New Method of Sparse Parameter Estimation in Separable Models and Its Use for Spectral Analysis of Irregularly Sampled Data

Petre Stoica, *Fellow, IEEE*, Prabhu Babu, and Jian Li, *Fellow, IEEE*

***Abstract*—Separable models occur frequently in spectral analysis, array processing, radar imaging and astronomy applications. Statistical inference methods for these models can be categorized in three large classes: parametric, nonparametric (also called "dense") and semiparametric (also called "sparse"). We begin by discussing the advantages and disadvantages of each class. Then we go on to introduce a new semiparametric/sparse method called SPICE (a *s*emi*p*arametric/*sp*arse *i*terative *c*ovariance-based *e*stimation method). SPICE is computationally quite efficient, enjoys global convergence properties, can be readily used in the case of replicated measurements and, unlike most other sparse estimation methods, does not require any subtle choices of user parameters. We illustrate the statistical performance of SPICE by means of a line-spectrum estimation study for irregularly sampled data.**

***Index Terms*—Irregular sampling, separable models, sparse parameter estimation, spectral analysis.**

I. Introduction and Problem Formulation

LET $\boldsymbol{y} \in \mathbb{C}^{N\times 1}$ denote the available data vector (or snapshot), and consider the following model for $\boldsymbol{y}$:

$$\boldsymbol{y} = \sum_{c=1}^{C} \boldsymbol{a}(\tilde{\omega}_c)\tilde{s}_c + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{\epsilon} \in \mathbb{C}^{N\times 1}$ is a noise term, $\tilde{s}_c \in \mathbb{C}$ and $\tilde{\omega}_c \in \Omega \subset \mathbb{R}$ are the unknown parameters of the cth signal component, $\boldsymbol{a}(\cdot) : \Omega \to \mathbb{C}^{N\times 1}$ is a known function, and C is the unknown number of components. This type of model is a frequent occurrence in numerous applications such as spectral analysis [1], [2], array processing [3], radar imaging [4], [5], astronomy [6], [7], and elsewhere [8], in all of which estimation of the unknown parameters in (1) is a basic goal. Note that in some of these applications, for instance in array processing, the number of available snapshots is larger than one. However, to keep the notation and explanations as simple as possible, we will first consider the single snapshot case, but afterwards will also discuss briefly the extension to the multisnapshot (also called replicated-measurement) case.

The estimation methods associated with (1) can be categorized in three large classes: parametric, nonparametric and semiparametric, see below for details.

A prominent member of the *parametric class* is the nonlinear least squares (NLS) method that consists of minimizing the following criterion:

$$\left\| \boldsymbol{y} - \sum_{c=1}^{C} \boldsymbol{a}(\tilde{\omega}_c)\tilde{s}_c \right\|^2 \tag{2}$$

with respect to $\{\tilde{\omega}_c, \tilde{s}_c\}$ (for given C). Observe that the criterion above depends quadratically on $\{\tilde{s}_c\}$, which means that it can be minimized explicitly with respect to these parameters (for fixed $\{\tilde{\omega}_c\}$). In other words, $\{\tilde{s}_c\}$ can be separated out—and this along with the fact that the signal components enter (1) through separated terms give the name of *separable models* to (1) [8], [9]. The NLS enjoys excellent statistical properties; in particular, under the normal white-noise assumption the minimization of (2) produces the maximum-likelihood estimate that is asymptotically statistically efficient. However, this is true only if C used in (2) is the "true" number of components and if (2) can be globally minimized; and both these conditions are difficult to meet in practice (especially the global minimization of (2) is a hard task).

