} = \left|\boldsymbol{\mu}_i^{(q)}\right|^2/(1-(\boldsymbol{\Sigma}_{\bar{\mathbf{p}}}^{(q)})_{i,i}/\gamma_i^{(q-1)}) + \zeta$ to speed up the convergence of the EM algorithm, with ζ being a small positive value, such as 10^{-10} [21]. The initialization and termination criteria of the EM algorithm are also set similarly as those in [24]. When a predefined stopping criterion is satisfied, the iteration process can be terminated to export the parameter estimates, which approach the global minimum.

So far we have obtained the hyperparameter set $\boldsymbol{\gamma}$, which describes the spatial power distribution of the incident signals on the pre-defined direction grid, in the above iteration. However, since the discreteness of Θ, no matter how finely it is chosen, causes high-precision DOA estimates hardly available from the peak locations of $\boldsymbol{\gamma}$. Therefore further processing of the reconstruction result should be introduced to eliminate the quantization error by estimating the off-grid distance vector $\boldsymbol{\beta}$ directly.

For $\boldsymbol{\beta} \in [-r/2, r/2]^I$, its estimate can be obtained by minimizing

$$\begin{aligned}\mathrm{E}\left\{\|\tilde{\mathbf{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\bar{\mathbf{p}}\|_2^2\right\} &= \|\tilde{\mathbf{z}} - \left[\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\mathrm{diag}(\boldsymbol{\beta}) + \tilde{\mathbf{C}}\mathrm{diag}(\boldsymbol{\alpha})\right]\boldsymbol{\mu}\|_2^2 \\ &\quad + \mathrm{Tr}\left\{\left[\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\mathrm{diag}(\boldsymbol{\beta}) + \tilde{\mathbf{C}}\mathrm{diag}(\boldsymbol{\alpha})\right]\right. \\ &\quad \left.\boldsymbol{\Sigma}_{\bar{\mathbf{p}}}\left[\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\mathrm{diag}(\boldsymbol{\beta}) + \tilde{\mathbf{C}}\mathrm{diag}(\boldsymbol{\alpha})\right]^H\right\} \\ &= \boldsymbol{\beta}^T\mathbf{P}_1\boldsymbol{\beta} - 2\mathbf{V}_1^T\boldsymbol{\beta} + \boldsymbol{\alpha}^T\mathbf{P}_2\boldsymbol{\alpha} - 2\mathbf{V}_2^T\boldsymbol{\alpha} \\ &\quad + 2\mathrm{Re}\left(\boldsymbol{\beta}^T\boldsymbol{\omega}\boldsymbol{\alpha}\right) + c_0\end{aligned} \tag{32}$$

where $\tilde{\mathbf{A}} = \mathbf{W}^{-1/2}\mathbf{J}\bar{\mathbf{A}}$, $\tilde{\mathbf{B}} = \mathbf{W}^{-1/2}\mathbf{J}\mathbf{B}$, $\tilde{\mathbf{C}} = \mathbf{W}^{-1/2}\mathbf{J}\mathbf{C}$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_I]^T = \boldsymbol{\beta}^2$ (using the MATLAB notations), c_0 is a constant term independent of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, $\mathbf{P}_1(\mathbf{P}_2)$ is a positive semi-definite matrix and

$$\mathbf{P}_1 = \mathrm{Re}\left[\left(\tilde{\mathbf{B}}^H\tilde{\mathbf{B}}\right)^* \oplus \left(\boldsymbol{\mu}\boldsymbol{\mu}^H + \boldsymbol{\Sigma}_{\bar{\mathbf{p}}}\right)\right] \tag{33}$$

$$\mathbf{P}_2 = \mathrm{Re}\left[\left(\tilde{\mathbf{C}}^H\tilde{\mathbf{C}}\right)^* \oplus \left(\boldsymbol{\mu}\boldsymbol{\mu}^H + \boldsymbol{\Sigma}_{\bar{\mathbf{p}}}\right)\right] \tag{34}$$

$$\mathbf{V}_1 = \mathrm{Re}\left[\mathrm{diag}(\boldsymbol{\mu}^*)\tilde{\mathbf{B}}^H\left(\tilde{\mathbf{z}} - \tilde{\mathbf{A}}\boldsymbol{\mu}\right)\right] - \mathrm{Re}\left[\mathrm{diag}\left(\tilde{\mathbf{B}}^H\tilde{\mathbf{A}}\boldsymbol{\Sigma}_{\bar{\mathbf{p}}}\right)\right] \tag{35}$$

$$\mathbf{V}_2 = \mathrm{Re}\left[\mathrm{diag}(\boldsymbol{\mu}^*)\tilde{\mathbf{C}}^H\left(\tilde{\mathbf{z}} - \tilde{\mathbf{A}}\boldsymbol{\mu}\right)\right] - \mathrm{Re}\left[\mathrm{diag}\left(\tilde{\mathbf{C}}^H\tilde{\mathbf{A}}\boldsymbol{\Sigma}_{\bar{\mathbf{p}}}\right)\right] \tag{36}$$

$$\boldsymbol{\omega} = \left[\boldsymbol{\Sigma}_{\bar{\mathbf{p}}} \oplus \left(\tilde{\mathbf{C}}^H\tilde{\mathbf{B}}\right)^*\right] + \left[\left(\tilde{\mathbf{B}}^H\tilde{\mathbf{C}}\right)^* \oplus \left(\boldsymbol{\mu}\boldsymbol{\mu}^H\right)\right] \tag{37}$$

Similar to [38], the detailed derivation of (32) is provided in Appendix for simplicity of exposition. As a result, we have

$$\begin{aligned}\boldsymbol{\beta}^{new} = \arg\min_{\boldsymbol{\beta}\in[-r/2,r/2]^I}\left\{\boldsymbol{\beta}^T\mathbf{P}_1\boldsymbol{\beta} - 2\mathbf{V}_1^T\boldsymbol{\beta} + \boldsymbol{\alpha}^T\mathbf{P}_2\boldsymbol{\alpha} - 2\mathbf{V}_2^T\boldsymbol{\alpha}\right. \\ \left. + 2\mathrm{Re}\left(\boldsymbol{\beta}^T\boldsymbol{\omega}\boldsymbol{\alpha}\right)\right\}\end{aligned} \tag{38}$$

In order to give an explicit expression of $\boldsymbol{\beta}^{new}$, a natural choice is the well-known multi-dimensional searching method, but the heavy computational load prohibits it from practical applications. To circumvent this issue, this paper aims at developing a search-free technique for grid matching with an affordable computational workload.

The refined DOA estimation procedure can be implemented in turn for all peak locations of the reconstructed spectrum. Take the kth peak for example, denote the signal DOA associated with it by θ_k, the corresponding grid mismatch by β_k. The variables irrelevant with this signal are identified with superscript $(\cdot)^{-k}$, e.g., $\Theta^{-k} = \Theta$ â§¹ θ_k, which means that removing θ_k from Θ yields Θ^{-k}, $\boldsymbol{\beta}^{-k} = \boldsymbol{\beta}$ â§¹ β_k. Furthermore, let

$$\mathbf{H} = \left[\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \ldots, \tilde{\mathbf{e}}_{k-1}, \tilde{\mathbf{e}}_I, \tilde{\mathbf{e}}_{k+1}, \ldots, \tilde{\mathbf{e}}_k\right] \tag{39}$$

be an $I \times I$ permutation matrix with $\tilde{\mathbf{e}}_i (i = 1, 2, \ldots, I)$ is an $I \times 1$ column vector with 1 at the ith position and 0 elsewhere. Then one can note that when left-multiplying a matrix with $\mathbf{H}$, the operation swaps the kth row with the Ith row of that matrix and when right-multiplying a matrix with $\mathbf{H}$, the operation swaps the kth column with the Ith column of that matrix. We also note an interesting property of $\mathbf{H}$, that is, its self-inverse and symmetric

