

This document is downloaded from DR-NTU, Nanyang Technological University Library, Singapore.

Title	A discretization-free sparse and parametric approach for linear array signal processing
Author(s)	Yang, Zai; Xie, Lihua; Zhang, Cishen
Citation	Yang, Z., Xie, L., & Zhang, C. (2014). A discretization-free sparse and parametric approach for linear array signal processing. IEEE transactions on signal processing, 62(19), 4959-4973.
Date	2014
URL	http://hdl.handle.net/10220/24496
Rights	© 2014 IEEE. This is the author created version of a work that has been peer reviewed and accepted for publication by IEEE Transactions on Signal Processing, IEEE. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [http://dx.doi.org/10.1109/TSP.2014.2339792].

A Discretization-Free Sparse and Parametric Approach for Linear Array Signal Processing

Zai Yang, Lihua Xie, *Fellow, IEEE*, and Cishen Zhang

Abstract—Direction of arrival (DOA) estimation in array processing using uniform/sparse linear arrays is concerned in this paper. While sparse methods via *approximate* parameter discretization have been popular in the past decade, the discretization may cause problems, e.g., modeling error and increased computations due to dense sampling. In this paper, an *exact* discretization-free method, named as sparse and parametric approach (SPA), is proposed for uniform and sparse linear arrays. SPA carries out parameter estimation in the continuous range based on well-established covariance fitting criteria and convex optimization. It guarantees to produce a sparse parameter estimate without discretization required by existing sparse methods. Theoretical analysis shows that the SPA parameter estimator is a large-snapshot realization of the maximum likelihood estimator and is statistically consistent (in the number of snapshots) under uncorrelated sources. Other merits of SPA include improved resolution, applicability to arbitrary number of snapshots, robustness to correlation of the sources and no requirement of user-parameters. Numerical simulations are carried out to verify our analysis and demonstrate advantages of SPA compared to existing methods.

Index Terms—Array processing, DOA estimation, sparse and parametric approach (SPA), continuous parameter estimation, compressed sensing.

I. INTRODUCTION

Farfield narrowband source localization based on observed snapshots of a sensor array is a major problem in array signal processing [1], [2], known also as direction of arrival (DOA) estimation. The difficulty of the problem arises from the fact that the observed snapshots are nonlinear functions of the directions of interest. According to estimation schemes adopted, existing methods for the DOA estimation can be classified into three categories: parametric, nonparametric and semiparametric, which are described as follows.

Parametric methods explicitly carry out parameter estimation using optimization or other methods. A prominent example is the nonlinear least squares (NLS) method (see, e.g., [2]), which adopts the least squares criterion and has a strong statistical motivation. However, NLS admits the following two shortcomings: 1) it requires the knowledge of the source number which is typically unavailable, and 2) its global optimum cannot be guaranteed with a practically efficient

algorithm due to nonconvexity caused by the nonlinearity nature. Either of them may greatly degrade the parameter estimation performance. MUSIC [3], [4] does the parameter estimation by studying the subspace of the data covariance matrix and is a large-snapshot realization of the maximum likelihood (ML) method in the case of uncorrelated sources. But it also requires the source number and does not have reliable performance when the number of snapshots is small or correlation exists between the sources. Other subspace-based methods include ESPRIT and their variants (see [2] for a complete review). In contrast, a *nonparametric* method typically produces a *dense* spectrum whose peaks are interpreted as source directions. So an alternative name *dense* is also used. Examples include conventional beamformer and MVDR (or Capon's method) (see e.g., [2]) and recently introduced iterative adaptive approach (IAA) [5] which eliminates to a large extent leakage of the beamformer and is robust to correlation of the sources. But it will be shown in this paper that IAA suffers from resolution limit, especially in the moderate/low SNR regime. While the classification above is consistent with [2], MUSIC is sometimes categorized in the literature as a nonparametric/dense method by interpreting the plot of its objective function as its power spectrum whose peaks are the optima of the directions.

Semiparametric methods have been popular in the past decade which remove the direction variables in the observation model by parameter discretization and transform the nonlinear parameter estimation problem into a sparse signal recovery problem under a linear model, followed by some sparse signal recovery technique and support detection of the sparse solution. In particular, the continuous direction range is approximated by a set of discrete grid points under the assumption that the grid is fine enough such that any of the true sources lies on (practically, close to) some grid point. After that, the knowledge is exploited that each of the expanded vector of source signals [composed of (virtual) source signals from candidate directions on the grid] is a sparse signal since the grid size is greatly larger than the source number, and a sparse solution is sought after. Finally, the directions are retrieved from the support of the sparse solution. Consequently, the semiparametric methods are also named as *sparse* methods in contrast to *dense*. In principle, a source will be detected once the estimated source signal/source power from some direction is nonzero. Following from the literature of sparse signal representation (SSR) and later developed compressed sensing (CS) [6]–[8], ℓ_1 norm minimization and other sparse signal recovery techniques have been widely used in the semiparametric methods (see, e.g., [9]–[19]). However, their theoretical

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Z. Yang and L. Xie are with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore (e-mail: {yangzai, elhxie}@ntu.edu.sg).

C. Zhang is with the Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Hawthorn VIC 3122, Australia (e-mail: cishenzhang@swin.edu.au).

support based on SSR and CS cannot be applied due to dense discretization. Moreover, the sparse recovery techniques usually require one or more practically unknown parameters, e.g., the noise statistics, the source number or regularization parameters, etc. The semiparametric iterative covariance-based estimation (SPICE) method [20]–[22] is a breakthrough of the semiparametric methods, in which covariance fitting criteria are adopted with sound statistical motivation, the source power and noise variance(s) are estimated in a natural manner, no user-parameters are required and a connection to the ℓ_1 norm minimization is shown. In principle, the semiparametric methods are approximation methods due to the discretization scheme adopted. Furthermore, grid selection remains a major problem since 1) a coarse grid leads to a high modeling error and 2) too dense a grid is computationally prohibitive and might result in computational instability (see, e.g., [23], [24]). To alleviate the drawbacks of the discretization, preliminary results have been obtained in [13], [16]–[18], [25], [26]. Overall, because of the inherent discretization scheme, the performance of semiparametric methods is dependent on the trade off between the discretization grid size and the computational workload.

This paper aims at developing discretization-free techniques for DOA estimation with an affordable computational workload in a common scenario of linear arrays, more specifically, uniform linear arrays (ULAs) and sparse linear arrays (SLAs). We consider stochastic source signals and the same covariance fitting criteria as in SPICE. By exploiting the Hermitian Toeplitz structure in the data covariance matrix, the covariance fitting problem is cast as semidefinite programming (SDP) and solved using off-the-shelf SDP solvers, e.g., SDPT3 [27], in a polynomial time. A postprocessing technique is presented in this paper to retrieve, from the data covariance estimate, the parameters of interest including source locations, source powers and noise variance(s). The proposed method is a *sparse* method because it utilizes the same covariance fitting criteria of the semiparametric SPICE method and guarantees to produce a sparse parameter estimate. At the same time, it is a *parametric* method since it is proven to be equivalently solving a covariance fitting problem parameterized by the aforementioned parameters. Therefore, the proposed method is named as sparse and parametric approach (SPA). The SPA method differs from the existing sparse/semiparametric methods by no need of discretization. Unlike the existing parametric methods, SPA is based on convex optimization and does not require the source number. Theoretical analysis shows that SPA is a large-snapshot realization of the ML estimation and is statistically consistent (in the number of snapshots) under uncorrelated sources. Other merits of SPA include improved resolution, applicability to arbitrary number of snapshots, robustness to correlation of the sources and no requirement of user-parameters. Numerical simulations are carried out to verify our theoretical results and demonstrate the superior performance of SPA compared to existing methods.

Though most, if not all, of the existing sparse methods can be applied to sensor arrays with arbitrary geometry, problems may occur when applied to the ULA and SLA cases as studied in this paper. In particular, their parameter estimates might

suffer from some identifiability problem and not be truly sparse as demonstrated in this paper, which conflicts with the name *sparse* and decreases their resolution considerably. As a byproduct, a modified SPICE method, named as SPICE-PP, is presented to rectify the problem by incorporating the postprocessing technique presented in this paper.

In the single-snapshot case, the parameter estimation problem studied in this paper is mathematically equivalent to spectral analysis [2], for which discretization-free methods have been developed recently in [28]–[30]. It is noted that the results of [28]–[30] are based on very different techniques. Furthermore, [28] and [30] are mainly focused on the noise-free case (the former on the ULA case and the latter on the SLA case in the language of this paper). [29] studies the ULA case in the presence of i.i.d. Gaussian noise with known noise statistics. The SPA method proposed in this paper can deal with all of the above scenarios without the knowledge of noise statistics. While this paper is focused on the array processing applications in which the multisnapshot case is of the main interest, it is still not clear whether the techniques in [28]–[30] specialized for the single-snapshot case can be extended to other cases.¹

Notations used in this paper are as follows. \mathbb{R} , \mathbb{R}_+ and \mathbb{C} denote the sets of real numbers, nonnegative real numbers and complex numbers, respectively. Boldface letters are reserved for vectors and matrices. For an integer N , $[N]$ is defined as the set $\{1, \dots, N\}$. $|\cdot|$ denotes the absolute value of a scalar or cardinality of a set. $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the ℓ_1 , ℓ_2 and Frobenius norms, respectively. \mathbf{x}^T , \mathbf{x}^H and $\bar{\mathbf{x}}$ are the matrix transpose, conjugate transpose and complex conjugate of \mathbf{x} , respectively. x_j is the j th entry of a vector \mathbf{x} and \mathbf{A}_{jk} is the jk th entry of a matrix \mathbf{A} . Unless otherwise stated, \mathbf{x}_T is a subvector of \mathbf{x} with the index set T . For a vector \mathbf{x} , $\text{diag}(\mathbf{x})$ is a diagonal matrix with \mathbf{x} being its diagonal. $\mathbf{x} \succeq \mathbf{0}$ means $x_j \geq 0$ for all j . $\text{tr}(\mathbf{A})$ denotes the trace of a matrix \mathbf{A} . For positive semidefinite matrices \mathbf{A} and \mathbf{B} , $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. $E[\cdot]$ denotes expectation and $\hat{\theta}$ is an estimator of θ . For notational simplicity, a random variable and its numerical value will not be distinguished.

The rest of this paper is organized as follows. Section II describes the problem of array processing with linear arrays. Section III presents the proposed SPA method in the ULA case followed by Section IV on the SLA case. Section V introduces theoretical properties of SPA. Section VI discusses its connections to existing methods. Section VII presents our numerical simulations. Section VIII concludes this paper.

II. PROBLEM DESCRIPTION

A. Observation Model

Consider K narrowband farfield sources s_k , $k \in [K]$, impinging on a linear array of omnidirectional sensors from directions $d_k \in [-90^\circ, 90^\circ]$, $k \in [K]$. We are interested in estimation of the direction vector $\mathbf{d} = [d_1, \dots, d_K]^T$, known

¹After submission of this paper, we have investigated connections between the proposed statistical inference method and the deterministic atomic norm technique in [28]–[30] and also extended the latter to the multisnapshot case in [31]–[33].

as the direction of arrival (DOA) estimation problem [1], [2]. Denote $\theta_k = \frac{\sin(d_k)+1}{2} \in [0, 1)$, $k \in [K]$. Then we can equivalently estimate $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]^T$ since the relation $\mathbf{d} \leftrightarrow \boldsymbol{\theta}$ is one-to-one. $\boldsymbol{\theta}$ is called the frequency parameter and its meaning will be clarified later. In this paper, we concern only with the estimation of $\boldsymbol{\theta}$ for convenience. According to [1], [2], the time delays at different sensors can be represented by simple phase shifts, leading to the observation model:

$$\mathbf{y}(t) = \sum_{k=1}^K \mathbf{a}(\theta_k) s_k(t) + \mathbf{e}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{e}(t), \quad t \in [N], \quad (1)$$

where t indexes the snapshot and N is the snapshot number, $\mathbf{y}(t) \in \mathbb{C}^M$, $\mathbf{s}(t) \in \mathbb{C}^K$ and $\mathbf{e}(t) \in \mathbb{C}^M$ denote the observed snapshot, the vector of source signals and the vector of measurement noise at snapshot t , respectively, and M is the number of sensors (some notations will be redefined in the SLA case). $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$ is the so-called array manifold matrix and $\mathbf{a}(\theta_k)$ is the steering vector of the k th source which is determined by the geometry of the sensor array and will be given later. More compactly, (1) can be written into

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\theta})\mathbf{S} + \mathbf{E}, \quad (2)$$

where $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(N)]$, and \mathbf{S} and \mathbf{E} are similarly defined. The number of sources K is assumed unknown in this paper. So, the objective of the array processing is to estimate the unknown parameter $\boldsymbol{\theta}$ (or equivalently, the vector of DOAs \mathbf{d}) given the sensor measurements \mathbf{Y} and the mapping $\boldsymbol{\theta} \rightarrow \mathbf{A}(\boldsymbol{\theta})$.