At the other end of the method spectrum we find the *nonparametric class*. The most basic method of this class is the single-frequency least-squares (SFLS) method (which is also known under other names, such as the periodogram or beamforming method, depending on the application). To explain this method in general terms, let $\{\omega_k\}_{k=1}^{K}$ denote a fine grid that covers Ω, and assume that $\{\tilde{\omega}_c\}$ lie on (practically, close to) the grid. This means that there exist $k_1, \ldots, k_C$ such that $\tilde{\omega}_c = \omega_{k_c}$ $(c = 1, \ldots, C)$. Let

$$\boldsymbol{a}_k = \boldsymbol{a}(\omega_k) \quad k = 1, \ldots, K \tag{3}$$

and also let

$$s_k = \begin{cases} \tilde{s}_c & k = k_c \quad (c = 1, \ldots, C) \\ 0 & \text{elsewhere} \end{cases} \tag{4}$$

Using this notation we can re-write (1) as:

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{a}_k s_k + \boldsymbol{\epsilon} \tag{5}$$

Manuscript received May 20, 2010; accepted September 25, 2010. Date of publication October 14, 2010; date of current version December 17, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alfred Hanssen. This work was supported in part by the Swedish Research Council (VR), the European Research Council (ERC), the Office of Naval Research (ONR) by Grants N00014-09-1-0211 and N00014-10-1-0054, and by the National Science Foundation (NSF) by Grant ECCS-0729727.

P. Stoica and P. Babu are with the Department of Information Technology, Uppsala University, Uppsala, SE 75105, Sweden (e-mail: prabhu.babu@it.uu.se).

J. Li is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA.

Digital Object Identifier 10.1109/TSP.2010.2086452

where typically $K \gg N$. The SFLS method estimates s_k, in a one-by-one manner, simply ignoring the presence of the other possible signal components in (5)

$$\hat{s}_k = \frac{\boldsymbol{a}_k^* y}{\|\boldsymbol{a}_k\|^2} \tag{6}$$

where the superscript $*$ denotes the conjugate transpose and $\|\cdot\|$ denotes the Euclidean (ℓ_2) norm. Evidently, (6) does not have the problems (indicated earlier) that affect the parametric method of NLS. However, the price for the elimination of these problems is poor statistical accuracy: SFLS suffers from local and global leakage problems; the local leakage reduces the resolution by making it hard to distinguish between signal components with closely spaced values of $\{\tilde{\omega}_c\}$, whereas global leakage leads to false-alarms, that is to large values of $\{|\hat{s}_k|\}$ for nonexistent components. Note that most (if not all) estimated values $\{\hat{s}_k\}$ in (6) will usually be different from zero, which motivates the alternative name of *dense* sometimes used to designate the nonparametric methods. In fact in some applications (e.g., radar imaging), in which the number of expected components in (5) is rather large, a dense estimate of $\{s_k\}$ can be preferable to a sparse estimate such as one provided by a parametric model. Consequently, there has been a significant interest in devising nonparametric (or dense) estimation methods for (1) that possess superior performance to the SFLS method. In Section II we will describe briefly one such enhanced method, called IAA (the **i**terative **a**daptive **a**pproach) [10], [11], which eliminates almost completely the leakage problems of the SFLS method in a fully data-adaptive manner (i.e., without requiring the selection of user parameters).

An intermediate category of methods is the *semiparametric class*. The sparse estimation methods form an important subgroup of this class. These methods use the nonparametric data model in (5) but, reminiscent of the parametric approach, they seek to exploit the information that the vector

$$\boldsymbol{s} = [s_1, \ldots, s_K]^T \tag{7}$$

in (5) is sparse (i.e., it has only a few nonzero elements). An archetypical sparse method consists of estimating $\{s_k\}$ by solving the following ℓ_1-norm constrained LS problem [12], [13]:

$$\min_{\{s_k\}} \left\| \boldsymbol{y} - \sum_{k=1}^{K} \boldsymbol{a}_k s_k \right\|^2 \quad \text{s.t. } \|\boldsymbol{s}\|_1 \triangleq \sum_{k=1}^{K} |s_k| \le \eta \tag{8}$$