$$\mathbf{H}^{-1} = \mathbf{H}^T = \mathbf{H} \tag{40}$$

Plugging (39) and (40) into the first term of (32) leads to

$$\begin{aligned}&\boldsymbol{\beta}^T\mathbf{P}_1\boldsymbol{\beta} = \boldsymbol{\beta}^T\mathbf{H}\mathbf{H}^T\mathbf{P}_1\mathbf{H}^T\mathbf{H}\boldsymbol{\beta} \\ &= \begin{bmatrix}\left(\boldsymbol{\beta}^{-k}\right)^T & \beta_k\end{bmatrix}\begin{bmatrix}\mathbf{P}_1^{-k} & (\mathbf{P}_1)_{:,k}^{-k} \\ (\mathbf{P}_1)_{k,:}^{-k} & (\mathbf{P}_1)_{k,k}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}^{-k} \\ \beta_k\end{bmatrix} \\ &= \beta_k(\mathbf{P}_1)_{k,:}^{-k}\boldsymbol{\beta}^{-k} + \left(\boldsymbol{\beta}^{-k}\right)^T(\mathbf{P}_1)_{:,k}^{-k}\beta_k + \beta_k(\mathbf{P}_1)_{k,k}\beta_k + \mathrm{res.1}\end{aligned} \tag{41}$$

where we use the MATLAB notations to define

$$\begin{cases}(\mathbf{P}_1)_{:,k}^{-k} = \mathbf{P}_1(1\!:\!k-1\ I\ k+1\!:\!I-1, k) \\ (\mathbf{P}_1)_{k,:}^{-k} = \mathbf{P}_1(k, 1\!:\!k-1\ I\ k+1\!:\!I-1)\end{cases} \tag{42}$$

$\mathbf{P}_1^{-k}$ equals $\mathbf{P}_1$ by removing the rows and columns corresponding to θ_k, and res.1 can be deemed to be the residual item of $\boldsymbol{\beta}^T\mathbf{P}_1\boldsymbol{\beta}$ by removing the k th signal component. This form of $\boldsymbol{\beta}^T\mathbf{P}_1\boldsymbol{\beta}$ is introduced to facilitate the estimation of the kth signal off-grid distance.

In a similar way to derive (41), one obtains

$$\mathbf{V}_1^T\boldsymbol{\beta} = \left(\mathbf{V}_1^T\right)_k\boldsymbol{\beta}_k + \mathrm{res.2} \tag{43}$$

$$\boldsymbol{\alpha}^T\mathbf{P}_2\boldsymbol{\alpha} = \alpha_k(\mathbf{P}_2)_{k,:}^{-k}\boldsymbol{\alpha}^{-k} + \left(\boldsymbol{\alpha}^{-k}\right)^T(\mathbf{P}_2)_{:,k}^{-k}\alpha_k + \alpha_k(\mathbf{P}_2)_{k,k}\alpha_k + \mathrm{res.3} \tag{44}$$

$$\mathbf{V}_2^T\boldsymbol{\alpha} = \left(\mathbf{V}_2^T\right)_k\boldsymbol{\alpha}_k + \text{res.4} \tag{45}$$

$$\boldsymbol{\beta}^T\boldsymbol{\omega}\boldsymbol{\alpha} = \beta_k\boldsymbol{\omega}_{k,:}^{-k}\boldsymbol{\alpha}^{-k} + \left(\boldsymbol{\beta}^{-k}\right)^T\boldsymbol{\omega}_{:,k}^{-k}\alpha_k + \beta_k\omega_{k,k}\alpha_k + \text{res.5} \tag{46}$$

where res.2, res.3, res.4 and res.5 represent the residual items of $\mathbf{V}_1^T\boldsymbol{\beta}$, $\boldsymbol{\alpha}^T\mathbf{P}_2\boldsymbol{\alpha}$, $\mathbf{V}_2^T\boldsymbol{\alpha}$ and $\boldsymbol{\beta}^T\boldsymbol{\omega}\boldsymbol{\alpha}$, respectively.

Combining (41)–(45) along with (38), the grid mismatch of the kth signal, denoted by β_k, can be estimated by minimizing the following objective function:

$$\begin{aligned} f(\beta_k) = {} & (\mathbf{P}_2)_{k,k}\beta_k^4 + 2\text{Re}(\boldsymbol{\omega}_{k,k})\beta_k^3 \\ & + \left\{(\mathbf{P}_1)_{k,k} + (\mathbf{P}_2)_{k,:}^{-k}\boldsymbol{\alpha}^{-k} + \left(\boldsymbol{\alpha}^{-k}\right)^T(\mathbf{P}_2)_{:,k}^{-k} + 2\text{Re}\left[\left(\boldsymbol{\beta}^{-k}\right)^T\boldsymbol{\omega}_{:,k}^{-k}\right]\right. \\ & \left. - 2\left(\mathbf{V}_2^T\right)_k\right\}\beta_k^2 \\ & + \left[2\text{Re}\left(\boldsymbol{\omega}_{k,:}^{-k}\boldsymbol{\alpha}^{-k}\right) - 2\left(\mathbf{V}_1^T\right)_k + (\mathbf{P}_1)_{k,:}^{-k}\boldsymbol{\beta}^{-k} + \left(\boldsymbol{\beta}^{-k}\right)^T(\mathbf{P}_1)_{:,k}^{-k}\right]\beta_k \\ & + \text{res} \end{aligned} \tag{47}$$

Here res = res.1 − 2res.2 + res.3 − 2res.4 + 2Re(res.5), and the relationship between α_k and β_k, viz., $\alpha_k = \beta_k^2$, has been used. Clearly, we obtain from (47) that the discretization error β_k can be alternatively estimated from the roots of the following polynomial:

$$\begin{aligned} f'(\beta_k) = {} & 4(\mathbf{P}_2)_{k,k}\beta_k^3 + 6\text{Re}(\boldsymbol{\omega}_{k,k})\beta_k^2 \\ & + 2\left\{(\mathbf{P}_1)_{k,k} + (\mathbf{P}_2)_{k,:}^{-k}\boldsymbol{\alpha}^{-k} + \left(\boldsymbol{\alpha}^{-k}\right)^T(\mathbf{P}_2)_{:,k}^{-k}\right. \\ & \left. + 2\text{Re}\left[\left(\boldsymbol{\beta}^{-k}\right)^T\boldsymbol{\omega}_{:,k}^{-k}\right] - 2\left(\mathbf{V}_2^T\right)_k\right\}\beta_k \\ & + \left[2\text{Re}\left(\boldsymbol{\omega}_{k,:}^{-k}\boldsymbol{\alpha}^{-k}\right) - 2\left(\mathbf{V}_1^T\right)_k + (\mathbf{P}_1)_{k,:}^{-k}\boldsymbol{\beta}^{-k} + \left(\boldsymbol{\beta}^{-k}\right)^T(\mathbf{P}_1)_{:,k}^{-k}\right] \end{aligned} \tag{48}$$

It is easy to see that the estimate of β_k, denoted by $\hat{\beta}_k$, can be derived by testing the equivalence of the item on the right-hand-side of (48) to 0. Similarly, setting $k = 1, 2, \ldots, K$ in (48) obtains the grid mismatch estimates of all the signals. Additionally, we should note the fact that $\boldsymbol{\beta}$ is jointly sparse with $\bar{\mathbf{p}}$ whose K nonzero entries correspond to the locations of the K sources. However, we do not need to know the number of signals. We calculate only entries of $\boldsymbol{\beta}$ that correspond to the locations of the maximum $M_2(M_1+1)-1$ (i.e., the so-called maximal separable signal number, and the proof of it will be postponed to Section 3.3 to avoid disruption to reading flow) entries of $\boldsymbol{\gamma}$ and set others to zeros. As a consequence, $\boldsymbol{\beta}$ can be truncated into dimension of $M_2(M_1+1)-1$. We still use $\boldsymbol{\beta}$ hereafter to denote their truncated version for simplicity. Then by constraining $\beta_k \in [-r/2, r/2]$, we update $\boldsymbol{\beta}$ element wise

$$\beta_k^{new} = \begin{cases} \hat{\beta}_k, & \text{if } \hat{\beta}_k \in \left[-r/2, r/2\right]; \\ -r/2, & \text{if } \hat{\beta}_k < -r/2; \\ r/2, & \text{otherwise.} \end{cases} \tag{49}$$

Combining the above derivation and analysis, it is now obvious that a simple procedure to estimate the signal off-grid distance one by one from the root of a polynomial has been developed. The MATLAB-based calculation of these roots is via the function roots($f'(\beta_k)$). Computationally, the proposed polynomial rooting technique is more efficient than the conventional multi-dimensional searching strategy.

