Signal Processing 132 (2017) 155–164

Contents lists available at ScienceDirect

Signal Processing

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

Grid-less T.V minimization for DOA estimation☆

Kaushik Mahata*, Md Mashud Hyder

Department of Electrical Engineering, University of Newcastle, Callaghan, NSW 2308, Australia

ARTICLE INFO

Keywords:
DOA estimation
Sparse recovery
Grid-less methods
Total variation
Arbitrary geometry
Atomic norm

ABSTRACT

We present a grid-less version of the L1-SVD algorithm for direction of arrival estimation. The resulting semidefinite programming approach is a globally convergent, fully parametric method capable of working with two dimensional arrays with any arbitrary sensor configurations. It is computationally efficient, and shows improved performance when compared with other popular alternatives. The analysis also allows us to formulate the SPICE algorithm in gridless manner.

1. Introduction

Direction of arrival (DOA) estimation is a major functional requirement in numerous applications like radar, sonar, etc [1]. The classical methods are often classified in two broad categories. The first category consists of algorithms like Capon [2] and APES [3], which pose the DOA estimation problem as a beamforming problem. A popular alternative to this is the class of subspace algorithms like MUSIC [4], ESPRIT [5] or weighted subspace fitting [6,7], which exploit the low-rank structure of the noise-free signal. Recently some sparse recovery methods like L1-SVD, SPICE, etc, [8–11] have gained popularity due to their capability of producing better results from shorter data records. Some of the above methods are parametric, and others are non-parametric. Traditionally, the parametric methods are preferred due to ease of use, and their capability of producing clean spatial spectrum at low signal to noise ratio (SNR). However, the parametric algorithms demand very specific, regular array geometries leading to nice mathematical structure of the problem. In this paper we propose a method to alleviate this difficulty. This technique allows us to formulate the sparse recovery methods like L1-SVD or SPICE in a gridless manner. This formulation holds for any two dimensional array configuration. Consequently the resulting algorithm can detect and localize multiple sources within 0–360° azimuth range.

Linear arrays (LA) where the sensors are located on a regular grid have been very attractive in parametric DOA estimation. Most parametric subspace algorithms need uniform linear array. The tradition continues in recent developments of co-prime array [12–14], minimum-redundancy linear arrays [15], and with the developments in grid-less sparse recovery methods [16–22]. However, with a LA one can cover at most 180° azimuth range without ambiguity. This makes 2D geometries interesting. To the best of our knowledge, 'L' shaped arrays [22] and uniform circular array [23–25] are the only 2-D geometries supporting parametric algorithms. In addition, in our opinion, it might be possible to extend some of above methods to arrays with sensors arranged on a regular rectangular grid. In this context it should be noted that there are sparsity based semi-parametric methods, discussed below, applicable to arbitrary array geometries.

DOA estimation methods applicable to arbitrary sensor configurations are either non-parametric, or require a non-parametric front-end. The classical among these are CAPON, MUSIC and APES [26]. These methods are outperformed by some of the recently developed techniques [8,9,27,10,11], particularly when the data record is not long. L1-SVD [8] combines the singular value decomposition (SVD) step of subspace algorithms with the ℓ_1 norm minimization of sparse recovery methods. IAA [27] is a non parametric weighted least square type algorithm, which is efficient for DOA estimation at moderate/high SNR. However, IAA solves a non-convex problem. Thus the global convergence is not guaranteed. SPICE [10,11,28] is a globally convergent, hyper-parameter-free method that can also estimate the noise variance.

The above techniques [8,9,27,10,11] must discretize the range of interest into a grid. Off-grid targets can lead to mismatches in the model and deteriorate the performance significantly [29]. Some semi-parametric methods have been proposed to reduce these problems. For instance, in [30] the joint sparsity between the original signal and the grid mismatch is exploited. A first-order approximation is used in [30] to handle the grid mismatch problem. Hence, the DOA estimation

☆ Research is supported by the Australian Research Council under the grant number DP130103909.
* Corresponding author.
E-mail addresses: Kaushik.Mahata@newcastle.edu.au (K. Mahata), mdmashud.hyder@newcastle.edu.au (M.M. Hyder).

http://dx.doi.org/10.1016/j.sigpro.2016.09.018
Received 27 June 2016; Received in revised form 8 September 2016; Accepted 27 September 2016
Available online 06 October 2016

performance is still limited by higher-order modelling mismatches. It is also common to increase the grid density to reduce the grid mismatch errors [31,32]. However, a dense grid increases computational load, and the correlation between adjacent atoms of the dictionary. Some argue that this increase in correlation causes performance degradation of the grid based sparse recovery algorithms [32].

In this paper we present a semidefinite programming algorithm for the total variation minimization approach (TVMA) [33] for DOA estimation. Our algorithm can be seen as the grid-less pathway to L1-SVD [8]. TVMA has been the key to the theory of super resolution [33], and has been adopted by several authors in spectral analysis. Similar results are also derived using the atomic norm evaluation paradigm [34] for DOA estimation using LAs [35,17,18]. In [35] atomic norm evaluation in presence of multiple measurement vectors is considered. This can be viewed as a continuous counterpart of the $\ell_{2,1}$ norm minimization method. The SPICE-like method of [36] can work with sparse linear arrays. The theory of super-resolution has been extended for co-prime arrays in [19], and a matrix completion based method has appeared in [21]. The main advantage of our solution is its applicability in arbitrary 2D array configurations. To the best of our knowledge it is the only fully parametric, globally convergent method applicable to arbitrary 2D array configurations. The other contribution of this paper is to make TVMA hyper-parameter free, which is a great practical advantage. Finally, we show how we can make SPICE gridless using the mathematical results presented herein.

2. Main results

We use $\mathbb{R}$ and $\mathbb{C}$ to denote the set of all real and complex numbers, respectively. We consider the direction of arrival (DOA) estimation problem, where K narrow band waves from K far field sources are incident on an array of P narrow band sensors located at points (r_p, θ_p), $p = 1, 2, \ldots, P$, *expressed in polar coordinates*. The unit of distance is taken as the half-wavelength of the waves. Let ξ_k be the azimuth of the k th source. We collect the P dimensional measurements over L sampling instants to build a $P \times L$ matrix $\boldsymbol{Y}$, which satisfies

$$\boldsymbol{Y} = \sum_{k=1}^{K} \boldsymbol{\gamma}(\xi_k)\boldsymbol{s}_k^{\top} + \boldsymbol{E}, \tag{1}$$

$$[\boldsymbol{\gamma}(\xi)]_p = \exp\{\mathrm{i}\pi r_p \cos(\xi - \theta_p)\}, \tag{2}$$

where $\boldsymbol{s}_k$ is the L dimensional signal of source k, and $\boldsymbol{E}$ is the additive noise. We wish to estimate $\{\xi_k\}_{k=1}^{K}$ from $\boldsymbol{Y}$.

2.1. Noise-free case

We start by considering the noise-free case where $\boldsymbol{E} = 0$. For this we present the TVMA solution. Later we extend our analysis for $\boldsymbol{E} \neq 0$. In TVMA we solve

$$\text{minimize} \quad \sum_{k=1}^{\mathring{K}} \|\mathring{\boldsymbol{s}}_k\|_2 \text{subject to} \quad \boldsymbol{Y} = \sum_{k=1}^{\mathring{K}} \boldsymbol{\gamma}(\mathring{\xi}_k)\mathring{\boldsymbol{s}}_k^{\top} \tag{3}$$

for some $\mathring{K}$, and $\mathring{\xi}_k \in [0, 2\pi]$, $k = 1, 2, \ldots, \mathring{K}$. In other words, among all possible choices of $\mathring{K}$, $\{\mathring{\xi}_k\}_{k=1}^{\mathring{K}}$ and $\{\mathring{\boldsymbol{s}}_k\}_{k=1}^{\mathring{K}}$ consistent with $\boldsymbol{Y}$, we pick the one for which the cost function in (3) is minimum. The corresponding minimum value, called the atomic norm of $\boldsymbol{Y}$, is denoted as $\|\boldsymbol{Y}\|_{\mathcal{A}}$.

Lemma 1 Below is the first step towards a tractable finite dimensional characterization of (3). It is a bit more general than some analogous results presented before [36]. When compared with the corresponding results in literature, the main difference is that the following result can be applied for any general γ. In that way this is the multidimensional extension to the corresponding scalar version presented in [37]. The proof appears in Appendix A.

Lemma 1. *Let $\boldsymbol{e}$ be the first column of the $P \times P$ identity matrix, and $\boldsymbol{\Gamma}: [0, 2\pi] \to \mathbb{H}_P$ be such that $\boldsymbol{\Gamma}(\xi) = \boldsymbol{\gamma}(\xi)\boldsymbol{\gamma}^*(\xi)$. Let $\mathbb{K}$ be the closed conic hull of the set $\{\boldsymbol{\Gamma}(\xi) \in \mathbb{H}_P: \xi \in [0, 2\pi]\}$. Then $\|\boldsymbol{Y}\|_{\mathcal{A}}$ is the optimum value of*

$$\underset{\boldsymbol{W}\in\mathbb{H}_L,\ \boldsymbol{Q}\in\mathbb{K}}{\text{minimize}} \quad \{\mathrm{Tr}(\boldsymbol{W}) + \boldsymbol{e}^*\boldsymbol{Q}\boldsymbol{e}\}/2 \text{subject to} \quad \begin{bmatrix} \boldsymbol{W} & \boldsymbol{Y}^* \\ \boldsymbol{Y} & \boldsymbol{Q} \end{bmatrix} \geqslant 0. \tag{4}$$

In addition if a decomposition $\boldsymbol{Y} = \sum_k \boldsymbol{\gamma}(\mathring{\xi}_k)\mathring{\alpha}_k\mathring{\boldsymbol{\phi}}_k^*$ is such that $\|\mathring{\boldsymbol{\phi}}_k\|_2 = 1$, $\forall\, k$, and $\|\boldsymbol{Y}\|_{\mathcal{A}} = \sum_k \mathring{\alpha}_k$, then $\boldsymbol{W} = \sum_k \mathring{\boldsymbol{\phi}}_k\mathring{\alpha}_k\mathring{\boldsymbol{\phi}}_k^*$, $\boldsymbol{Q} = \sum_k \mathring{\alpha}_k\boldsymbol{\Gamma}(\mathring{\xi}_k)$ is a solution to (4).