where $\|\cdot\|_1$ stands for the ℓ_1 norm, and η is a threshold that must be chosen by the user. The ℓ_1-norm constraint in (8) is what induces the sparsity of the solution to (8) [12], which is a potentially useful feature of this type of estimation methods, provided that their user parameters [e.g., η in (8)] are well selected. However this selection is by no means a simple task [clearly it is related to the task of estimating C in the parametric model (1)], and quite typically the rules proposed for performing it depend on quantities that are unavailable in applications (such as the noise power, or the "true" value of C); see e.g., [13] for a recent critical discussion on this aspect. As alluded to above, most sparse estimation methods share this drawback with the parametric methods. However, there are a few sparse methods that are fully data adaptive—we will review one such method called SLIM (**s**parse **l**earning via **i**terative **m**inimization) [14], [15] in Section II. These user parameter-free sparse methods have an edge over the parametric ones (which require the selection of C), despite the fact that they cannot be possibly more accurate statistically nor are they necessarily more efficient computationally than the best methods of the parametric class. Additionally, many sparse estimation methods are numerically more reliable than the theoretically more accurate parametric methods (whose global convergence can rarely be guaranteed).

In this paper we will introduce a semiparametric/sparse estimation method for the separable model in (1) [see also (5)]. This method will be obtained using a covariance-based fitting criterion, which was apparently never employed before to derive sparse estimation methods, and will be designated by means of the acronym SPICE (**s**emi**p**arametric/**sp**arse **i**terative **c**ovariance-based **e**stimation). SPICE has several useful features that are shared by very few (if any) sparse estimation methods : i) it is fully data adaptive (i.e., its operation does not require the subtle selection of any user parameters); ii) it enjoys global convergence properties; and iii) its use in the multisnapshot case is straightforward. Following the theoretical derivation and analysis of SPICE, we describe the use of this method for the spectral analysis of irregularly sampled data and make use of a numerical example to illustrate the performance achievable by SPICE in the said application and compare this performance with that of SFLS, IAA, and SLIM.

II. The Competing Methods: IAA and SLIM

Let us assume that

$$E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^*) = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_N \end{bmatrix} \tag{9}$$

and that the phases of $\{s_k\}$ are independently and uniformly distributed in $[0, 2\pi]$. Then the covariance matrix of $\boldsymbol{y}$ has the following expression:

$$\boldsymbol{R} = E(\boldsymbol{y}\boldsymbol{y}^*) = \sum_{k=1}^{K} |s_k|^2 \boldsymbol{a}_k \boldsymbol{a}_k^* + \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_N \end{bmatrix} \triangleq \boldsymbol{A}^* \boldsymbol{P} \boldsymbol{A} \tag{10}$$

where

$$\boldsymbol{A}^* = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K \boldsymbol{I}] \triangleq [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{K+N}] \tag{11}$$

$$\boldsymbol{P} = \begin{bmatrix} |s_1|^2 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & |s_2|^2 & 0 & \cdots & \cdots & \vdots \\ \vdots & 0 & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \sigma_1 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \sigma_N \end{bmatrix}$$
$$\triangleq \begin{bmatrix} p_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & p_2 & 0 & \cdots & \cdots & \vdots \\ \vdots & 0 & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & p_{K+1} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & p_{K+N} \end{bmatrix}. \quad (12)$$

The assumption that the noise components in different measurements are uncorrelated to one another, which led to (9), is quite reasonable in most applications. On the other hand, the assumption that s_k and $s_{\bar{k}}$ are uncorrelated for $k \neq \bar{k}$ does not always hold. However, all methods considered in this paper are *robust* to this assumption, and thus they work well even in the case when some signal components are correlated to one another, as we explain in the next subsection.