Table 1
Scheme of the proposed algorithm.

Input data: Observed snapshots $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)$, $t = 1, 2, \ldots, N$.
Step 1: Compute the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{x}}$ from the array observed snapshots $\mathbf{x}(t)$.
Step 2: Vectorize $\hat{\mathbf{R}}_{\mathbf{x}}$, get $\hat{\mathbf{y}}$ according to (6), then remove the repeated rows of $\hat{\mathbf{y}}$ and sort the remaining elements, get $\hat{\mathbf{z}}$ according to (7).
Step 3: Construct the overcomplete formulation of $\hat{\mathbf{z}}$ according to (10).
Step 4: Eliminate the unknown noise component of $\hat{\mathbf{z}}$, get $\bar{\mathbf{z}}$ according to (20).
Step 5: Calculate the weighted matrix $\mathbf{W}^{-1/2}$ according to (25), then utilize it to normalize the vectorized covariance matrix perturbation incorporated in $\bar{\mathbf{z}}$, get $\tilde{\mathbf{z}}$ according to (24).
Step 6: Initialize the hyperparameters $\boldsymbol{\gamma}$ and the off-grid distances $\boldsymbol{\beta}$.
Step 7: Build $\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})$ and $\boldsymbol{\Gamma}$ based on the current values of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively.
Step 8: Calculate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_{\bar{\mathbf{p}}}$ according to (27) and (28), respectively.
Step 9: Update $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ according to (31) (or its fixed-point homotopy version) and (48) to (49), respectively.
Step 10: Terminate the iteration process if the final convergence criterion is satisfied or the maximum number of iterations is reached. Otherwise, return to step 7 and continue the whole process again.
Output data: Parameter estimator $(\boldsymbol{\gamma}, \boldsymbol{\beta}^{new})$.

3.3. Outline of the proposed method and some complementary issues

In order to provide an overall perspective of the proposed off-grid DOA estimation method, we tabulate its scheme in Table 1.

Some supplements listed as follows should be added to complete the proposed method.

Remark 1. During initialization, $\boldsymbol{\gamma}^{(0)}$ can be set according to the least squares estimate of the source power, i.e., $\boldsymbol{\mu}^{(0)} = \tilde{\boldsymbol{\Phi}}^H(\boldsymbol{\beta})[\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\tilde{\boldsymbol{\Phi}}^H(\boldsymbol{\beta})]^{-1}\tilde{\mathbf{z}}$ and $\gamma_i^{(0)} = \left|\mu_i^{(0)}\right|^2$, and $\boldsymbol{\beta}$ only needs to be set equal to 0.

Remark 2. We terminate the updating process of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ if $\|\boldsymbol{\gamma}^{(q+1)} - \boldsymbol{\gamma}^{(q)}\|_2/\|\boldsymbol{\gamma}^{(q)}\|_2 \le \tau$ (i.e., the so-called final convergence criteria) or the maximum number of iterations is reached, where τ is a user-defined tolerance. In the simulation Section, we set $\tau = 10^{-4}$ and the maximum number of iterations to 500.

Remark 3. The maximal separable signal number is also an important issue that should be considered for sparse recovery based methods. This problem naturally relates to the unique solution of the vector $\bar{\mathbf{p}}$ since as is shown in (24), the DOA estimation problem relies on the sparse recovery of this vector. The following proposition provides the key result on the separable source number.

Proposition. *For a nested array with M_1 and M_2 $(M_1 + M_2 = M)$ elements respectively in each sub-configuration, the maximal separable signal number is $M_2(M_1+1)-1$.*

Proof. Consider the single measurement vector (SMV) problem (24), i.e., $\tilde{\mathbf{z}} = \tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\bar{\mathbf{p}} + \tilde{\boldsymbol{\varepsilon}}$, where $\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta}) = \mathbf{W}^{-1/2}\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})$ and $\boldsymbol{\Phi}(\boldsymbol{\beta})$ is an overcomplete basis matrix, $\bar{\mathbf{p}}$ denotes the sparse vector to be recovered. Based on corollary 1 of [39], there exists a unique sparsest representation $\bar{\mathbf{p}}$ such that $\tilde{\mathbf{z}} = \tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\bar{\mathbf{p}} + \tilde{\boldsymbol{\varepsilon}}$ if and only if

$$\|\bar{\mathbf{p}}\|_0 \le \text{Krank}\left[\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\right]/2 \tag{50}$$

where the Kruskal rank of $\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})$, denoted as Krank$[\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})]$ is defined as the largest integer m such that arbitrary m columns of $\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})$ are linearly independent. The following analyses show the Krank of $\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})$ used in the proposed method.

Address Krank$[\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})]$ first. According to the co-array principle [7], the maximum DOF achievable in $\boldsymbol{\Phi}(\boldsymbol{\beta})$ is $\text{DOF}_{max}^{\boldsymbol{\Phi}(\boldsymbol{\beta})} = 2M_2(M_1+1)-1$. Recall the construction procedure of $\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})$, and it shows that the $2M_2(M_1+1)-2$ rows of $\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})$ also belong to the collection of $2M_2(M_1+1)-1$ rows of Vandermonde

matrix $\boldsymbol{\Phi}(\boldsymbol{\beta})$. Thus pre-multiplying $\mathbf{J}$ to $\boldsymbol{\Phi}(\boldsymbol{\beta})$ will give rise to a degree 1 deficiency, namely, the maximum DOF achievable of $\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})$ is $\mathrm{DOF}_{max}^{\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})} = 2M_2(M_1+1)-2$. With nonambiguity in the array structure, we can deduce that $\mathrm{Krank}[\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})] = 2M_2(M_1+1)-2$.

Address $\mathrm{Krank}[\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})]$ second. According to the construction procedure of $\mathbf{W}^{-1/2}$, it is obvious that $\mathbf{W}^{-1/2}$ is of full rank, which implies the following equation:

$$\mathrm{Krank}\left[\tilde{\boldsymbol{\Phi}}(\boldsymbol{\beta})\right] = \mathrm{Krank}\left[\mathbf{W}^{-1/2}\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})\right] = \mathrm{Krank}\left[\mathbf{J}\boldsymbol{\Phi}(\boldsymbol{\beta})\right] = 2M_2(M_1+1)-2 \tag{51}$$

By combining (50) and (51), a conclusion can be derived: the maximal number of sources that can be resolved by the proposed method is equal to $M_2(M_1+1)-1$. This completes the proof.

It follows from the above Proposition that it is possible to tackle the underdetermined case of $K > M$ if we exploit the configuration of nested array to increase the DOF. The idea behind this DOF enhancement property is to span large aperture by using the concept of "difference set" of the original samples of the data, and build a larger positive semi-definite matrix from the original covariance matrix, on which super-resolution spectral estimation algorithms can be directly applied.

Remark 4. Like other spectral-based methods, in our proposed algorithm, the DOAs are estimated using the locations of the highest peaks of the reconstructed spectrum. Suppose that the grid indices of the highest K peaks of $\boldsymbol{\gamma}$ are n_k, $k = 1, 2, \ldots, K$. The estimated K DOAs will be $\hat{\theta}_k = \tilde{\theta}_{n_k} + \beta_{n_k}^{new}$, $k = 1, 2, \ldots, K$.