Conversely, if $\boldsymbol{Q} = \sum_k \mathring{\alpha}_k\boldsymbol{\Gamma}(\mathring{\xi}_k)$ is a solution to (4) for some $\mathring{\alpha}_k > 0$, then the trace of the corresponding optimum $\boldsymbol{W}$ is $\sum_k \mathring{\alpha}_k = \|\boldsymbol{Y}\|_{\mathcal{A}}$, and there are unit norm complex vectors $\mathring{\boldsymbol{\phi}}_k$ satisfying the atomic decomposition $\boldsymbol{Y} = \sum_k \boldsymbol{\gamma}(\mathring{\xi}_k)\mathring{\alpha}_k\mathring{\boldsymbol{\phi}}_k^*$.

Thus, we can solve (3) by solving the equivalent problem (4). To be able to do so, we need a convenient characterization of $\mathbb{K}$, which we derive next. Let ρ_{jk} and ϕ_{jk} be defined via the complex addition:

$$\rho_{j,k}\mathrm{e}^{\mathrm{i}\phi_{j,k}} = \pi(r_j\ \mathrm{e}^{\mathrm{i}\theta_j} - r_k\ \mathrm{e}^{\mathrm{i}\theta_k}). \tag{5}$$

In addition, we use ρ to denote the array aperture:

$\rho = \max_{j,k}\rho_{jk}$.

From (2) and (5) and the definition of $\boldsymbol{\Gamma}$ it follows that

$$[\boldsymbol{\Gamma}(\xi)]_{jk} = \exp\{\mathrm{i}\rho_{jk}\cos(\xi - \phi_{jk})\}. \tag{6}$$

Let J_n denote the order n Bessel function of first kind. We write (6) using Jacobi-Anger expansion [38] as

$$[\boldsymbol{\Gamma}(\xi)]_{jk} = \sum_{n=-\infty}^{\infty} \mathrm{i}^n J_n(\rho_{jk})\mathrm{e}^{-\mathrm{i}\phi_{jk}n}\mathrm{e}^{\mathrm{i}\xi n}. \tag{7}$$

Although (7) is an infinite series expansion, it has only a finite number of non-zero terms for all practical purposes. The number of non-zero terms depends on ρ. Given any $r > 0$, it is wellknown that $|J_n(r)|$ decays quite rapidly with $|n|$. We define $N(\rho, \epsilon)$ as the smallest positive integer such that for every n satisfying $|n| \leq N(\rho, \epsilon)$ it holds that $|J_n(r)| < \epsilon$ uniformly for all $r \in [0, \rho]$. In Fig. 1 we plot $N(\rho, \epsilon)$ as a function of ρ for few different values of ϵ. Set ϵ to the precision of the underlying computational platform. Since $|J_n(\rho)| = |J_{-n}(\rho)|$ for all ρ and n, by (7) we can approximate $\boldsymbol{\Gamma}(\xi)$ without any loss of precision as

$$\boldsymbol{\Gamma}(\xi) = \sum_{n=-N(\rho,\epsilon)}^{N(\rho,\epsilon)} \boldsymbol{C}_n\mathrm{e}^{\mathrm{i}\xi n}, \tag{8}$$

where $\boldsymbol{C}_n$ is given element-wise as

$$[\boldsymbol{C}_n]_{j,k} = \mathrm{i}^n J_n(\rho_{j,k})\mathrm{e}^{-\mathrm{i}\phi_{j,k}n}. \tag{9}$$

In the following we drop the arguments of $N(\rho, \epsilon)$ for convenience, and simply write N.

It is a classical result [39] that $\boldsymbol{Q} \in \mathbb{K}$ if and only if there is a positive measure μ on $[0, 2\pi]$ such that

$$\boldsymbol{Q} = \int_0^{2\pi} \boldsymbol{\Gamma}(\xi)\, \mathrm{d}\mu(\xi). \tag{10}$$

Using (8) we conclude that $\boldsymbol{Q} \in \mathbb{K}$ if and only if

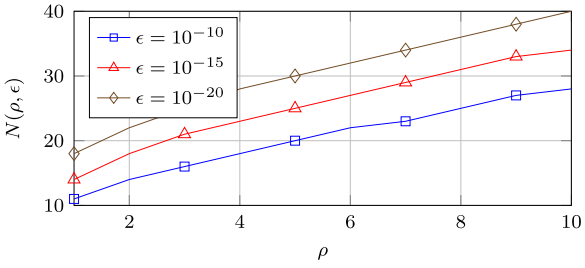

Fig. 1. The plot of $N(\rho, \epsilon)$ against ρ for three different values of ϵ.

$$\boldsymbol{Q} = \boldsymbol{S}(\boldsymbol{\nu}) := \nu_0 \boldsymbol{C}_0 + \sum_{n=1}^{N} (\boldsymbol{C}_n \nu_n + \boldsymbol{C}_{-n} \nu_n^*) \tag{11}$$

for some $\boldsymbol{\nu} = [\nu_0\ \nu_1\ \cdots\ \nu_N]^T \in \mathbb{R} \times \mathbb{C}^N$ satisfying

$$\nu_n = \int_0^{2\pi} e^{i\xi n}\, d\mu(\xi), \quad n = 0, 1, \ldots, N \tag{12}$$

for some positive measure μ on $[0, 2\pi]$. Next we use the celebrated theorem of Carathéodory and Fejér for the trigonometric moment problem [39]. Here u denotes the unit step function, i.e., $u(\xi) = 1$ if $\xi \geq 0$, and $u(\xi) = 0$ if $\xi < 0$.

Theorem 1. *Given $\boldsymbol{\nu} \in \mathbb{R} \times \mathbb{C}^N$ there exists some positive measure μ satisfying (12) if and only if the Hermitian Toeplitz matrix*

$$\boldsymbol{T}(\boldsymbol{\nu}) = \begin{bmatrix} \nu_0 & \nu_1^* & \cdots & \nu_N^* \\ \nu_1 & \nu_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \nu_1^* \\ \nu_N & \cdots & \nu_1 & \nu_0 \end{bmatrix}$$

is non - negative definite. In addition, if $K = \operatorname{rank}\{\boldsymbol{T}(\boldsymbol{\nu})\} \leq N$, then there is a unique positive measure μ satisfying (12), and it is of the form

$$\mu(\xi) = \sum_{k=1}^{K} \alpha_k u(\xi - \xi_k) \tag{13}$$

for some $\alpha_k > 0$ and $\xi_k \in [0, 2\pi]$, where $\{\alpha_k, \xi_k\}_{k=1}^K$ are obtained from the unique Vandermonde decomposition

$$\boldsymbol{T}(\boldsymbol{\nu}) = \sum_{k=1}^{K} \alpha_k \boldsymbol{\omega}(\xi_k) \boldsymbol{\omega}^*(\xi_k), \tag{14}$$

where $\boldsymbol{\omega}(\xi) := [1\ e^{i\xi} \cdots e^{i\xi N}]^T$.

From (11) and (12), and Theorem 1 we conclude that $\boldsymbol{Q} \in \mathbb{K}$ if and only if $\boldsymbol{Q} = \boldsymbol{S}(\boldsymbol{\nu})$ for some $\boldsymbol{\nu}$ in the set

$$\mathbb{G} = \{\boldsymbol{\nu} \in \mathbb{R} \times \mathbb{C}^N : \boldsymbol{T}(\boldsymbol{\nu}) \geqslant 0\}.$$

Here $\boldsymbol{A} \geqslant 0$ means $\boldsymbol{A}$ is a non-negative definite matrix.

Since $[\boldsymbol{\Gamma}(\xi)]_{1,1} = 1$, by using (10), and (12) we get $\boldsymbol{e}^* \boldsymbol{Q} \boldsymbol{e} = \nu_0$. Hence (4) is equivalent to

$$\underset{\boldsymbol{W} \in \mathbb{H}_L,\, \boldsymbol{\nu} \in \mathbb{G}}{\text{minimize}} \quad \{\nu_0 + \operatorname{Tr}(\boldsymbol{W})\}/2, \text{ subject to } \begin{bmatrix} \boldsymbol{W} & \boldsymbol{Y}^* \\ \boldsymbol{Y} & \boldsymbol{S}(\boldsymbol{\nu}) \end{bmatrix} \geqslant 0. \tag{15}$$

Let $\boldsymbol{\nu}_*$ be a solution to (15). Based on the value of $\boldsymbol{\nu}_*$ two cases can arise:

1. $\boldsymbol{T}(\boldsymbol{\nu}_*)$ is singular: In this case the second half of Theorem 1 applies, and $\boldsymbol{T}(\boldsymbol{\nu}_*)$ admits a unique Vandermonde decomposition
$$\boldsymbol{T}(\boldsymbol{\nu}_*) = \sum_{k=1}^{\mathring{K}} \mathring{\alpha}_k \boldsymbol{\omega}(\mathring{\xi}_k) \boldsymbol{\omega}^*(\mathring{\xi}_k), \tag{16}$$
where $\operatorname{rank}\{\boldsymbol{T}(\boldsymbol{\nu}_*)\} = \mathring{K}$. To compute $\{\mathring{\alpha}_k, \mathring{\xi}_k\}_{k=1}^{\mathring{K}}$, we write the individual elements of (16) in terms of the optimal value ν_{n*} of ν_n:
$$\nu_{n*} = \sum_{k=1}^{\mathring{K}} \mathring{\alpha}_k z_k^n, \quad z_k = e^{i\mathring{\xi}_k}, \tag{17}$$
and these equations can be solved for $\{\mathring{\alpha}_k, \mathring{\xi}_k\}$ via Prony's method [26], [40, Appendix A].
2. $\boldsymbol{T}(\boldsymbol{\nu}_*)$ is non-singular: When $\boldsymbol{Y}$ does not represent signals received from some targets (e.g. when the entries of $\boldsymbol{Y}$ are random numbers, then (15) may result a non-singular $\boldsymbol{T}(\boldsymbol{\nu}_*)$. In this case the Vandermonde decomposition (16) is no longer unique. When $\boldsymbol{Y}$ does represent some signals received from some targets, $\boldsymbol{T}(\boldsymbol{\nu}_*)$ is rarely non-singular. Nevertheless, to obtain a desirable solution from a non-singular $\boldsymbol{T}(\boldsymbol{\nu}_*)$, the previous authors on TVMA have suggested computing the Vandermonde decomposition of $\boldsymbol{T}(\boldsymbol{\nu}_*) - \delta \boldsymbol{I}$ instead, where δ is the smallest eigenvalue of $\boldsymbol{T}(\boldsymbol{\nu}_*)$ [40].

