Signal Processing 142 (2018) 87–95

Contents lists available at ScienceDirect

Signal Processing

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

Two sparse-based methods for off-grid direction-of-arrival estimation

Xiaohuan Wu[a,*], Wei-Ping Zhu[a,b], Jun Yan[a], Zeyun Zhang[a]

[a] The Key Lab of Broadband Wireless Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

[b] Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada

ARTICLE INFO

Article history:
Received 26 April 2017
Revised 15 June 2017
Accepted 7 July 2017
Available online 8 July 2017

Keywords:
Direction-of-arrival (DOA) estimation
Off-grid model
Sparse-based method
Covariance matrix

ABSTRACT

Recently, many sparse-based methods have been proposed for direction-of-arrival (DOA) estimation. However, these methods often suffer from the grid mismatch problem caused by the discretization of the potential angle space. Most of them employ the iterative grid refinement (IGR) method to alleviate this problem. Nevertheless, IGR requires a high computational load and may not comply with the restricted isometry property (RIP) condition in the compressed sensing (CS) theory. This paper aims to overcome the grid mismatch limitation inherent in conventional sparse-based techniques. In particular, we first introduce an off-grid model by incorporating the bias parameter into the signal model, then propose a two-step iterative method named off-grid ℓ_1 Cholesky covariance decomposition (OGL1CCD) to solve the DOA estimation problem. Our method can be accelerated to save computations and the proposed algorithm framework can be extended for any other sparse-based method to improve their estimation accuracy. We then propose another off-grid method named off-grid ℓ_1 covariance matrix reconstruction approach (OGL1CMRA) based on the covariance matrix model. Compared to OGL1CCD, OGL1CMRA is more computationally efficient and accurate, but requires sufficient snapshots and uncorrelated sources. Our proposed methods are superior to many other methods in estimation performance, which is verified by extensive numerical simulations.

1. Introduction

Direction finding, also known as direction-of-arrival (DOA) estimation, is a fundamental problem in array signal processing and has been intensively studied in the past few decades for various applications, e.g., radar [1], sonar and wireless communications [2]. Since the 1970s when Pisarenko found the DOAs can be estimated from the second order statistics of the received signal, a large number of methods have been proposed for DOA estimation, e.g., MUSIC [3], ESPRIT [4] and their variants [5–8]. Although these methods show super resolution ability in some certain scenarios, they often suffer from several limitations. For example, the so-called subspaced-based methods highly depend on the number of sources and the estimation accuracy of the covariance matrix of the array output and are, in general, sensitive to the number of snapshots and signal-to-noise ratio (SNR) as well. Furthermore, the correlation between the impinged signals may cause a rank deficiency problem in the sample covariance matrix, which may lead to a poor estimation performance.

Recently, by exploring the underlying connection between the sparse signal reconstruction (SSR) [9,10] and the DOA estimation, many sparse methods have been proposed by dividing the potential angle space with a predefined set of dense grid points and assuming the fact that the true DOAs exactly lie on the predefined grid [11–20]. With the help of SSR theory, these sparse methods are usually insensitive to the number of snapshots (most of them can correctly locate sources with a single snapshot) and the correlation between the sources. However, the set of grid points usually contains a finite number of atoms while the DOAs of the impinged sources are continuously valued, which leads to infinitely many atoms. The sparse methods then assume that the DOAs of the sources exactly lie on the grids and are also called on-grid method. This assumption, although widely employed by many sparse methods, holds true only when the number of grid points is infinitely large. Consequently, there always exists an unavoidable bias between the true DOA and its nearest grid point, namely, the bias mismatch issue is inherent in the on-grid methods. Using a set of dense grids is able to mitigate this basis mismatch issue but is not an appropriate choice since it not only increases the computational cost but also gives rises to the correlation issue. A high complexity may reduce the practicability of the algorithm while a high correlation may conflict with the restrict isometry property

* Corresponding author.
E-mail address: 2013010101@njupt.edu.cn (X. Wu).

http://dx.doi.org/10.1016/j.sigpro.2017.07.004
0165-1684/

(RIP) in the compressive sensing (CS) theory [21]. Therefore, almost all the sparse methods choose the iterative grid refinement (IGR) procedure to find a balance between the accuracy and the efficiency. The IGR uses an initial set of coarse grid points along with a sparse-based algorithm to obtain a coarse DOA estimate. It then refines the angle space division around the coarse DOA estimate with a new set of finer grid points and recalls the sparse-based algorithm to achieve a better DOA estimate. However, the use of a set of dense grid points in IGR would contradict the RIP condition required by the sparse algorithm. Hence, further improvement of the estimation performance of IGR is impossible.

To address the basis mismatch issue, some so-called off-grid sparse methods have been proposed [22–26]. Compared to the on-grid methods, the discretization is still required in these methods but the DOAs of the sources are not restricted to lie on the grid. Instead, the bias is parameterized into the signal model by using the first order approximation of the manifold matrix. But the new model usually is nonconvex and hence is not easy to solve. For example, the nonconvex model in [22,23] has been solved from the sparse Bayesian learning (SBL) perspective. In [24–26], an alternating procedure is adopted to solve the potential sources and biases. However, a difficult problem in [24] is the parameter tuning while the methods in [25,26] are time-consuming. In conclusion, the main focus of these methods has been to determine the biases. Recently, Tang et al. proposed a gridless method based on the atomic norm theory without requiring grid division [27] and showed that this method can be regarded as the sparse method with an infinite size of the grids, provided that the sources are sufficiently separated. Nevertheless, this precondition prohibits commonly known high resolution of the DOA estimation requirement. We proposed a gridless method named as covariance matrix reconstruction approach (CMRA) by exploiting the Toeplitz structure of the covariance matrix of the array output [28,29]. Since it explores the maximum number of degrees of freedom (DOF), CMRA is able to locate more sources than sensors.

In this paper, we address the DOA estimation problem in an off-grid mode under the sparse framework with an objective of aiming to overcome the grid mismatch limitation inherent in conventional sparse techniques. In particular, we start with a set of coarse grid points and then introduce a parameter to describe the bias between the true DOA and its nearest grid point. The unknown DOAs are determined by its nearest grid point together with the corresponding bias. We propose a two-step iterative technique for joint sparse signal recovery and bias estimation. In the first step, we fix the bias and obtain the peak indices of the recovered sparse signal, while in the second step, we fix the sparse signal and determine the bias by deriving a closed-form solution. To formulate exactly the problem, we propose two methods based on the output of the array and its covariance matrix, named as off-grid ℓ_1 Cholesky covariance decomposition (OGL1CCD) and off-grid ℓ_1 covariance matrix reconstruction approach (OGL1CMRA), respectively. In the first method, the complexity of the problem in step one is reduced by the dimensionality reduction technique. An update rule for the regularization parameter is also proposed. In the second method, we derive the cramer-rao lower bound (CRLB) of the bias to show that the covariance-based model is able to achieve more accurate bias estimates if sufficient snapshots are collected. We also point out that, the dimensionality reduction and the parameter updating are not required in the second method. Furthermore, the proposed algorithm can be extended for any other sparse methods to improve their estimation accuracy.

Notations used in this paper are as follows. $\boldsymbol{A}^*$, $\boldsymbol{A}^T$, $\boldsymbol{A}^H$ and $\boldsymbol{A}^\dagger$ denote the conjugate, transpose, conjugate transpose and pseudo-inverse of matrix $\boldsymbol{A}$, respectively. $\boldsymbol{A}_{(n)}$ denotes the nth row of $\boldsymbol{A}$. vec($\boldsymbol{A}$) denotes the vectorization operator that stacks matrix $\boldsymbol{A}$ column by column. $\boldsymbol{A}\circ\boldsymbol{B}$, $\boldsymbol{A}\otimes\boldsymbol{B}$ and $\boldsymbol{A}\odot\boldsymbol{B}$ are the Hadamard, Kronecker and Khatri-Rao products of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. tr($\bullet$) is the trace operator. $\boldsymbol{I}_N$ denotes the identity matrix of size $N\times N$. $\|\boldsymbol{A}\|_1$ and $\|\boldsymbol{A}\|_F$ denote the ℓ_1-norm and Frobenius norm of $\boldsymbol{A}$, respectively. $\boldsymbol{A}\succeq\boldsymbol{0}$ means that matrix $\boldsymbol{A}$ is positive semidefinite. For a vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|_2$ denotes the ℓ_2-norm of $\boldsymbol{x}$. $\boldsymbol{x}\succeq\boldsymbol{0}$ means that every entry of $\boldsymbol{x}$ is nonnegative, diag($\boldsymbol{x}$) denotes a diagonal matrix with its diagonal entries being the entries of vector $\boldsymbol{x}$ in turn.

The rest of the paper is organized as follows. Section 2 revisits the on-grid and off-grid models, followed by a CRLB analysis on the bias estimate. Section 3 presents the proposed iterative methods. A new covariance-based off-grid method is proposed in Section 4. Simulations are carried out in Section 5 to demonstrate the performance of our methods. Finally, Section 6 concludes the whole paper.