4. Numerical simulations

For illustrative purpose, in this Section, we carry out numerical simulations to show the merits of implementing our proposed algorithm to nested array under various scenarios. Five other existing methods are also listed for performance comparison, including SS-MUSIC [10], L1-SVD [15], L1-SRACV [17], CMSR [16,20] and SPICE [18,19]. To provide a benchmark for evaluating the effectiveness of these methods, the CRLB given in [36] is also included though it is a lower bound for unbiased estimators. Multiple spatially uncorrelated narrowband source signals, which are samples from an uncorrelated complex Gaussian process with zero mean, impinge on this sensor array, while the number and DOAs of these sources vary in the following different simulations. The measurements are corrupted by temporally and spatially uncorrelated white Gaussian noises. As some of the aforementioned DOA estimation methods require the prior information of the incident signal number, while the others are able to achieve joint detection and estimation, we assume that such information is available for all of them when comparing their DOA estimation precision to make the comparison more fair.

In the following simulations, the input $\mathrm{SNR} = 10\log_{10}(\sigma_k^2/\sigma_n^2)$, where σ_k^2 is the power of the kth signal and all signals are assumed to have identical powers. Monte Carlo simulation with 500 runs are performed for each parameter that is used for comparison, and the corresponding RMSE is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{500K}\sum_{t=1}^{500}\sum_{k=1}^{K}\left(\hat{\theta}_{k,t}-\theta_k\right)^2} \tag{52}$$

where $\hat{\theta}_{k,t}$ is the estimate of DOA of the kth signal in the tth Monte Carlo trial. Additionally, in the sparsity-inducing DOA estimators, the $[-90^\circ\ 90^\circ]$ space is sampled uniformly with interval $\Delta\theta = 1^\circ$ to obtain the direction set Θ, and the proposed algorithm realizes refined DOA estimation via directly calculating the unknown grid mismatches based on the reconstructed spectrum, while the adaptive grid refinement method in [15] is adopted for the other methods to alleviate the limitation of the direction grid: use a rough grid (1° resolution) for initialization and then form a refined grid with refinement factor of 3 around K dominant peaks for the next iteration (5 iterations are needed). Nonetheless, as a small neighborhood around the DOA estimates derived from the previous iteration is retained for the upcoming construction, the total computational complexity of these methods increase slightly due to the grid refinement process. All experiments are carried out in MATLAB v.2010a on a PC with a Windows 7 system and a 2.70 GHz dual-core CPU.

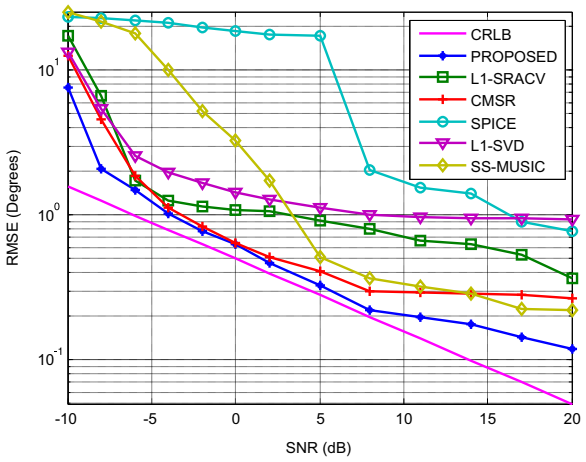

Fig. 2. RMSE versus SNR for two equal-power independent sources in the overdetermined case, with 500 snapshots are collected.

4.1. Overdetermined DOA estimation

In this Subsection, we consider a nested array with 4 physical sensors to perform DOA estimation. The first layer of sensors is located at positions $[0, d]$, and the second layer of sensors is located at positions $[2d, 5d]$, with d taken as half of the wavelength. Suppose that two equal-power independent narrowband signals impinge onto this nested array from directions of $-4^\circ + \delta\theta$ and $4^\circ + \delta\theta$, respectively, with $\delta\theta$ chosen randomly and uniformly within $[-0.5^\circ\ 0.5^\circ]$ in each trial to remove the possible prior information contained in the predefined direction set. In view of the requirement of noise variance in L1-SVD and L1-SRACV, the estimate of noise variance is given by the arithmetic mean of the $M-K$ smallest eigenvalues of $\hat{\mathbf{R}}_x$, provided that the source number K is known a priori. However, this requirement is not needed in our proposed approach. We fix the snapshot number at 500 and vary the SNR of each signal from -10 dB to 20 dB, then the RMSE of six methods, including L1-SRACV, CMSR, SPICE, L1-SVD, SS-MUSIC and the proposed method, are given in Fig. 2.

The simulation results in Fig. 2 indicate that, excluding SPICE, the other four sparsity-inducing methods succeed in separating the two sources at lower SNR, but among them, L1-SVD and L1-SRACV have larger DOA estimation RMSE than the subspace-based method S-S-MUSIC when the SNR is higher than 3 dB, and the SNR threshold for SS-MUSIC to achieve a higher DOA estimation precision than CMSR is 14 dB. This superior resolution performance of sparse recovery applied in low SNR regime is natural, which has been widely shown in the literature [13–20,24–26]. Meanwhile, the proposed method owns the best estimation performance w.r.t. variation of SNR. Interpretation of this outperformance is twofold. One aspect is that our proposed method entirely uses the extended aperture of the sparse array, while L1-SVD and L1-SRACV only use the raw array output data. The other aspect is that,

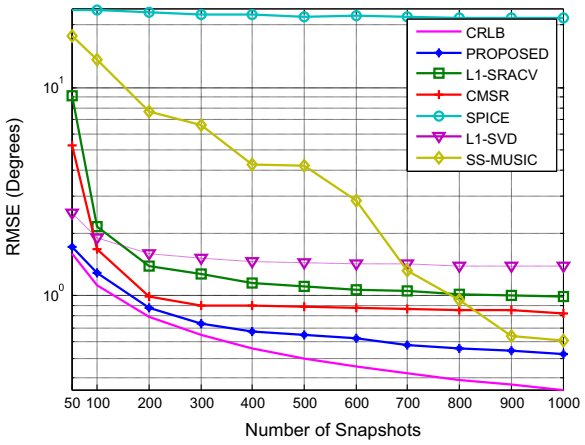

Fig. 3. RMSE versus number of snapshots for two equal-power independent sources in the overdetermined case, with SNR=0 dB.

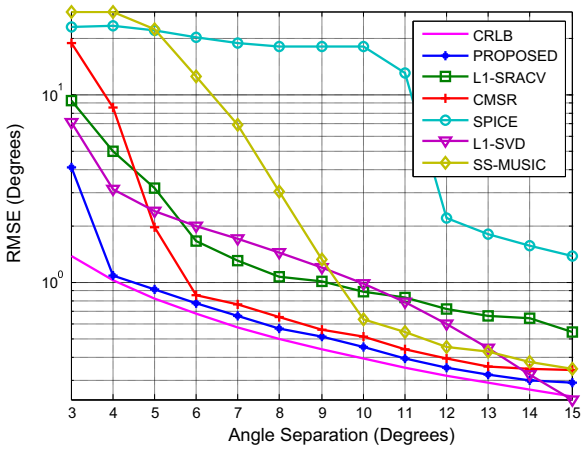

Fig. 4. RMSE versus angle separation of two equal-power independent sources in the overdetermined case, with SNR=0 dB and 500 snapshots are collected.

Table 2
Average computation time comparison of various algorithms in the overdetermined case (seconds).