Regardless of whether $\boldsymbol{T}(\boldsymbol{\nu}_*)$ is singular or not, the optimal value of $\boldsymbol{Q}$ in (4) is $\boldsymbol{Q}_* = \boldsymbol{S}(\boldsymbol{\nu}_*)$, which by using (8), (11), and (17) gives

$$\boldsymbol{Q}_* = \boldsymbol{S}(\boldsymbol{\nu}_*) \overset{(11)}{=} \nu_{0*} \boldsymbol{C}_0 + \sum_{n=1}^{N} (\boldsymbol{C}_n \nu_{n*} + \boldsymbol{C}_{-n} \nu_{n*}^*) \overset{(17)}{=} \sum_{k=1}^{\mathring{K}} \mathring{\alpha}_k \sum_{n=-N}^{N} \boldsymbol{C}_n e^{i\mathring{\xi}_k n}$$
$$\overset{(18)}{=} \sum_{k=1}^{\mathring{K}} \mathring{\alpha}_k \boldsymbol{\Gamma}(\mathring{\xi}_k).$$

The above equality, according to the converse part of Lemma 1, leads to the required atomic decomposition of $\boldsymbol{Y}$ in terms of $\{\mathring{\alpha}_k, \mathring{\xi}_k\}_{k=1}^{\mathring{K}}$. When $\boldsymbol{T}(\boldsymbol{\nu}_*)$ is singular then $\{\mathring{\alpha}_k, \mathring{\xi}_k\}_{k=1}^{\mathring{K}}$ are unique, and the optimal atomic decomposition given by (3) is also unique. Otherwise, the optimal atomic decomposition is non-unique.

Recall that we truncate of the infinite sum (7) by omitting the negligible terms to obtain a finite sum (8). For this reason (15) is an approximation of (4). Nevertheless, when N chosen appropriately then approximation errors are not noticeable even when $\boldsymbol{E} = 0$. In practice $\boldsymbol{E} \neq 0$. Then it is enough to ensure that the approximation errors are below the noise floor. Taking N too small results systematic bias in the estimates due to approximation errors.

2.2. Noisy data

So far we have assumed $\boldsymbol{E} = 0$. Now we turn to the case where $\boldsymbol{E} \neq 0$. We assume that the rows of $\boldsymbol{E}$ are mutually uncorrelated, and the variance of any element of $\boldsymbol{E}$ is σ^2. Typically LP is large enough such that the law of large numbers holds:

$$\operatorname{Tr}(\boldsymbol{E}\boldsymbol{E}^*) \approx \sigma^2 LP. \tag{18}$$

The natural way to adapt the approach developed so far is to use a basis pursuit denoising (BPDN), where we estimate the noise-free component $\mathring{\boldsymbol{Y}}$ of $\boldsymbol{Y}$ together with $\boldsymbol{\nu}$ by solving

$$\underset{\boldsymbol{W} \in \mathbb{H}_L,\, \boldsymbol{\nu} \in \mathbb{G},\, \hat{\boldsymbol{Y}} \in \mathbb{C}^{P \times L}}{\text{minimize}} \quad \{\operatorname{Tr}(\boldsymbol{W}) + \nu_0\}/2,$$
$$\text{subject to } \begin{bmatrix} \boldsymbol{W} & \hat{\boldsymbol{Y}}^* \\ \hat{\boldsymbol{Y}} & \boldsymbol{S}(\boldsymbol{\nu}) \end{bmatrix} \geqslant 0 \quad \operatorname{Tr}\{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})(\boldsymbol{Y} - \hat{\boldsymbol{Y}})^*\} \leq \zeta \tag{19}$$

for some ζ that is a monotonic function of σ^2.

The main numerical difficulty associated with solving (21) is that the size of the matrix $\boldsymbol{W}$ is $L \times L$. Often L is large. Then the matrix in the LMI in (21) can be quite big. Nevertheless, by Lemma 2 [41, Proposition 1], [42] we can use an orthogonal factorization of $\boldsymbol{Y}$ to work with a matrix with a much smaller dimension.

Lemma 2. *Let P_1 be the rank of $\boldsymbol{Y}$. Consider an orthogonal factorization $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{V}^*$, where $\boldsymbol{V}$ is a $L \times P_1$ matrix with mutually orthogonal columns. For any given non-negative definite matrix $\boldsymbol{Q}$ the problems*

$$\underset{\boldsymbol{W}_1 \in \mathbb{H}_{P_1},\, \hat{\boldsymbol{Z}} \in \mathbb{C}^{P \times P_1}}{\text{minimize}} \quad \operatorname{Tr}(\boldsymbol{W}_1), \text{ subject to } \begin{bmatrix} \boldsymbol{W}_1 & \hat{\boldsymbol{Z}}^* \\ \hat{\boldsymbol{Z}} & \boldsymbol{Q} \end{bmatrix} \geqslant 0,$$
$$\operatorname{Tr}\{(\boldsymbol{Z} - \hat{\boldsymbol{Z}})(\boldsymbol{Z} - \hat{\boldsymbol{Z}})^*\} \leq \zeta \tag{20}$$

and

$$\underset{\boldsymbol{W} \in \mathbb{H}_L,\, \hat{\boldsymbol{Y}} \in \mathbb{C}^{P \times L}}{\text{minimize}} \quad \operatorname{Tr}(\boldsymbol{W}), \text{ subject to } \begin{bmatrix} \boldsymbol{W} & \hat{\boldsymbol{Y}}^* \\ \hat{\boldsymbol{Y}} & \boldsymbol{Q} \end{bmatrix} \geqslant 0 \operatorname{Tr}\{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})(\boldsymbol{Y} - \hat{\boldsymbol{Y}})^*\} \leq \zeta. \tag{21}$$

are equivalent. In particular, if $\boldsymbol{W}_{1*}, \hat{\boldsymbol{Z}}_*$ constitute the solution to (20) then the solution to (21) is $\boldsymbol{W}_* = \boldsymbol{V}\boldsymbol{W}_{1*}\boldsymbol{V}^*$ and $\hat{\boldsymbol{Y}}_* = \hat{\boldsymbol{Z}}_*\boldsymbol{V}^*$.

Using Lemma 2 we note that (21) is equivalent to

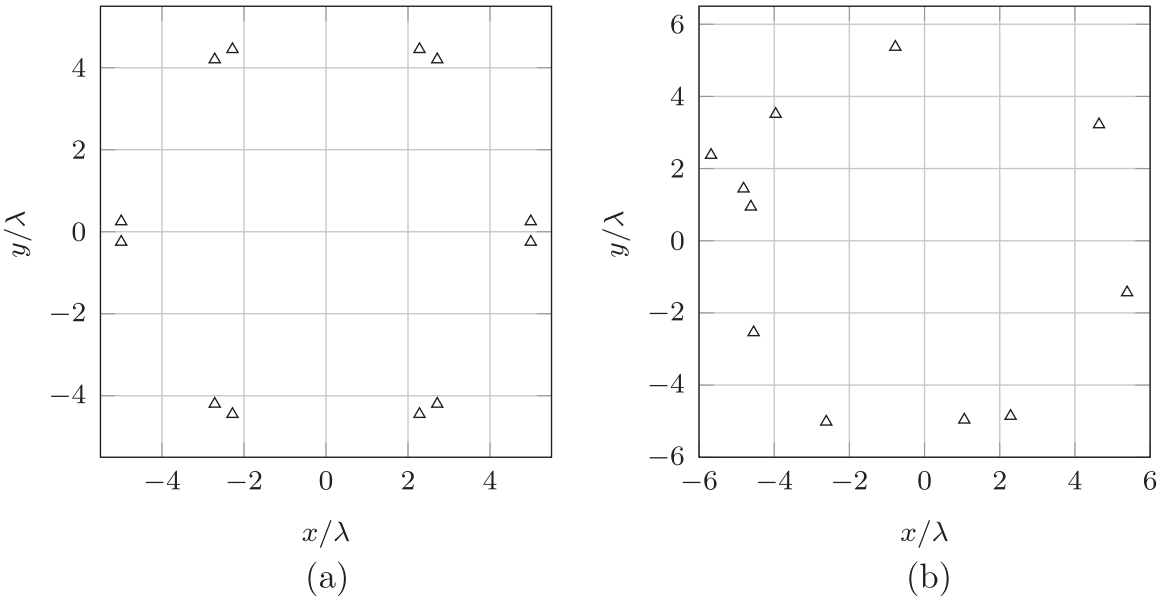

Fig. 2. (a) Array arrangement for the simulations in Figs. 3–6. (b) Array arrangement for the simulations in Fig. 7.

$$\underset{\boldsymbol{W}_1\in\mathbb{H}_{P_1},\,\boldsymbol{\nu}\in\mathbb{G},\,\mathring{\boldsymbol{Z}}\in\mathbb{C}^{P\times P_1}}{\text{minimize}} \{\nu_0+\text{Tr}(\boldsymbol{W}_1)\}/2, \text{ subject to } \begin{bmatrix}\boldsymbol{W}_1 & \mathring{\boldsymbol{Z}}^*\\ \mathring{\boldsymbol{Z}} & \boldsymbol{S}(\boldsymbol{\nu})\end{bmatrix}\succcurlyeq 0,$$

$$\text{Tr}\{(\boldsymbol{Z}-\mathring{\boldsymbol{Z}})(\boldsymbol{Z}-\mathring{\boldsymbol{Z}})^*\}\le\zeta. \tag{22}$$

The above semidefinite program can be solved via any of the popular solvers like SeDuMi [43], SDPT3 [44], etc. These solvers require us to represent (22) as a cone linear program, which is then solved using some primal dual path following scheme [45]. The LMI in (22) is converted into an inequality over a semidefinite cone, while the second inequality is converted into an inequality of over a second order cone. Since the size of the LMI is $N+1+P_1$, a path following method takes $O(\sqrt{N+1+P_1})$ iterations [45]. The underlying matrices associated with such cone linear programs are highly sparse [46]. When this sparsity is exploited, the worst case computational complexity per iteration typically depends on the cube of the number of real valued variables [46]. In (22) we have P_1^2 independent real valued variables in $\boldsymbol{W}_1$, $2P_1P$ real valued variables in $\mathring{\boldsymbol{Z}}$, and $2N+1$ real valued variables in $\boldsymbol{\nu}$. Hence the per iteration complexity is $O\{(P_1^2+2P_1P+2N+1)^3\}$. Next, we argue that a good choice for ζ is $\text{Tr}(\boldsymbol{E}\boldsymbol{E}^*)\approx\sigma^2LP$.