2. Off-grid signal model

2.1. On-grid signal model

Suppose that K narrowband far-field signals impinge onto an array with M omnidirectional sensors from directions of $\boldsymbol{\theta}=\{\theta_1,\cdots,\theta_K\}$ simultaneously. The array output at time t, which is corrupted by additive circular complex Gaussian white noise, can be expressed as,

$$\boldsymbol{x}(t)=\sum_{k=1}^{K}\boldsymbol{a}(\theta_k)s_k(t)+\boldsymbol{e}(t)=\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t)+\boldsymbol{e}(t), \tag{1}$$

where $\boldsymbol{x}(t)=[x_1(t),\cdots,x_M(t)]^T$ is the array output, $\boldsymbol{s}(t)=[s_1(t),\cdots,s_K(t)]^T$ is the signal waveform, $\boldsymbol{A}(\boldsymbol{\theta})=[\boldsymbol{a}(\theta_1),\cdots,\boldsymbol{a}(\theta_K)]$ is the array manifold matrix, $\boldsymbol{a}(\theta_k)=[e^{j2\pi f_0\tau_{k,1}},\cdots,e^{j2\pi f_0\tau_{k,M}}]^T$ contains the time-delay of the kth signal received at each sensor relative to the reference sensor, and $\boldsymbol{e}(t)$ is the complex independent white Gaussian noise with zero mean and variance $\sigma\boldsymbol{I}$.[1] When L snapshots are collected, the array output can be given by

$$\boldsymbol{X}=\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}+\boldsymbol{E}, \tag{2}$$

where $\boldsymbol{X}=[\boldsymbol{x}(t_1),\cdots,\boldsymbol{x}(t_L)]$, $\boldsymbol{S}=[\boldsymbol{s}(t_1),\cdots,\boldsymbol{s}(t_L)]$ and $\boldsymbol{E}=[\boldsymbol{e}(t_1),\cdots,\boldsymbol{e}(t_L)]$. The goal is to determine the unknown DOAs $\boldsymbol{\theta}$ given the array output $\boldsymbol{X}$.

Uniformly sampling the angle space generates a fixed finite set of N potential angles $\boldsymbol{\vartheta}=\{\vartheta_1,\cdots,\vartheta_N\}$ and a manifold dictionary $\bar{\boldsymbol{A}}=[\boldsymbol{a}(\vartheta_1),\cdots,\boldsymbol{a}(\vartheta_N)]$. Generally speaking, N is much greater than the incident sources number K and the sensors number M. We first assume $\boldsymbol{\theta}\subset\boldsymbol{\vartheta}$, i.e., all interested unknown DOAs exactly lie on the predefined grid. Thus, the array output in (2) can be reformulated as an multiple measurement vectors (MMV) model,

$$\boldsymbol{X}=\bar{\boldsymbol{A}}\bar{\boldsymbol{S}}+\boldsymbol{E}, \tag{3}$$

where $\bar{\boldsymbol{S}}$ is the extension of $\boldsymbol{S}$ from $\boldsymbol{\theta}$ to $\boldsymbol{\vartheta}$ with non-zero entries denoting the true sources locations. Since $N\gg K$, the signal $\bar{\boldsymbol{S}}$ is row-sparse. The problem of DOA estimation is now transformed into an SSR problem and the joint sparsity can be exploited for signal recovery. Usually, following the ℓ_1-norm optimization in a single measurement vector (SMV) problem, the $\ell_{2,1}$-norm, which is defined as $\|\boldsymbol{X}\|_{2,1}=\sum_n\|\boldsymbol{X}_{(n)}\|_2$, is expected to be an appropriate penalty. The problem is then formulated as,

$$\min_{\bar{\boldsymbol{S}}}\ \lambda\left\|\bar{\boldsymbol{S}}\right\|_{2,1}+\frac{1}{2}\left\|\boldsymbol{X}-\bar{\boldsymbol{A}}\bar{\boldsymbol{S}}\right\|_F^2, \tag{4}$$

where λ is the regularization parameter controlling the tradeoff between the sparsity of the solution and the data fitting error.

[1] We have assumed that the noise variances σ_m $(m=1,\cdots,M)$ are equivalent, which is true in most scenarios.

There exist many efficient algorithms to solve the MMV problem (4), and the DOAs can then be determined by the positions of the K largest peaks in the spectrum of $\bar{\boldsymbol{p}} = \mathrm{diag}(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^H)$.

2.2. Off-grid signal model

Now we consider the case when the DOAs of the sources are not on the grid, i.e., $\boldsymbol{\theta} \not\subset \boldsymbol{\vartheta}$. To circumvent this issue, it is natural to assume the overcomplete dictionary $\bar{\boldsymbol{A}}$ is known except that the parameter $\boldsymbol{\vartheta}$ can be treated unknown. Then the DOA estimation problem can be formulated as an SSR problem with an unknown parametric dictionary, which is shown as follows,

$$\min_{\bar{\boldsymbol{S}},\boldsymbol{\vartheta}} \ \lambda\left\|\bar{\boldsymbol{S}}\right\|_{2,1} + \frac{1}{2}\left\|\boldsymbol{X} - \sum_{n=1}^{N}\boldsymbol{a}(\vartheta_n)\boldsymbol{s}_n\right\|_F^2, \tag{5}$$

where $\boldsymbol{s}_n = [s_n(t_1), \cdots, s_n(t_L)] = \bar{\boldsymbol{S}}_{(n)}$. Unfortunately, model (5) is nonconvex, making it very difficult to obtain the global solution.

Denoting ϑ_{n_k} as the grid point nearest to the direction of the kth signal θ_k with $n_k \in \{1, \cdots, N\}$, the bias between the true position θ_k and ϑ_{n_k} may always exist, which can be denoted as $\delta_{n_k} = \theta_k - \vartheta_{n_k}$. The bias corresponding to the potential signal in each grid can be denoted as,

$$\delta_n = \begin{cases} \theta_k - \vartheta_{n_k} & \text{if } n = n_k \text{ for any } k \in \{1, \cdots, K\} \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

Letting r be the grid interval, obviously we have $-\frac{r}{2} \le \delta_n \le \frac{r}{2}$. Then the steering vector $\boldsymbol{a}(\theta_k)$ can be approximated using the first order approximation as,

$$\boldsymbol{a}(\theta_k) = \boldsymbol{a}(\vartheta_{n_k}) + \boldsymbol{b}(\vartheta_{n_k})\delta_{n_k}, \tag{7}$$

where $\boldsymbol{b}(\vartheta_{n_k})$ is the derivative of $\boldsymbol{a}(\boldsymbol{\vartheta})$ with respect to ϑ_{n_k}. By denoting $\bar{\boldsymbol{B}} = [\boldsymbol{b}(\vartheta_1), \cdots, \boldsymbol{b}(\vartheta_N)]$, $\boldsymbol{\Delta} = \mathrm{diag}(\bar{\boldsymbol{\delta}})$, $\bar{\boldsymbol{\delta}} = [\delta_1, \cdots, \delta_N]^T$, model (3) can be reformulated as

$$\boldsymbol{X} = (\bar{\boldsymbol{A}} + \bar{\boldsymbol{B}}\boldsymbol{\Delta})\bar{\boldsymbol{S}} + \boldsymbol{E}. \tag{8}$$

Further, problem (5) can be reformulated as

$$\min_{\bar{\boldsymbol{S}},\bar{\boldsymbol{\delta}}} \ \lambda\left\|\bar{\boldsymbol{S}}\right\|_{2,1} + \frac{1}{2}\left\|\boldsymbol{X} - (\bar{\boldsymbol{A}} + \bar{\boldsymbol{B}}\boldsymbol{\Delta})\bar{\boldsymbol{S}}\right\|_F^2. \tag{9}$$

Comparing problems (5) and (9), it can be seen that, instead of estimating the whole grid parameter $\boldsymbol{\vartheta}$ in (5), problem (9) solves the bias parameter $\bar{\boldsymbol{\delta}}$, which is restricted in a certain region and has the same sparse pattern as $\bar{\boldsymbol{S}}$. The structural information of $\bar{\boldsymbol{\delta}}$ can be exploited to enhance the estimation speed or accuracy of the algorithm. Except for some special cases, the problem (9) is nonconvex. Moreover, the regularization parameter λ, which is usually chosen proportional to the standard deviation of the noise, is also difficult to determine since the noise power is usually unknown.

We then derive the CRLB for the biases $\bar{\boldsymbol{\delta}}$ to theoretically evaluate the performance of model (8) from the Bayesian perspective. Mathematically, we have the following proposition,

Proposition 1. *For model (8), the CRLB of DOA estimation error at each grid satisfies,*

$$\mathrm{CRLB}(\delta_n) = \frac{3\sigma}{\pi^2\cos^2(\vartheta_n)M(M-1)(2M-1)p_n}, \quad n = 1, \cdots, N \tag{10}$$

where p_n denotes the power of the nth signal.

Proof. Please see Appendix A. □

We can conclude from Proposition 1 that (i) the estimation error variance of each bias is inversely proportional to SNR and M. Hence, the estimation performance can be improved by increasing the signal power or the number of sensors; (ii) the resolution varies from direction to direction, and the normal direction with $\theta = 0^\circ$ enjoys the highest resolution.