	SS-MUSIC	L1-SVD	SPICE	PROPOSED	CMSR	L1-SRACV
$N=50$	0.003	0.598	0.101	0.114	0.462	0.686
$N=300$	0.003	0.601	0.107	0.121	0.465	0.713
$N=500$	0.004	0.608	0.109	0.123	0.471	0.735
$N=700$	0.005	0.647	0.112	0.145	0.474	0.740
$N=1000$	0.005	0.700	0.144	0.249	0.482	0.743

compared with other on-grid based sparsity-driven methods, which employ the ℓ_1-norm instead of ℓ_0-norm to enforce sparsity and have a poor performance on describing the true observation model, our proposed approach utilizes a theoretically more promising probabilistic hierarchical prior to approximate the ℓ_0-norm and overcomes the basis mismatch effect to a large extent by the off-grid model used in this paper. It is also observed from Fig. 2 that the DOA estimation precision of SPICE is hardly able to attain a satisfying level even at high SNRs, which shows that its resolution performance is apt to be influenced by the scarcity of angle separation between sources.

In the second experiment, we keep the SNR at 0 dB, and increase the number of snapshots from 50 to 1000. The remaining parameters of the experiment are the same as before. Fig. 3 depicts RMSE versus the number of snapshots for the same six methods, and their average computation time are given in Table 2. The results in Fig. 3 indicate that CMSR and the proposed algorithm require only 200 snapshots to achieve super-resolution, and the proposed method owns smaller RMSE than CMSR all through. SS-MUSIC fails to separate the two signals until as many as 800 snapshots are collected, and its DOA estimation precision is still worse than the proposed method above this threshold. As the snapshot number increases, the DOA estimation precision of L1-SVD and L1-SRACV improves only slightly and L1-SRACV surpasses L1-SVD with a small margin. However, these two methods provide somewhat inferior performance than CMSR and our proposed approach, mainly due to the underestimated estimate of the noise variance. Also we notice that SPICE is not able to separate the two signals in the considered scenarios even as many as 1000 snapshots are collected. One can then conclude from Table 2 that, the six algorithms witness a monotonous increase in their average computation time with increasing snapshot number. Moreover, excluding SPICE method, our proposed technique gains significant predominance over state-of-the-art sparsity-inducing methods in computational efficiency as the number of snapshots gets larger. Although SS-MUSIC has the highest computational efficiency among the six methods, but this makes no sense as it performs much worse than the other methods in small snapshots adaptation according to Fig. 3.

Finally, in order to compare the super-resolution of those aforementioned methods, we fix the snapshot number at 500 and the SNR of the two incident signals at 0 dB, the directions of the two signals are $-|\theta_2-\theta_1|/2+\delta\theta$ and $|\theta_2-\theta_1|/2+\delta\theta$, respectively, where $\delta\theta$ is defined in the same way as that in the previous simulations. When the angle separation of the two signals, i.e., $|\theta_2-\theta_1|$, varies from 3° to 15°, the RMSE of the six DOA estimators are given in Fig. 4. The results shown in Fig. 4 demonstrate the superiority of the sparsity-inducing methods over SS-MUSIC in separating spatially adjacent signals. Note that the angle separation threshold of the proposed method to achieve super-resolution is 4°, which is much smaller than the 6°, 8°, 9° and 10° thresholds of CMSR, L1-SRACV, L1-SVD and SS-MUSIC, respectively. The DOA estimation precision of the proposed method exceeds the other algorithms in almost all the angle separations considered, with rare exception happen when two signals are located farther away than 14°, that is, L1-SVD surpasses the proposed method in DOA estimation precision when the angle separation is larger than 14°. The reason for such exception is that, only one measurement vector is available in the proposed method, thus the local minima in its objective function cannot be smoothed away effectively as that in the multi-measurement case [23], and its RMSE departs from the CRLB even when the angular separation is as large as 15°. Moreover, it can also be seen that SPICE fails completely in the

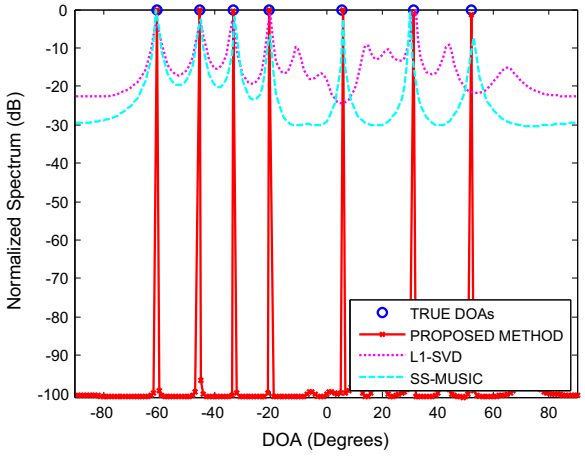

Fig. 5. Normalized spatial spectra for L1-SVD, SS-MUSIC and the proposed method, with SNR=0 dB and 100 snapshots are collected.

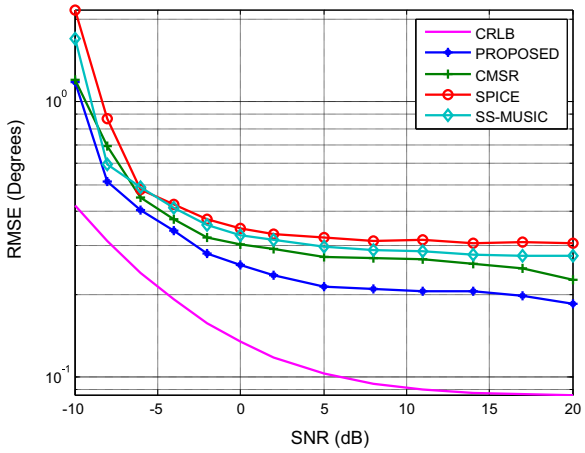

Fig. 6. RMSE versus SNR for seven equal-power independent sources in the underdetermined case, with 500 snapshots are collected.

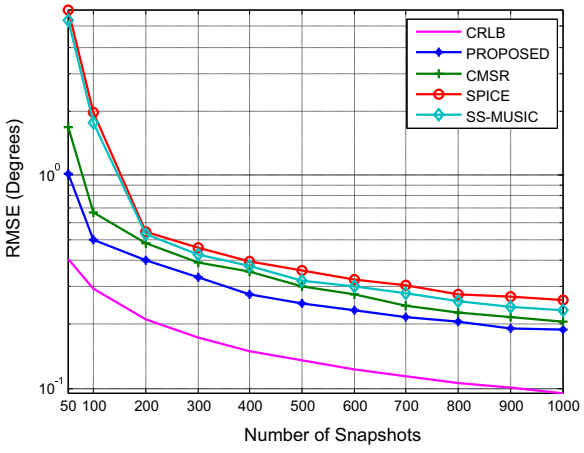

Fig. 7. RMSE versus number of snapshots for seven equal-power independent sources in the underdetermined case, with SNR=0 dB.

demanding scenarios.

4.2. Underdetermined DOA estimation

In order to demonstrate the ability of the proposed method in separating more sources than the number of physical sensors, in what follows, we investigate the performance of the proposed method for the underdetermined DOA estimation environment. To this end, we suppose that seven narrowband stationary Gaussian signals with equal powers impinge onto a 6-element nested array simultaneously, the array elements are located at 0, 1, 2, 3, 7 and 11 times the half-wavelength of the narrowband signals, and the following five scenarios are considered.

In the first experiment of this section, we intend to give a brief description of the spectra of the proposed method, and compare the result with that of the SS-MUSIC and L1-SVD algorithms. Suppose that seven 0 dB signals located off the grids at the randomly generated angular positions $-60.7°$, $-45.4°$, $-33.3°$, $-20.6°$, $5.8°$, $31.5°$ and $52.2°$. The number of snapshots is set to be 100. Fig. 5 depicts the spectra of those aforementioned three methods in a randomly chosen trial. The simulation result implies that L1-SVD misses two targets from the directions of $5.8°$ and $52.2°$. This is due to the fact that for L1-SVD, resolving five sources by using this 6-element array arrives at its extreme in resolution, which may be an aspect of why it cannot give excellent estimates like the other simulated methods that employ aperture extension in simulation of Fig. 5. On the contrary, SS-MUSIC and the proposed technique succeed to localize all seven sources which is more than the number of physical sensors, but some of the signal peaks of SS-MUSIC are biased and have very small amplitudes. This confirms the superior resolution ability of our algorithm than the SS-MUSIC counterpart.