A. IAA

Let $p_k(i)$ denote the estimate of p_k at the ith iteration, and let $\boldsymbol{R}(i)$ be the matrix $\boldsymbol{R}$ made from $\{p_k(i)\}$. Then IAA updates the powers by means of the following iterative process [10] and [11]:

$$p_k(i+1) = \frac{\left|\boldsymbol{a}_k^* \boldsymbol{R}^{-1}(i)\boldsymbol{y}\right|^2}{[\boldsymbol{a}_k^* \boldsymbol{R}^{-1}(i)\boldsymbol{a}_k]^2} \quad k = 1, \ldots, K+N. \quad (13)$$

The initial estimates $\{p_k(0)\}$ can be obtained, for example, using the SFLS method [see (6)]

$$p_k(0) = \frac{|\boldsymbol{a}_k^* \boldsymbol{y}|^2}{||\boldsymbol{a}_k||^4}. \quad (14)$$

The power estimation formula (13) has satisfactory properties even in those cases in which the covariance matrix of $\boldsymbol{y}$ does not have the assumed structure due to, for example, coherent signal components (i.e., components with the same phase). To explain why this is so, let us assume that two components, $\boldsymbol{a}_1 s_1$ and $\boldsymbol{a}_2 s_2$, in $\boldsymbol{y}$ are coherent. Then their covariance matrix is not

$$|s_1|^2 \boldsymbol{a}_1 \boldsymbol{a}_1^* + |s_2|^2 \boldsymbol{a}_2 \boldsymbol{a}_2^* \quad (15)$$

as assumed in (10), but

$$[|s_1|\boldsymbol{a}_1 + |s_2|\boldsymbol{a}_2][|s_1|\boldsymbol{a}_1 + |s_2|\boldsymbol{a}_2]^*. \quad (16)$$

However this difference between (15) and (16) does not cause any serious problem to (13). Observe that (13) can be rewritten as (omitting the iteration index for simplicity)

$$p_k = |\boldsymbol{h}_k^* \boldsymbol{y}|^2; \quad \boldsymbol{h}_k^* = \frac{\boldsymbol{a}_k^* \boldsymbol{R}^{-1}}{\boldsymbol{a}_k^* \boldsymbol{R}^{-1} \boldsymbol{a}_k}. \quad (17)$$

The "filter" (or linear combiner) $\boldsymbol{h}_k$ in (17) passes without distortion the signal component corresponding to $\boldsymbol{a}_k$. At the same time it attenuates (or even annihilates, depending on their powers) the signal components corresponding to $\boldsymbol{a}_1 \neq \boldsymbol{a}_k$ and $\boldsymbol{a}_2 \neq \boldsymbol{a}_k$, as it should (see [1] for details on this aspect). For $\boldsymbol{a}_k = \boldsymbol{a}_1$ the filter will attenuate any other component with $\boldsymbol{a}_{\bar{k}} \neq \boldsymbol{a}_1$, including $\boldsymbol{a}_2$ (and similarly, for $\boldsymbol{a}_k = \boldsymbol{a}_2$). Note that a similar argument also applies to the methods of SLIM and SPICE which are yet to be discussed: indeed, as we will see, the estimation formulas of these methods comprise a filter that is proportional to the $\boldsymbol{h}_k$ in (17). This fact explains why these methods as well are *robust* to the assumed structure of $\boldsymbol{R}$ in (10).

B. SLIM

This method operates under the assumption that $\sigma_1 = \cdots = \sigma_N \triangleq \sigma$ (which is a reasonable assumption in some applications). The updated estimates for SLIM are iteratively obtained as follows [14] and [15]:

$$\begin{aligned} s_k(i+1) &= p_k(i)\boldsymbol{a}_k^* \boldsymbol{R}^{-1}(i)\boldsymbol{y} \quad k = 1, \ldots, K \\ p_k(i+1) &= |s_k(i+1)|^2 \quad k = 1, \ldots, K \\ \sigma(i+1) &= \frac{1}{N}\left\|\boldsymbol{y} - \sum_{k=1}^{K} \boldsymbol{a}_k s_k(i+1)\right\|^2. \end{aligned} \quad (18)$$

The initial estimates for $\{p_k\}_{k=1}^K$ are obtained as for IAA, whereas $\sigma(0)$ is typically chosen as a small positive number (e.g., $\sigma(0) = 10^{-5}$). Even though derived in the cited papers in a different way, SLIM is similar to the regularized FOCUSS (**foc**al **u**nderdetermined **s**ystem **s**olver) algorithm introduced in [16]. The main difference between these two methods consists in the way they estimate σ: SLIM computes the estimate of σ iteratively as in (18), while FOCUSS uses a fixed estimate of σ in all iterations that is obtained by one of several possible heuristical methods (see [16]).