3. Proposed iterative algorithm

We now develop an iterative algorithm for joint sparse signal $\bar{\boldsymbol{S}}$ recovery and dictionary parameter $\bar{\boldsymbol{\delta}}$ learning. In particular, A two-step iterative technique is employed to solve the problem (9). In the first step, we use a sparse method to determine $\bar{\boldsymbol{S}}$ while keeping $\bar{\boldsymbol{\delta}}$ fixed,

$$\bar{\boldsymbol{S}}^{(q+1)} = \arg\min_{\bar{\boldsymbol{S}}} \ \lambda\left\|\bar{\boldsymbol{S}}\right\|_{2,1} + \frac{1}{2}\left\|\boldsymbol{X} - (\bar{\boldsymbol{A}} + \bar{\boldsymbol{B}}\boldsymbol{\Delta}^{(q)})\bar{\boldsymbol{S}}\right\|_F^2, \tag{11}$$

where the superscripts denote the algorithm iteration. The second step optimizes over $\bar{\boldsymbol{\delta}}$ while fixing $\bar{\boldsymbol{S}}$,

$$\bar{\boldsymbol{\delta}}^{(q+1)} = \arg\min_{\bar{\boldsymbol{\delta}}} \ \left\|\boldsymbol{X} - \bar{\boldsymbol{A}}\bar{\boldsymbol{S}}^{(q+1)} - \bar{\boldsymbol{B}}\boldsymbol{\Delta}\bar{\boldsymbol{S}}^{(q+1)}\right\|_F^2. \tag{12}$$

In the first step, the dimensionality of the problem depends on the number of snapshots, which could be very large. As a consequence, the computational workload can be significantly increased. A dimensionality reduction technique using singular value decomposition (SVD) was proposed in [11] to reduce the dimensionality from L to K. Mathematically, we compute the SVD of $\boldsymbol{X}$ as

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H. \tag{13}$$

Then let $\boldsymbol{X}_1 = \boldsymbol{X}\boldsymbol{V}\boldsymbol{D}_K^T$ and $\bar{\boldsymbol{S}}_1 = \bar{\boldsymbol{S}}\boldsymbol{V}\boldsymbol{D}_K^T$, where $\boldsymbol{D}_K \triangleq [\boldsymbol{I}_K, \boldsymbol{0}]$. Model (11) can be rewritten as,

$$\min_{\bar{\boldsymbol{S}}_1} \ \lambda\left\|\bar{\boldsymbol{S}}_1\right\|_{2,1} + \frac{1}{2}\left\|\boldsymbol{X}_1 - (\bar{\boldsymbol{A}} + \bar{\boldsymbol{B}}\boldsymbol{\Delta}^{(q)})\bar{\boldsymbol{S}}_1\right\|_F^2, \tag{14}$$

However, since the main purpose of model (14) is to keep only the K-dimensional signal subspace rather than the observation space, solving model (14) is slightly different from solving model (11). Furthermore, since the SVD procedure changes the data structure in the model, parameter tuning in (14) turns out to be difficult. An alternative dimensionality reduction technique was provided in [30], which is to keep both the signal and noise subspaces. This technique replaces $\boldsymbol{D}_K$ with $\boldsymbol{D}_M = [\boldsymbol{I}_M, \boldsymbol{0}]$ by following the idea of L1-SVD. Letting $\boldsymbol{X}_2 = \boldsymbol{X}\boldsymbol{V}\boldsymbol{D}_M^T$ and $\bar{\boldsymbol{S}}_2 = \bar{\boldsymbol{S}}\boldsymbol{V}\boldsymbol{D}_M^T$, the model (11) can be rewritten as,

$$\bar{\boldsymbol{S}}_2^{(q+1)} = \arg\min_{\bar{\boldsymbol{S}}_2} \ \lambda\left\|\bar{\boldsymbol{S}}_2\right\|_{2,1} + \frac{1}{2}\left\|\boldsymbol{X}_2 - (\bar{\boldsymbol{A}} + \bar{\boldsymbol{B}}\boldsymbol{\Delta}^{(q)})\bar{\boldsymbol{S}}_2\right\|_F^2, \tag{15}$$

Compared to the method in [11], the solution of (15) is equivalent to that of model (11) in the sense that the corresponding rows of the two solutions have the same power spectrum [30,31], which can be obtained as

$$\hat{\boldsymbol{p}}_2^{(q+1)} = \mathrm{diag}\left(\bar{\boldsymbol{S}}_2^{(q+1)}\bar{\boldsymbol{S}}_2^{(q+1)^H}\right). \tag{16}$$

Hence by solving (15) we may always find the same peak indices of the power spectrum as that from (11). In this sense, the dimensionality of the problem is reduced from L to M.

In the second step, we first note that $\bar{\boldsymbol{\delta}}$ is expected to show the same sparse pattern as $\hat{\boldsymbol{p}}_2$, whose maximum K peaks correspond to the locations of the K sources. Hence it is reasonable to calculate entries of $\bar{\boldsymbol{\delta}}$ only related to the maximum K peaks of the spectrum of $\hat{\boldsymbol{p}}_2$ and set other entries to zeros.

By the fact that $\bar{\boldsymbol{S}}$ is row-sparse whose K nonzero rows correspond to the locations of the K sources, we can update these nonzero rows of $\bar{\boldsymbol{S}}$ as,

$$\hat{\boldsymbol{S}}_K^{(q+1)} = \boldsymbol{A}^\dagger(\hat{\boldsymbol{\theta}}_K^{(q+1)})\boldsymbol{X} = \left(\boldsymbol{A}_K^{(q+1)}\right)^\dagger\boldsymbol{X}, \tag{17}$$

where $\boldsymbol{A}_K^{(q+1)} \triangleq \boldsymbol{A}(\hat{\boldsymbol{\theta}}_K^{(q+1)})$, $\hat{\boldsymbol{\theta}}_K^{(q+1)}$ denotes the directions corresponding to the maximum K peaks of $\hat{\boldsymbol{p}}_2^{(q+1)}$. Hence, the dimensionality of problem (12) reduces from N to K. We then reformulate problem (12) as

$$\min_{\boldsymbol{\delta}_K} \ \left\|\boldsymbol{A}_K^{(q+1)}\hat{\boldsymbol{S}}_K^{(q+1)} - \boldsymbol{X} + \boldsymbol{B}_K^{(q+1)}\boldsymbol{\Delta}_K\hat{\boldsymbol{S}}_K^{(q+1)}\right\|_F^2, \tag{18}$$

where $\boldsymbol{\Delta}_K = \mathrm{diag}(\boldsymbol{\delta}_K)$ with $\boldsymbol{\delta}_K = [\delta_1, \cdots, \delta_K]^T$ being the biases of the K sources, $\boldsymbol{B}_K^{(q+1)}$ is defined in a manner similar to $\boldsymbol{A}_K^{(q+1)}$. Following the derivation in [23], the objective function of (18) can be rewritten as,[2]

$$\begin{aligned}
&\|\boldsymbol{A}_K\hat{\boldsymbol{S}}_K - \boldsymbol{X} + \boldsymbol{B}_K\boldsymbol{\Delta}_K\hat{\boldsymbol{S}}_K\|_F^2 \\
&\simeq \mathrm{tr}\left\{\hat{\boldsymbol{S}}_K^H\boldsymbol{\Delta}_K\boldsymbol{B}_K^H\boldsymbol{B}_K\boldsymbol{\Delta}_K\hat{\boldsymbol{S}}_K - 2\Re\left[(\boldsymbol{X} - \boldsymbol{A}_K\hat{\boldsymbol{S}}_K)^H\boldsymbol{B}_K\boldsymbol{\Delta}_K\hat{\boldsymbol{S}}_K\right]\right\} \\
&= \mathrm{tr}\left\{\boldsymbol{\Delta}_K\boldsymbol{B}_K^H\boldsymbol{B}_K\boldsymbol{\Delta}_K\hat{\boldsymbol{S}}_K\hat{\boldsymbol{S}}_K^H\right\} - 2\Re\left\{\mathrm{tr}\left[\hat{\boldsymbol{S}}_K(\boldsymbol{X} - \boldsymbol{A}_K\hat{\boldsymbol{S}}_K)^H\boldsymbol{B}_K\boldsymbol{\Delta}_K\right]\right\} \\
&= \boldsymbol{\delta}_K^T\Big(\underbrace{\boldsymbol{B}_K^H\boldsymbol{B}_K \circ (\hat{\boldsymbol{S}}_K\hat{\boldsymbol{S}}_K^H)^*}_{\boldsymbol{G}}\Big)\boldsymbol{\delta}_K - 2\Re\Big\{\underbrace{\mathrm{diag}\big[\hat{\boldsymbol{S}}_K(\boldsymbol{X} - \boldsymbol{A}_K\hat{\boldsymbol{S}}_K)^H\boldsymbol{B}_K\big]}_{\boldsymbol{h}^T}\boldsymbol{\delta}_K\Big\}.
\end{aligned} \tag{19}$$

Then by letting the derivative of (19) with respect to $\boldsymbol{\delta}_K$ be zero, the solution of (18) can be given as

$$\hat{\boldsymbol{\delta}}_K^{(q+1)} = \Re\{\boldsymbol{G}^{-1}\boldsymbol{h}\}. \tag{20}$$

We then constrain the estimated biases $\hat{\boldsymbol{\delta}}_K^{(q+1)}$ in $[-\frac{r}{2}, \frac{r}{2}]$ as,

$$\hat{\delta}_K^{(q+1)} = \begin{cases} -\frac{r}{2} & \hat{\delta}_K^{(q+1)} < -\frac{r}{2} \\ \frac{r}{2} & \hat{\delta}_K^{(q+1)} > \frac{r}{2} \\ \hat{\delta}_K^{(q+1)} & \text{otherwise} \end{cases}. \tag{21}$$

The bias $\bar{\boldsymbol{\delta}}^{(q+1)}$ can be obtained by expanding $\hat{\boldsymbol{\delta}}_K^{(q+1)}$ with zero filled at the positions of other than the K maximum peaks of $\hat{\boldsymbol{p}}_2^{(q+1)}$.

Finally, when the iteration converges, the DOA estimates can be given by

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_K^{\text{final}} + \hat{\boldsymbol{\delta}}_K^{\text{final}}. \tag{22}$$

As mentioned before, the regularization parameter λ controls the tradeoff between the sparsity of the solution and the data fitting error. When the knowledge of the noise level is known *a priori*, λ can be chosen easily. For the case of unknown noise level, one of the famous approach is the L-curve, which, however, is computationally expensive since it requires to solve the optimization problem (5) for a large number of different values of λ. In our method, according to the signal model (2), the noise power can be iteratively updated as

$$\hat{\sigma}^{(q+1)} = \frac{\left\|\boldsymbol{X} - \boldsymbol{A}\big(\hat{\boldsymbol{\theta}}_K^{(q+1)} + \hat{\boldsymbol{\delta}}_K^{(q+1)}\big)\hat{\boldsymbol{S}}_K^{(q+1)}\right\|_F^2}{ML}, \tag{23}$$

from which we update λ by,

$$\hat{\lambda}^{(q+1)} = c\sqrt{\hat{\sigma}^{(q+1)}}, \tag{24}$$

where c is a weighting factor. Note that by using the update rule for λ, the influences of SNR, L and array geometry can be integrated in $\hat{\sigma}^{(q+1)}$ while c just introduces fine tuning to the threshold, hence c has a much smaller dynamic range than λ and is easier to choose. Finally, by incorporating the update rule of λ (24) into the iterations, we complete our algorithm called OGL1CCD as summarized in Algorithm 1.