Lemma 3. *Suppose* $\text{Tr}(\boldsymbol{Z}\boldsymbol{Z}^*)^*>\zeta$, *and* $\mathring{\boldsymbol{Z}}_*$ *is the optimal value of* $\mathring{\boldsymbol{Z}}$ *in* (20). *Then* $\text{Tr}\{(\boldsymbol{Z}-\mathring{\boldsymbol{Z}}_*)(\boldsymbol{Z}-\mathring{\boldsymbol{Z}}_*)^*\}=\zeta$.

The proof of Lemma 3 appears in Appendix B. Unless we take $\zeta<\text{Tr}(\boldsymbol{Z}\boldsymbol{Z}^*)$, the denoising method (22) treats the observed data as noise. When $\zeta<\text{Tr}(\boldsymbol{Z}\boldsymbol{Z}^*)^*$, then $\text{Tr}\{(\boldsymbol{Z}-\mathring{\boldsymbol{Z}}_*)(\boldsymbol{Z}-\mathring{\boldsymbol{Z}}_*)^*\}=\zeta$. From Lemma 2 we know that the estimated noise $\boldsymbol{Y}-\mathring{\boldsymbol{Y}}_*=(\boldsymbol{Z}-\mathring{\boldsymbol{Z}}_*)\boldsymbol{V}^*$. Since the columns of $\boldsymbol{V}$ are mutually orthogonal, we get $\text{Tr}\{(\boldsymbol{Y}-\mathring{\boldsymbol{Y}}_*)(\boldsymbol{Y}-\mathring{\boldsymbol{Y}}_*)^*\}=\zeta$. Typically, LP is sufficiently large enough for the 'true noise' $\boldsymbol{E}$ to satisfy (18) rather closely. These motivate the choice $\zeta=LP\sigma^2$. In order to be able to do so, we must have an estimate of σ^2. But that can be done using classical techniques. Choose the orthogonal factorization $\boldsymbol{Y}=\boldsymbol{Z}\boldsymbol{V}^*$ to be the singular value decomposition (SVD) so that

$$\boldsymbol{Z}=\boldsymbol{U}\boldsymbol{\Sigma}, \tag{23}$$

i.e. $\boldsymbol{U}$ consists of the mutually orthogonal left singular vectors of $\boldsymbol{Y}$, while $\boldsymbol{\Sigma}$ is a diagonal matrix of singular values. From (23) we can estimate σ^2 using well known methods, some classical [47–49], and some more modern [50,51].

3. Gridless SPICE

In a spirit similar to [37,40], our results are also useful in deriving a gridless version of the SPICE algorithm [10,28]. Define the sample covariance matrix

$$\widehat{\boldsymbol{R}}=\boldsymbol{Y}\boldsymbol{Y}^*/L=\boldsymbol{Z}\boldsymbol{Z}^*/L.$$

SPICE estimates the 'true covariance matrix'

$$\boldsymbol{R}=\sum_{k=1}^{K}\boldsymbol{\gamma}(\xi_k)\boldsymbol{\gamma}^*(\xi_k)\mathsf{E}\{\|\boldsymbol{s}_k\|_2^2\}/L+\text{diag}(\sigma_1^2,\sigma_2^2,\ldots,\sigma_P^2) \tag{24}$$

by minimizing

$$\text{Tr}(\boldsymbol{R}^{-1}\widehat{\boldsymbol{R}}+\boldsymbol{R}\widehat{\boldsymbol{R}}^{-1}).$$

A grid based parameterization of $\boldsymbol{R}$ is typically employed to carry out this minimization. However, note that the first term in the right hand side of (24) resides in $\mathbb{K}$. Thus, the set of all admissible $\boldsymbol{R}$ is

$$\{\boldsymbol{Q}+\text{diag}(\sigma_1^2,\sigma_2^2,\ldots,\sigma_P^2):\,\boldsymbol{Q}\in\mathbb{K}\},$$

and we can parameterize the set of all admissible $\boldsymbol{R}$ via $\boldsymbol{\nu}\in\mathbb{G}$. Then after some standard manipulations using Schur complements we can cast SPICE as

$$\underset{\boldsymbol{W}\in\mathbb{H}_L,\,\boldsymbol{\nu}\in\mathbb{G},\,\sigma_1,\ldots,\sigma_P}{\text{minimize}} \text{Tr}[\boldsymbol{W}+\widehat{\boldsymbol{R}}^{-1}\{\boldsymbol{S}(\boldsymbol{\nu})+\boldsymbol{D}\}],$$

$$\text{subject to } \begin{bmatrix}\boldsymbol{W} & \boldsymbol{Z}^*\\ \boldsymbol{Z} & \boldsymbol{S}(\boldsymbol{\nu})+\boldsymbol{D}\end{bmatrix}\succcurlyeq 0 \quad \boldsymbol{D}=\text{diag}(\sigma_1^2,\sigma_2^2,\ldots,\sigma_P^2) \tag{25}$$

Using the solution $\boldsymbol{\nu}_*$ we can obtain a Vandermonde decomposition of $\boldsymbol{T}(\boldsymbol{\nu}_*)$, which then gives K and $\{\xi_k\}_{k=1}^K$.

4. Simulation results

In simulations we take P=12. Unless stated otherwise, we use the array configuration in Fig. 2(a). Here $r_p=5\lambda,\ p=1,2,\ldots,12$, where λ is the wavelength of propagation. In addition, $\theta_1=-\sin^{-1}(1/20)$ and $\theta_2=\sin^{-1}(1/20)$ to ensure that the distance between (r_1,θ_1) and (r_2,θ_2) is $\lambda/2$. To construct the 12 element array we repeat this strategy 6 times. For $p=0,1,\ldots,5$ we take

$$\theta_{2p+1}=2\pi p/6-\sin^{-1}(1/20),\quad \theta_{2p+2}=2\pi p/6+\sin^{-1}(1/20)$$

to ensure that the nearest neighbour of each sensor is at a distance $\lambda/2$. The configuration in Fig. 2(b) is used later to demonstrate the utility of

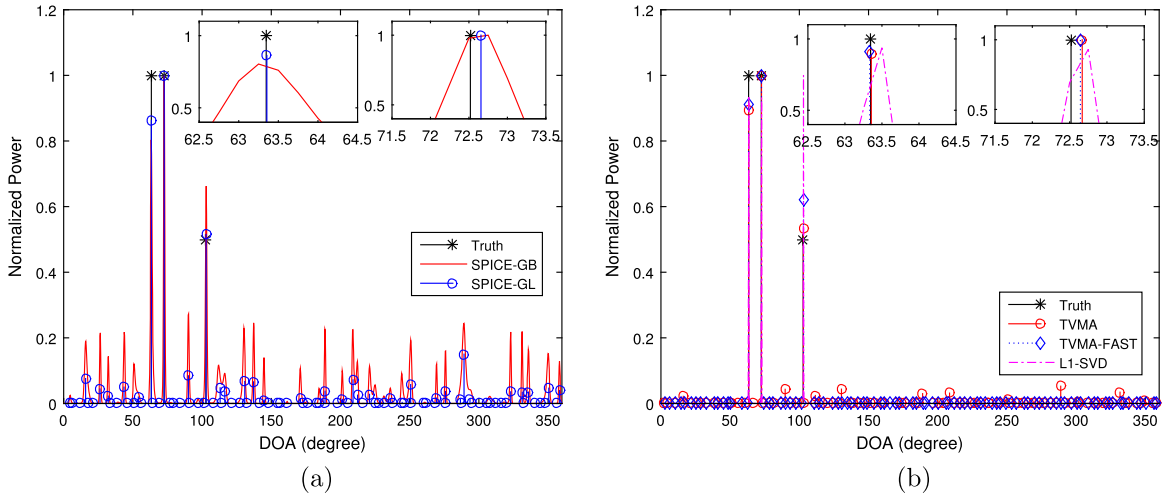

Fig. 3. Spatial spectrum obtained by different algorithms to resolve three uncorrelated sources with power [1, 1, 0.5] and azimuth angle [63.3466, 72.5189, 102.7684] degree and SNR=−2.5 dB. Total number of snapshots L=200. The areas around first and second sources are zoomed for better visualization. (a) SPICE-GB and SPICE-GL; (b) TVMA, TVMA-FAST, L1-SVD.

our algorithm when the sensor arrangement is rather irregular. We assume that all signals are impinging the array along the same elevation angle 90° [24]. Note that for any source, the azimuth angle $\xi \in [0, 360°)$ We compare the performance of proposed grid-less methods with other existing grid-based algorithms. We consider two versions of the TVMA algorithm. The first version, referred to as TVMA, solves (22) as described. The second version is motivated by L1-SVD algorithm [8]. Here we apply Akaike Information Criterion [47,48] to first estimate K. Then from the singular value decomposition $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$ we construct

$$\boldsymbol{Z}_1 = \boldsymbol{U}(:,1{:}\,K)\,\boldsymbol{\Sigma}(1{:}\,K,\,1{:}\,K). \tag{26}$$