B. Some Standard Assumptions

We introduce some standard assumptions for the problem formulation and solution. $\mathbf{e}(t)$, $t \in [N]$, are assumed to be spatially and temporarily white, i.e.,

$$E[\mathbf{e}(t_1)\mathbf{e}^H(t_2)] = \text{diag}(\boldsymbol{\sigma})\delta_{t_1, t_2}, \quad (3)$$

where $\boldsymbol{\sigma} = [\sigma_1, \dots, \sigma_M]^T \in \mathbb{R}_+^M$ is the noise variance parameter and δ_{t_1, t_2} is a delta function that equals 1 if $t_1 = t_2$ or 0 otherwise. The source signals and the noise are assumed to be uncorrelated with each other. Moreover, assume that the source signals are uncorrelated spatially and temporarily, i.e.,

$$E[\mathbf{s}(t_1)\mathbf{s}^H(t_2)] = \text{diag}(\mathbf{p})\delta_{t_1, t_2}, \quad (4)$$

where $\mathbf{p} = [p_1, \dots, p_M]^T \in \mathbb{R}_+^M$ denotes the source power parameter. Under the assumptions above, the data snapshots $\{\mathbf{y}(1), \dots, \mathbf{y}(N)\}$ are uncorrelated with each other and have the covariance matrix

$$\mathbf{R} = E[\mathbf{y}(t)\mathbf{y}^H(t)] = \mathbf{A}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta}) + \text{diag}(\boldsymbol{\sigma}). \quad (5)$$

It is worthy noting that the spatial uncorrelatedness of the sources may not be satisfied in practice, i.e., two sources $\{s_{k_1}(t)\}$ and $\{s_{k_2}(t)\}$ can be correlated or even coherent (i.e., completely correlated). However, the method proposed in this paper will be shown to be robust to this assumption.

C. Uniform and Sparse Linear Arrays

Both uniform and sparse linear arrays will be studied in this paper. In the ULA case, the sensors are uniformly spaced with a spacing of $\frac{\lambda}{2}$, where λ denotes the wavelength of the sources. For an M -element ULA, the steering vector $\mathbf{a}(\theta_k)$ of source k has the following form with $i = \sqrt{-1}$:

$$\mathbf{a}(\theta_k) = \left[1, e^{i2\pi\theta_k}, \dots, e^{i2(M-1)\pi\theta_k}\right]^T. \quad (6)$$

It is clear that θ_k is the frequency of the uniformly sampled complex sinusoid $\mathbf{a}(\theta_k)$, a reason for why $\boldsymbol{\theta}$ is called the frequency parameter. A well-known result is that up to $M-1$ sources can be detected using an M -element ULA (see, e.g., [34]).

An SLA can be considered as a ULA with ‘‘missing’’ sensors, i.e., an SLA takes only a subset, say $\boldsymbol{\Omega}$, of the sensors of a ULA. Thus an SLA can be represented by its sensor index set $\boldsymbol{\Omega} \subset [M]$. Without loss of generality, we assume that $\boldsymbol{\Omega}$ is sorted ascendingly with $\Omega_1 = 1$ and $\Omega_L = M$, where $L = |\boldsymbol{\Omega}|$ denotes the array size (otherwise, we can redefine $\boldsymbol{\Omega}$, say $\bar{\boldsymbol{\Omega}}$, such that $\bar{\Omega}_j = \Omega_j - \Omega_1 + 1$, $j \in [L]$, and let $M = \bar{\Omega}_L$). Then the steering vector of the SLA $\boldsymbol{\Omega}$ for source k , denoted by $\mathbf{a}_{\boldsymbol{\Omega}}(\theta_k)$, is

$$\mathbf{a}_{\boldsymbol{\Omega}}(\theta_k) = \left[e^{i2\pi(\Omega_1-1)\theta_k}, \dots, e^{i2\pi(\Omega_L-1)\theta_k}\right]^T. \quad (7)$$

Denote $\boldsymbol{\Gamma}_{\boldsymbol{\Omega}} \in \{0, 1\}^{L \times M}$ a selection matrix such that the j th row of $\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}$ contains all 0s but a single 1 at the Ω_j th position. It is clear that

$$\mathbf{a}_{\boldsymbol{\Omega}}(\theta_k) = \boldsymbol{\Gamma}_{\boldsymbol{\Omega}}\mathbf{a}(\theta_k). \quad (8)$$

Let

$$\mathcal{D} = \{m_1 - m_2 + 1 : m_1, m_2 \in \boldsymbol{\Omega}, m_1 \geq m_2\} \subset [M]. \quad (9)$$

An SLA defined by \mathcal{D} is called the coarray of $\boldsymbol{\Omega}$. $\boldsymbol{\Omega}$ is called a redundancy array if \mathcal{D} defines a ULA, i.e., $\mathcal{D} = [M]$. It is obvious that a ULA is a redundancy array. The maximum number of sources detectable using the array $\boldsymbol{\Omega}$ is determined by its coarray \mathcal{D} [34], [35]. In particular, a redundancy array can detect up to $M-1$ sources like a ULA. For a non-redundancy SLA, the maximum number of sources detectable will be less than $M-1$. It is noted that redundancy SLAs are quite common and there generally exists such an array satisfying that $L^2 \leq 3(M-1)$ according to [35].

Example 1: The SLA $\boldsymbol{\Omega} = \{1, 2, 5, 7\}$ is a redundancy array, where $L = 4$ and $M = 7$. So maximally 6 sources can be detected using this 4-element array.

III. SPA WITH A ULA

A. Covariance Fitting Criteria

Since a ULA can detect up to $M-1$ sources, the condition $K \leq M-1$ can be considered as *a priori* knowledge while the exact value of K is unavailable. The covariance of the data snapshots \mathbf{Y} is given in (5) under the assumptions specified in Subsection II-B. Denote by $\tilde{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ the sample covariance. When $\tilde{\mathbf{R}}$ and \mathbf{R} are both invertible we consider the following covariance fitting criterion for the purpose of

parameter estimation (see [20]–[22], [34] and the references therein):

$$f_1(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\sigma}) = \left\| \mathbf{R}^{-\frac{1}{2}} (\tilde{\mathbf{R}} - \mathbf{R}) \tilde{\mathbf{R}}^{-\frac{1}{2}} \right\|_{\text{F}}^2. \quad (10)$$

\mathbf{R}^{-1} exists in the presence of noise, i.e., $\sigma_j > 0$ for $j \in [M]$. Then $\tilde{\mathbf{R}}^{-1}$ exists with probability one if $N \geq M$. According to [34], [36], the minimization of the criterion in (10) is a large-snapshot realization of the ML estimator. In the case where $\tilde{\mathbf{R}}$ is singular (it happens when $N < M$) we consider an alternative criterion:

$$f_2(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\sigma}) = \left\| \mathbf{R}^{-\frac{1}{2}} (\tilde{\mathbf{R}} - \mathbf{R}) \right\|_{\text{F}}^2 \quad (11)$$

which has been studied in [20], [21]. We defer the latter case to Subsection III-B2.

Remark 1: The covariance fitting criteria above exploit the assumption that the sources are uncorrelated which results in the expression of \mathbf{R} in (5). However, a theoretical explanation is provided in [20], [21] to show that the criteria are robust to correlations of the sources. Moreover, according to [21], [37] they are connected to the ℓ_1 norm minimization which is known to be robust to the correlations. Consequently, such robustness property is maintained in the method proposed in this paper which utilizes the same criteria.

A simple calculation shows that

$$\begin{aligned} f_1 &= \text{tr} \left[\mathbf{R}^{-\frac{1}{2}} (\tilde{\mathbf{R}} - \mathbf{R}) \tilde{\mathbf{R}}^{-1} (\tilde{\mathbf{R}} - \mathbf{R}) \mathbf{R}^{-\frac{1}{2}} \right] \\ &= \text{tr} \left[(\mathbf{R}^{-1} \tilde{\mathbf{R}} - \mathbf{I}) (\mathbf{I} - \tilde{\mathbf{R}}^{-1} \mathbf{R}) \right] \\ &= \text{tr} (\mathbf{R}^{-1} \tilde{\mathbf{R}}) + \text{tr} (\tilde{\mathbf{R}}^{-1} \mathbf{R}) - 2M. \end{aligned} \quad (12)$$

It is challenging to minimize f_1 with respect to the unknown parameters $\boldsymbol{\theta}$, \mathbf{p} and $\boldsymbol{\sigma}$ due to their nonlinear relation to \mathbf{R} by (5). We propose to estimate \mathbf{R} firstly by reparameterization and then determine the parameters of interest. Let

$$\mathbf{C}(\boldsymbol{\theta}, \mathbf{p}) = \mathbf{A}(\boldsymbol{\theta}) \text{diag}(\mathbf{p}) \mathbf{A}^H(\boldsymbol{\theta}). \quad (13)$$

It is easy to see that $\mathbf{C} \geq \mathbf{0}$ and $\text{rank}(\mathbf{C}) = K \leq M - 1$. Moreover,

$$C_{jl} = \sum_{k=1}^K p_k a_j(\theta_k) \bar{a}_l(\theta_k) = \sum_{k=1}^K p_k e^{i2\pi(j-l)\theta_k}. \quad (14)$$

\mathbf{C} is thus a (Hermitian) Toeplitz matrix which is determined by M complex numbers and can be written as $\mathbf{C} = T(\mathbf{u})$ for some $\mathbf{u} \in \mathbb{C}^M$, where

$$T(\mathbf{u}) = \begin{bmatrix} u_1 & u_2 & \cdots & u_M \\ \bar{u}_2 & u_1 & \cdots & u_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{u}_M & \bar{u}_{M-1} & \cdots & u_1 \end{bmatrix}. \quad (15)$$

B. SDP Formulations in the Case of Different $\{\sigma_j\}$

It follows from (15) that the characterization

$$\mathbf{R}(\mathbf{u}, \boldsymbol{\sigma}) = T(\mathbf{u}) + \text{diag}(\boldsymbol{\sigma}) \quad (16)$$

captures the structure of \mathbf{R} under the constraints $T(\mathbf{u}) \geq \mathbf{0}$ and $\boldsymbol{\sigma} \succeq \mathbf{0}$. Further observations reveal that \mathbf{R} is inherently

determined by $2M - 1$ numbers (M for the diagonal and the other $M - 1$ for the off diagonal). However, $2M$ numbers (M in \mathbf{u} and the other M in $\boldsymbol{\sigma}$) have been used in the above expression of \mathbf{R} ($M + 1$ for the diagonal and $M - 1$ for the off diagonal). As a result, redundancy exists along the diagonal of \mathbf{R} . The effect of the redundancy is twofold. On one hand, it will cause an identifiability problem, which will be tackled in Subsection III-D, that in general \mathbf{u} , $\boldsymbol{\sigma}$ cannot be uniquely identified from \mathbf{R} without accounting for additional information except $T(\mathbf{u}) \geq \mathbf{0}$ and $\boldsymbol{\sigma} \succeq \mathbf{0}$. It means that the solution $\boldsymbol{\sigma}^*$ of the SDPs to be presented cannot be directly used as the final estimate of the noise variance. On the other hand, the adoption of one redundant variable enables us not to impose the nonconvex rank-deficiency constraint on $T(\mathbf{u})$ to characterize \mathbf{R} (it will be clarified in Subsection III-D). However, the redundancy problem will not affect the estimation of \mathbf{R} which is our current focus.

1) *The Case of $N \geq M$:* With the characterization of \mathbf{R} in (16) the minimization of f_1 is equivalent to

$$\begin{aligned} \min_{\mathbf{u}, \{\boldsymbol{\sigma} \succeq \mathbf{0}\}} \quad & \text{tr}(\mathbf{R}^{-1} \tilde{\mathbf{R}}) + \text{tr}(\tilde{\mathbf{R}}^{-1} \mathbf{R}), \\ \text{subject to} \quad & T(\mathbf{u}) \geq \mathbf{0}. \end{aligned} \quad (17)$$

Then we can show the following equivalences:

$$\begin{aligned} (17) &\Leftrightarrow \min_{\mathbf{u}, \{\boldsymbol{\sigma} \succeq \mathbf{0}\}} \text{tr} \left(\tilde{\mathbf{R}}^{\frac{1}{2}} \mathbf{R}^{-1} \tilde{\mathbf{R}}^{\frac{1}{2}} \right) + \text{tr}(\tilde{\mathbf{R}}^{-1} \mathbf{R}), \\ &\text{subject to } T(\mathbf{u}) \geq \mathbf{0} \\ &\Leftrightarrow \min_{\mathbf{X}, \mathbf{u}, \{\boldsymbol{\sigma} \succeq \mathbf{0}\}} \text{tr}(\mathbf{X}) + \text{tr}(\tilde{\mathbf{R}}^{-1} \mathbf{R}), \\ &\text{subject to } T(\mathbf{u}) \geq \mathbf{0} \text{ and } \mathbf{X} \geq \tilde{\mathbf{R}}^{\frac{1}{2}} \mathbf{R}^{-1} \tilde{\mathbf{R}}^{\frac{1}{2}} \\ &\Leftrightarrow \min_{\mathbf{X}, \mathbf{u}, \{\boldsymbol{\sigma} \succeq \mathbf{0}\}} \text{tr}(\mathbf{X}) + \text{tr}(\tilde{\mathbf{R}}^{-1} \mathbf{R}), \\ &\text{subject to } \begin{bmatrix} \mathbf{X} & \tilde{\mathbf{R}}^{\frac{1}{2}} \\ \tilde{\mathbf{R}}^{\frac{1}{2}} & \mathbf{R} \\ & & T(\mathbf{u}) \end{bmatrix} \geq \mathbf{0}. \end{aligned} \quad (18)$$

So the problem in (17) can be formulated as an SDP and thus is convex. As the result, \mathbf{R} can be estimated by solving the SDP with its estimate given by $\hat{\mathbf{R}} = T(\mathbf{u}^*) + \text{diag}(\boldsymbol{\sigma}^*)$, where $(\mathbf{u}^*, \boldsymbol{\sigma}^*)$ is the solution of the SDP.