Remark 1. It is worth mentioning that, the idea of two-step iterative algorithm has been used for single snapshot case in [24,32]. Although its extension to multiple snapshots case is straightforward, the high computational complexity caused by the large dimensionality of the signal model has been a major issue for multiple snapshot applications. In this paper, we have carried out dimensionality reduction and obtained a closed-form solution to tackle the complexity issue.

[2] We omit the superscript in (19) for brevity.

Algorithm 1 OGL1CCD.

Input: $\boldsymbol{X}$, K, ϑ.
  Initialization: $\bar{\boldsymbol{\delta}}^{(0)} = \boldsymbol{0}$.
  repeat
    Update $\hat{\boldsymbol{S}}_K$ by using (15)-(17);
    Update $\hat{\boldsymbol{\delta}}_K$ by using (19)-(21);
    Update $\hat{\lambda}$ by using (23), (24);
  until Convergence
Output: $\hat{\boldsymbol{\theta}}$ by using (22).

Remark 2. In each iteration of OGL1CCD, it requires to solve an MMV optimization problem which is time-consuming, hence it is essential to modify the algorithm framework for speed consideration. We emprically observed that when the SNR is moderate or large, the positions of the maximum K peaks of $\hat{\boldsymbol{p}}_2^{(q+1)}$ remains unchanged after q_0 iterations, where q_0 is a small integer (e.g., one or less than 10). Hence it is reasonable to bypass model (15) in the rest of the iterations. In this spirit, The update rule for $\hat{\boldsymbol{S}}_K^{(q+1)}$ in the qth $(q > q_0)$ iteration can be directly using (17) in which $\hat{\boldsymbol{\theta}}_K^{(q+1)} = \hat{\boldsymbol{\theta}}_K^{(q_0)} + \hat{\boldsymbol{\delta}}_K^{(q)}$.

Remark 3. Algorithm 1 can be applied to any other sparse method such as ℓ_1 reconstruction after singular value decomposition (L1SVD) [11] or ℓ_1 sparse representation of array covariance vectors (L1SRACV) [33] by replacing the peak-finding technique in the first step with other sparse methods while other operations remain unchanged. Moreover, the fast implementation in Remark 2 can be also extended to other sparse methods. Examples are provided in Section 5 to show the extendibility of our algorithm framework.

4. Covariance matrix reconstruction approach

Since the covariance matrix of the array output also contains the interesting DOA information, the DOAs can be estimated by reconstructing the covariance matrix. It has been shown that by using the covariance matrix of the array output the performance of the sparse DOA estimation methods can be improved since the second order statistics model has more degrees of freedom and higher array output SNR (ASNR) [34]. In this section, we first formulate the off-grid covariance-based DOA estimation model and then propose an iterative method for DOA estimation.

4.1. Off-grid covariance-based model

When sufficient snapshots are collected, the covariance matrix of the array output can be given as

$$\boldsymbol{R} = E[\boldsymbol{X}\boldsymbol{X}^H] = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{R}_{\boldsymbol{S}}\boldsymbol{A}^H(\boldsymbol{\theta}) + \sigma\boldsymbol{I}, \tag{25}$$

where $\boldsymbol{R}_{\boldsymbol{S}}$ denotes the covariance matrix of the signal $\boldsymbol{S}$. When the impinged sources are uncorrelated with each other, Eq. (25) can be rewritten as

$$\boldsymbol{r} = (\boldsymbol{A}^*(\boldsymbol{\theta}) \odot \boldsymbol{A}(\boldsymbol{\theta}))\boldsymbol{p} = \tilde{\boldsymbol{A}}(\boldsymbol{\theta})\boldsymbol{p}, \tag{26}$$

where $\boldsymbol{r} = \mathrm{vec}(\boldsymbol{R} - \sigma\boldsymbol{I})$, $\tilde{\boldsymbol{A}}(\boldsymbol{\theta}) \triangleq \boldsymbol{A}^*(\boldsymbol{\theta}) \odot \boldsymbol{A}(\boldsymbol{\theta})$ and $\boldsymbol{p} = [p_1, \cdots, p_K]^T$ denotes the signal power.

In practical applications, the covariance matrix $\boldsymbol{R}$ can only be estimated using limited snapshots as follows,

$$\hat{\boldsymbol{R}} = \frac{1}{L}\boldsymbol{X}\boldsymbol{X}^H. \tag{27}$$

Note that the noise power can be well estimated as the minimum eigenvalue of $\hat{\boldsymbol{R}}$. Similar to Section 2, we expand model (26) into an

off-grid SMV model,

$$\tilde{\boldsymbol{r}} = (\tilde{\boldsymbol{A}} + \tilde{\boldsymbol{B}}\boldsymbol{\Delta})\bar{\boldsymbol{p}} + \boldsymbol{\epsilon}, \tag{28}$$

where $\tilde{\boldsymbol{r}} = \mathrm{vec}(\hat{\boldsymbol{R}} - \sigma\boldsymbol{I})$, $\tilde{\boldsymbol{A}} = \tilde{\boldsymbol{A}}(\boldsymbol{\vartheta})$, $\tilde{\boldsymbol{B}} = [\tilde{\boldsymbol{b}}(\vartheta_1), \cdots, \tilde{\boldsymbol{b}}(\vartheta_N)]$ denotes the derivative of $\tilde{\boldsymbol{A}}$ with respect to $\boldsymbol{\vartheta}$, $\bar{\boldsymbol{p}}$ is the extension of $\boldsymbol{p}$ with zero filled at the positions of other than the grids corresponding to the true DOAs, and $\boldsymbol{\epsilon}$ contains the perturbations due to finite snapshots effect and satisfies the following Gaussian distribution [34,35],

$$\boldsymbol{\epsilon} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{Q}), \tag{29}$$

where $\boldsymbol{Q} = \frac{1}{L}\boldsymbol{R}^T \otimes \boldsymbol{R}$.

Similar to Section 2.2, we derive the CRLB of the biases with respect to model (28) as follows,

Proposition 2. *For model (28), the CRLB of DOA estimation error at each grid satisfies,*

$$\mathrm{CRLB}(\delta_n') = \frac{1}{2p_n^2\tilde{\boldsymbol{b}}^H(\vartheta_n)\boldsymbol{Q}^{-1}\tilde{\boldsymbol{b}}(\vartheta_n)}, \quad n = 1, \cdots, N. \tag{30}$$

Proof. Please see Appendix B. □

It is not easy to directly analysis (28). Instead, by comparing (10) (or (A.6)) and (30) we can have

$$\frac{\mathrm{CRLB}(\delta_n')}{\mathrm{CRLB}(\delta_n)} = \frac{\boldsymbol{b}^H(\vartheta_n)\boldsymbol{b}(\vartheta_n)}{\sigma L p_n \tilde{\boldsymbol{b}}^H(\vartheta_n)(\boldsymbol{R}^{-T} \otimes \boldsymbol{R}^{-1})\tilde{\boldsymbol{b}}(\vartheta_n)}. \tag{31}$$

It is hard to calculate the explicit value of (31). However, it is evident that this ratio decreases as the number of snapshots increases, indicating that compared to the model (8), using the covariance-based model (28) is able to achieve more accurate bias estimates if sufficient snapshots are collected.

4.2. OGL1CMRA method

We consider the generalized least squares method to obtain the DOA estimates. Since the perturbation error $\boldsymbol{\epsilon}$ satisfies a complex Gaussian distribution with zero mean and covariance $\boldsymbol{Q}$, we have

$$\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{\epsilon} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}), \tag{32}$$

from which it can be concluded that

$$\left\|\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{\epsilon}\right\|_2^2 \sim \chi^2(M^2), \tag{33}$$

where $\chi^2(M^2)$ denotes the asymptotic chi-square distribution with M^2 degrees of freedom. Further since each entry of $\bar{\boldsymbol{p}}$ is nonnegative, we propose the following optimization problem,

$$\min_{\bar{\boldsymbol{p}}\geq\boldsymbol{0},\tilde{\boldsymbol{\delta}}} \|\bar{\boldsymbol{p}}\|_1 \text{ s.t. } \left\|\hat{\boldsymbol{Q}}^{-\frac{1}{2}}\left(\tilde{\boldsymbol{r}} - (\tilde{\boldsymbol{A}} + \tilde{\boldsymbol{B}}\boldsymbol{\Delta})\bar{\boldsymbol{p}}\right)\right\|_2 \leq \beta, \tag{34}$$

where $\hat{\boldsymbol{Q}} = \frac{1}{L}\hat{\boldsymbol{R}}^T \otimes \hat{\boldsymbol{R}}$ is the estimate of $\boldsymbol{Q}$, β can be uniquely determined by $\chi^2(M^2)$ and a predefined parameter η, in which η is usually set to a very small value, e.g., 10^{-4}. It should be noted that, $\hat{\boldsymbol{Q}}$ in equation (34) is explicitly assumed to be invertible, which is true if and only if $\hat{\boldsymbol{R}}$ is invertible, meaning that the number of snapshots should be larger than that of sensors.