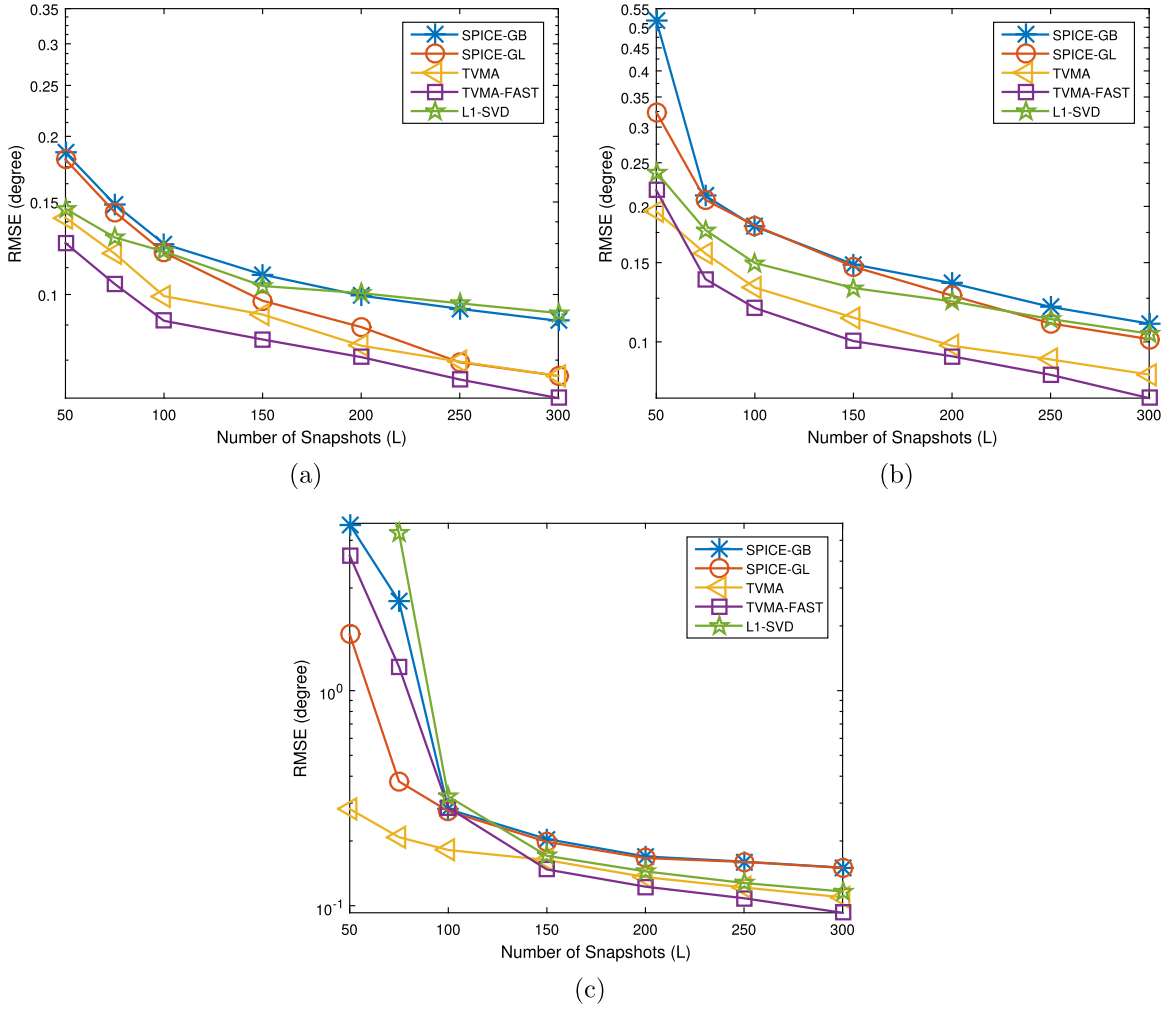

Fig. 4. RMSE versus number of snapshots for estimating uncorrelated DOAs by different algorithms. (a) SNR =0 dB, (b) SNR =−2.5 dB. (c) SNR =−5 dB.

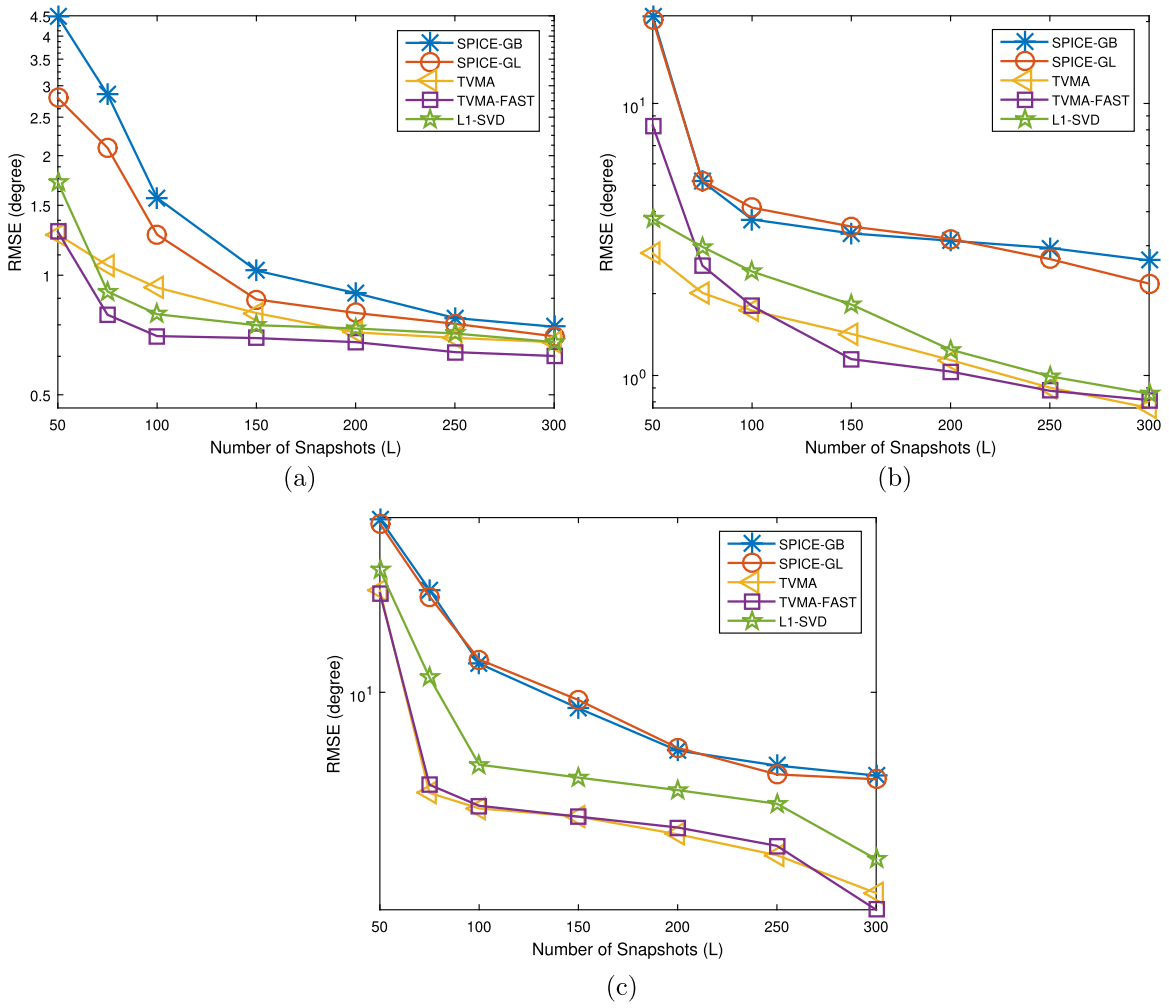

Fig. 5. RMSE versus number of snapshots for estimating correlated DOAs by different

Here we assume that $[\mathbf{\Sigma}]_{1,1} > [\mathbf{\Sigma}]_{2,2} > \cdots$. The algorithm obtained using $\boldsymbol{Z}_1$ instead of $\boldsymbol{Z}$ in (22) is referred to as TVMA-FAST. The construction of $\boldsymbol{Z}_1$ is similar to the preprocessing step in L1-SVD, and is known to result a faster algorithm with better noise rejection property [8]. We also implement grid-less SPICE as described in Section 3, and call it SPICE-GL. We refer to grid based SPICE [10] as SPICE-GB, and implement it as in [28]. SPICE-GB terminates when the relative change of objective function in two consecutive iterations is below 10^{-5}, or the maximum number of iterations, 500, is reached. For SPICE-GB and L1-SVD, we discretize the azimuth angle range $\theta \in [0, 360^\circ)$ into G=4×360=1440 uniform grid points. The SNR is defined as

$$\text{SNR} = 20\log_{10}(\|\mathring{\boldsymbol{Y}}\|_F/\|\boldsymbol{E}\|_F).$$

To implement SPICE-GL, TVMA and TVMA-FAST, we need an appropriate N (see the discussion above (8)). Using (5) we can compute $\rho = \max_{j,k}\rho_{j,k}$. We set $\epsilon = 10^{-10}$. Then we take N as the smallest number such that $|J_n| < \epsilon$ uniformly on the interval $[0, \rho]$ for all $|n| < N$. For the configuration in Fig. 2(a) we get N=92. We use the SeDuMi [43] solver via CVX modeller [52] to solve the proposed optimization problems..

4.1. Spectral comparison

Fig. 3(a) shows typical power spectra for 3 sources. Here $\xi_1 = 63.3466^\circ$, $\xi_2 = 75.5189^\circ$, and $\xi_3 = 102.7684^\circ$. The components of s_1 are drawn independently from a Gaussian distribution of mean zero and unit variance. We say this briefly as "the power of source 1 is unity". Similarly, the powers of source 2 and source 3 are 1 and 0.5, respectively. The number of snapshots L=200. The SNR is −2.5 dB. Fig. 3(a) shows the results for SPICE-GB and SPICE-GL. The areas around first and second sources are zoomed for better visualization. Both SPICE-GB and SPICE-GL produce many noise peaks, while the amplitudes of the noise peaks of SPICE-GL are smaller compared to SPICE-GB. Hence, the identification of DOAs is somewhat easier using SPICE-GL compared to SPICE-GB. Note in Fig. 3(a) that the locations of SPICE-GL's peaks almost coincide with SPICE-GB's local maxima.