2) *The Case of $N < M$:* When $N < M$, $\tilde{\mathbf{R}}$ ceases to be nonsingular and the criterion in (11) is used which can be written into

$$\begin{aligned} f_2(\mathbf{u}, \boldsymbol{\sigma}) &= \text{tr} \left[(\tilde{\mathbf{R}} - \mathbf{R}) \mathbf{R}^{-1} (\tilde{\mathbf{R}} - \mathbf{R}) \right] \\ &= \text{tr}(\tilde{\mathbf{R}} \mathbf{R}^{-1} \tilde{\mathbf{R}}) + \text{tr}(\mathbf{R}) - 2\text{tr}(\tilde{\mathbf{R}}) \end{aligned} \quad (19)$$

Hence, an SDP similar to (18) can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{u}, \{\boldsymbol{\sigma} \succeq \mathbf{0}\}} \quad & \text{tr}(\mathbf{X}) + \text{tr}(\mathbf{R}), \\ \text{subject to} \quad & \begin{bmatrix} \mathbf{X} & \tilde{\mathbf{R}} \\ \tilde{\mathbf{R}} & \mathbf{R} \\ & & T(\mathbf{u}) \end{bmatrix} \geq \mathbf{0}. \end{aligned} \quad (20)$$

C. SDP Formulations in the Case of Equal $\{\sigma_j\}$

It is reasonable to assume that the noise variances $\sigma_j, j \in [M]$, are equal in some scenarios. Though the formulations in the last subsection can be applied to such a case by imposing the constraint $\sigma_1 = \dots = \sigma_M = \sigma$, simpler formulations in fact exist. In such a case, $\text{diag}(\boldsymbol{\sigma}) = \sigma \mathbf{I}$ with $\sigma \in \mathbb{R}_+$. Therefore, the covariance \mathbf{R} has a Toeplitz structure and the characterization

$$\mathbf{R} = T(\mathbf{u}) \quad (21)$$

captures its structure without redundancy under the constraint $T(\mathbf{u}) \geq \mathbf{0}$ for some $\mathbf{u} \in \mathbb{C}^M$. Note that the $T(\mathbf{u})$ here is the same as that in (16) off the diagonal but different on the diagonal, i.e., the two \mathbf{u} 's are the same except the first entry. Following from the similar procedures as in the last subsection, similar SDPs can be formulated. In particular, when $N \geq M$ we have

$$\begin{aligned} & \min_{\mathbf{X}, \mathbf{u}} \text{tr}(\mathbf{X}) + \text{tr}(\tilde{\mathbf{R}}^{-1} T(\mathbf{u})), \\ & \text{subject to } \begin{bmatrix} \mathbf{X} & \tilde{\mathbf{R}}^{\frac{1}{2}} \\ \tilde{\mathbf{R}}^{\frac{1}{2}} & T(\mathbf{u}) \end{bmatrix} \geq \mathbf{0}. \end{aligned} \quad (22)$$

When $N < M$,

$$\begin{aligned} & \min_{\mathbf{X}, \mathbf{u}} \text{tr}(\mathbf{X}) + \text{tr}(T(\mathbf{u})), \\ & \text{subject to } \begin{bmatrix} \mathbf{X} & \tilde{\mathbf{R}} \\ \tilde{\mathbf{R}} & T(\mathbf{u}) \end{bmatrix} \geq \mathbf{0}. \end{aligned} \quad (23)$$

Note that the constraint $T(\mathbf{u}) \geq \mathbf{0}$ has been implicitly included in the constraints in (22) and (23). The simplified SDP formulations will lead to faster computations in practice. By solving one of the SDPs, $\hat{\mathbf{R}}$ is obtained as $\hat{\mathbf{R}} = T(\mathbf{u}^*)$ given the solution \mathbf{u}^* . For notational convenience, we say $\boldsymbol{\sigma}^* = \mathbf{0}$ in this case. In fact, note that (21), (22) and (23) can be obtained from (16), (18) and (20), respectively, by setting $\boldsymbol{\sigma} = \mathbf{0}$.

D. Postprocessing of $\hat{\mathbf{R}}$

After obtaining $\hat{\mathbf{R}}$, the following task is to estimate the parameters $\boldsymbol{\theta}$, \mathbf{p} and $\boldsymbol{\sigma}$ by writing it back into the form of (5). To do this, we decompose $\hat{\mathbf{R}}$ into

$$\hat{\mathbf{R}} = T(\hat{\mathbf{u}}) + \text{diag}(\hat{\boldsymbol{\sigma}}), \quad (24)$$

where $T(\hat{\mathbf{u}}) = \mathbf{A}(\hat{\boldsymbol{\theta}}) \text{diag}(\hat{\mathbf{p}}) \mathbf{A}^H(\hat{\boldsymbol{\theta}}) \geq \mathbf{0}$ is the estimate of $\mathbf{C}(\boldsymbol{\theta}, \mathbf{p})$ in (13) and $\hat{\boldsymbol{\sigma}} \succeq \mathbf{0}$ is the noise covariance estimate. Such $(\hat{\mathbf{u}}, \hat{\boldsymbol{\sigma}})$ always exists according to the way $\hat{\mathbf{R}}$ is given, however, it is generally not unique. In particular, for any $\delta \geq -\min(\boldsymbol{\sigma}^*)$ satisfying that $T(\mathbf{u}^*) - \delta \mathbf{I} \geq \mathbf{0}$, $(\hat{\mathbf{u}}, \hat{\boldsymbol{\sigma}}) = \left(\mathbf{u}^* - \begin{bmatrix} \delta \\ \mathbf{0} \end{bmatrix}, \boldsymbol{\sigma}^* + \delta \mathbf{1} \right)$ leads to one realization of the decomposition, which in fact also enumerates all possible realizations. We utilize the prior knowledge that $K \leq M - 1$ to make the decomposition unique as follows. It follows from $K \leq M - 1$ that $\text{rank}(\mathbf{C}) = K \leq M - 1$. Therefore, it is natural to impose that $\text{rank}(T(\hat{\mathbf{u}})) \leq M - 1$, i.e., $T(\hat{\mathbf{u}}) = T(\mathbf{u}^*) - \delta \mathbf{I}$ is rank-deficient. A direct result is that

δ is an eigenvalue of $T(\mathbf{u}^*)$. Then by $T(\mathbf{u}^*) - \delta \mathbf{I} \geq \mathbf{0}$ we see that

$$\delta = \lambda_{\min}(T(\mathbf{u}^*)) \quad (25)$$

and thus the decomposition is unique, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue.²

Remark 2: The postprocessing is used to separate sources and noise in the estimated covariance matrix such that the source part can be represented by as few sources as possible based on the minimum description length principle [38] (see the next subsection for clarity). In fact, the idea of postprocessing has been studied in the literature for a special case where $\hat{\mathbf{R}}$ is a Toeplitz matrix as in the case of equal $\{\sigma_j\}$ (see, e.g., [2, Section 4.9.2]). It is noted that the postprocessing is very important in SPA, without which the final parameter estimate is generally not unique and does not have the statistical properties that will be shown in Section V. Even worse, the frequency estimate of SPA can cease to be sparse. To see this, suppose that $T(\mathbf{u}^*)$ has full rank (this is generally the case in the presence of noise). Choose $\theta_1 \in [0, 1)$ arbitrarily and let $p_1 < (\mathbf{a}^H(\theta_1) T^{-1}(\mathbf{u}^*) \mathbf{a}(\theta_1))^{-1}$. It follows that the residue $T(\mathbf{u}^*) - p_1 \mathbf{a}(\theta_1) \mathbf{a}^H(\theta_1) > \mathbf{0}$ which still has the Toeplitz structure. Then we can choose θ_2 and p_2 similarly based on the residue. The process can be repeated infinitely many times and results in infinitely long vectors $\boldsymbol{\theta}$ and \mathbf{p} .

E. Frequency and Power Solutions

The remaining task is to retrieve the frequency estimate $\hat{\boldsymbol{\theta}}$ and the power estimate $\hat{\mathbf{p}}$ given $T(\hat{\mathbf{u}})$, which is based on the following classical Vandermonde decomposition lemma for positive semidefinite Toeplitz matrices (see, e.g., [2], [39]).

Lemma 1: Any positive semidefinite Toeplitz matrix $T(\mathbf{u}) \in \mathbb{C}^{M \times M}$ can be represented as

$$T(\mathbf{u}) = \mathbf{V} \mathbf{P} \mathbf{V}^H, \quad (26)$$

where

$$\mathbf{V} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_r)], \quad (27)$$

$$\mathbf{P} = \text{diag}(p_1, \dots, p_r), \quad (28)$$

$\theta_j \in [0, 1)$, $p_j > 0$ for $j \in [r]$, and $r = \text{rank}(T(\mathbf{u}))$. Moreover, the representation is unique up to permutation of elements of $\boldsymbol{\theta}$ and \mathbf{p} if $r \leq M - 1$. ■

It follows from Lemma 1 that $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{p}}$ can be uniquely determined given $T(\hat{\mathbf{u}})$ since $\text{rank}(T(\hat{\mathbf{u}})) \leq M - 1$. In practice, $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{p}}$ can be obtained as follows. Given $T(\hat{\mathbf{u}}) = \mathbf{A}(\hat{\boldsymbol{\theta}}) \text{diag}(\hat{\mathbf{p}}) \mathbf{A}^H(\hat{\boldsymbol{\theta}})$, it is easy to show that

$$\begin{bmatrix} \mathbf{A}(\hat{\boldsymbol{\theta}}) \\ \overline{\mathbf{A}}_{\{2, \dots, M\}}(\hat{\boldsymbol{\theta}}) \end{bmatrix} \hat{\mathbf{p}} = \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{u}}_{\{2, \dots, M\}} \end{bmatrix} \quad (29)$$

since $\hat{\mathbf{p}} \succ \mathbf{0}$, where $\overline{\mathbf{A}}_{\{2, \dots, M\}}(\hat{\boldsymbol{\theta}})$ takes all but the first rows of the matrix $\overline{\mathbf{A}}(\hat{\boldsymbol{\theta}})$ which denotes the complex conjugate of

²The postprocessing as in (24) can be done if we have only $\hat{\mathbf{R}}$ rather than \mathbf{u}^* and $\boldsymbol{\sigma}^*$. To see this, let $\tilde{\mathbf{u}}$ be the transpose of the first row of $\hat{\mathbf{R}}$. Then $\hat{\mathbf{u}} = \tilde{\mathbf{u}} - \begin{bmatrix} \lambda_{\min}(T(\hat{\mathbf{u}})) \\ \mathbf{0} \end{bmatrix}$ and $\hat{\boldsymbol{\sigma}}$ is obtained as the diagonal of $\hat{\mathbf{R}}$ minus $\hat{u}_1 \mathbf{1}$.

$\mathbf{A}(\hat{\boldsymbol{\theta}})$. So we build a system of $2M - 1$ equations that is linear in $\hat{\mathbf{p}}$ whose length is maximally $M - 1$, where each column of the coefficient matrix corresponds to a uniformly sampled sinusoid (after permutation of rows). According to [40], Prony's method can be applied to this type of systems to efficiently solve $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{p}}$. In particular, $\hat{\boldsymbol{\theta}}$ is firstly obtained from zeros of a polynomial which is obtained by solving a linear system involving only $\hat{\mathbf{u}}$. After that, $\hat{\mathbf{p}}$ is solved from (29). Readers are referred to [40] for the detailed procedure.

The proposed SPA algorithm for the array processing with a ULA is presented in Algorithm 1. The covariance matrix \mathbf{R} is firstly estimated by solving an SDP. Then the postprocessing procedure is carried out for $\hat{\mathbf{R}}$ to resolve an identifiability problem that exists in the solution of the SDP (the noise variance is estimated at this step). Finally, the frequency and the power are solved using Prony's method. To carry out the parameter estimation, SPA requires only the data snapshots \mathbf{Y} of the sensor array without any other user parameters. More properties of SPA will be presented in Section V.

Algorithm 1 Sparse and parametric approach (SPA)

Input: observed snapshots \mathbf{Y} .

- 1) Estimate \mathbf{R} by solving the solution $(\mathbf{u}^*, \boldsymbol{\sigma}^*)$ of an SDP;
- 2) Postprocess $\hat{\mathbf{R}}$ to obtain $\hat{\mathbf{u}}$ and $\hat{\boldsymbol{\sigma}}$ via $T(\hat{\mathbf{u}}) = T(\mathbf{u}^*) - \lambda_{\min}(T(\mathbf{u}^*))\mathbf{I}$ and $\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}^* + \lambda_{\min}(T(\mathbf{u}^*))\mathbf{1}$;
- 3) Solve $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{p}}$ from (29) using Prony's method.

Output: parameter estimator $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}}, \hat{\boldsymbol{\sigma}})$.

Remark 3: In general, SPA cannot correctly determine the true number of sources K but provides additionally spurious sources. This is because that we do not assume in SPA any knowledge of the source number or the noise variance(s). Consequently, SPA solves the problem as in the worst case where up to $M - 1$ sources can be present. In fact, the source number estimation problem itself is very difficult which is known as model order selection [41], and lacking guarantees on the estimated source number seems to be a common feature of sparse methods, especially when no prior knowledge mentioned above is available (see, e.g., [9], [19], [21], [22]). A positive side of SPA is that it has strong statistical properties which will be shown in Section V, and theoretically the powers of the spurious sources can be very close to zero under sufficient snapshots or appropriately low SNR, which enables the spurious sources to be distinguished from the real ones. Currently, methods have been proposed to remove spurious sources by modifying the solution of a sparse method. For example, intuitive thresholding is used in [19] and information criteria-based methods are presented in [5], [23]. In future studies, we may seek to incorporate model order selection in SPA such that the source number can be estimated automatically together with the parameters but it is beyond the scope of this paper.