Similar to OGL1CCD, we employ the two-step iterative technique to solve the problem. In particular, in the first step, we fix $\boldsymbol{\delta}$ and solve the following optimization problem,

$$\bar{\boldsymbol{p}}^{(q+1)} = \arg\min_{\bar{\boldsymbol{p}}\geq\boldsymbol{0}} \|\bar{\boldsymbol{p}}\|_1 \text{ s.t. } \left\|\hat{\boldsymbol{Q}}^{-\frac{1}{2}}\left(\tilde{\boldsymbol{r}} - (\tilde{\boldsymbol{A}} + \tilde{\boldsymbol{B}}\boldsymbol{\Delta}^{(q)})\bar{\boldsymbol{p}}\right)\right\|_2 \leq \beta, \tag{35}$$

which can be solved using CVX toolbox [36], while in the second step, we fix $\bar{\boldsymbol{p}}$ and update $\boldsymbol{\delta}$ by

$$\boldsymbol{\delta}^{(q+1)} = \arg\min_{\boldsymbol{\delta}} \left\|\hat{\boldsymbol{Q}}^{-\frac{1}{2}}\left(\tilde{\boldsymbol{r}} - (\tilde{\boldsymbol{A}} + \tilde{\boldsymbol{B}}\boldsymbol{\Delta})\bar{\boldsymbol{p}}^{(q+1)}\right)\right\|_2^2. \tag{36}$$

For speed consideration, we reduce its dimensionality similar to problem (18) as,

$$\min_{\boldsymbol{\delta}} \left\|(\boldsymbol{r}_1 - \tilde{\boldsymbol{A}}_K^{(q+1)}\bar{\boldsymbol{p}}_K^{(q+1)}) - \tilde{\boldsymbol{B}}_K^{(q+1)}\boldsymbol{\Delta}_K\bar{\boldsymbol{p}}_K^{(q+1)}\right\|_2^2, \tag{37}$$

where $\boldsymbol{r}_1 = \hat{\boldsymbol{Q}}^{-\frac{1}{2}}\tilde{\boldsymbol{r}}$, $\tilde{\boldsymbol{A}}_K^{(q+1)} = \hat{\boldsymbol{Q}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}(\hat{\boldsymbol{\theta}}_K^{(q+1)})$, $\tilde{\boldsymbol{B}}_K^{(q+1)} = \hat{\boldsymbol{Q}}^{-\frac{1}{2}}\tilde{\boldsymbol{B}}(\hat{\boldsymbol{\theta}}_K^{(q+1)})$ and $\bar{\boldsymbol{p}}_K^{(q+1)}$ consists of the value of K maximum peaks of $\bar{\boldsymbol{p}}^{(q+1)}$.

Problem (37) can also be efficiently solved with the following equalities,

$$\begin{aligned}
&\left\|(\boldsymbol{r}_1 - \tilde{\boldsymbol{A}}_K\bar{\boldsymbol{p}}_K) - \tilde{\boldsymbol{B}}_K\boldsymbol{\Delta}_K\bar{\boldsymbol{p}}_K^{(q+1)}\right\|_2^2 \\
&\simeq \bar{\boldsymbol{p}}_K^H\boldsymbol{\Delta}_K\tilde{\boldsymbol{B}}_K^H\tilde{\boldsymbol{B}}_K\boldsymbol{\Delta}_K\bar{\boldsymbol{p}}_K - 2\Re\left\{(\boldsymbol{r}_1 - \tilde{\boldsymbol{A}}_K\bar{\boldsymbol{p}}_K)^H\tilde{\boldsymbol{B}}_K\boldsymbol{\Delta}_K\bar{\boldsymbol{p}}_K\right\} \\
&= \boldsymbol{\Delta}_K\tilde{\boldsymbol{B}}_K^H\tilde{\boldsymbol{B}}_K\boldsymbol{\Delta}_K\bar{\boldsymbol{p}}_K\bar{\boldsymbol{p}}_K^H - 2\Re\left\{\bar{\boldsymbol{p}}_K(\boldsymbol{r}_1 - \tilde{\boldsymbol{A}}_K\bar{\boldsymbol{p}}_K)^H\tilde{\boldsymbol{B}}_K\boldsymbol{\Delta}_K\right\} \\
&= \boldsymbol{\delta}_K^T\Big(\underbrace{\tilde{\boldsymbol{B}}_K^H\tilde{\boldsymbol{B}}_K \circ (\bar{\boldsymbol{p}}_K\bar{\boldsymbol{p}}_K^H)^*}_{\tilde{\boldsymbol{G}}}\Big)\boldsymbol{\delta}_K - 2\Re\left\{\underbrace{\mathrm{diag}[\bar{\boldsymbol{p}}_K(\boldsymbol{r}_1 - \tilde{\boldsymbol{A}}_K\bar{\boldsymbol{p}}_K)^H\tilde{\boldsymbol{B}}_K]}_{\tilde{\boldsymbol{h}}^T}\boldsymbol{\delta}_K\right\}.
\end{aligned} \tag{38}$$

From (38), the update rule for $\boldsymbol{\delta}_K$ can be easily obtained as,

$$\boldsymbol{\delta}_K = \Re\left\{\tilde{\boldsymbol{G}}^{-1}\tilde{\boldsymbol{h}}\right\}. \tag{39}$$

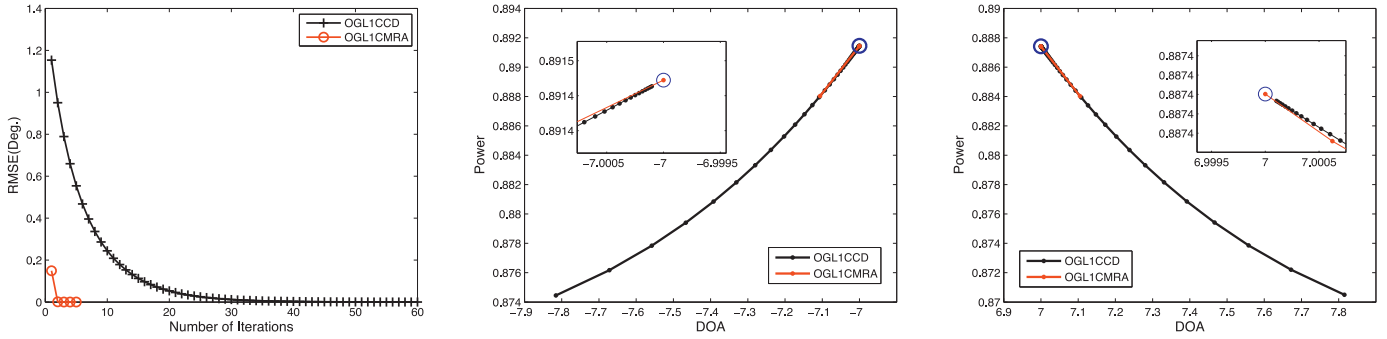
Remark 4. Compared to OGL1CCD, OGL1CMRA has three advantages, (1) the dimensionality reduction procedure is unnecessary the first step; (2) unlike OGL1CCD, parameter updating is not required; (3) OGL1CMRA is more computationally efficient than OGL1CCD.

Remark 5. Moreover, OGL1CMRA is able to correctly locate more sources than the number of sensors since it explores the maximum number of DOFs. Readers are referred to our recent results in [28,29] for more information.

5. Numerical results

In this section, we will evaluate the performance of our proposed methods in convergence speed and estimation performance with comparison to several other state-of-the-art methods, including MUSIC [3], L1SVD [11], L1SRACV [33] and SPA [37]. The fast version of OGL1CCD, as mentioned in Remark 2, is called fast OGL1CCD (FOGL1CCD) in simulations for comparison. For simplicity, we only solve the model (15) in the first iteration of FOGL1CCD, while in the rest of the iterations, only the second step is carried out. All experiments are carried out on a 7-element ULA. It should be noted that, our methods are also applicable to the sparse uniform array (SLA). In our simulations, the origin is set at the middle point of the ULA to reduce the approximation error in (7) or (28). Unless otherwise stated, we use an uniform sampling grid $\{-90^\circ, -88^\circ, \cdots, 88^\circ, 90^\circ\}$ to divide the whole angle space for all of the compared methods except SPA. For our methods, the maximum number of iterations is empirically set as 60, the stopping criterion is defined as $\|\tilde{\boldsymbol{\delta}}^{(q+1)} - \tilde{\boldsymbol{\delta}}^{(q)}\|/\|\tilde{\boldsymbol{\delta}}^{(q)}\| < 10^{-4}$. Specifically, in OGL1CCD, we set $c = 1$ and $\hat{\sigma}^{(0)}$ as the minimum eigenvalue of $\hat{\boldsymbol{R}}$ while in OGL1CMRA, we set $\gamma = 10^{-4}$ and calculate β using MATLAB routine `chi2inv`$(1-\gamma, M^2)$. In L1SVD, we set $\lambda = 0.625$ and employ the IGR procedure to refine the grid gradually. The IGR settings are the same as that in [11]. The root mean square error (RMSE) is defined as,

$$\mathrm{RMSE} = \sqrt{\frac{1}{\bar{N}K}\sum_{n=1}^{\bar{N}}\left\|\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}^{(n)}\right\|_2^2}, \tag{40}$$



(a) Performance in each iteration (b) DOA and power estimates of signal 1 (c) DOA and power estimates of signal 2

Fig. 1. Illustration of our methods OGL1CCD and OGL1CMRA with two noiseless signals impinging onto a 7-ULA from $[-7^\circ, 7^\circ]$ and $L = 100$.

where $\overline{N}$ denotes the number of trials, $\hat{\boldsymbol{\theta}}^{(n)}$ and $\boldsymbol{\theta}^{(n)}$ are the sets of the estimated and true directions of signals in the nth trial, respectively. All simulations are carried out in Matlab R2012a on a PC with a Windows 7 system and a dual-core 3.2 GHz CPU.