Next, we test the estimation accuracy of our proposed algorithm and other existing methods via Monte Carlo simulations. Note that L1-SVD and L1-SRACV are not simulated here as resolving more sources than the number of sensor elements is beyond their resolution capabilities. In this experiment, we fix the snapshot number at 500, and the signal directions are $-57° + \delta\theta$, $-44° + \delta\theta$, $-28° + \delta\theta$, $3° + \delta\theta$, $12° + \delta\theta$, $26° + \delta\theta$ and $41° + \delta\theta$, respectively, with $\delta\theta$ set in the same way as that in the simulations corresponding to Fig. 2. The SNR of each signal is increased from -10 dB to 20 dB, and the RMSE of the DOA estimates of the seven sources are given in Fig. 6. From Fig. 6, we see that our technique still has a similar performance to that of the example shown in Fig. 2, and can still yield more competitive estimate than others, which becomes readily apparent at high SNRs. The other methods, however, provide a larger error. One thing to note is that, the RMSE curves of the proposed method does not approach the CRLB well with the increment of SNR. This is because the assumption of Gaussian distribution to $\bar{\mathbf{p}}$ is deviated from the actual signal model, thus great efforts should be made to work out a way to modify the likelihood model of SBL strategies by taking the intra-correlation of $\bar{\mathbf{p}}$ into account, such a work is in progress by the same authors of this paper.

Table 3
Average computation time comparison of various algorithms in the underdetermined case (seconds).

	SS-MUSIC	PROPOSED	SPICE	CMSR
$N=50$	0.005	0.586	0.311	1.213
$N=300$	0.006	0.598	0.379	1.219
$N=500$	0.008	0.627	0.385	1.227
$N=700$	0.009	0.642	0.419	1.251
$N=1000$	0.012	0.703	0.432	1.255

Experiment 3 studies the performance of the four DOA estimators w.r.t. the snapshot number. We repeat the previous experiment but set SNR=0 dB and vary the number of the collected snapshots from 50 to 1000. The simulation results are depicted in Fig. 7. Similar to Table 2, we tabulate the computation burdens of the four methods related to this experiment in Table 3. Remarkably, from Fig. 7, the proposed approach performs the best over all the range of snapshot number values. The SS-MUSIC algorithm is somewhat inferior to the proposed method mainly because the spatial smoothing processing will lead to aperture loss, and CMSR owns a middle estimation performance between them. Also we observe that SPICE provide relatively poor performance mainly due to the substantial bias for closely spaced sources. This experiment again supports the predominance of our proposed method over the other state-of-the-art counterparts in DOA estimation precision. In addition, as can be seen in Table 3, SS-MUSIC is fastest among all, while CMSR turns out to be the slowest. For the listed algorithms, their average run times increase with increasing snapshot number. Meanwhile, the proposed method needs much more time to complete a single simulation than SPICE all along. However, the computational superiority of SS-MUSIC and SPICE becomes secondary when taking their inferior performance at small snapshots into consideration. So in situations where speed of execution and accuracy of estimation are some concerns, the proposed method might be the preferred choice.

Finally, we close this Subsection by examining the impact of SNR or the size of data samples on the probability of the four

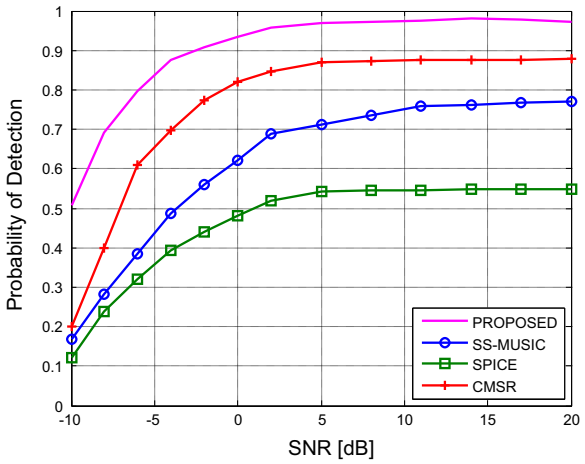

Fig. 8. Detection probability versus SNR for seven equal-power independent sources in the underdetermined case, with 500 snapshots are collected.

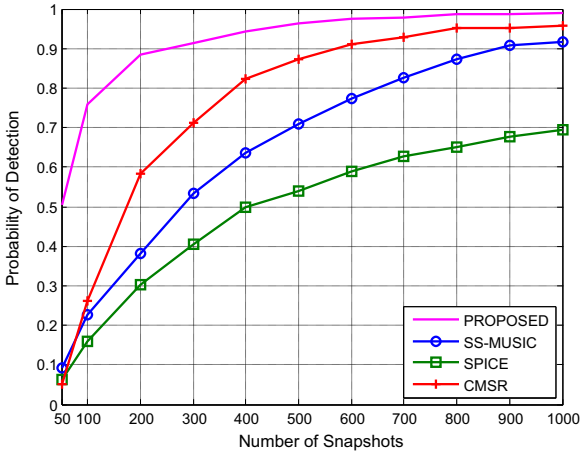

Fig. 9. Detection probability versus number of snapshots for seven equal-power independent sources in the underdetermined case, with SNR=5 dB.

algorithms to detect K sources within φ_θ degrees from the true DOA, i.e., $p_d \triangleq p\{\left|\theta_i - \hat{\theta}_i\right| < \varphi_\theta, \forall\, i\}$ and φ_θ is half of the distance between the two closely-spaced DOAs. In the remaining two experiments, we repeated the simulations with a 6-element nested array. Here, seven uncorrelated incident signals are simulated with directions $-40^\circ + \delta\theta$, $-30^\circ + \delta\theta$, $-20^\circ + \delta\theta$, $10^\circ + \delta\theta$, $14^\circ + \delta\theta$, $30^\circ + \delta\theta$ and $40^\circ + \delta\theta$, and $\delta\theta$ is randomly chosen as that in the above three experiments. Accordingly, φ_θ equals 2° by its definition. For each value of SNR or snapshot number, the probability estimates are computed by averaging over 500 Monte Carlo runs. First, we keep the number of snapshots fixed at 500 and vary the SNR from -10 dB to 20 dB. The corresponding results are shown in Fig. 8. From this figure, it is obvious that the detection performance of respective algorithm improves dramatically with increasing SNR and approaches a consistent level asymptotically. More specifically, the incorrect detection rate of CMSR falls below 15% when the SNR is higher than 5 dB, which is much lower than the 25% incorrect detection rate of SS-MUSIC. The detection probability of SPICE is significantly inferior to that of CMSR and SS-MUSIC all along, and it achieves only 55% when the SNR is higher than 5 dB. Fortunately, the proposed algorithm can provide nearly 100% correct detection for SNR $\geq$ 5 dB, and the other three algorithms have poorer performances than ours at all the SNR values. This is evidently because correct detection relies heavily on the primary representation result of overcomplete basis: good representation leads to correct detection, while inferior representation leads to false detection. Therefore, by taking a mathematically more precise off-grid model into consideration, our method yields a superior detection capability. A similar plot depicting the impact of the number of snapshots on the detection probabilities of various algorithms is shown in Fig. 9, where we fix the SNR at 5 dB and vary the snapshot number from 50 to 1000. Other experimental parameters are the same as those in Fig. 8. In this case, we notice that our proposed method outperforms the other approaches by a considerable margin again.