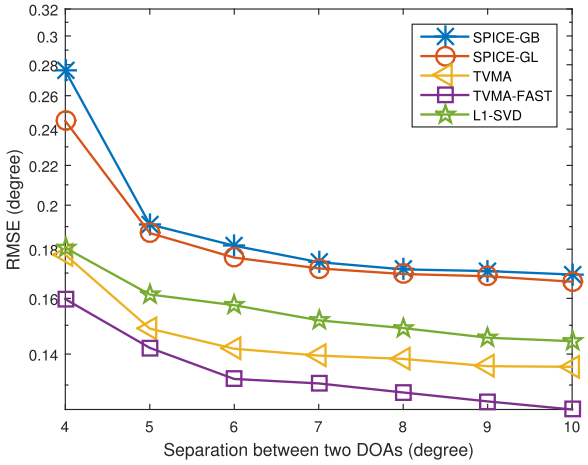

Fig. 6. RMSE of estimating closely spaced uncorrelated DOAs by different algorithms. Total snapshots L=200 and SNR =−5 dB.

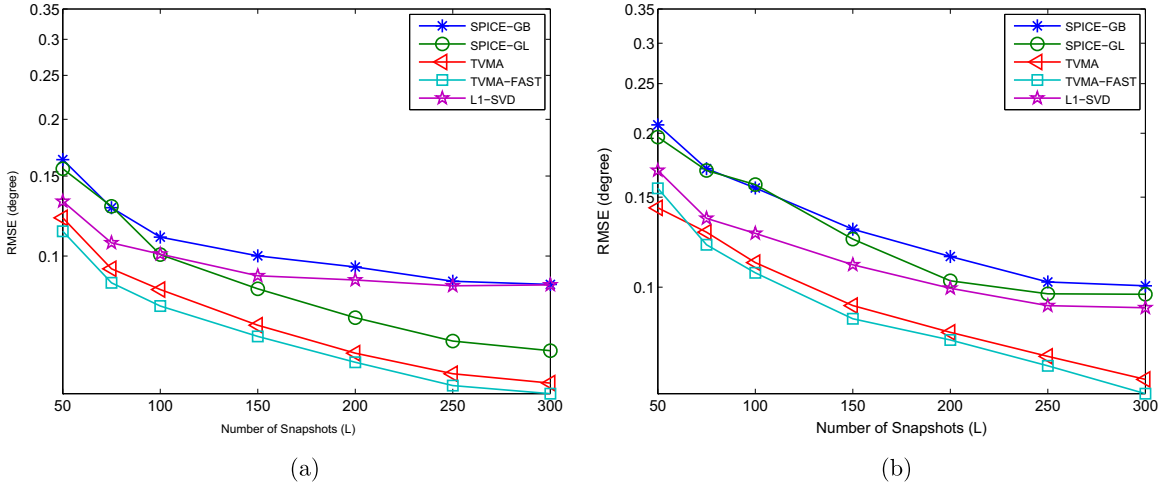

Fig. 7. RMSE versus number of snapshots for estimating uncorrelated DOAs with the array arrangement in Fig. 2(b). (a) SNR =0 dB, (b) SNR =−2.5 dB.

As expected, we have found that the relative heights of the noise peaks increase with decreasing the SNR value. Fig. 3(b) shows the spectra for TVMA, TVMA-FAST and L1-SVD. All three algorithms generate sparse spectra. The zoomed view in the figure shows that the algorithms can locate the first source more accurately compared to the second source. The average computation times (on a 2.8 GHz, 4 GB RAM PC) are as follows: SPICE-GB: 34.7 s, SPICE-GL: 3.99 s, L1-SVD: 2.1 s, TVMA: 4.65 s and TVMA-FAST: 3.24 s.

4.2. Estimation accuracy

To measure the DOA estimation accuracy of different algorithms we simulate k=2 unit power sources. We maintain $\xi_2 - \xi_1 = 10^{\circ}$, where ξ_1 takes values uniformly at random in $[0, 360^{\circ})$ in each simulation. The following results are based on 200 independent Monte Carlo simulation runs. To compute DOA estimation accuracy, we must obtain an estimate of the number of sources in each realization. As illustrated, SPICE generates many noise peaks. Therefore, for a fair comparison, we follow the procedure proposed in [11,36]. We assume the number of sources is known. For SPICE-GB and L1-SVD, the three largest peaks of the spectra are taken to estimate RMSE while it corresponds to the largest 3 components for TVMA and SPICE-GL.

In Fig. 4 we plot the root mean squared estimation error of different algorithms as a function of L for different values of SNR. Here the sources are uncorrelated. Note that SPICE-GL and SPICE-GB perform similarly. On average, L1-SVD performs bit better than SPICE-GB and SPICE-GL, while TVMA and TVMA-FAST outperform L1-SVD. TVMA-FAST performs better than TVMA for moderate SNR and moderate number of snapshots due to its noise rejection step. For the same reason L1-SVD outperforms TVMA in some cases when SNR is moderate. However this approach of truncating SVD in (26) does not work well with lower SNR and smaller L due to frequent erroneous estimation of K, when TVMA outperforms the other algorithms. Similar results for correlated sources are shown in Fig. 5. The simulation setting is same as above apart from the fact that the sources are now correlated with a correlation coefficient 0.9. The findings for correlated source case is similar as those of the uncorrelated sources apart from the fact the performance of all the algorithms deteriorate when the sources are correlated. It is noted that the performance of SPICE deteriorates more compared to other algorithms. This is due to the covariance fitting criteria used in SPICE (see (24)). SPICE [28] assumes that the phases of $\{s_k(j)\}$ are independent and uniformly distributed in $[0, 2\pi]$ which results in a diagonal structure of signal covariance matrix. In correlated case, this assumption is not valid. Although it has been claimed in [28, Section-II-A] that the inherent filtering effect of SPICE optimization can yield robust DOA estimation in correlated case. However, we find that at low SNR when correlated sources are closed to each other, the assumption of diagonal signal covariance matrix structure results in high leakage effect affecting the resolution performance...

In Fig. 6 we investigate the performance of different algorithms of estimating closely spaced uncorrelated DOAs at −5 dB SNR and L=200 snapshots. As before, we generated the first DOA uniformly at random in $[0, 360^{\circ})$. We vary $|\theta_1 - \theta_2|$ in different setups. As can be seen, the estimation accuracy of TVMA-FAST is the best compared to other algorithms. The performance of SPICE-GB and SPICE-GL deteriorate when the sources are closely spaced. TVMA performs better than L1-SVD. This is caused by frequent underestimation of K by L1-SVD when the sources are close to each other..

To further demonstrate the utility of the proposed algorithm for arbitrary sensor configurations, we consider the arrangement in Fig. 2(b). As before P=12. However the sensor locations are picked randomly within the region between two concentric circles with radii 5λ and 6.5λ. Fig. 7 shows the accuracy results for uncorrelated case. The findings are similar to those in Fig. 4..

5. Conclusions

We have presented a grid-less algorithm for TVMA based DOA estimation. The advantage of our method is that it can be applied to 2D arrays with arbitrary sensor configurations. To the best of our knowledge this is the only fully parametric algorithm capable of doing so. The proposed method has been tested using numerical simulations. It is faster than other grid based algorithms. When compared to the state of the art approaches, it shows superior performance when the number of snapshots is small. It also performs better when the sources are close to each other.

Appendix A. Proof of Lemma 1

We need the following result, which has been used also in several other papers on atomic norm minimization based methods, e.g., [40].

Proposition 1. *Suppose that the LMI*

$$\begin{bmatrix} \boldsymbol{W} & \boldsymbol{Y}^* \\ \boldsymbol{Y} & \boldsymbol{Q} \end{bmatrix} \succcurlyeq 0. \tag{27}$$

holds. Then for every $\overline{\boldsymbol{Q}}$ *satisfying* $\boldsymbol{Q} = \overline{\boldsymbol{Q}}\,\overline{\boldsymbol{Q}}^*$ *there exists a corresponding vector* $\overline{\boldsymbol{Y}}$ *such that* $\boldsymbol{Y} = \overline{\boldsymbol{Q}}\,\overline{\boldsymbol{Y}}$. *In addition, the minimum value of* $\mathrm{Tr}(\boldsymbol{W})$ *subject to the LMI (27) is achieved when* $\boldsymbol{W} = \boldsymbol{Y}^*\boldsymbol{Q}^\dagger\boldsymbol{Y} = \inf_{\overline{\boldsymbol{Y}}}\{\overline{\boldsymbol{Y}}^*\overline{\boldsymbol{Y}}\colon\ \boldsymbol{Y} = \overline{\boldsymbol{Q}}\,\overline{\boldsymbol{Y}}\}$.

Let $\boldsymbol{W}_*$, $\boldsymbol{Q}_*$ be the solutions to (4). Since $\boldsymbol{Q}_* \in \mathbb{K}$ we note by definition of $\mathbb{K}$ that there are strictly positive numbers $\mathring{\alpha}_k$ and frequencies $\mathring{\xi}_k \in [0, 2\pi]$ such that

$$\boldsymbol{Q}_* = \sum_{k=1}^{P} \mathring{\alpha}_k \boldsymbol{\Gamma}(\mathring{\xi}_k) = \sum_{k=1}^{P} [\boldsymbol{\gamma}(\mathring{\xi}_k)\sqrt{\mathring{\alpha}_k}][\boldsymbol{\gamma}(\mathring{\xi}_k)\sqrt{\mathring{\alpha}_k}]^*.$$

Here we don't require the above decomposition to be unique, and the value of P can be more than M. The above equation can be re-written as $\boldsymbol{Q}_* = \overline{\boldsymbol{Q}}\,\overline{\boldsymbol{Q}}^*$, with

$$\overline{\boldsymbol{Q}} = [\boldsymbol{\gamma}(\xi_1)\sqrt{\mathring{\alpha}_1} \ \cdots \ \boldsymbol{\gamma}(\xi_P)\sqrt{\mathring{\alpha}_P}]$$

Since $\boldsymbol{W}_*$, $\boldsymbol{Q}_*$ are the solutions to (4), the linear matrix inequality (LMI)

$$\begin{bmatrix} \boldsymbol{W}_* & \boldsymbol{Y}^* \\ \boldsymbol{Y} & \boldsymbol{Q}_* \end{bmatrix} \succcurlyeq 0$$

must hold. Hence Proposition 2 ensures the existence of complex numbers $\{\boldsymbol{\beta}_k\}_{k=1}^{P}$ such that

$$\boldsymbol{Y} = \sum_{k=1}^{P} \boldsymbol{\gamma}(\mathring{\xi}_k)\boldsymbol{\beta}_k^*\sqrt{\mathring{\alpha}_k}, \quad \boldsymbol{W}_* = \sum_{k=1}^{P} \boldsymbol{\beta}_k\boldsymbol{\beta}_k^*. \tag{28}$$