5.1. Convergence of our methods

We first provide a simple example to illustrate the iterative process of our proposed methods. Without loss of generality, we consider the noiseless case and assume the signals impinge onto the ULA from $[-7^\circ, 7^\circ]$ with unit power. The number of collected snapshots is set to 100. The RMSEs of our proposed methods during the iteration procedure are given in Fig. 1(a), from which we can see that OGL1CMRA enjoys a much smaller estimation error and faster convergence speed. In particular, OGL1CCD requires more than 30 iterations to converge, while OGL1CMRA is able to converge in 5 iterations. We then show the DOA and power estimates of the signals from $[-7^\circ, 7^\circ]$ from our methods in Fig. 1(b) and (c), respectively. the subfigures inside show the details around the true DOAs and powers. It can be seen that both of our methods are able to converge to the true points. Specifically, OGL1CMRA starts with a much more accurate location than OGL1CCD, and converges to the true points after few iterations, while OGL1CCD requires more iterations to converge. Furthermore, we observe from the subfigures that OGL1CMRA is able to provide a more accurate estimate than OGL1CCD. Hence, it can be concluded that OGL1CMRA is superior to OGL1CCD in estimation performance and convergence speed, provided that the number of snapshots is moderate or large.

5.2. Performance comparison

Here, we evaluate the performance of our proposed methods with comparison to other state-of-the-art methods. First, we consider the special case of single snapshot.[3] In particular, we assume two signals impinge onto the 7-element ULA from $[-5^\circ, 5^\circ]$ and set $L = 1$, SNR= 30 dB. The spectra of the compared methods are shown in Fig. 2, from which it can be observed that, all the methods are able to locate the sources except MUSIC. Furthermore, the subfigures in the lower left and right corners of Fig. 2 zoom in the areas around the true DOAs of signals 1 and 2, respectively. We can observe that OGL1CCD enjoys the best estimation performance among these methods.

Then, we assume that two narrowband uncorrelated signals impinge onto the 7-element ULA from $[-5^\circ + \nu, 5^\circ + \nu]$, where ν is a random variable, uniformly chosen from the interval $[-\frac{r}{2}, \frac{r}{2}] = [-1^\circ, 1^\circ]$. The variable ν is introduced to avoid the influence of the predefined grid set to the performance of the grid-based methods. We show the RMSEs of the DOAs estimated from the compared methods in Fig. 3 with SNR varying from −6dB to 14dB where we consider two cases: $L = 100$ and $L = 200$. It can be seen that with the increment of L, the estimation performance of our three methods and SPA is improved. In particular, when $L = 100$, OGL1CCD and FOGL1CCD are able to coincide with the CRLB curve if SNR is large enough while there still exists a noticeble gap between the OGL1CMRA and the CRLB curves. The gap vanishes in Fig. 3(b) since sufficient snapshots are collected. Furthermore, the performance of OGL1CMRA the low SNR region is also improved, achieving the smallest RMSE among the compared methods. FOGL1CCD fails to provide a satisfactory performance when SNR is lower than 0 dB. In contrast, L1SVD and OGSBI show worse performance than other methods and there is little improvement when the SNR grows. We have found in our experiment, SPA is able to approach the CRLB curve when the SNR is large enough, even though this could not be shown in Fig. 3, due to the middle range of SNR set in the figure.

We have also compared the CPU time of these methods and show the results in Fig. 4. Since our proposed methods involve the iteration procedure, more computations are required. By comparing the two subfigures we can see that increasing the number of snapshots is able to reduce the computations of OGL1CMRA, especially in the small SNR region, while those of OGL1CCD and FOGL1CCD remain unchanged.

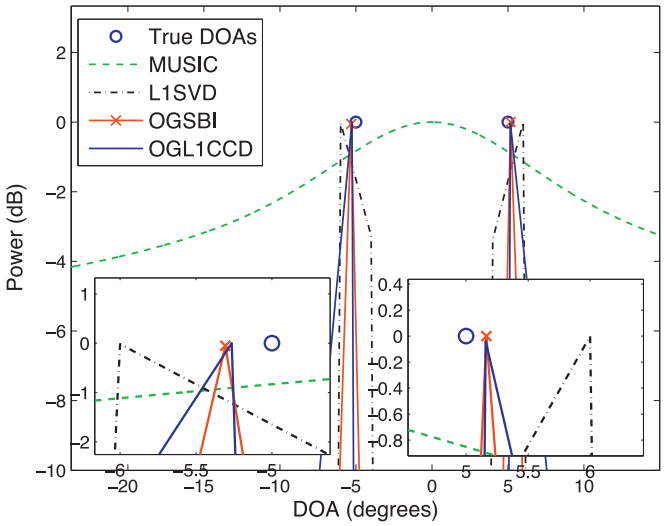
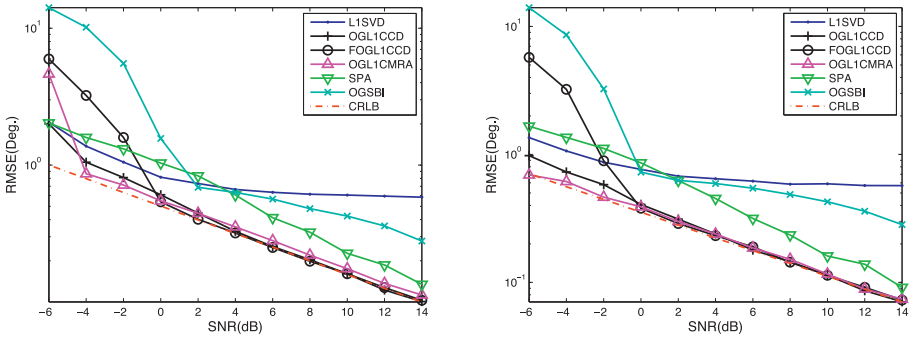


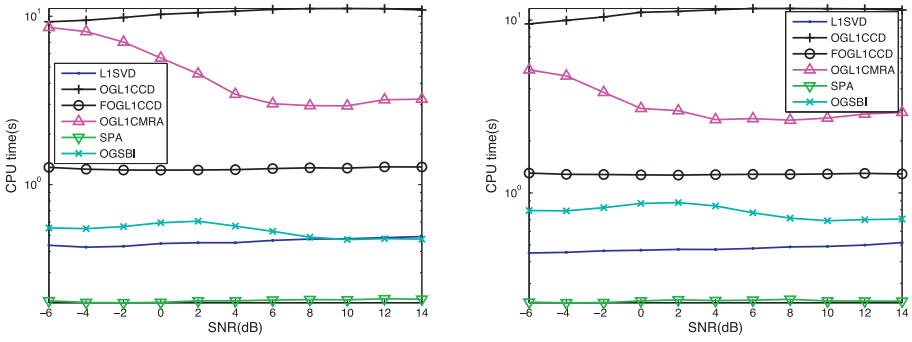
Fig. 2. spectra comparisons with two signals impinging onto a 7-ULA from $[-5^\circ, 5^\circ]$. $L = 1$. SNR = 30 dB.

[3] Since OGL1CMRA requires multiple snapshots, we omit it in this experiment.



(a) $L = 100$ (b) $L = 200$

Fig. 3. RMSE comparisons with two uncorrelated signals impinging onto a 7-ULA from $[-5^\circ + v, 5^\circ + v]$. The SNR varies from −6 dB to 14 dB.



(a) $L = 100$ (b) $L = 200$

Fig. 4. CPU time comparisons with two uncorrelated signals impinging onto a 7-ULA from $[-5^\circ + v, 5^\circ + v]$. The SNR varies from −6 dB to 14 dB.

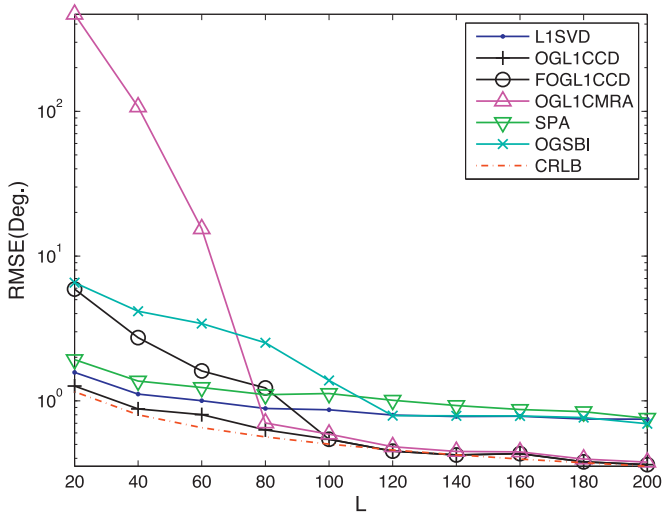


Fig. 5. RMSE comparisons with two uncorrelated signals impinging onto a 7-ULA from $[-5^\circ + v, 5^\circ + v]$. The number of snapshots varies from 20 to 200. SNR = 0dB.

We then compare the RMSE of these methods with respect to different values of L. We repeat the previous simulation except that the SNR is set to 0 dB and L varies from 20 to 200. The simulation results are shown in Fig. 5, from which we can see that the curves of our three methods coincide with the CRLB curve when sufficient number of snapshots are collected. However, OGL1CMRA almost fails when $L < 80$, revealing the high importance of the number of snapshots in the method OGL1CMRA. On the other hand, the performance of FOGL1CCD is deteriorated when $L < 100$. In contrast, SPA, L1SVD and OGSBI fail to coincide with the CRLB curve in the compared region.