IV. SPA WITH AN SLA

In this section we extend SPA to the SLA case. With respect to an SLA $\boldsymbol{\Omega}$, denote the steering matrix by $\mathbf{A}_{\boldsymbol{\Omega}}(\boldsymbol{\theta}) =$

$[\mathbf{a}_{\boldsymbol{\Omega}}(\theta_1), \dots, \mathbf{a}_{\boldsymbol{\Omega}}(\theta_K)] \in \mathbb{C}^{L \times K}$, the data snapshots by $\mathbf{Y}_{\boldsymbol{\Omega}} = [\mathbf{y}_{\boldsymbol{\Omega}}(1), \dots, \mathbf{y}_{\boldsymbol{\Omega}}(N)] = \mathbf{A}_{\boldsymbol{\Omega}}(\boldsymbol{\theta})\mathbf{S} + \mathbf{E}_{\boldsymbol{\Omega}} \in \mathbb{C}^{L \times N}$, the covariance matrix by $\mathbf{R}_{\boldsymbol{\Omega}} = E\{\mathbf{y}_{\boldsymbol{\Omega}}(t)\mathbf{y}_{\boldsymbol{\Omega}}^H(t)\} = \mathbf{A}_{\boldsymbol{\Omega}}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}_{\boldsymbol{\Omega}}^H(\boldsymbol{\theta}) + \text{diag}(\boldsymbol{\sigma}_{\boldsymbol{\Omega}}) \in \mathbb{C}^{L \times L}$ and the sample covariance by $\tilde{\mathbf{R}}_{\boldsymbol{\Omega}} = \frac{1}{N}\mathbf{Y}_{\boldsymbol{\Omega}}\mathbf{Y}_{\boldsymbol{\Omega}}^H \in \mathbb{C}^{L \times L}$, where $\boldsymbol{\sigma}_{\boldsymbol{\Omega}} \in \mathbb{R}_+^L$ denotes the noise variance parameter. It follows from (8) that $\mathbf{A}_{\boldsymbol{\Omega}}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}_{\boldsymbol{\Omega}}\mathbf{A}(\boldsymbol{\theta})$ and then $\mathbf{A}_{\boldsymbol{\Omega}}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}_{\boldsymbol{\Omega}}^H(\boldsymbol{\theta}) = \boldsymbol{\Gamma}_{\boldsymbol{\Omega}}T(\mathbf{u})\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}^T \triangleq T_{\boldsymbol{\Omega}}(\mathbf{u})$, where the Toeplitz matrix $T(\mathbf{u}) = \mathbf{A}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta})$ for some $\mathbf{u} \in \mathbb{C}^M$. A careful study of $T_{\boldsymbol{\Omega}}(\mathbf{u})$ reveals that

$$T_{\boldsymbol{\Omega}}(\mathbf{u})_{jl} = \sum_{k=1}^K p_k a_{\Omega_j}(\theta_k) \bar{a}_{\Omega_l}(\theta_k) = \sum_{k=1}^K p_k e^{i2\pi(\Omega_j - \Omega_l)\theta_k}, \quad (30)$$

i.e., the entries of $T_{\boldsymbol{\Omega}}(\mathbf{u})$ are specified by the coarray \mathcal{D} defined in (9).

In this paper we are mainly interested in the case where the SLA $\boldsymbol{\Omega}$ is a redundancy array. As in the ULA case, $K \leq M - 1$ is considered as *a priori* knowledge according to Subsection II-C. The matrix $T_{\boldsymbol{\Omega}}(\mathbf{u})$ contains all elements of \mathbf{u} explicitly by (30). It implies that the relation $T(\mathbf{u}) \leftrightarrow T_{\boldsymbol{\Omega}}(\mathbf{u})$ is one-to-one.

Example 2: Given a redundancy SLA $\boldsymbol{\Omega} = \{1, 2, 5, 7\}$, we have $\mathbf{u} \in \mathbb{C}^7$ and $T_{\boldsymbol{\Omega}}(\mathbf{u}) = \boldsymbol{\Gamma}_{\boldsymbol{\Omega}}T(\mathbf{u})\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}^T = \begin{bmatrix} u_1 & u_2 & u_5 & u_7 \\ \bar{u}_2 & u_1 & u_4 & u_6 \\ \bar{u}_5 & \bar{u}_4 & u_1 & u_3 \\ \bar{u}_7 & \bar{u}_6 & \bar{u}_3 & u_1 \end{bmatrix}$, where all elements of \mathbf{u} are contained.

It follows that the covariance matrix $\mathbf{R}_{\boldsymbol{\Omega}}$ can be characterized as

$$\begin{aligned} \mathbf{R}_{\boldsymbol{\Omega}} &= T_{\boldsymbol{\Omega}}(\mathbf{u}) + \text{diag}(\boldsymbol{\sigma}_{\boldsymbol{\Omega}}) \\ &= \boldsymbol{\Gamma}_{\boldsymbol{\Omega}}T(\mathbf{u})\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}^T + \text{diag}(\boldsymbol{\sigma}_{\boldsymbol{\Omega}}) \end{aligned} \quad (31)$$

under the constraints $T(\mathbf{u}) \geq \mathbf{0}$ and $\boldsymbol{\sigma}_{\boldsymbol{\Omega}} \geq \mathbf{0}$. Consider similar covariance fitting criteria as in the ULA case, i.e., the minimization of $\left\| \mathbf{R}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} \left(\tilde{\mathbf{R}}_{\boldsymbol{\Omega}} - \mathbf{R}_{\boldsymbol{\Omega}} \right) \tilde{\mathbf{R}}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} \right\|_{\text{F}}^2$ when $N \geq L$ and $\left\| \mathbf{R}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} \left(\tilde{\mathbf{R}}_{\boldsymbol{\Omega}} - \mathbf{R}_{\boldsymbol{\Omega}} \right) \right\|_{\text{F}}^2$ otherwise. Then SPA can be extended to the SLA case. In particular, when $N \geq L$ we obtain the following SDP:

$$\begin{aligned} &\min_{\mathbf{X}, \mathbf{u}, \{\boldsymbol{\sigma}_{\boldsymbol{\Omega}} \geq \mathbf{0}\}} \text{tr}(\mathbf{X}) + \text{tr} \left(\tilde{\mathbf{R}}_{\boldsymbol{\Omega}}^{-1} \mathbf{R}_{\boldsymbol{\Omega}} \right), \\ &\text{subject to} \quad \begin{bmatrix} \mathbf{X} & \tilde{\mathbf{R}}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} \\ \tilde{\mathbf{R}}_{\boldsymbol{\Omega}}^{-\frac{1}{2}} & \mathbf{R}_{\boldsymbol{\Omega}} \\ & & T(\mathbf{u}) \end{bmatrix} \geq \mathbf{0}. \end{aligned} \quad (32)$$

When $N < L$, it is

$$\begin{aligned} &\min_{\mathbf{X}, \mathbf{u}, \{\boldsymbol{\sigma}_{\boldsymbol{\Omega}} \geq \mathbf{0}\}} \text{tr}(\mathbf{X}) + \text{tr}(\mathbf{R}_{\boldsymbol{\Omega}}), \\ &\text{subject to} \quad \begin{bmatrix} \mathbf{X} & \tilde{\mathbf{R}}_{\boldsymbol{\Omega}} \\ \tilde{\mathbf{R}}_{\boldsymbol{\Omega}} & \mathbf{R}_{\boldsymbol{\Omega}} \\ & & T(\mathbf{u}) \end{bmatrix} \geq \mathbf{0}. \end{aligned} \quad (33)$$

In the case of equal noise variances, $\mathbf{R}_{\boldsymbol{\Omega}}$ can be characterized as $\mathbf{R}_{\boldsymbol{\Omega}} = \boldsymbol{\Gamma}_{\boldsymbol{\Omega}}T(\mathbf{u})\boldsymbol{\Gamma}_{\boldsymbol{\Omega}}^T$. Then similar SDPs can be formulated as in (32) and (33) by simply setting $\boldsymbol{\sigma}_{\boldsymbol{\Omega}} = \mathbf{0}$. Note that the problem dimension in such a case cannot be

further reduced as in the ULA case since $\mathbf{\Gamma}_\Omega T(\mathbf{u}) \mathbf{\Gamma}_\Omega^T \geq \mathbf{0}$ does not imply $T(\mathbf{u}) \geq \mathbf{0}$.

Note that the same identifiability problem exists in the solution $(\mathbf{u}^*, \boldsymbol{\sigma}_\Omega^*)$ of the SDPs above. To resolve the problem, the same procedures can be carried out as in the ULA case by exploiting that $\text{rank}(T(\hat{\mathbf{u}})) \leq M - 1$. Then the parameter estimate $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}}, \hat{\boldsymbol{\sigma}}_\Omega)$ can be obtained.

Remark 4: When the SLA Ω is not a redundancy array, i.e., the coarray \mathcal{D} is not a ULA, SPA presented above can be applied straightforwardly. In such a case, it is natural to require that $K \leq \overline{M} - 1 < M - 1$, where $\overline{M} - 1$ denotes the maximum number of sources detectable using the non-redundancy SLA. However, the introduced implementation of SPA can only use the information up to $K \leq M - 1$, which should be a waste of knowledge. A thorough study on exploitation of the full information should be investigated in the future but is beyond the scope of this paper.

V. PROPERTIES OF SPA

A. Sparse and Parametric Method

As its name suggests, SPA is a *sparse* and *parametric* method. It carries out the optimization by reparameterization and thus is a parametric method. The length of its frequency estimator $\hat{\boldsymbol{\theta}}$ is maximally $M - 1$. Since this is an important result of this paper, we state it formally in the following theorem.

Theorem 1: The length of the frequency estimator $\hat{\boldsymbol{\theta}}$ of SPA is maximally $M - 1$. ■

B. Statistical Properties

We consider ULAs and redundancy SLAs in this subsection, or collectively, redundancy arrays. Notice that dimensions of the parameter estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{p}}$ may be different from the true dimension K . To discuss statistical properties of the SPA estimator, the dimension problem should be resolved first. Notice that the proposed SPA is equivalently assuming $K = M - 1$ (the worst case) with the knowledge $K \leq M - 1$. Then we expand both the true parameter and its estimator to the same dimension. Define

$$\mathcal{E}_p(\mathbf{v}) = \begin{bmatrix} \mathbf{v} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{M-1}, \quad (34)$$

$$\mathcal{E}_f(\mathbf{v}) = \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix} \in \mathbb{R}^{M-1} \quad (35)$$

for a vector \mathbf{v} of dimension no more than $M - 1$, where $\mathbf{w} \in [0, 1]^{M-1-|\mathbf{v}|}$ is arbitrary. It is easy to see that $(\mathcal{E}_f(\boldsymbol{\theta}), \mathcal{E}_p(\mathbf{p}))$ and $(\boldsymbol{\theta}, \mathbf{p})$ are physically equivalent since all of the added virtual sources have zero powers. Then we have the following results.

Theorem 2: Assume that $\{s(t)\}$ and $\{e(t)\}$ are uncorrelated, $E[s(t_1)s^H(t_2)] = \text{diag}(\mathbf{p}) \delta_{t_1, t_2}$ and $E[e(t_1)e^H(t_2)] = \text{diag}(\boldsymbol{\sigma}) \delta_{t_1, t_2}$ for $t_1, t_2 \in [N]$. Moreover, assume that the sensor array is a redundancy array and $K \leq M - 1$. Then the

parameter estimator $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}}, \hat{\boldsymbol{\sigma}})$ of SPA is statistically consistent (in N).³

Proof: Without loss of generality, we consider only the ULA case. The proof is based on the observation that the parameter estimator $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}}, \hat{\boldsymbol{\sigma}})$ can be uniquely determined given the covariance estimator $\hat{\mathbf{R}}$. Suppose that the true parameter value is $(\boldsymbol{\theta}^o, \mathbf{p}^o, \boldsymbol{\sigma}^o)$. As $N \rightarrow \infty$, $\hat{\mathbf{R}}$ approaches the true covariance $\mathbf{R}^o = \mathbf{A}(\boldsymbol{\theta}^o) \text{diag}(\mathbf{p}^o) \mathbf{A}^H(\boldsymbol{\theta}^o) + \text{diag}(\boldsymbol{\sigma}^o)$. Then by (10) the SDP of SPA admits a unique minimizer \mathbf{R}^o of \mathbf{R} , which determines the unique parameter estimate $(\boldsymbol{\theta}^o, \mathbf{p}^o, \boldsymbol{\sigma}^o)$ given $K \leq M - 1$. ■

Theorem 3: Assume that $\{s(t)\}$ and $\{e(t)\}$ are uncorrelated and both are i.i.d. circular Gaussian with means zero and covariance matrices $\text{diag}(\mathbf{p})$ and $\text{diag}(\boldsymbol{\sigma})$, respectively. Moreover, assume that the sensor array is a redundancy array and $K \leq M - 1$. Then $(\mathcal{E}_f(\hat{\boldsymbol{\theta}}), \mathcal{E}_p(\hat{\mathbf{p}}), \hat{\boldsymbol{\sigma}})$ is asymptotically an ML estimator of $(\mathcal{E}_f(\boldsymbol{\theta}), \mathcal{E}_p(\mathbf{p}), \boldsymbol{\sigma})$.

Proof: Consider first the ULA case. Since $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}}, \hat{\boldsymbol{\sigma}})$ can be uniquely determined given $\hat{\mathbf{R}}$, the SPA estimator is equivalent to solving the following (nonconvex) optimization problem:⁴

$$\min_{\boldsymbol{\theta} \in [0, 1]^{M-1}, \mathbf{p} \in \mathbb{R}_+^{M-1}, \boldsymbol{\sigma} \geq \mathbf{0}} \left\| \mathbf{R}^{-\frac{1}{2}} \left(\tilde{\mathbf{R}} - \mathbf{R} \right) \tilde{\mathbf{R}}^{-\frac{1}{2}} \right\|_{\text{F}}^2, \quad (36)$$

subject to $\mathbf{R} = \mathbf{A}(\boldsymbol{\theta}) \text{diag}(\mathbf{p}) \mathbf{A}^H(\boldsymbol{\theta}) + \text{diag}(\boldsymbol{\sigma})$.