Both IAA and SLIM are known to converge *locally* to the minimum value of their corresponding criterion (see the proof of local convergence for IAA in [17] and for SLIM in [14]). However little is known about the global convergence of these algorithms, or in effect about the convergence of their associated sequences.

The main difference between these two algorithms is that IAA is a nonparametric method (which provides a dense power estimate), whereas SLIM is a semiparametric method (whose result is a sparse power estimate, due to the use of sparsity-inducing parameter priors that lead to an implicit norm constraint similar to the one in (8), see [14] for details). In particular, the semiparametric character of SLIM makes its extension to the multisnapshot case a bit more difficult than that of IAA for which the extension is more or less straightforward (this difference is due to the fact that for a sparse method, unlike for a dense one, the estimates of $\{s_k\}_{k=1}^K$ for different snapshots should maintain the same sparsity pattern versus the snapshot index); we refer to [17] for details on these extensions.

III. The Proposed Method: SPICE

The following is a weighted covariance fitting criterion that can be used for the purpose of parameter estimation (see, e.g., [18], [19], and the references therein; also see [20] and [21])

$$f = ||\boldsymbol{R}^{-1/2}(\boldsymbol{y}\boldsymbol{y}^* - \boldsymbol{R})||^2 \tag{19}$$

where $||\cdot||$ denotes the Frobenius norm for matrices, and $\boldsymbol{R}^{-1/2}$ is the Hermitian positive definite square root of $\boldsymbol{R}^{-1}$. Admittedly, the use of (19) makes more sense in the multisnapshot case than in the single snapshot one [in which the sample covariance matrix is just $\boldsymbol{y}\boldsymbol{y}^*$, as used in (19)], but the minimization of f can yield satisfactory estimates even in the latter case (as we explain later on in this section). A simple calculation shows that

$$f = -2||\boldsymbol{y}||^2 + ||\boldsymbol{y}||^2\boldsymbol{y}^*\boldsymbol{R}^{-1}\boldsymbol{y} + \mathrm{tr}(\boldsymbol{R}) \tag{20}$$

where [see (10)–(12)]

$$\mathrm{tr}(\boldsymbol{R}) = E(||\boldsymbol{y}||^2) = \sum_{k=1}^{K+N} ||\boldsymbol{a}_k||^2 p_k \tag{21}$$

(and where E is the expectation operator). It follows from (20) and (21) that the minimization of f is equivalent to the minimization of the function:

$$g = \boldsymbol{y}^*\boldsymbol{R}^{-1}\boldsymbol{y} + \sum_{k=1}^{K+N} w_k p_k; \quad w_k = \frac{||\boldsymbol{a}_k||^2}{||\boldsymbol{y}||^2}. \tag{22}$$

As shown next, this is a convex problem.

A. Convexity of the Problem

The following equivalences can be readily verified:

$$\begin{aligned}\min_{\{p_k\}} g &\Leftrightarrow \min_{x,\{p_k\geq 0\}} x + \sum_{k=1}^{K+N} w_k p_k \quad \text{s.t. } x \geq \boldsymbol{y}^*\boldsymbol{R}^{-1}\boldsymbol{y} \\ &\Leftrightarrow \min_{x,\{p_k\geq 0\}} x + \sum_{k=1}^{K+N} w_k p_k \quad \text{s.t.} \begin{bmatrix} x & \boldsymbol{y}^* \\ \boldsymbol{y} & \boldsymbol{R} \end{bmatrix} \geq 0. \end{aligned} \tag{23}$$

The minimization problem in (23) [where $\boldsymbol{R} = \boldsymbol{A}^*\boldsymbol{P}\boldsymbol{A}$, see (10)] is a semidefinite program (SDP) [22] which is well known to be convex. The convexity of the original problem therefore follows.