Next we show $\|\boldsymbol{\beta}_k\|_2 = \sqrt{\mathring{\alpha}_k}$ by contradiction. Suppose $\|\boldsymbol{\beta}_k\|_2 \neq \sqrt{\mathring{\alpha}_k}$. Take

$$\widehat{\boldsymbol{W}} = \sum_{k=1}^{P} \frac{\boldsymbol{\beta}_k}{\|\boldsymbol{\beta}_k\|_2}\|\boldsymbol{\beta}_k\|_2\sqrt{\mathring{\alpha}_k}\frac{\boldsymbol{\beta}_k^*}{\|\boldsymbol{\beta}_k\|_2},$$

$$\widehat{\boldsymbol{Q}} = \sum_{k=1}^{P} \boldsymbol{\gamma}(\mathring{\xi}_k)\|\boldsymbol{\beta}_k\|_2\sqrt{\mathring{\alpha}_k}\boldsymbol{\gamma}^*(\mathring{\xi}_k),$$

so that $\widehat{\boldsymbol{Q}} \in \mathbb{K}$. Verify that

$$\begin{bmatrix} \widehat{\boldsymbol{W}} & \boldsymbol{Y}^* \\ \boldsymbol{Y} & \widehat{\boldsymbol{Q}} \end{bmatrix} = \sum_{k=1}^{P} \begin{bmatrix} \frac{\boldsymbol{\beta}_k}{\|\boldsymbol{\beta}_k\|_2} \\ \boldsymbol{\gamma}(\mathring{\xi}_k) \end{bmatrix} \|\boldsymbol{\beta}_k\|_2\sqrt{\mathring{\alpha}_k} \begin{bmatrix} \frac{\boldsymbol{\beta}_k}{\|\boldsymbol{\beta}_k\|_2} \\ \boldsymbol{\gamma}(\mathring{\xi}_k) \end{bmatrix}^*$$

is non-negative definite, and thereby $\widehat{\boldsymbol{W}}$, $\widehat{\boldsymbol{Q}}$ belong to the feasible set of the optimization problem (4). In addition,

$$\mathrm{Tr}(\boldsymbol{W}_* - \widehat{\boldsymbol{W}}) + \boldsymbol{e}^*(\boldsymbol{Q}_* - \widehat{\boldsymbol{Q}})\boldsymbol{e} \quad = \sum_{k=1}^{P} \{\mathring{\alpha}_k + \|\boldsymbol{\beta}_k\|_2^2 - 2\|\boldsymbol{\beta}_k\|_2\sqrt{\mathring{\alpha}_k}\} > 0,$$

but that leads to a contradiction since $\boldsymbol{W}_*$, $\boldsymbol{Q}_*$ are the solutions to (4). Hence $\|\boldsymbol{\beta}_k\|_2 = \sqrt{\mathring{\alpha}_k}$, and

$$\mathrm{Tr}(\boldsymbol{W}_*) = \boldsymbol{e}^*\boldsymbol{Q}_*\boldsymbol{e} = \{\mathrm{Tr}(\boldsymbol{W}_*) + \boldsymbol{e}^*\boldsymbol{Q}_*\boldsymbol{e}\}/2 = \sum_k \mathring{\alpha}_k.$$

Also we can write $\boldsymbol{\beta}_k = \mathring{\boldsymbol{\phi}}_k\mathring{\alpha}_k$ where $|\mathring{\boldsymbol{\phi}}_k| = 1$. Hence the first equation in (28) gives us an atomic decomposition of $\boldsymbol{Y}$:

$$\boldsymbol{Y} = \sum_{k=1}^{P} \boldsymbol{\gamma}(\mathring{\xi}_k)\mathring{\alpha}_k\mathring{\boldsymbol{\phi}}_k^*, \tag{29}$$

and therefore using (3) we infer that

$$\|\boldsymbol{Y}\|_{\mathcal{A}} \leq \sum_{k=1}^{P} \mathring{\alpha}_k = \{\mathrm{Tr}(\boldsymbol{W}_*) + \boldsymbol{e}^*\boldsymbol{Q}_*\boldsymbol{e}\}/2. \tag{30}$$

Now consider a solution to (3). Such a solution constitute some $\check{P}$ positive numbers $\{\check{\alpha}_k\}_{k=1}^{\check{P}}$, associated numbers $\{\check{\xi}_k\}_{k=1}^{\check{P}}$ with each $\check{\xi}_k \in [0, 2\pi]$, and unit norm vectors $\{\check{\boldsymbol{\phi}}_k\}_{k=1}^{\check{P}}$ such that

$$\|\boldsymbol{Y}\|_{\mathcal{A}} = \sum_{k=1}^{\check{P}} \check{\alpha}_k, \tag{31}$$

and in addition, the atomic decomposition

$$\boldsymbol{Y} = \sum_{k=1}^{P} \boldsymbol{\gamma}(\check{\xi}_k)\check{\alpha}_k\check{\boldsymbol{\phi}}_k^* \tag{32}$$

holds. Take

$$\check{\boldsymbol{W}} = \sum_{k=1}^{\check{P}} \check{\boldsymbol{\phi}}_k \check{\alpha}_k \check{\boldsymbol{\phi}}_k^*, \quad \widehat{\boldsymbol{Q}} = \sum_{k=1}^{P} \boldsymbol{\gamma}(\check{\xi}_k)\check{\alpha}_k \boldsymbol{\gamma}^*(\check{\xi}_k),$$

so that $\check{\boldsymbol{Q}} \in \mathbb{K}$, and in addition

$$\begin{bmatrix} \check{\boldsymbol{W}} & \boldsymbol{Y}^* \\ \boldsymbol{Y} & \check{\boldsymbol{Q}} \end{bmatrix} = \sum_{k=1}^{\check{P}} \begin{bmatrix} \check{\boldsymbol{\phi}}_k \\ \boldsymbol{\gamma}(\check{\xi}_k) \end{bmatrix} \check{\alpha}_k \begin{bmatrix} \check{\boldsymbol{\phi}}_k \\ \boldsymbol{\gamma}(\check{\xi}_k) \end{bmatrix}^*$$

is non-negative definite. Thus $\check{\boldsymbol{W}}$, $\check{\boldsymbol{Q}}$ belong to the feasible set of the optimization problem (4). Since $\boldsymbol{W}_*$, $\boldsymbol{Q}_*$ are the solutions to (4) we conclude that

$$\{\mathrm{Tr}(\boldsymbol{W}_*) + \boldsymbol{e}^*\boldsymbol{Q}_*\boldsymbol{e}\}/2 = \sum_k \mathring{\alpha}_k \quad \le \{\mathrm{Tr}(\check{\boldsymbol{W}}) + \boldsymbol{e}^*\check{\boldsymbol{Q}}\boldsymbol{e}\}/2 = \sum_k \check{\alpha}_k = \|\boldsymbol{Y}\|_{\mathcal{A}}.$$

This observation and (30) imply that

$$\|\boldsymbol{Y}\|_{\mathcal{A}} = \{\mathrm{Tr}(\boldsymbol{W}_*) + \boldsymbol{e}^*\boldsymbol{Q}_*\boldsymbol{e}\}/2.$$

Consequently, it follows that

- The atomic decomposition (29) obtained from the solution of (4) gives an optimal solution to (3); and
- An atomic decomposition (32) achieving the optimality criterion (31) gives an optimal solution $\check{\boldsymbol{W}}$ and $\check{\boldsymbol{Q}}$ to (4).

Appendix B. Proof of Lemma 3

Proof. We need the following result, which has been used also in several other papers on atomic norm minimization based methods, e.g., [40].

Proposition 2. *Suppose a fixed $\boldsymbol{Q} \succcurlyeq 0$ is given, and $\overline{\boldsymbol{Q}}$ is any matrix satisfying $\boldsymbol{Q} = \overline{\boldsymbol{Q}}\overline{\boldsymbol{Q}}^*$. Then the LMI*

$$\begin{bmatrix} \boldsymbol{W}_1 & \mathring{\boldsymbol{Z}}^* \\ \mathring{\boldsymbol{Z}} & \boldsymbol{Q} \end{bmatrix} \succcurlyeq 0. \tag{33}$$

holds if and only if there exists $\overline{\boldsymbol{Z}}$ such that $\mathring{\boldsymbol{Z}} = \overline{\boldsymbol{Q}}\overline{\boldsymbol{Z}}^*$, and $\boldsymbol{W}_1 \succcurlyeq \overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^*$. Given such $\overline{\boldsymbol{Z}}$ the value of $\mathrm{Tr}(\boldsymbol{W}_1)$ is minimized $\boldsymbol{W}_1 = \overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^*$.