On the other hand, the data snapshots $\{\mathbf{y}(t)\}$ are i.i.d. Gaussian with mean zero and covariance \mathbf{R} under the assumptions. According to the extended invariance principle (EXIP) [36] and the derivations in [34], the global minimizer of (36) is a large-snapshot realization of the ML estimator of $(\mathcal{E}_f(\boldsymbol{\theta}), \mathcal{E}_p(\mathbf{p}), \boldsymbol{\sigma})$.

Similarly, the SPA estimator in the redundancy SLA case is the global minimizer of the following problem:

$$\min_{\boldsymbol{\theta} \in [0, 1]^{M-1}, \mathbf{p} \in \mathbb{R}_+^{M-1}, \boldsymbol{\sigma}_\Omega \geq \mathbf{0}} \left\| \mathbf{R}_\Omega^{-\frac{1}{2}} \left(\tilde{\mathbf{R}}_\Omega - \mathbf{R}_\Omega \right) \tilde{\mathbf{R}}_\Omega^{-\frac{1}{2}} \right\|_{\text{F}}^2,$$

subject to $\mathbf{R}_\Omega = \mathbf{A}_\Omega(\boldsymbol{\theta}) \text{diag}(\mathbf{p}) \mathbf{A}_\Omega^H(\boldsymbol{\theta}) + \text{diag}(\boldsymbol{\sigma}_\Omega)$. (37)

Then the same result follows. ■

Theorem 3 states that $(\mathcal{E}_f(\hat{\boldsymbol{\theta}}), \mathcal{E}_p(\hat{\mathbf{p}}), \hat{\boldsymbol{\sigma}})$ is asymptotically an ML estimator of $(\mathcal{E}_f(\boldsymbol{\theta}), \mathcal{E}_p(\mathbf{p}), \boldsymbol{\sigma})$ under some technical assumptions. Theorem 2 shows that this estimator is also consistent. Since, asymptotically, the power estimator $\mathcal{E}_p(\hat{\mathbf{p}})$ may lie on the boundary of \mathbb{R}_+^{M-1} and thus the asymptotic normality of $(\mathcal{E}_f(\hat{\boldsymbol{\theta}}), \mathcal{E}_p(\hat{\mathbf{p}}), \hat{\boldsymbol{\sigma}})$ does not hold directly. However, it indeed holds in the case of $K = M - 1$ where the true parameter \mathbf{p}^o is an interior point of \mathbb{R}_+^{M-1} . So we have the following theorem.

Theorem 4: Under the assumptions of Theorem 3 and $K = M - 1$ with $\mathbf{p} \succ \mathbf{0}$ and $\boldsymbol{\sigma} \succ \mathbf{0}$, the SPA estimator

³Without ambiguity, $\hat{\boldsymbol{\sigma}}$ and $\boldsymbol{\sigma}$ denote respectively $\hat{\boldsymbol{\sigma}}_\Omega$ and $\boldsymbol{\sigma}_\Omega$ in the SLA case, or $\hat{\boldsymbol{\sigma}}$ and $\boldsymbol{\sigma}$ in the case of equal noise variances. It is the same for Theorems 3 and 4.

⁴The zero entries, if any, of the solution of \mathbf{p} are removed in SPA as well as the corresponding entries of the solution of $\boldsymbol{\theta}$.

$(\hat{\theta}, \hat{\mathbf{p}}, \hat{\sigma})$ is asymptotically normal with asymptotic unbiased mean $(\theta^o, \mathbf{p}^o, \sigma^o)$ and covariance \mathbf{F}^{-1} , where \mathbf{F} denotes the Fisher information matrix of the parameter. It follows that $(\hat{\theta}, \hat{\mathbf{p}}, \hat{\sigma})$ is statistically asymptotically efficient.

Proof: The result follows from Theorems 2 and 3 and the properties of consistent ML estimators. ■

Remark 5: The asymptotic covariance matrix \mathbf{F}^{-1} is usually referred to as the Crammer-Rao lower bound (CRLB), which can be computed following from [42] and will be omitted in this paper. Alternatively, its numerical results will be presented in Section VII.

VI. CONNECTION TO PRIOR ARTS

A. Connection to SPICE

The SPA method proposed in this paper is closely connected to SPICE in [21]. The two methods adopt the same covariance fitting criteria. Roughly speaking, SPICE can be considered as a discretized version of SPA when applied to ULAs and SLAs. However, the following two main differences make their performances very different. One is that SPA is a parametric method while SPICE is a semiparametric method. In particular, the covariance \mathbf{R} is approximated in SPICE by discretizing the continuous range $[0, 1)$ of the frequency. Consider the ULA case as an example. Denote $\tilde{\theta} \in [0, 1)^{\tilde{N}}$ the discretized sampling grid of the frequency and $\tilde{\mathbf{p}} \in \mathbb{R}_+^{\tilde{N}}$ the corresponding power vector, where \tilde{N} denotes the grid size. Then \mathbf{R} is expressed as

$$\mathbf{R}(\tilde{\mathbf{p}}, \sigma) = \mathbf{A}(\tilde{\theta}) \text{diag}(\tilde{\mathbf{p}}) \mathbf{A}^H(\tilde{\theta}) + \text{diag}(\sigma) \quad (38)$$

which is a linear function of $(\tilde{\mathbf{p}}, \sigma)$. With (38), the covariance fitting criterion (10) or (11) becomes a convex function of $(\tilde{\mathbf{p}}, \sigma)$ and is optimized in [20], [21] via an iterative algorithm, named as SPICE. Note that the characterization of \mathbf{R} in (38) is only an approximation since there is no guarantee that the true frequencies lie on the grid $\tilde{\theta}$. Thus the approximation error (or modeling error), which depends on the grid density, is one potential reason causing estimation inaccuracy of SPICE. The other difference is that the parameter estimate of SPICE is obtained directly from the solution $(\tilde{\mathbf{p}}^*, \sigma^*)$ of the covariance fitting optimization problem while an additional postprocessing procedure is carried out in SPA. In particular, a source at θ_j is detected in SPICE once $\tilde{p}_j^* > 0$ and the frequency estimate is constrained on the grid, which will be referred to as the on-grid issue hereafter and becomes a second potential reason causing inaccuracy of SPICE. Furthermore, the parameter estimate without the postprocessing is generally not unique according to Remark 2 due to an identifiability problem which also exists in the solution $(\tilde{\mathbf{p}}^*, \sigma^*)$ of SPICE. This can be a third potential reason causing inaccuracy of SPICE. In fact, the SPICE estimate might not be *sparse* since the support of $\tilde{\mathbf{p}}^*$ can be as large as its dimension \tilde{N} according to Remark 2, which will be numerically verified in Section VII.

In summary, there are three potential reasons that may cause inaccuracy to the parameter estimation of SPICE, including the modeling error, the on-grid issue and the identifiability problem, rendering that SPICE does not possess the sparse and

the statistical properties of SPA presented in Section V. The first reason corresponds to the first difference and is introduced by the discretization, which can be alleviated by adopting a dense sampling grid but at the cost of more expensive computations. The last two are related to the second difference. Inspired by SPA, we present SPICE-PP to resolve them using the postprocessing technique in the next subsection.⁵

Before proceeding to SPICE-PP, we compare computational costs of SPICE and SPA. We consider only the dominant part in the following comparison, i.e., the computation of $\hat{\mathbf{R}}^{\frac{1}{2}}$ or $\hat{\mathbf{R}}_{\Omega}^{\frac{1}{2}}$ is excluded for both SPICE and SPA which takes $O(M^2N + M^3)$ or $O(L^2N + L^3)$ flops, respectively, and the postprocessing and parameter solving are also excluded for SPA which take $O(M^3)$ flops. SPICE is an iterative algorithm whose computational complexity equals the complexity per iteration times the number of iterations, i.e., $O(M^3\tilde{N}T)$ in the ULA case and $O(L^3\tilde{N}T)$ in the SLA case, where T denotes the number of iterations which is hard to quantify and empirically observed to vary in different scenarios. For SPA we adopt an off-the-shelf SDP solver, SDPT3 [27], where the interior-point method is implemented to solve the SDP. Denote by n_1 and $n_2 \times n_2$ the variable size and dimension of the positive semidefinite matrix in the semidefinite constraint of an SDP, respectively. Then the SDP can be solved in $O(n_1^2n_2^{2.5})$ flops in the worst case according to [43]. In the ULA case, n_1 is on the order of M^2 and n_2 is proportional to M . It follows that the complexity of SPA is $O(M^{6.5})$. In the SLA case, n_1 and n_2 are on the order of $L^2 + M$ and $L + M$, respectively. Then the complexity is $O(L^4M^{2.5} + M^{4.5})$ since $L \leq M$. Without surprise, the order on M or L for SPA is higher than that for SPICE. But the positive side is that the complexity of SPA does not depend on the grid size \tilde{N} which is typically much greater than M in array processing. As a result, SPA can be possibly faster than SPICE if a dense sampling grid (large \tilde{N}) is adopted in SPICE for obtaining high accuracy, and vice versa. Note that the order on M or L might be decreased in the future for both SPA and SPICE if there are more sophisticated algorithms but the linear dependence on \tilde{N} for SPICE probably cannot due to the discretization (similar results hold for other discretization-based methods).

B. Proposed SPICE-PP

Like SPA, we modify SPICE and obtain the parameter estimate within three steps. Firstly, \mathbf{R} is estimated as

$$\hat{\mathbf{R}} = \mathbf{A}(\tilde{\theta}) \text{diag}(\tilde{\mathbf{p}}^*) \mathbf{A}^H(\tilde{\theta}) + \text{diag}(\sigma^*) \quad (39)$$

given the solution $(\tilde{\mathbf{p}}^*, \sigma^*)$ of the original SPICE. Then the postprocessing is applied to decompose $\hat{\mathbf{R}}$ into $\hat{\mathbf{R}} = T(\hat{\mathbf{u}}) + \text{diag}(\hat{\sigma})$ by exploiting $K \leq M - 1$. Finally, the parameter estimate is obtained via the Vandermonde decomposition of $T(\hat{\mathbf{u}})$. The modified SPICE algorithm is named as SPICE with the postprocessing, abbreviated as SPICE-PP. Note that the frequency estimate of SPICE-PP is no long constrained on

⁵Note that the on-grid issue is not necessarily a consequence of the discretization but can be solved using the postprocessing.

the grid. It is also noted that this postprocessing technique can be possibly applied to other covariance-based methods.

C. Connection to Existing Discretization-Free Methods in the Case of $N = 1$

In the limiting single-snapshot case, i.e., when $N = 1$, the array processing problem is mathematically equivalent to spectral analysis [2]. For the latter topic, discretization-free techniques have been recently proposed in [28]–[30] based on atomic norm minimization (see [44]) which, also called total variation norm in [28], is a continuous version of the ℓ_1 norm. It is noted that the SPA method proposed in this paper can also be applied to the single-snapshot case. In fact, the SDPs in (20) and (23) can be further simplified in such case. Given (23) as an example. It follows from $\tilde{\mathbf{R}} = \mathbf{y}\mathbf{y}^H$ that $\text{tr}(\tilde{\mathbf{R}}\mathbf{R}^{-1}\tilde{\mathbf{R}}) = \|\mathbf{y}\|_2^2 \mathbf{y}^H \mathbf{R}^{-1} \mathbf{y}$. So, an alternative formulation of (23) will be

$$\begin{aligned} & \min_{x, \mathbf{u}} x + \text{tr}(T(\mathbf{u})), \\ & \text{subject to } \begin{bmatrix} x & \|\mathbf{y}\|_2 \mathbf{y}^H \\ \|\mathbf{y}\|_2 \mathbf{y} & T(\mathbf{u}) \end{bmatrix} \geq \mathbf{0}. \end{aligned} \quad (40)$$

It is interesting to note that, though obtained from a very different technique, (40) is quite similar to the SDP formulations in [28]–[30]. A detailed investigation of the relation is beyond the scope of this paper and will be posed as a future work since this paper is focused on the array processing applications in which the multisnapshot case is of the main interest.

VII. NUMERICAL SIMULATIONS

A. Simulation Setups

In this section we illustrate the performance of the proposed SPA method and compare it with existing methods via numerical simulations. The methods that we consider include SPICE [21], SPICE-PP, IAA [5], MUSIC and OGSBI-SVD [13]. SPICE is a semiparametric method which can be roughly considered as a discretized version of SPA as described in Subsection VI-A. SPICE-PP is a modified version of SPICE by incorporating the postprocessing technique presented in this paper. IAA is an enhanced nonparametric method. MUSIC is a classical subspace-based parametric method. OGSBI-SVD is a semiparametric method for off-grid DOA estimation. The information of source number, K , is required in MUSIC but not in the other methods. Since SPA and SPICE operate in different manners in the cases of equal/different noise variances, ‘+’ is used to indicate the case when the equal noise variances assumption is imposed. For example, SPA+ refers to SPA with the assumption. Without ambiguity, ‘SPA’ can refer to either the collectively called SPA technique or the SPA method with different noise variances (in contrast to SPA+) hereafter (similarly for the use of ‘SPICE’). Some setups of the algorithms above are as follows. The SDPs of SPA are implemented using CVX with the SDPT3 solver [27], [45]. SPICE is implemented as in [21] and terminated when the relative change of the objective function value in two consecutive iterations falls below 1×10^{-6} or the maximum number of iterations, set to 500, is reached. IAA is terminated

if the relative change of the ℓ_2 norm of the power vector in two consecutive iterations falls below 1×10^{-6} or the maximum number of iterations, set to 500, is reached. OGSBI-SVD is implemented as in [13] except that the source number is unknown (see details in [46]).