There are several well-documented software packages for solving an SDP such as (23) (see, e.g., [23]). However SDP solvers are in general rather computationally intensive (as an example, using such a solver for (23) with $N = 50$ and $K = 1000$ takes about 1 hour on a relatively powerful PC). Consequently, we do not recommend solving the SDP in (23) as the preferred method for estimating the powers $\{p_k\}$, and suggest a different line of attack in the next subsection.

B. Derivation of SPICE

The literature on optimal experiment design contains a host of results on the minimization of functions of the form of our $\boldsymbol{y}^*\boldsymbol{R}^{-1}\boldsymbol{y}$, under the constraints that $p_k \geq 0$ and $\sum_{k=1}^{K+N} w_k p_k = 1$; see, e.g., [24], [25], and the many references therein. In order to make use of these results, particularly those of [24] (which inspired the derivation of SPICE below), we consider a reformulation of the problem introduced above. Specifically, it follows from (21) that $||\boldsymbol{y}||^2$ is an unbiased and consistent (in N) estimate of $\sum_{k=1}^{K+N} ||\boldsymbol{a}_k||^2 p_k$. Based on this observation we will estimate the powers by solving the following linearly constrained minimization problem, instead of minimizing the g in (22):

$$\min_{\{p_k\geq 0\}} \boldsymbol{y}^*\boldsymbol{R}^{-1}\boldsymbol{y} \quad \text{s.t.} \sum_{k=1}^{K+N} w_k p_k = 1. \tag{24}$$

It can be shown ([20]) that the problems (22) and (24) are equivalent in the sense that the solution to (22) is a scaled version of the solution of (24). Note that (24) is also a convex problem, see the previous subsection. Additionally note that the constraint in (24) is of the (weighted) ℓ_1-norm type, and therefore it is expected to be sparsity inducing for the solution to (24) (see, e.g., [12], [13], and [16]).

Let $\boldsymbol{Q} \in \mathbb{C}^{(K+N)\times N}$ be such that $\boldsymbol{Q}^*\boldsymbol{A} = \boldsymbol{I}$, and let (assuming $\{p_k > 0\}$ to simplify the explanations; to accommodate the case of $p_k = 0$, for some values of k, a pseudoinverse has to be used instead of the matrix inverse below, see [24])

$$h = \boldsymbol{y}^*\boldsymbol{Q}^*\boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{y}. \tag{25}$$

The minimization of h with respect to $\boldsymbol{Q}$, under the constraint $\boldsymbol{Q}^*\boldsymbol{A} = \boldsymbol{I}$, yields (see Appendix A for a proof)

$$\boldsymbol{Q}_0 = \boldsymbol{P}\boldsymbol{A}\boldsymbol{R}^{-1} \tag{26}$$

and

$$\begin{aligned} h_0 \triangleq h|_{\boldsymbol{Q}=\boldsymbol{Q}_0} &= \boldsymbol{y}^*\boldsymbol{R}^{-1}\boldsymbol{y} \\ &= \text{the original objective in (24)}. \end{aligned} \tag{27}$$

The important consequence of this result is that the minimization of h with respect to $\{p_k\}$ and $\boldsymbol{Q}(\text{s.t. } \boldsymbol{Q}^*\boldsymbol{A} = \boldsymbol{I})$ results in the same $\{p_k\}$ as the minimization of (24). The usefulness of this observation lies in the fact that the minimization of the augmented function h can be conveniently done by means of a cyclic (aka alternating) algorithm that consists of iterating the following steps until convergence:

Step 0. Compute initial estimates of $\{p_k\}$, e.g., by using (14).

Step 1. With $\{p_k\}$ fixed at their most recent estimates, minimize h with respect to $\boldsymbol{Q}$, s.t. $\boldsymbol{Q}^*\boldsymbol{A} = \boldsymbol{I}$. The minimizing matrix $\boldsymbol{Q}$ is given by (26).