Take any $\overline{\boldsymbol{Q}}$ satisfying $\boldsymbol{Q} = \overline{\boldsymbol{Q}}\overline{\boldsymbol{Q}}^*$. Using Proposition 2 we see that (20) is equivalent to the problem

$$\underset{\overline{\boldsymbol{Z}} \in \mathbb{C}^{P \times P_1}}{\text{minimize}} \quad \mathrm{Tr}\{\overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^*\}, \text{ subject to } \quad \mathrm{Tr}\{(\boldsymbol{Z} - \overline{\boldsymbol{Q}}\overline{\boldsymbol{Z}}^*)(\boldsymbol{Z} - \overline{\boldsymbol{Q}}\overline{\boldsymbol{Z}}^*)^*\} \le \zeta \tag{34}$$

in the sense that the solution $\overline{\boldsymbol{Z}}_*$ to (34) gives the solutions $\boldsymbol{W}_{1*} = \overline{\boldsymbol{Z}}_*\overline{\boldsymbol{Z}}_*^*$ and $\mathring{\boldsymbol{Z}}_* = \overline{\boldsymbol{Q}}\overline{\boldsymbol{Z}}_*^*$ to (20). The Lagrangian associated with (34) is given by

$$L(\overline{\boldsymbol{Z}}, \lambda) = \mathrm{Tr}\{\overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^*\} - \lambda[\zeta - \mathrm{Tr}\{(\boldsymbol{Z} - \overline{\boldsymbol{Q}}\overline{\boldsymbol{Z}}^*)(\boldsymbol{Z} - \overline{\boldsymbol{Q}}\overline{\boldsymbol{Z}}^*)^*\}],$$

where the Lagrange multiplier $\lambda \ge 0$. We denote the dual optimal point by λ_*. The KKT complementary slackness condition associated with (34) asserts either $\lambda_* = 0$, or

$$\mathrm{Tr}\{(\boldsymbol{Z} - \overline{\boldsymbol{Q}}\overline{\boldsymbol{Z}}_*^*)(\boldsymbol{Z} - \overline{\boldsymbol{Q}}\overline{\boldsymbol{Z}}_*^*)^*\} = \zeta. \tag{35}$$

Suppose (35) does not hold, i.e. $\lambda_* = 0$. Then, since the derivative of L with respect to any element of $\overline{\boldsymbol{Z}}$ must vanish when evaluated at $(\overline{\boldsymbol{Z}}_*, \lambda_*)$, we get $\overline{\boldsymbol{Z}}_* = 0$. Consequently, the inequality constraint in (34) must hold with $\overline{\boldsymbol{Z}}_* = 0$, while (35) cannot hold. Together these imply $\mathrm{Tr}(\boldsymbol{Z}\boldsymbol{Z}^*)^* < \zeta$. But that contradicts the assumption of the Lemma, meaning (35) must hold. Since $\mathring{\boldsymbol{Z}}_* = \overline{\boldsymbol{Q}}\overline{\boldsymbol{Z}}_*^*$, the proof is complete.□

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] J. Capon, High-resolution frequency-wavenumber spectrum analysis, Proc. IEEE 57 (8) (1969) 1408–1418.
[3] P. Stoica, H. Li, J. Li, A new derivation of the apes filter, IEEE Signal Process. Lett. 6 (1999) 205–206.
[4] R. Schmidt, Multiple emitter location and signal parameter estimation, Antennas Propag., IEEE Trans. on 34 (3) (1986) 276–280.
[5] R. Roy, T. Kailath, Esprit-estimation of signal parameters via rotational invariance techniques, Acoust., Speech Signal Process., IEEE Trans. on 37 (7) (1989) 984–995.
[6] M. Viberg, B. Ottersten, Sensor array processing based on subspace fitting, IEEE Trans. Signal Process. 39 (5) (1991) 1110–1121.
[7] P. Stoica, K.C. Sharman, Maximum likelihood methods for direction-of-arrival estimation, IEEE Trans. Acoust., Speech Signal Process. 38 (7) (1990) 1132–1143.
[8] D. Malioutov, M. Cetin, A. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, Signal Process., IEEE Trans. 53 (8) (2005) 3010–3022.
[9] M. Hyder, K. Mahata, Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation, Signal Process., IEEE Trans. 58 (2010) 4646–4655.
[10] P. Stoica, P. Babu, J. Li, SPICE: a sparse covariance-based estimation method for array processing, Signal Process., IEEE Trans. 59 (2) (2011) 629–638.
[11] P. Stoica, P. Babu, SPICE and LIKES two hyperparameter-free methods for sparse-parameter estimation, Signal Process. 92 (7) (2012) 1580–1590.
[12] 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.
[13] P.P. Vaidyanathan, P. Pal, Sparse sensing with co-prime samplers and arrays, IEEE Trans. Signal Process. 59 (2) (2011) 573–586.
[14] P. Pal, P.P. Vaidyanathan, Coprime sampling and the music algorithm, in: Proceedings of Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE, pp. 289–294.
[15] A. Moffet, Minimum-redundancy linear arrays, IEEE Trans. Antennas Propag. 16 (2) (1968) 172–175.
[16] B. Bhaskar, G. Tang, B. Recht, Atomic norm denoising with applications to line spectral estimation, Signal Process., IEEE Trans. 61 (23) (2013) 5987–5999.
[17] Y. Li, Y. Chi, Off-the-grid line spectrum denoising and estimation with multiple measurement vectors, IEEE Trans. Signal Process. 64 (5) (2016) 1257–1269.
[18] B. Lin, J. Liu, M. Xie, J. Zhu, Super-resolution doa estimation using single snapshot via compressed sensing off the grid, in: Proceedings of IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2014, pp. 825–829.
[19] Z. Tan, Y.C. Eldar, A. Nehorai, Direction of arrival estimation using co-prime arrays: a super resolution viewpoint, IEEE Trans. Signal Process. 62 (21) (2014) 5565–5576.
[20] 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.
[21] P. Pal, P.P. Vaidyanathan, A grid-less approach to underdetermined direction of arrival estimation via low rank matrix denoising, IEEE Signal Process. Lett. 21 (6) (2014) 737–741.
[22] Y. Pan, H. Zhu, N. Tai, X. Zhang, N. Yuan, 2-d off-grid doa estimation using sparse bayesian learning with l-shape array, in: Proceedings of IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC),

2015, pp. 1–6.
[23] C.P. Mathews, M.D. Zoltowski, Eigenstructure techniques for 2-d angle estimation with uniform circular arrays, IEEE Trans. Signal Process. 42 (9) (1994) 2395–2407.
[24] R. Goossens, H. Rogier, S. Werbrouck, UCA root-music with sparse uniform circular arrays, Signal Process., IEEE Trans. 56 (8) (2008) 4095–4099.
[25] M. Pesavento, J.F. Bohme, Direction of arrival estimation in uniform circular arrays composed of directional elements, in: Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002, pp. 503–507.
[26] P. Stoica, R. Moses, Spectral Analysis of Signals, Prentice-Hall, Upper Saddle River, NJ, 2005.
[27] T. Yardibi, J. Li, P. Stoica, M. Xue, A. Baggeroer, Source localization and sensing: a nonparametric iterative adaptive approach based on weighted least squares, Aerosp. Electron. Syst., IEEE Trans. 46 (1) (2010) 425–443.
[28] 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, Signal Process., IEEE Trans. 59 (1) (2011) 35–47.
[29] Y. Chi, L.L. Scharf, A. Pezeshki, A.R. Calderbank, Sensitivity to basis mismatch in compressed sensing, IEEE Trans. Signal Process. 59 (5) (2011) 2182–2195.
[30] Z. Tan, P. Yang, A. Nehorai, Joint sparse recovery method for compressed sensing with structured dictionary mismatches, IEEE Trans. Signal Process. 62 (19) (2014) 4997–5008.
[31] D. Malioutov, M. Cetin, A. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, Signal Process., IEEE Trans. 53 (8) (2005) 3010–3022.
[32] M.F. Duarte, R.G. Baraniuk, Spectral compressive sensing, Appl. Comput. Harmon. Anal. 35 (1) (2013) 111–129.
[33] E.J. Candès, C. Fernandez-Granda, Towards a mathematical theory of super-resolution, Commun. Pure Appl. Math. 67 (6) (2014) 906–956 [Online][. Available: ⟨http://dx.doi.org/10.1002/cpa.21455⟩].
[34] V. Chandrasekaran, B. Recht, P.A. Parrilo, A.S. Willsky, The convex geometry of linear inverse problems, Found. Comput. Math. 12 (6) (2012) 805–849 [Online][. Available: ⟨http://dx.doi.org/10.1007/s10208-012-9135-7⟩].
[35] Z. Yang, L. Xie, Exact joint sparse frequency recovery via optimization methods, IEEE Trans. Signal Process. 64 (19) (2016) 5145–5157.
[36] Z. Yang, L. Xie, C. Zhang, A discretization-free sparse and parametric approach for linear array signal processing, Signal Process., IEEE Trans. on 62 (19) (2014) 4959–4973.
[37] K. Mahata, M. Hyder, Frequency estimation from arbitrary time samples, 2015, to appear in IEEE Transactions on Signal Processing.
[38] M. Abramowitz, I.A. Stegun (Eds.), Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover, New York, 1965.
[39] M.G. Krein, A.A. Nudelman, The Markov moment problem and extremal problems: Ideas and problems of P. L. Chebyshev and A. A. Markov and their further development, ser. Translations of Mathematical Monographs. American Mathematical Society, Providence, R.I., 1977, vol. 50, translated from the Russian by D. Louvish.
[40] Z. Yang, L. Xie, On gridless sparse methods for line spectral estimation from complete and incomplete data, Signal Process., IEEE Trans. 63 (12) (2015) 3139–3153.
[41] Z. Yang, L. Xie, Enhancing sparsity and resolution via reweighted atomic norm minimization, IEEE Trans. Signal Process. 64 (4) (2016) 995–1006.
[42] S. Haghighatshoar, G. Caire, Channel vector subspace estimation from low-dimensional projections, arXiv preprint arxiv:1509.07469 [cs.IT], 2016.
[43] J.F. Sturm, Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones, Optim. Methods Softw. 11–12 (1999) 625–653 [version 1.05 available from ⟨http://fewcal.kub.nl/sturm⟩.].
[44] R.H. Tutuncu, K. Toh, M. Todd, Solving semidefinite-quadratic-linear programs using sdpt3, Math. Program., Ser. B 95 (2003) 189–217.
[45] A. Ben-Tal, A.S. Nemirovskiaei, Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2001.
[46] J.F. Sturm, Implementation of interior-point methods for mixed semidefinite and second order cone optimization problems, Optim. Methods Softw. 17 (6) (2002) 1105–1154.
[47] M. Wax, T. Kailath, Detection of signals by information theoretic criteria, IEEE Trans. Acoust., Speech, Signal Process. 33 (2) (1985) 387–392.
[48] M. Kaveh, H. Wang, H. Hung, On the theoretical performance of a class of estimators of the number of narrow-band sources, IEEE Trans. Acoust., Speech, Signal Process. 35 (9) (1987) 1350–1352.
[49] W. Xu, M. Kaveh, Analysis of the performance and sensitivity of eigendecomposition-based detectors, IEEE Trans. Signal Process 43 (6) (1995) 1413–1426.
[50] B. Nadler, A. Kontorovich, Model selection for sinusoids in noise: statistical analysis and a new penalty term, IEEE Trans. Signal Process. 59 (4) (2011) 1333–1345.
[51] A. Mariani, A. Giorgetti, M. Chiani, Model order selection based on information theoretic criteria: design of the penalty, IEEE Trans. Signal Process. 63 (11) (2015) 2779–2789.
[52] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, ⟨http://cvxr.com/cvx⟩, March 2014.