B. Spectra Comparisons

We compare spectra of the aforementioned methods in this subsection. In our simulation, we consider $K = 3$ uncorrelated/coherent sources with power $\mathbf{p}^o = [5, 5, 1]^T$ from directions specified by the frequency vector $\boldsymbol{\theta}^o = [0.1014, 0.1532, 0.5077]^T$. A ULA with $M = 10$ is used to receive the signals. Each of the source signals is randomly generated with constant amplitude and random phase, which is usually the situation in communications applications [21]. Zero-mean white circular Gaussian noise is added with the noise variance σ^o . The signal to noise ratio (SNR) is defined as the ratio of the minimum source power to the noise variance (in dB), i.e., $\text{SNR} = 10 \log_{10} \frac{\min(p_j^o)}{\sigma^o}$. The grid number is set to 500 for SPICE and IAA.

Our simulation results of uncorrelated sources are presented in Fig. 1 with respect to different settings of (N, SNR) , where some curves are omitted for better visual effects. In the case of large snapshots, e.g., $N = 200$ as shown in Figs. 1(a) and 1(b), we observe the following phenomena. IAA produces a dense spectrum and exhibits significant resolution degradation in the moderate/low SNR regime. It cannot separate the first two sources in Fig. 1(b). SPICE+ produces a sparse spectrum which contains only a few spikes in the high SNR regime, e.g., $\text{SNR} = 20\text{dB}$. However, it produces a dense spectrum due to the identifiability problem as addressed in Subsection VI-A in the moderate/low SNR regime, e.g., $\text{SNR} = 0\text{dB}$, where it returns a zero estimate of the noise variance. Unlike SPICE+, SPICE always produces a dense spectrum whenever the SNR is (its spectra are omitted in Fig. 1). Without surprise, the SPA method proposed in this paper always produces a sparse spectrum. Small spurious spikes exhibit due to the presence of noise and absence of the knowledge of source number and the noise level. By using the postprocessing technique presented in this paper both SPICE+-PP and SPICE-PP produce sparse spectra, and their differences from SPA+ (or SPA) are caused by the modeling error of SPICE as described in Subsection VI-A. Moreover, SPICE+ and SPICE+-PP have the same spectrum in the high SNR regime since the postprocessing does not alter the spectrum when SPICE+ has already produced a sparse spectrum. This point will be revisited later.

The limiting case $N = +\infty$ with a very low SNR (-20dB) is studied in Fig. 1(c), where the true covariance matrix is adopted to implement SPA, SPICE and MUSIC. IAA is not considered in this scenario. It is shown that SPA (similarly for MUSIC) can exactly localize the three sources, which verifies the conclusion of Theorem 2 that SPA is statistically consistent. SPICE+ has low resolution in the low SNR regime due to the identifiability problem and cannot separate the first two sources. Due to the modeling error of SPICE, SPICE+-PP is not consistent as well. Though this paper is mainly focused on the case of moderate/large snapshots, it is noted that SPA

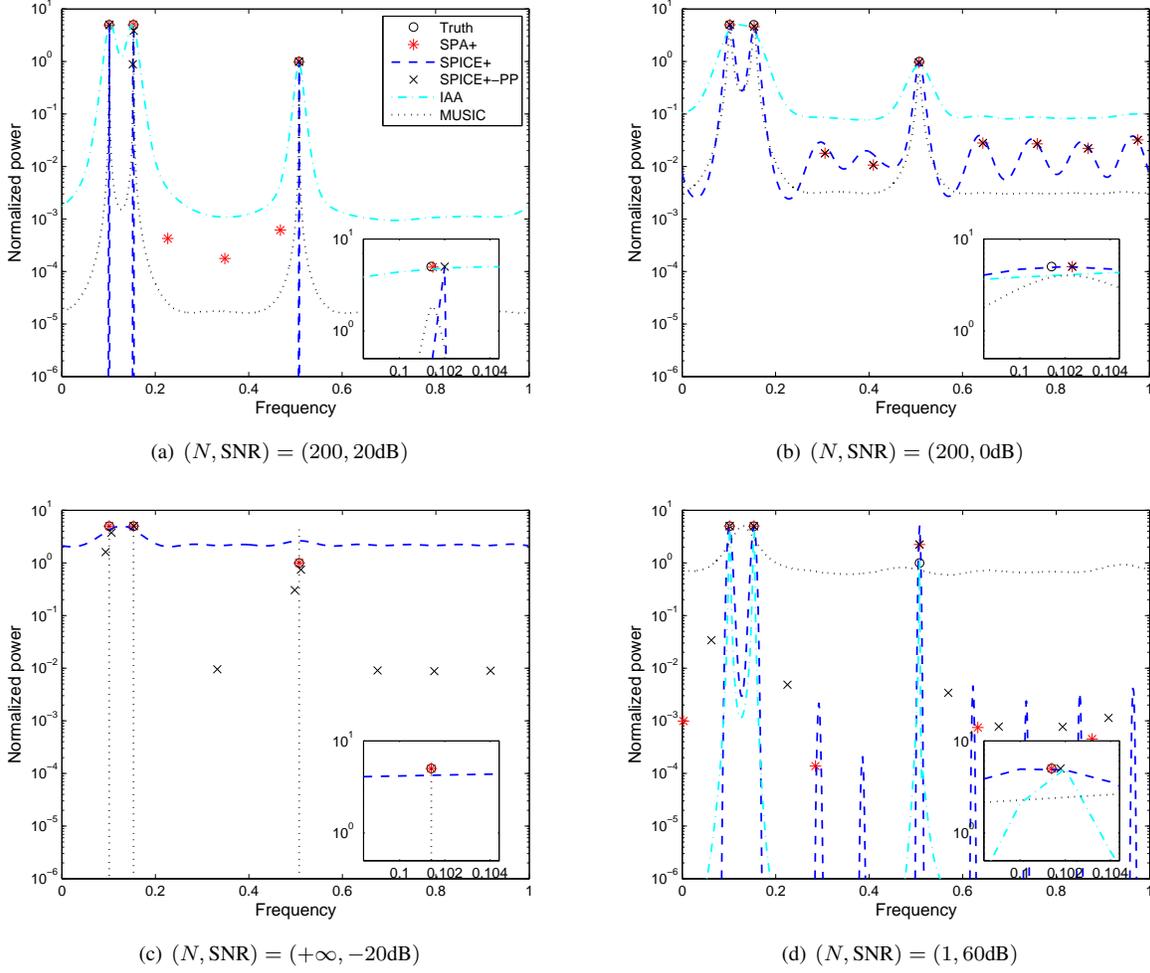


Fig. 1. Spectra of SPA+, SPICE+, SPICE+PP, IAA and MUSIC for uncorrelated sources with respect to (N, SNR) . Other settings include: $\theta^o = [0.1014, 0.1532, 0.5077]^T$, $p^o = [5, 5, 1]^T$ and ULA with $M = 10$. The subfigure in the lower right corner zooms in the area around source 1.

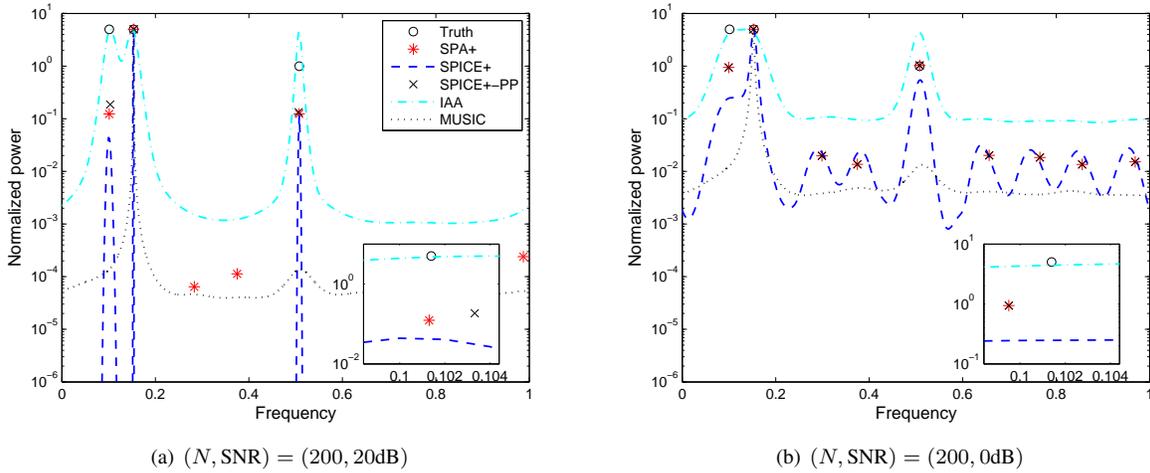


Fig. 2. Spectra of SPA+, SPICE+, SPICE+PP, IAA and MUSIC for coherent sources with respect to (N, SNR) , where source 3 is a replica of source 1. Other settings include: $\theta^o = [0.1014, 0.1532, 0.5077]^T$, $p^o = [5, 5, 1]^T$ and ULA with $M = 10$. The subfigure in the lower right corner zooms in the area around source 1.

can be applied to the case of small or even a single snapshot as shown in Fig. 1(d) where MUSIC fails.

Fig. 2 presents simulation results of coherent sources, where source 3 is exactly a replica of source 1. It is shown that the proposed SPA method has consistently good performance in the presence of complete correlation due to the adopted covariance fitting criterion as shown in [20], [21]. In contrast, IAA has low resolution as in the case of uncorrelated sources. MUSIC typically misses coherent sources. SPICE tends to miss coherent sources as well in the moderate/low SNR regime. Finally, note that the power estimates of the coherent sources are attenuated in SPA, the reason of which should be investigated in the future.

C. Quantitative Comparisons with SPICE

Quantitative comparisons will be carried out in this subsection to demonstrate advantages of our discretization-free SPA method compared to SPICE. Readers are referred to [21] for performance comparisons of SPICE with IAA and MUSIC. Unlike SPA, SPICE might produce a dense spectrum as the nonparametric methods. The frequency estimate of SPICE is obtained using the peaks of the spectrum following from [21]. To illustrate effects of the discretization adopted in SPICE, three discretization levels are considered with the grid size $\tilde{N} = 200, 500, 1000$, respectively. For convenience, the SPICE algorithm adopting the three discretization schemes will be referred to as SPICE1, SPICE2 and SPICE3, respectively. The ‘+’ symbol will be used as before. Metrics recorded include mean squared error (MSE) and CPU time usage. The MSE of the frequency estimation is computed as $\frac{1}{K} \left\| \hat{\theta}_K - \theta^o \right\|_2^2$ and then averaged over a number of Monte Carlo runs, where $\hat{\theta}_K$ denotes the frequency estimate which is obtained by keeping the associated largest K entries of the power estimate \hat{p} . The CRLB is commonly used as a benchmark when evaluating the performance of various estimators though it is a lower bound for only unbiased estimators. Note also that to compute the CRLB requires the knowledge of K which is not used in SPA and SPICE. So, there might exist a gap between the CRLB and the performance of SPA or SPICE that we study. The simulations study both the ULA and SLA cases and are focused on uncorrelated sources.

1) *The ULA Case: Experiment 1* studies performance variation with respect to the SNR. We consider a ULA with $M = 10$. Without loss of generality, $K = 2$ uncorrelated sources impinge on the array with (off-grid) frequencies $\frac{1}{6}$ and $\frac{4}{15}$ and unit powers. Notice that each frequency is a third grid interval away from the nearest grid point for SPICE. Since the best frequency estimate for a given source is the nearest grid point, the MSE of the frequency estimation of SPICE is lower bounded by $\frac{1}{9\tilde{N}^2}$ regardless of the SNR.⁶ The number of snapshots is set to $N = 200$ and the SNR varies in $\{-20, -15, \dots, 25\}$ dB. 200 Monte Carlo runs are used for each algorithm to obtain the metrics, where the source signals and the noise are both i.i.d. Gaussian. Fig. 3 plots MSEs

of the simulation results, where the two cases of equal and different noise variances are separately presented to provide a better illustration. SPICE has a better performance with a finer discretization but lower bounded by some constant as mentioned above. To the contrary, the MSEs of the SPA methods improve constantly with the SNR and gradually approach the CRLBs. Both the SPA methods and the SPICE methods have similar performance trends in the two cases of equal and different noise variances. However, SPICE-PPs perform differently. In the former case, SPICE+PPs and SPICE+s coincide in the high SNR regime since they produce the same sparse spectra as shown in Subsection VII-B. In the latter case, SPICE-PPs outperform the associated SPICES when the SNR is larger than some threshold thanks to the postprocessing technique presented in this paper. But when the SNR is sufficiently high, the modeling error caused by the discretization dominates the total uncertainties and further performance improvement is impossible. On the other hand, in the moderate/low SNR regime where the measurement noise dominates the uncertainties, SPICE-PP coincides with SPA as expected. Note that the bad performances of SPICE2+PP and SPICE3+PP at SNR = 5dB are caused by very few outliers (1 and 4 trials, respectively, out of 200), where SPICE+ might not converge within 500 iterations and result in a less accurate \hat{R} that is used for parameter estimation. Finally, notice that SPICE can possibly outperform SPA in the SNR range $[-10, 10]$ dB (depending on the discretization level), which will be discussed in Subsection VII-E.