5. Concluding remarks and future work

In this paper, we have extended the mathematical theory of sparse support recovery to DOA estimation using nested array. For nested array, a single snapshot condition arises in the process of formulating the virtual aperture with extended steering vectors. A new algorithm from a SBL perspective was proposed to estimate the DOAs of off-grid targets based on the difference co-array with a single snapshot, thereby increasing the DOF from $O(M)$ to $O(M^2)$. We showed that existing sparsity-promoting algorithms have poor performance when grid mismatch is present, and thus they have limited ability in practice. To solve this problem, we approximated the true measurement matrix as a summation of a presumed matrix and a structured parameterized matrix via the second-order Taylor expansion, therefore alleviating the perils of basis mismatch. Based on this novel off-grid model and the statistical property of sample covariance matrix perturbation, we proposed a SBL framework, which allows for easily eliminating the unknown noise variances and normalizing the covariance matrix estimation error by imposing a linear transformation. After recovering the sparse support sets by retrieving the peak locations of the reconstructed spatial spectrum, a low complexity one-dimensional root finding procedure was introduced to calculate the unknown off-grid distances one by one. Via numerical simulation experiments, we have verified that, without requiring the number of signals, the proposed method has much enhanced adaptation than many well-assessed state-of-the-art algorithms to demanding scenarios with low SNR, limited snapshots and spatially adjacent signals, it also obtains considerably accurate DOA estimates. Meanwhile, the developed approach also makes good use of the nested array geometry to enhance its ability in separating more simultaneous signals than the number of physical sensors.

Future researches, currently under investigation and out-of-the-scope of this paper, will concentrate on performing the proposed strategies in terms of planar (two-dimensional) nested array configuration. It is also important to point out that, one major assumption made in current research on nested array is that the sources, which have the same variance, are uncorrelated. Thus incorporating correlations among sources and extending our algorithm to the nonuniform signal environment (i.e., strong sources and weak sources coexist) are some important topics for future work. In a nutshell, this paper provides a first step towards explicit use of correlation priors in sparse recovery which can lead to many future directions for correlation-aware sparse recovery techniques.

Acknowledgments

This research is supported partially by the National Natural Science Foundation of China (Nos. 61231027, 61271292 and 61431016) and the National Basic Research Program of China (973 program) (No. 2011CB707001).

Appendix

Derivation of (32)

Denote $\mathbf{\Delta} = \mathrm{diag}(\boldsymbol{\beta})$ and $\mathbf{\Lambda} = \mathrm{diag}(\boldsymbol{\alpha})$. Eq. (32) is based on the following two equalities

$$\begin{aligned}\|\tilde{\mathbf{z}} - \left(\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\mathbf{\Delta} + \tilde{\mathbf{C}}\mathbf{\Lambda}\right)\boldsymbol{\mu}\|_2^2 &= \|\left(\tilde{\mathbf{z}} - \tilde{\mathbf{A}}\boldsymbol{\mu}\right) - \tilde{\mathbf{B}}\mathrm{diag}(\boldsymbol{\mu})\boldsymbol{\beta} - \tilde{\mathbf{C}}\mathrm{diag}(\boldsymbol{\mu})\boldsymbol{\alpha}\|_2^2 \\ &= \boldsymbol{\beta}^T\left[\left(\tilde{\mathbf{B}}^H\tilde{\mathbf{B}}\right)^* \oplus \left(\boldsymbol{\mu}\boldsymbol{\mu}^H\right)\right]\boldsymbol{\beta} \\ &\quad - 2\mathrm{Re}\left\{\mathrm{diag}(\boldsymbol{\mu}^*)\tilde{\mathbf{B}}\left(\tilde{\mathbf{z}} - \tilde{\mathbf{A}}\boldsymbol{\mu}\right)\right\}^T\boldsymbol{\beta} \\ &\quad - 2\mathrm{Re}\left\{\mathrm{diag}(\boldsymbol{\mu}^*)\tilde{\mathbf{C}}^H\left(\tilde{\mathbf{z}} - \tilde{\mathbf{A}}\boldsymbol{\mu}\right)\right\}^T\boldsymbol{\alpha} \\ &\quad + 2\mathrm{Re}\left\{\boldsymbol{\beta}^T\left[\left(\tilde{\mathbf{B}}^H\tilde{\mathbf{C}}\right)^* \oplus \left(\boldsymbol{\mu}\boldsymbol{\mu}^H\right)\right]\boldsymbol{\alpha}\right\} \\ &\quad + \boldsymbol{\alpha}^T\left[\left(\tilde{\mathbf{C}}^H\tilde{\mathbf{C}}\right)^* \oplus \left(\boldsymbol{\mu}\boldsymbol{\mu}^H\right)\right]\boldsymbol{\alpha} + c_1\end{aligned} \tag{53}$$

$$\begin{aligned}&\mathrm{Tr}\left\{\left(\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\mathbf{\Delta} + \tilde{\mathbf{C}}\mathbf{\Lambda}\right)\mathbf{\Sigma}_{\bar{\mathbf{p}}}\left(\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\mathbf{\Delta} + \tilde{\mathbf{C}}\mathbf{\Lambda}\right)^H\right\} = \\ &\qquad 2\mathrm{Re}\left\{\mathrm{Tr}\left(\tilde{\mathbf{B}}^H\tilde{\mathbf{A}}\mathbf{\Sigma}_{\bar{\mathbf{p}}}\mathbf{\Delta}\right)\right\} \\ &\qquad + \mathrm{Tr}\left\{\tilde{\mathbf{B}}\mathbf{\Delta}\mathbf{\Sigma}_{\bar{\mathbf{p}}}\mathbf{\Delta}^H\tilde{\mathbf{B}}^H\right\} \\ &\qquad + 2\mathrm{Re}\left\{\mathrm{Tr}\left(\tilde{\mathbf{C}}^H\tilde{\mathbf{A}}\mathbf{\Sigma}_{\bar{\mathbf{p}}}\mathbf{\Lambda}\right)\right\} \\ &\qquad + 2\mathrm{Re}\left\{\mathrm{Tr}\left(\tilde{\mathbf{B}}\mathbf{\Delta}\mathbf{\Sigma}_{\bar{\mathbf{p}}}\mathbf{\Lambda}\tilde{\mathbf{C}}^H\right)\right\} \\ &\qquad + \mathrm{Tr}\left\{\tilde{\mathbf{C}}\mathbf{\Lambda}\mathbf{\Sigma}_{\bar{\mathbf{p}}}\mathbf{\Lambda}\tilde{\mathbf{C}}^H\right\} + c_2 \\ &\qquad = 2\mathrm{Re}\left\{\mathrm{diag}\left(\tilde{\mathbf{B}}^H\tilde{\mathbf{A}}\mathbf{\Sigma}_{\bar{\mathbf{p}}}\right)\right\}^T \\ &\qquad \boldsymbol{\beta} + \boldsymbol{\beta}^T\left[\mathbf{\Sigma}_{\bar{\mathbf{p}}} \oplus \left(\tilde{\mathbf{B}}^H\tilde{\mathbf{B}}\right)^*\right]\boldsymbol{\beta} \\ &\qquad + 2\mathrm{Re}\left\{\mathrm{diag}\left(\tilde{\mathbf{C}}^H\tilde{\mathbf{A}}\mathbf{\Sigma}_{\bar{\mathbf{p}}}\right)\right\}^T\boldsymbol{\alpha} \\ &\qquad + 2\mathrm{Re}\left\{\boldsymbol{\beta}^T\left[\mathbf{\Sigma}_{\bar{\mathbf{p}}} \oplus \left(\tilde{\mathbf{C}}^H\tilde{\mathbf{B}}\right)^*\right]\boldsymbol{\alpha}\right\} \\ &\qquad + \boldsymbol{\alpha}^T\left[\mathbf{\Sigma}_{\bar{\mathbf{p}}} \oplus \left(\tilde{\mathbf{C}}^H\tilde{\mathbf{C}}\right)^*\right]\boldsymbol{\alpha} + c_2\end{aligned} \tag{54}$$

where c_1, c_2 are constants independent of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, and the equality