Finally, we evaluate the RMSEs of these methods in the case of correlated signals. Since OGL1CMRA cannot deal with the correlated sources, we omit it in this experiment. We assume that the two correlated sources impinge onto the 7-ULA from $[-15^\circ + v, 5^\circ + v]$. The correlation coefficient ρ varies from 0 to 1 where $\rho = 0$ denotes the uncorrelated case while $\rho = 1$ denotes the coherent sources. We set SNR= 10dB, $L = 100$ and show the results in Fig. 6. It is easy to see that our methods OGL1CCD and FOGL1CCD have similar performance and coincide with the CRLB. Note that our methods still provide a satisfactory performance when the two sources are coherent while the performance of SPA and L1SVD is severely deteriorated in this case.

5.3. Extension to L1SVD and L1SRACV

In Remark 3 we stated that the framework in Algorithm 1 can be extended to other sparse methods for performance improvement, which will be verified in this section. Without loss of generality, we choose L1SVD [11] and L1SRACV [33] as two representative methods. In particular, the original L1SVD/L1SRACV employs the IGR procedure to refine the estimation results. Instead, we modify the L1SVD/L1SRACV output by replacing the peak-finding technique in the first step of Algorithm 1 with L1SVD/L1SRACV while other operations remain unchanged. We carry out simulation with the same setting as that in Fig. 3(a) to evaluate the RMSE and CPU time of the two methods when the SNR varies from −10 dB to 10 dB and show the results in Fig. 7. From Fig. 7 (a), it can be seen that the modified versions of these two methods are superior to the original ones. All the four proposed methods are able to coincide with the CRLB curve while L1SVD and L1SRACV are

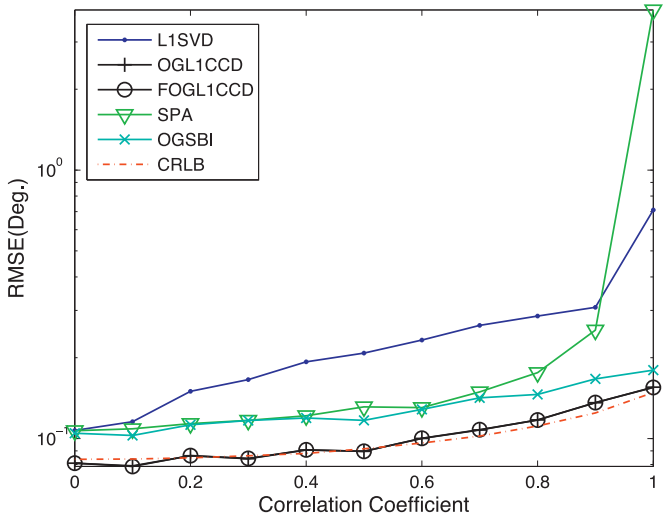


Fig. 6. RMSE comparisons with two correlated signals impinging onto a 7-ULA from $[-15^\circ+v, 5^\circ+v]$ with the correlation coefficient varying from 0 to 1. SNR = 10dB, $L = 100$.

biased. The CPU times of these methods are shown in Fig. 7 (b), indicating that the fast version FOGL1SVD/FOGL1SRACV is more efficient than L1SVD/L1SRACV and OGL1SVD/OGL1SRACV, respectively. Finally, we can also observe that OGL1SVD/OGL1SRACV and its fast version have very similar estimation performance, indicating that by using FOGL1SVD/FOGL1SRACV, a higher computational efficiency can be achieved while the estimation accuracy is retained simultaneously.

6. Conclusions

In this paper, we proposed two off-grid sparse methods OGL1CCD and OGL1CMRA to overcome the grid mismatch limitation inherent in the traditional on-grid methods. We first introduced two off-grid models involving the bias parameter and then proposed a two-step iterative method to solve the models. Our methods are superior to many other methods in estimation performance, which is verified by numerical simulations. Nevertheless, since OGL1CMRA is proposed based on sufficient snapshots and the special Toeplitz Hermitian structure, which has been true only when the impinged signals are uncorrelated with each other, the applicable scenarios of OGL1CMRA are limited. Furthermore, the proposed algorithm can help other sparse methods to improve their estimation accuracy.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under grant No. 61372122, No. 61471205; the Innovation Program for Postgraduate in Jiangsu Province under grant No. KYLX 0813.

Appendix A. Proof of Proposition 1

By definition, the CRLB of $\boldsymbol{\delta}$ is given by

$$\begin{aligned}\mathrm{CRLB}^{-1}(\boldsymbol{\delta}) &= -E_{\boldsymbol{X},\boldsymbol{\delta}}\left\{\frac{\partial}{\partial\boldsymbol{\delta}}\left[\frac{\partial \log p(\boldsymbol{X},\boldsymbol{\delta})}{\partial\boldsymbol{\delta}}\right]^T\right\}\\ &= -E_{\boldsymbol{X},\boldsymbol{\delta}}\left\{\frac{\partial}{\partial\boldsymbol{\delta}}\left[\frac{\partial \log p(\boldsymbol{X}|\boldsymbol{\delta})}{\partial\boldsymbol{\delta}}\right]^T\right\}.\end{aligned} \tag{A.1}$$

By noting that

$$p(\boldsymbol{X}|\boldsymbol{\delta}) = |\pi\sigma I_M|^{-L}\exp\left\{-\frac{1}{\sigma}\left\|\boldsymbol{X}-(\bar{\boldsymbol{A}}+\bar{\boldsymbol{B}}\boldsymbol{\Delta})\bar{\boldsymbol{S}}\right\|_F^2\right\}, \tag{A.2}$$

equation (A.1) can be rewritten as,

$$\mathrm{CRLB}^{-1}(\boldsymbol{\delta}) = \frac{1}{\sigma}\left\{\frac{\partial}{\partial\boldsymbol{\delta}}\left[\frac{\partial E\left[\left\|\boldsymbol{X}-(\bar{\boldsymbol{A}}+\bar{\boldsymbol{B}}\boldsymbol{\Delta})\bar{\boldsymbol{S}}\right\|_F^2\right]}{\partial\boldsymbol{\delta}}\right]^T\right\}. \tag{A.3}$$

Recall that $p(\bar{\boldsymbol{S}}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{P})$, where $\boldsymbol{P} = \mathrm{diag}(\boldsymbol{p})$, in which $\boldsymbol{p}$ denotes the signal power, the term $E\left[\left\|\boldsymbol{X}-(\bar{\boldsymbol{A}}+\bar{\boldsymbol{B}}\boldsymbol{\Delta})\bar{\boldsymbol{S}}\right\|_F^2\right]$ can be deduced as,

$$\begin{aligned}&E\left[\left\|\boldsymbol{X}-(\bar{\boldsymbol{A}}+\bar{\boldsymbol{B}}\boldsymbol{\Delta})\bar{\boldsymbol{S}}\right\|_F^2\right]\\ &= E\left[\left\|(\boldsymbol{X}-\bar{\boldsymbol{A}}\bar{\boldsymbol{S}})-\bar{\boldsymbol{B}}\boldsymbol{\Delta}\bar{\boldsymbol{S}}\right\|_F^2\right]\\ &\simeq \mathrm{tr}\left[\boldsymbol{\Delta}\boldsymbol{P}\boldsymbol{\Delta}\bar{\boldsymbol{B}}^H\bar{\boldsymbol{B}}\right] + 2\Re\left\{\mathrm{tr}\left[\bar{\boldsymbol{B}}^H(\bar{\boldsymbol{A}}\boldsymbol{P}-\boldsymbol{X}\boldsymbol{\mu}_{\bar{\boldsymbol{s}}}^H)\boldsymbol{\Delta}\right]\right\}\\ &\simeq \boldsymbol{\delta}^T\Re\left\{\boldsymbol{P}\circ\bar{\boldsymbol{B}}^H\bar{\boldsymbol{B}}\right\}\boldsymbol{\delta} + 2\Re\left\{\mathrm{diag}\left[\bar{\boldsymbol{B}}^H(\bar{\boldsymbol{A}}\boldsymbol{P}-\boldsymbol{X}\boldsymbol{\mu}_{\bar{\boldsymbol{s}}}^H)\right]\right\}^T\boldsymbol{\delta},\end{aligned} \tag{A.4}$$

where $\boldsymbol{\mu}_{\bar{\boldsymbol{s}}} = E[\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^H]$. By substituting (A.4) into (A.3), the CRLB of $\boldsymbol{\delta}$ is given by

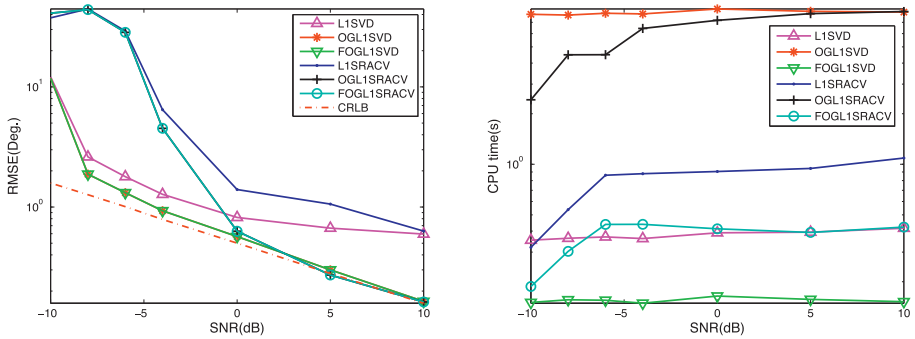
$$\mathrm{CRLB}^{-1}(\boldsymbol{\delta}) = \frac{2}{\sigma}\Re\left\{\boldsymbol{P}\circ(\bar{\boldsymbol{B}}^H\bar{\boldsymbol{B}})\right\}. \tag{A.5}$$

Based on (A.5), the CRLB of δ_n can be shown as,

$$\mathrm{CRLB}^{-1}(\delta_n) = \frac{2}{\sigma}p_n\boldsymbol{b}^H(\vartheta_n)\boldsymbol{b}(\vartheta_n),\ \ n = 1,\cdots,N. \tag{A.6}$$