Fig. 4 presents CPU times of *Experiment 1*. Since the postprocessing can be applied efficiently, the time usage of SPICE-PP is slightly longer than SPICE and is omitted. Both SPICE+s and SPICES have similar performance trends with different discretization levels because the number of iterations used is approximately the same. As a result, the time usage of SPICE is proportional to the grid size \tilde{N} while the proposed discretization-free SPA does not depend on \tilde{N} . Fig. 4 shows that when $\tilde{N} = 1000$ SPICE+ and SPICE are constantly slower than SPA+ and SPA, respectively. When $\tilde{N} = 500$, SPICE+ is also slower than SPA+ sometimes.

Experiment 2 studies the performance with respect to the array length M . We repeat *Experiment 1* but set SNR = 10dB and vary M in $\{5, 10, \dots, 40\}$. Moreover, we consider only the case $\tilde{N} = 1000$ for SPICE. Only the case where $M \leq 20$ is considered for SPA (excluding SPA+) due to time consideration. The simulation results are presented in Fig. 5. Since CRLB+ and CRLB are slightly different and almost undistinguishable, only CRLB+ is plotted. As before, SPICE3+, SPICE3 and SPICE3+PP share the same lower bound due to discretization and the property of the SPICE+ estimator, while SPA+, SPA and SPICE3-PP can outperform this bound. As M increases, e.g., when $M \geq 25$, the performances of SPA and SPICE3-PP hardly improve and the gaps between the algorithms and CRLB+ become larger. Two possible reasons are as follows: 1) Unlike CRLB+, SPA does not use the knowledge of K but $K \leq M - 1$, which becomes rougher when M increases and K keeps unaltered, and 2) The modeling error of SPICE increases as M increases. The figure at the bottom indicates that the speed of SPA+ scales well with

⁶Under the assumption that the source is randomly located in one or more grid intervals, the lower bound will be $\frac{1}{12\tilde{N}^2}$ [13].

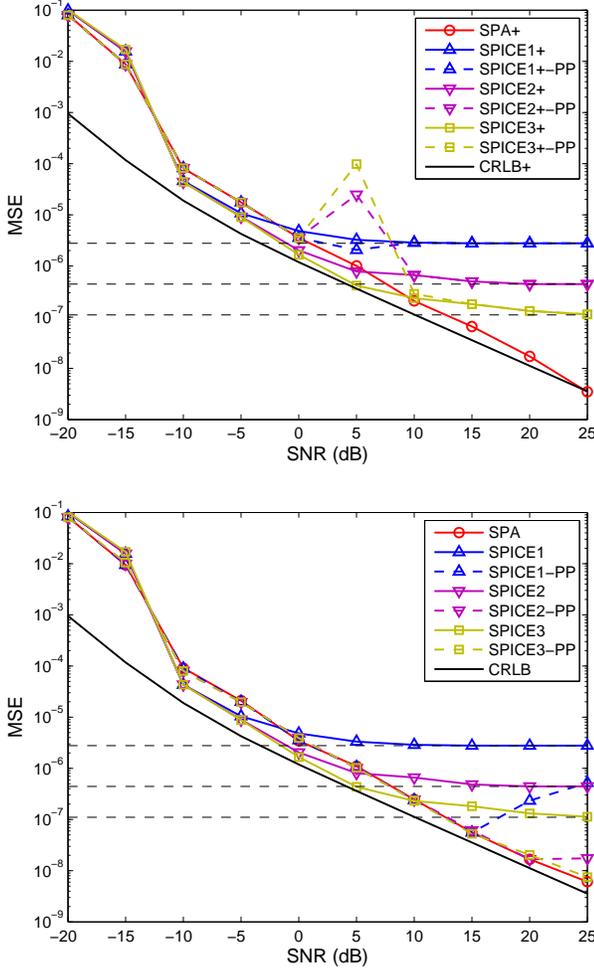


Fig. 3. MSEs of frequency estimates of SPA compared with SPICE and the CRLB. Some settings include: ULA with $M = 10$, $K = 2$ uncorrelated sources with $\theta^o = [\frac{1}{6}, \frac{4}{15}]^T$ and $p^o = [1, 1]^T$, and $N = 200$. The horizontal dashed lines are lower bounds of the SPICES due to the discretization.

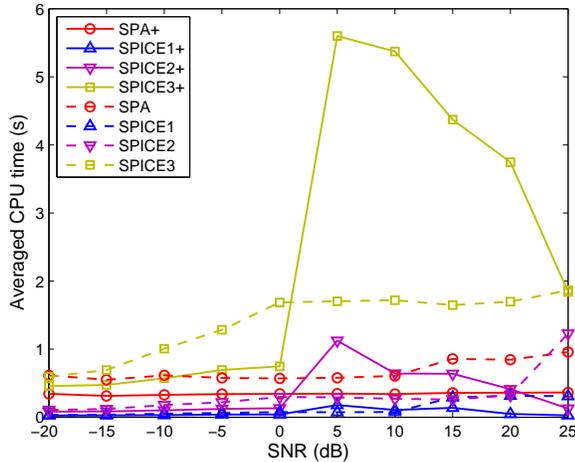


Fig. 4. CPU time usages of SPA compared with SPICE for estimating $K = 2$ uncorrelated sources. Some settings include: ULA with $M = 10$, $\theta^o = [\frac{1}{6}, \frac{4}{15}]^T$, $p^o = [1, 1]^T$ and $N = 200$.

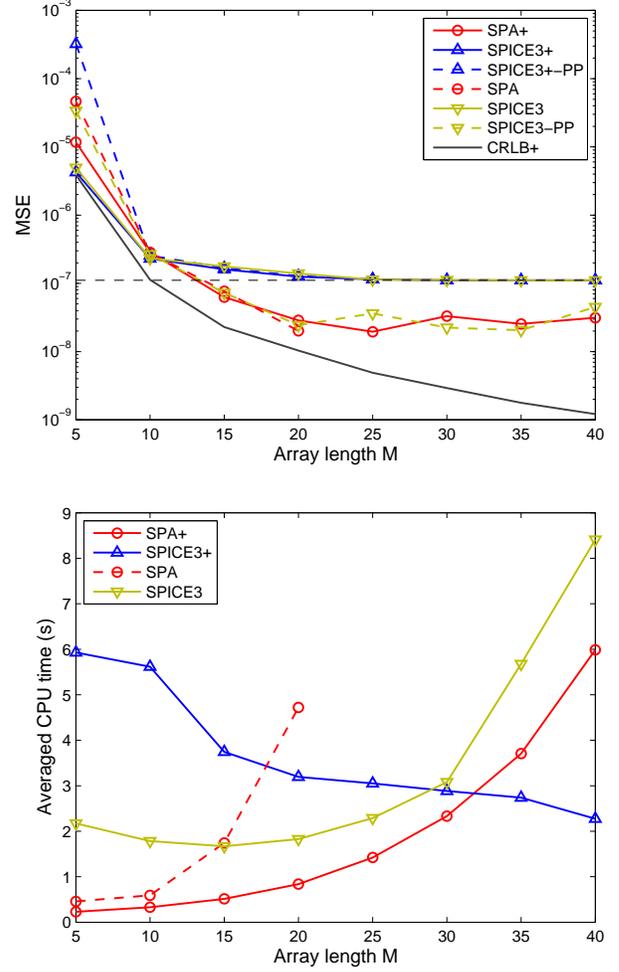


Fig. 5. MSEs of frequency estimates (top) and CPU time usages (bottom) of SPA with an M -element ULA compared with SPICE. Some settings include: $K = 2$ uncorrelated sources with $\theta^o = [\frac{1}{6}, \frac{11}{30}]^T$ and $p^o = [1, 1]^T$, SNR = 10dB and $N = 200$.

M in our considered scenario and is faster than SPICE3+ as $M \leq 30$ and constantly faster than SPICE3. The number of iterations of SPICE3+ is empirically observed to decrease with increasing array length in our considered scenario, leading to a seeming strange result that SPICE3+ gets faster when M increases.

2) *The SLA Case: Experiment 3* studies the SLA case where a 4-element redundant array $\Omega = \{1, 2, 5, 7\}$ is considered. We try to verify Theorem 4 and attempt to locate maximally $K = 6$ uncorrelated sources with the frequency vector $\theta^o = [0.1008, 0.1809, 0.4001, 0.5509, 0.7006, 0.8501]^T$ and power vector $p^o = [2, 2, 2, 1, 1, 1]^T$. Moreover, we set SNR = 10dB and vary the number of snapshots N in $\{20, 200, 2000, 20000\}$. 1000 Monte Carlo runs are used to obtain the metrics for each N . To evaluate the performance, we calculate the ratio $\frac{\text{CRLB}^+}{\text{MSE}}$ for each algorithm at each N . For an unbiased estimator, the ratio is called its efficiency. The larger the ratio is (≤ 1 for an unbiased estimator), the more accurate the estimator will be. Table. I presents the simulation results. Remarkably, SPA+ (or SPA) can outperform the CRLB+ (or

TABLE I
MSEs OF FREQUENCY ESTIMATES OF SPA COMPARED WITH SPICE AND CRLB+. EACH OF THE PRESENTED VALUES IS THE RATIO $\frac{\text{MSE}(\text{OR CRLB})}{\text{MSE}(\text{OR CRLB})}$ (THE LARGER THE BETTER).

	$N = 20$	$N = 200$	$N = 2000$	$N = 20000$
CRLB+	1	1	1	1
SPA+	0.6780	1.2558	1.1327	1.0370
SPICE3+	0.4730	1.0419	0.2286	0.0267
SPICE3+PP	0.6610	1.2528	0.6039	0.1049
CRLB	0.9281	0.9281	0.9281	0.9281
SPA	0.5615	1.2134	1.0671	1.0168
SPICE3	0.5019	0.9535	0.2250	0.0267
SPICE3-PP	0.5555	1.1693	0.5791	0.1048

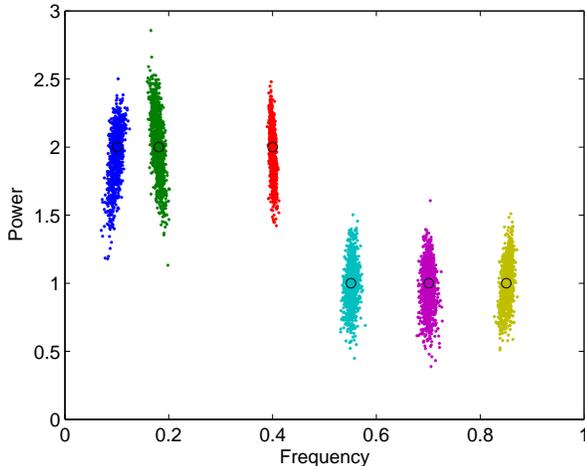


Fig. 6. Frequency and power estimates of SPA+ for estimating $K = 6$ uncorrelated sources with a 4-element SLA $\Omega = \{1, 2, 5, 7\}$ when $N = 200$ and SNR = 10dB (1000 Monte Carlo runs). Black circles indicate the true frequencies and corresponding powers.

CRLB) when $N \geq 200$ since the ratios are larger than 1 (or 0.9281). Though SPICE+ (or SPICE) can outperform the CRLB+ (or CRLB) as well at $N = 200$, its gap to the CRLB+ (or CRLB) becomes larger as N increases. Moreover, SPICE+PP (or SPICE-PP) is consistently better than SPICE+ (or SPICE) and worse than SPA+ (or SPA). Notice that as N gets larger, the gap between SPA+ (or SPA) and 1 (or 0.9281) becomes smaller, which is consistent with the conclusions of Theorem 4 that $\frac{\text{CRLB}^+}{\text{MSE}_{\text{SPA}^+}}, \frac{\text{CRLB}}{\text{MSE}_{\text{SPA}}} \rightarrow 1$, as $N \rightarrow +\infty$. It is also noted that the source power and noise variance estimates of the proposed SPA method have similar performances and their metrics are omitted. The frequency and power estimates of SPA+ at $N = 200$ are plotted in Fig. 6.

D. Comparison With OGSBI-SVD

We compare SPA with the off-grid method OGSBI-SVD in [13] in this subsection. OGSBI-SVD utilizes the sparse Bayesian learning (SBL) technique [47], [48], which mimics the ML estimation according to [22], and is based on an off-grid observation model which is a first-order approximation of the exact model with continuous frequencies and has a reduced modeling error (recall that the observation model of SPICE and many others is a zeroth-order approximation while SPA relies on the exact model). In OGSBI-SVD the grid offset

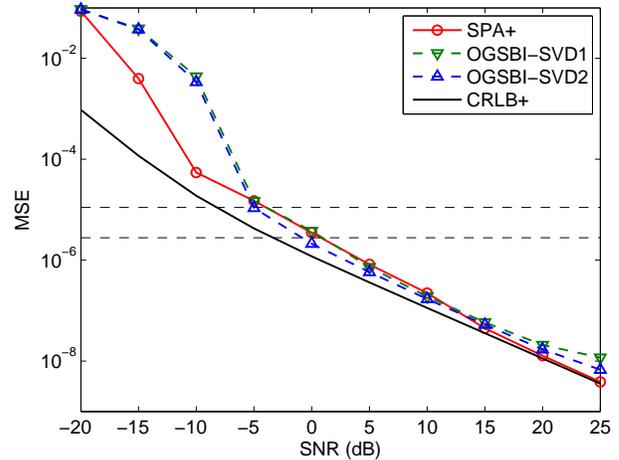


Fig. 7. MSEs of frequency estimates of $K = 2$ uncorrelated sources using SPA compared with OGSBI-SVD and the CRLB, with $N = 200$. The horizontal dashed lines refer to lower bounds of on-grid methods with $\tilde{N} = 100, 200$.