$$\mathrm{Tr}\left\{\mathrm{diag}^H(\mathbf{u})\mathbf{Q}\mathrm{diag}(\mathbf{v})\mathbf{R}^T\right\} = \mathbf{u}^H(\mathbf{Q} \oplus \mathbf{R})\mathbf{v} \tag{55}$$

for vectors $\mathbf{u}$, $\mathbf{v}$ and matrices $\mathbf{Q}$, $\mathbf{R}$ with proper dimensions is used. Note that $\boldsymbol{\beta}^T\mathbf{S}\boldsymbol{\beta} \in \mathbb{R}$ for a positive semi-definite matrix $\mathbf{S}$ with proper dimension and thus $\boldsymbol{\beta}^T\mathbf{S}\boldsymbol{\beta} = \mathrm{Re}(\boldsymbol{\beta}^T\mathbf{S}\boldsymbol{\beta}) = \boldsymbol{\beta}^T\mathrm{Re}(\mathbf{S})\boldsymbol{\beta}$ since $\boldsymbol{\beta}$ is real-valued. Then (32) is obtained by observing that $(\tilde{\mathbf{B}}^H\tilde{\mathbf{B}})^* \oplus (\boldsymbol{\mu}\boldsymbol{\mu}^H)$, $(\tilde{\mathbf{C}}^H\tilde{\mathbf{C}})^* \oplus (\boldsymbol{\mu}\boldsymbol{\mu}^H)$, $\mathbf{\Sigma}_{\bar{\mathbf{p}}} \oplus (\tilde{\mathbf{B}}^H\tilde{\mathbf{B}})^*$ and $\mathbf{\Sigma}_{\bar{\mathbf{p}}} \oplus (\tilde{\mathbf{C}}^H\tilde{\mathbf{C}})^*$ are all positive semi-definite.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.sigpro.2016.03.024.

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] S. Min, D. Seo, K.B. Lee, H.M. Kwon, Y.H. Lee, Direction-of-arrival tracking scheme for DS/CDMA systems: direction lock loop, IEEE Trans. Wirel. Commun. 3 (1) (2004) 191–202.
[3] P. Stoica, R.L. Moses, Introduction to Spectral Analysis, Prentice-Hall, Upper Saddle River, NJ, 1997.
[4] X. Huang, Y.J. Guo, J.D. Bunton, A hybrid adaptive antenna array, IEEE Trans. Wirel. Commun 9 (5) (2010) 1770–1779.
[5] A.B. Gershman, M. Rübsamen, M. Pesavento, One-and two-dimensional direction-of-arrival estimation: an overview of search-free techniques, Signal Process 90 (5) (2010) 1338–1349.
[6] Y. Wu, C. Hou, G. Liao, Q. Guo, Direction-of-arrival estimation in the presence of unknown nonuniform noise fields, IEEE J. Ocean. Eng. 31 (2) (2006) 504–510.
[7] R.T. Hoctor, S. Kassam, The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging, Proc. IEEE. 78 (4) (1990) 735–752.
[8] H.L. Van Trees, Optimum Array Processing: Part IV of Detection, Estimation and Modulation Theory, John Wiley and Sons, NY, 2004.
[9] A.T. Moffet, Minimum-redundancy linear arrays, IEEE Trans. Antennas Propag 16 (2) (1968) 172–175.
[10] P. Pal, P.P. Vaidyanathan, Nested arrays: a novel approach to array processing with enhanced degrees of freedom, IEEE Trans. Signal Process. 58 (8) (2010) 4167–4181.
[11] R.O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (3) (1986) 276–280.
[12] R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust. Speech Signal Process. 37 (7) (1989) 984–995.
[13] M.S. Lee, Y.H. Kim, Robust L1-norm beamforming for phased array with antenna switching, IEEE Commun. Lett. 12 (8) (2008) 566–568.
[14] I.F. Gorodnitsky, B.D. Rao, Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm, IEEE Trans. Signal Process. 45 (3) (1997) 600–616.
[15] D. Malioutov, M. Çetin, A.S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (2005) 3010–3022.
[16] Z. Liu, Z. Huang, Y. Zhou, Direction-of-arrival estimation of wideband signals via covariance matrix sparse representation, IEEE Trans. Signal Process. 59 (9) (2011) 4256–4270.
[17] J. Yin, T. Chen, Direction-of-arrival estimation using a sparse representation of array covariance vectors, IEEE Trans. Signal Process. 59 (9) (2011) 4489–4493.
[18] P. Stoica, P. Babu, SPICE and LIKES: two hyperparameter-free methods for sparse-parameter estimation, Signal Process. 92 (7) (2012) 1580–1590.
[19] P. Stoica, D. Zachariah, J. Li, Weighted SPICE: a unifying approach for hyperparameter-free sparse estimation, Digit. Signal Process. 33 (2014) 1–12.
[20] Z. Liu, Z. Huang, Y. Zhou, Array signal processing via sparsity-inducing representation of the array covariance matrix, IEEE Trans. Aerosp. Electron. Syst. 49 (3) (2013) 1710–1724.
[21] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (2001) 211–244.
[22] D.P. Wipf, B.D. Rao, Sparse Bayesian learning for basis selection, IEEE Trans. Signal Process. 52 (8) (2004) 2153–2164.
[23] D.P. Wipf, B.D. Rao, An empirical Bayesian strategy for solving the simultaneous sparse approximation problem, IEEE Trans. Signal Process. 55 (7) (2007) 3704–3716.
[24] Z. Liu, Z. Huang, Y. Zhou, An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning, IEEE Trans. Wirel. Commun. 11 (10) (2012) 1–11.
[25] Z. Liu, Z. Huang, Y. Zhou, Sparsity-inducing direction finding for narrowband and wideband signals based on array covariance vectors, IEEE Trans. Wireless Commun. 12 (8) (2013) 1–12.
[26] Z. Liu, 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.
[27] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Trans. Signal Process. 56 (6) (2008) 2346–2356.
[28] D.P. Wipf, B.D. Rao, S. Nagarajan, Latent variable Bayesian models for promoting sparsity, IEEE Trans. Inf. Theory 57 (9) (2011) 6236–6255.
[29] L. He, L. Carin, Exploiting structure in wavelet-based Bayesian compressive sensing, IEEE Trans. Signal Process 57 (9) (2009) 3488–3497.
[30] Y. Chi, L.L. Scharf, A. Pezeshki, Sensitivity to basis mismatch in compressed sensing, IEEE Trans. Signal Process. 59 (5) (2011) 2182–2195.
[31] C.D. Austin, R.L. Moses, J.N. Ash, E. Ertin, On the relation between sparse reconstruction and parameter estimation with model order selection, IEEE J. Sel. Top. Signal Process. 4 (3) (2010) 560–570.
[32] P. Stoica, P. Babu, Sparse estimation of spectral lines: Grid selection problems and their solutions, IEEE Trans. Signal Process. 60 (2) (2012) 962–967.
[33] A. Yeredor, Non-orthogonal joint diagonalization in the least-squares sense

with application in blind source separation, IEEE Trans. Signal Process. 50 (7) (2002) 1545–1553.
[34] B. Ottersten, P. Stoica, R. Roy, Covariance matching estimation techniques for array signal processing applications, Digit. Signal Process. 8 (3) (1998) 185–210.
[35] A.J. Laub, Matrix Analysis for Scientists and Engineers, SIAM, Philadephia, PA, USA, 2005.
[36] P. Stoica, N. Arye, MUSIC, maximum likelihood, and Cramer-Rao bound, IEEE Trans. Acoust. Speech Signal Process. 37 (5) (1989) 720–741.
[37] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. B (1977) 1–38.
[38] 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.
[39] D.L. Donoho, M. Elad, Optimally sparse representation in general (non-orthogonal) dictionaries via $\ell 1$ minimization, Proc. Nat. Acad. Sci. 100 (5) (2003) 2197–2202.