Finally, (10) can be easily drawn by noting that,

$$\boldsymbol{b}^H(\vartheta_n)\boldsymbol{b}(\vartheta_n) = \pi^2\cos^2(\vartheta_n)\sum_{n=1}^{M-1}n^2$$



(a) RMSE (b) CPU time

Fig. 7. RMSE and CPU time comparisons of L1SVD, L1SRACV and their modified versions by using Algorithm 1. Two uncorrelated signals impinge onto a 7-ULA from $[-5^\circ+v, 5^\circ+v]$. The SNR varies from −10 dB to 10 dB and 100 snapshots are collected for DOA estimation.

$$= \pi^2 \cos^2(\vartheta_n) \frac{M(M-1)(2M-1)}{6}. \quad \text{(A.7)}$$

Appendix B. Proof of Proposition 2

The derivation of Proposition 2 is similar to that of Proposition 1, except that (A.2) is replaced by the following equation,

$$p(\tilde{\boldsymbol{r}}|\boldsymbol{\delta}) = |\pi \boldsymbol{Q}|^{-L} \exp\left\{ -\mathrm{tr}\left[(\tilde{\boldsymbol{r}} - \tilde{\boldsymbol{C}}\bar{\boldsymbol{p}})^H \boldsymbol{Q}^{-1} (\tilde{\boldsymbol{r}} - \tilde{\boldsymbol{C}}\bar{\boldsymbol{p}})\right]\right\}, \quad \text{(B.1)}$$

where $\tilde{\boldsymbol{C}} = \tilde{\boldsymbol{A}} + \tilde{\boldsymbol{B}}\boldsymbol{\Delta}$. We then obtain the BCRB of $\boldsymbol{\delta}$ as given by

$$\mathrm{BCRB}^{-1}(\boldsymbol{\delta}) = 2\Re\left\{\boldsymbol{\Gamma}^2 \circ (\tilde{\boldsymbol{B}}^H \boldsymbol{Q}^{-1} \tilde{\boldsymbol{B}})\right\}, \quad \text{(B.2)}$$

from which (30) can be easily drawn.

References

[1] I. Bekkerman, J. Tabrikian, Target detection and localization using MIMO radars and sonars, IEEE Trans. on Signal Process. 54 (10) (2006) 3873–3883, doi:10.1109/TSP.2006.879267.

[2] J. Picheral, U. Spagnolini, Angle and delay estimation of space-time channels for TD-CDMA systems, IEEE Trans. Wireless Commun. 3 (3) (2004) 758–769, doi:10.1109/TWC.2004.826320.

[3] R. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (3) (1986) 276–280.

[4] R. Roy, T. Kailath, Esprit-estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust. Speech Signal Process. 37 (7) (1989) 984–995.

[5] B. Friedlander, The root-music algorithm for direction finding with interpolated arrays, Signal Process. 30 (1) (1993) 15–29.

[6] M.G. Porozantzidou, M.T. Chryssomallis, Azimuth and elevation angles estimation using 2-D music algorithm with an L-shape antenna, in: 2010 IEEE Antennas and Propagation Society International Symposium, 2010, pp. 1–4, doi:10.1109/APS.2010.5560953.

[7] M.D. Zoltowski, M. Haardt, C.P. Mathews, Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary esprit, IEEE Trans. Signal Process. 44 (2) (1996) 316–328, doi:10.1109/78.485927.

[8] H. Chen, C. Hou, W.-P. Zhu, W. Liu, Y.-Y. Dong, Z. Peng, Q. Wang, Esprit-like two-dimensional direction finding for mixed circular and strictly noncircular sources based on joint diagonalization, Signal Process. 141 (2017) 48–56.

[9] Y. Gu, N.A. Goodman, A. Ashok, Radar target profiling and recognition based on TSI-optimized compressive sensing kernel, IEEE Trans. Signal Process. 62 (12) (2014) 3194–3207.

[10] Y. Gu, N.A. Goodman, Information-theoretic compressive sensing kernel optimization and bayesian cramér-rao bound for time delay estimation, IEEE Trans. Signal Process. 65 (17) (2017) 4525–4537.

[11] D. Malioutov, M. Çetin, A.S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. on Signal Process. 53 (8) (2005) 3010–3022.

[12] Z. Shi, C. Zhou, Y. Gu, N.A. Goodman, F. Qu, Source estimation using coprime array: a sparse reconstruction perspective, IEEE Sensors J. 17 (3) (2017) 755–765, doi:10.1109/JSEN.2016.2637059.

[13] C. Zhou, Y. Gu, S. He, Z. Shi, A robust and efficient algorithm for coprime array adaptive beamforming, IEEE Trans. Veh. Technol. PP (99) (2017). 1–1, doi: 10.1109/TVT.2017.2704610.

[14] C. Zhou, Y. Gu, Y.D. Zhang, Z. Shi, T. Jin, X. Wu, Compressive sensing based coprime array direction-of-arrival estimation, IET Commun. (2017), doi:10.1049/iet-com.2016.1048.

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

[16] Z.-M. Liu, Z.-T. Huang, Y.-Y. Zhou, Array signal processing via sparsity-inducing representation of the array covariance matrix, IEEE Transactions on Aerospace and Electronic Systems 49 (3) (2013) 1710–1724, doi:10.1109/TAES.2013.6558014.

[17] P. Stoica, P. Babu, J. Li, New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data, IEEE Trans. on Signal Process. 59 (1) (2011) 35–47, doi:10.1109/TSP.2010.2086452.

[18] M.M. Hyder, K. Mahata, Direction-of-arrival estimation using a mixed $l_{2,0}$ norm approximation, IEEE Trans. on Signal Process. 58 (9) (2010) 4646–4655.

[19] J. Liu, W. Zhou, F.H. Juwono, D. Huang, Reweighted smoothed ℓ_0 -norm based DOA estimation for MIMO radar, Signal Process. 137 (2017) 44–51.

[20] N. Hu, B. Sun, J. Wang, J. Dai, C. Chang, Source localization for sparse array using nonnegative sparse bayesian learning, Signal Process. 127 (2016) 37–43.

[21] E.J. Candès, The restricted isometry property and its implications for compressed sensing, Comptes Rendus Mathematique 346 (9) (2008) 589–592.

[22] X. Wu, W.-P. Zhu, J. Yan, Direction of arrival estimation for off-grid signals based on sparse Bayesian learning, IEEE Sensors J. 16 (7) (2016) 2004–2016, doi:10.1109/JSEN.2015.2508059.

[23] Z. Yang, L. Xie, C. Zhang, Off-grid direction of arrival estimation using sparse Bayesian inference, IEEE Trans. on Signal Process. 61 (1) (2013) 38–43, doi:10.1109/TSP.2012.2222378.

[24] H. Zhu, G. Leus, G. Giannakis, Sparsity-cognizant total least-squares for perturbed compressive sampling, IEEE Trans. on Signal Process. 59 (5) (2011) 2002–2016, doi:10.1109/TSP.2011.2109956.

[25] J. Fang, J. Li, Y. Shen, H. Li, S. Li, Super-resolution compressed sensing: an iterative reweighted algorithm for joint parameter learning and sparse signal recovery, IEEE Signal Process. Lett. 21 (6) (2014) 761–765, doi:10.1109/LSP.2014.2316004.

[26] J. Fang, F. Wang, Y. Shen, H. Li, R.S. Blum, Super-resolution compressed sensing for line spectral estimation: an iterative reweighted approach, IEEE Trans. on Signal Process. 64 (18) (2016) 4649–4662, doi:10.1109/TSP.2016.2572041.

[27] G. Tang, B. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, IEEE Trans. Inf. Theory 59 (11) (2013) 7465–7490, doi:10.1109/TIT.2013.2277451.

[28] X. Wu, W.P. Zhu, J. Yan, A toeplitz covariance matrix reconstruction approach for direction-of-arrival estimation, IEEE Trans. Veh. Technol. PP (99) (2017). 1–1, doi: 10.1109/TVT.2017.2695226.

[29] X. Wu, W.-P. Zhu, J. Yan, Direction-of-arrival estimation based on Toeplitz covariance matrix reconstruction, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016.

[30] Z. Yang, L. Xie, Enhancing sparsity and resolution via reweighted atomic norm minimization, IEEE Trans. on Signal Process. 64 (4) (2016) 995–1006, doi:10.1109/TSP.2015.2493987.

[31] Z. Yang, J. Li, P. Stoica, L. Xie, Sparse methods for direction-of-arrival estimation, arXiv preprint, (2016), arXiv:1609.09596.

[32] C.D. Austin, J.N. Ash, R.L. Moses, Dynamic dictionary algorithms for model order and parameter estimation, IEEE Trans. on Signal Process. 61 (20) (2013) 5117–5130, doi:10.1109/TSP.2013.2276428.

[33] J. Yin, T. Chen, Direction-of-arrival estimation using a sparse representation of array covariance vectors, IEEE Trans. on Signal Process. 59 (9) (2011) 4489–4493, doi:10.1109/TSP.2011.2158425.

[34] Z.-M. Liu, Z.-T. Huang, Y.-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, doi:10.1109/TWC.2013.071113.121305.

[35] B. Ottersten, P. Stoica, R. Roy, Covariance matching estimation techniques for array signal processing applications, Digital Signal Process. 8 (3) (1998) 185–210.

[36] M. Grant, S. Boyd, Y. Ye, CVX: Matlab software for disciplined convex programming, 2008.

[37] Z. Yang, L. Xie, C. Zhang, A discretization-free sparse and parametric approach for linear array signal processing, IEEE Trans. Signal Process. 62 (19) (2014) 4959–4973, doi:10.1109/TSP.2014.2339792.