(the distance from a true frequency location to its nearest grid point) is estimated jointly with the sparse signal and singular value decomposition (SVD) is used for reducing dimension of the observed data and faster convergence. Since the covariance \mathbf{R} is also involved in OGSBI-SVD as in SPICE which cannot be appropriately separated into the source and noise parts, OGSBI-SVD usually produces a dense spectrum like SPICE.

In *Experiment 4*, we repeat *Experiment 1* for SPA+ and OGSBI-SVD and consider two discretization levels, $\tilde{N} = 100, 200$, for OGSBI-SVD (denoted by ‘1’, ‘2’ respectively). The MSEs are plotted in Fig. 7. As expected, OGSBI-SVD can exceed the lower bounds of on-grid methods (horizontal dashed lines). As SNR increases, e.g., at SNR = 25dB, the modeling error of OGSBI-SVD becomes nonnegligible. Then SPA is more accurate than OGSBI-SVD. Note also that SPA is more robust to noise. For example, SPA can produce satisfactory results at SNR = -10dB while OGSBI-SVD cannot. In computational speed, SPA is slightly slower than OGSBI-SVD with $\tilde{N} = 100$ and about 3 times faster with $\tilde{N} = 200$.

In *Experiment 5*, we repeat the simulation by fixing SNR = 10dB and varying the number of snapshots N from 1 to 200 at a step of 3. Simulation results presented in Fig. 8 show that SPA has consistently satisfactory performance though it is slightly worse than OGSBI-SVD (the MSEs of SPA are less than 1.5 times those of OGSBI-SVD in most of the scenarios). Since OGSBI-SVD mimics the ML estimation and SPA is a large-snapshot realization of the ML, the performance gap between them vanishes as N increases.

E. Discussion: Resolution vs. Robustness

If an infinitely fine discretization scheme is allowed in SPICE (to eliminate the modeling error), then SPICE and SPA will produce the same covariance matrix estimate $\hat{\mathbf{R}}$. However, SPA always has a sparse spectrum while SPICE may produce a dense one. The two spectra can be considered as

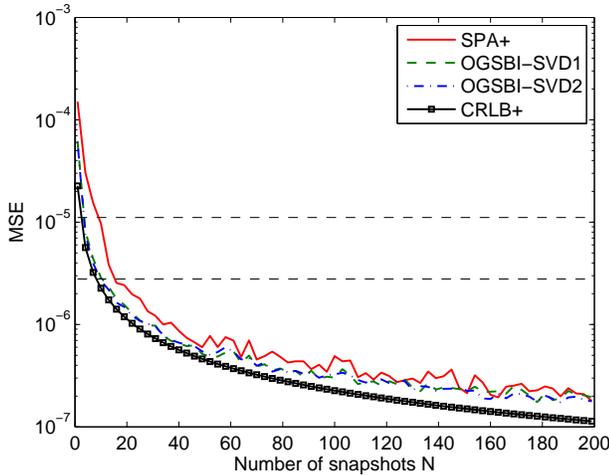


Fig. 8. MSEs of frequency estimates of $K = 2$ uncorrelated sources using SPA compared with OGSBI-SVD and the CRLB, with SNR = 10dB. The horizontal dashed lines refer to lower bounds of on-grid methods with $\tilde{N} = 100, 200$.

two different decomposition schemes applied to \hat{R} to estimate the source powers. The sparse decomposition of SPA has good statistical properties as presented in Subsection V-B while the dense decomposition of SPICE does not. Moreover, as shown in Figs. 1(c) and 2(b), the dense decomposition of SPICE has potentially inferior resolution. However, the simulation results of Fig. 3 show that SPICE can outperform SPA in terms of MSE in the middle range of the SNR (similar results are presented in Figs. 7 and 8 for OGSBI-SVD). It is because that 1) SPA is empirically observed to produce a more heavy-tailed frequency estimator than SPICE since the frequency of a source is determined by a single point unlike SPICE for which the frequency estimate is given by the peaks of the spectrum,⁷ and 2) the MSE metric is sensitive to heavy-tailed estimators (in fact, a careful study reveals that SPA+ is more accurate than SPICE3+ in over half of the Monte Carlo runs in the SNR range $[-10, 5]$ dB in Fig. 3).

VIII. CONCLUSION AND FUTURE WORK

In this paper, the linear array signal processing problem was studied and a discretization-free technique named as SPA was proposed. The new method adopts the covariance fitting criteria of SPICE and was formulated as an SDP followed by a postprocessing technique. SPA is a parametric method and guarantees to produce a sparse parameter estimate and in the mean time enjoys several other merits. Its asymptotic statistical properties were analyzed and practical performance was demonstrated via simulations compared to existing methods.

The following directions should be investigated in the future, some of which have been mentioned in the main context

⁷The use of peaks of a dense spectrum as the frequency estimates is inspired by spectral-based methods, e.g. MUSIC and beamformer, and has been adopted in many “sparse” methods (see, e.g., [13], [21]). However, it conflicts with the principle of sparse methods that any (significant) nonzero power estimate corresponds to a source (ruling out numerical effects). This problem has been neglected in the literature and should be investigated in the future.

of this paper: 1) connection of the proposed SPA method to the atomic norm-based discretization-free methods in [28]–[30] for spectral analysis, 2) postprocessing techniques for general but not redundancy SLAs to utilize the full information of the range of the source number, 3) performance analysis of SPA in the finite-snapshot case while this paper is mainly on its asymptotic statistical properties, 4) fast implementations of the SPA method via developing more computationally efficient algorithms for solving the SDPs involved, 5) modified SPA methods with automatic source number estimation, and 6) discretization-free methods for array processing with general array geometries.

REFERENCES

- [1] H. Krim and M. Viberg, “Two decades of array signal processing research: The parametric approach,” *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.
- [2] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson/Prentice Hall Upper Saddle River, NJ, 2005.
- [3] R. Schmidt, “A signal subspace approach to multiple emitter location spectral estimation,” Ph.D. dissertation, Stanford University, 1981.
- [4] P. Stoica and N. Arye, “MUSIC, maximum likelihood, and Cramer-Rao bound,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 5, pp. 720–741, 1989.
- [5] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, “Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 425–443, 2010.
- [6] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [7] E. Candès, “Compressive sampling,” in *Proceedings of the International Congress of Mathematicians*, vol. 3, 2006, pp. 1433–1452.
- [8] D. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [9] D. Malioutov, M. Cetin, and A. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays,” *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.
- [10] M. Hyder and K. Mahata, “Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation,” *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646–4655, 2010.
- [11] Z. Tang, G. Blaquiere, and G. Leus, “Aliasing-free wideband beamforming using sparse signal representation,” *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3464–3469, 2011.
- [12] X. Wei, Y. Yuan, and Q. Ling, “DOA estimation using a greedy block coordinate descent algorithm,” *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6382–6394, 2012.
- [13] Z. Yang, L. Xie, and C. Zhang, “Off-grid direction of arrival estimation using sparse Bayesian inference,” *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, 2013.
- [14] A. C. Gurbuz, V. Cevher, and J. H. McClellan, “Bearing estimation via spatial sparsity using compressive sensing,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1358–1369, 2012.
- [15] N. Hu, Z. Ye, X. Xu, and M. Bao, “DOA estimation for sparse array via sparse signal reconstruction,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 2, pp. 760–773, 2013.
- [16] H. Abeida, Q. Zhang, J. Li, and N. Merabine, “Iterative sparse asymptotic minimum variance based approaches for array processing,” *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 933–944, 2013.
- [17] M. F. Duarte and R. G. Baraniuk, “Spectral compressive sensing,” *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, 2013.
- [18] T. Huang, Y. Liu, H. Meng, and X. Wang, “Adaptive matching pursuit with constrained total least squares,” *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–12, 2012.
- [19] M. Carlin, P. Rocca, G. Oliveri, F. Viani, and A. Massa, “Directions-of-arrival estimation through Bayesian compressive sensing strategies,” vol. 61, no. 7, pp. 3828–3838, 2013.
- [20] P. Stoica, P. Babu, and J. Li, “New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data,” *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 35–47, 2011.

- [21] —, “SPICE: A sparse covariance-based estimation method for array processing,” *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, 2011.
- [22] P. Stoica and P. Babu, “SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation,” *Signal Processing*, vol. 92, no. 7, pp. 1580–1590, 2012.
- [23] C. D. Austin, R. L. Moses, J. N. Ash, and E. Ertin, “On the relation between sparse reconstruction and parameter estimation with model order selection,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 3, pp. 560–570, 2010.
- [24] P. Stoica and P. Babu, “Sparse estimation of spectral lines: Grid selection problems and their solutions,” *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 962–967, 2012.
- [25] Z. Yang, C. Zhang, and L. Xie, “Robustly stable signal recovery in compressed sensing with structured matrix perturbation,” *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4658–4671, 2012.
- [26] L. Hu, Z. Shi, J. Zhou, and Q. Fu, “Compressed sensing of complex sinusoids: An approach based on dictionary refinement,” *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3809–3822, 2012.
- [27] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, “SDPT3—a MATLAB software package for semidefinite programming, version 1.3,” *Optimization Methods and Software*, vol. 11, no. 1–4, pp. 545–581, 1999.
- [28] E. J. Candès and C. Fernandez-Granda, “Towards a mathematical theory of super-resolution,” *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.
- [29] B. N. Bhaskar, G. Tang, and B. Recht, “Atomic norm denoising with applications to line spectral estimation,” *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.
- [30] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, “Compressed sensing off the grid,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.
- [31] Z. Yang and L. Xie, “On gridless sparse methods for line spectral estimation from complete and incomplete data,” *Available online at <http://arxiv.org/abs/1407.2490>*, 2014.
- [32] —, “Continuous compressed sensing with a single or multiple measurement vectors,” in *IEEE Workshop on Statistical Signal Processing (SSP)*, 2014, pp. 308–311.
- [33] —, “Exact joint sparse frequency recovery via optimization methods,” *Available online at <http://arxiv.org/abs/1405.6585>*, 2014.
- [34] B. Ottersten, P. Stoica, and R. Roy, “Covariance matching estimation techniques for array signal processing applications,” *Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.
- [35] D. A. Linebarger, I. H. Sudborough, and I. G. Tollis, “Difference bases and sparse sensor arrays,” *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 716–721, 1993.
- [36] P. Stoica and T. Soderstrom, “On reparametrization of loss functions used in estimation and the invariance principle,” *Signal processing*, vol. 17, no. 4, pp. 383–387, 1989.
- [37] C. Rojas, D. Katselis, and H. Hjalmarsson, “A note on the SPICE method,” *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4545–4551, 2013.
- [38] P. D. Grünwald, *The minimum description length principle*. MIT press, 2007.
- [39] U. Grenander and G. Szegő, *Toeplitz forms and their applications*. Univ of California Press, 1958.
- [40] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, “Sparse sampling of signal innovations,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, 2008.
- [41] P. Stoica and Y. Selen, “Model-order selection: a review of information criterion rules,” *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.
- [42] P. Stoica and A. Nehorai, “Performance study of conditional and unconditional direction-of-arrival estimation,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, 1990.
- [43] L. Vandenberghe and S. Boyd, “Semidefinite programming,” *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.
- [44] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,” *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.
- [45] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming,” *Available online at <http://cvxr.com/cvx>*, 2008.
- [46] Z. Yang, “Analysis, algorithms and applications of compressed sensing,” Ph.D. dissertation, Nanyang Technological University, Available online at https://dl.dropboxusercontent.com/u/34897711/Thesis_YANGZai.pdf, 2013.
- [47] M. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
- [48] S. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.



Zai Yang was born in Anhui, China, in 1986. He received the B.S. degree in mathematics and M.S. degree in applied mathematics from Sun Yat-sen (Zhongshan) University, China, in 2007 and 2009 respectively, and the Ph.D degree in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2014. He is currently a post-doc at NTU. His research interests include compressed sensing, spectral analysis of signals, array processing, source localization, and medical imaging.



Lihua Xie (F'07) received the B.E. and M.E. degrees in electrical engineering from Nanjing University of Science and Technology in 1983 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Newcastle, Australia, in 1992. Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a professor and the Head of Control and Instrumentation Division. He held teaching appointments in the Department of Automatic Control,

Nanjing University of Science and Technology from 1986 to 1989 and Changjiang Visiting Professorship with South China University of Technology from 2006 to 2011.

Dr Xie's research interests include robust control, estimation theory, signal processing, sensor networks, networked control systems, and multi-agent networks. In these areas, he has published many journal papers and co-authored two patents and four books. He has served as an editor of IET Book Series in Control and an Associate Editor of a number of journals including *IEEE Transactions on Automatic Control*, *Automatica*, *IEEE Transactions on Control Systems Technology*, and *IEEE Transactions on Circuits and Systems-II*. Dr Xie is Fellow of IEEE and Fellow of IFAC.



Cishen Zhang received the B.Eng. degree in computer engineering from Tsinghua University, China, in 1982 and Ph.D. degree in electrical engineering from Newcastle University, Australia, in 1990. He was an Electrician with Changxindian (February Seven) Locomotive Manufactory, Beijing, China during 1970-1978, carried out research on control systems at Delft University of Technology, the Netherlands in 1983-1985, was with the Department of Electrical and Electronic Engineering at the University of Melbourne, Australia, as a Lecturer,

Senior Lecturer, and Associate Professor and Reader in 1989-2002 and with the School of Electrical and Electronic Engineering and School of Chemical and Biomedical Engineering at Nanyang Technological University, Singapore, as an Associate Professor during 2002-2010. Since November 2010, he has been the Professor and Head of Electrical and Electronic Engineering with the Faculty of Engineering and Industrial Sciences at Swinburne University of Technology in Melbourne, Australia. His research interests include control, signal processing and medical imaging.