Step 2. With $\boldsymbol{Q}$ fixed at its most recent value, minimize h with respect to $\{p_k \geq 0\}$ s.t. $\sum_{k=1}^{K+N} w_k p_k = 1$. The solution of this step can also be obtained in closed form, as detailed below.

Let

$$\boldsymbol{Q}(i) = \boldsymbol{P}(i)\boldsymbol{A}\boldsymbol{R}^{-1}(i) \tag{28}$$

be the $\boldsymbol{Q}$ matrix in (26) at the ith iteration of the algorithm, and let

$$\boldsymbol{\beta}(i) = \boldsymbol{Q}(i)\boldsymbol{y} = \boldsymbol{P}(i)\boldsymbol{A}\boldsymbol{R}^{-1}(i)\boldsymbol{y}. \tag{29}$$

TABLE I
THE SPICE ALGORITHM

Step	Computation/operation	Eq. no.
0) Initialization	$p_k = \|\boldsymbol{a}_k^*\boldsymbol{y}\|^2/\|\boldsymbol{a}_k\|^4 \ (k = 1, \ldots, K+N)$	(14)
1) Computation of $\boldsymbol{R}$	$\boldsymbol{R} = \sum_{k=1}^{K+N} p_k \boldsymbol{a}_k \boldsymbol{a}_k^*$	(10)
2) Power update	$\boldsymbol{z} = \boldsymbol{R}^{-1}\boldsymbol{y}$ $w_k^{1/2} = \|\boldsymbol{a}_k\|/\|\boldsymbol{y}\| \ (k = 1, \ldots, K+N)$ $r_k = \|\boldsymbol{a}_k^*\boldsymbol{z}\| \ (k = 1, \ldots, K+N)$ $\rho = \sum_{l=1}^{K+N} w_l^{1/2} p_l r_l$ $p_k \longleftarrow p_k r_k / w_k^{1/2} \rho \ (k = 1, \ldots, K+N)$	(34)
Iteration : iterate steps 1 and 2 until convergence.		

Then the minimization problem that needs to be solved in Step 2 of the above algorithm, at its ith iteration, is

$$\min_{\{p_k \geq 0\}} \sum_{k=1}^{K+N} \frac{|\beta_k(i)|^2}{p_k} \quad \text{s.t.} \quad \sum_{k=1}^{K+N} w_k p_k = 1 \tag{30}$$

where $\beta_k(i)$ is the kth element of the vector $\boldsymbol{\beta}(i)$ *viz.*

$$\beta_k(i) = p_k(i)\boldsymbol{a}_k^*\boldsymbol{R}^{-1}(i)\boldsymbol{y}. \tag{31}$$

By the Cauchy-Schwartz inequality we have that

$$\begin{aligned}\left[\sum_{k=1}^{K+N} w_k^{1/2}|\beta_k(i)|\right]^2 &= \left[\sum_{k=1}^{K+N} \frac{|\beta_k(i)|}{p_k^{1/2}} w_k^{1/2} p_k^{1/2}\right]^2 \\ &\leq \left[\sum_{k=1}^{K+N} \frac{|\beta_k(i)|^2}{p_k}\right]\left[\sum_{k=1}^{K+N} w_k p_k\right] \\ &= \left[\sum_{k=1}^{K+N} \frac{|\beta_k(i)|^2}{p_k}\right] \end{aligned} \tag{32}$$

It follows immediately from (32) that the solution to (30), and hence the solution to Step 2 of the cyclic algorithm, is given by

$$p_k(i+1) = \frac{|\beta_k(i)|}{w_k^{1/2}\sum_{l=1}^{K+N} w_l^{1/2}|\beta_l(i)|}. \tag{33}$$

Inserting (31) into (33) we get the following compact expression for the iterative *power updates of the SPICE algorithm* :

$$\begin{aligned} p_k(i+1) &= p_k(i)\frac{|\boldsymbol{a}_k^*\boldsymbol{R}^{-1}(i)\boldsymbol{y}|}{w_k^{1/2}\rho(i)} \\ \rho(i) &= \sum_{l=1}^{K+N} w_l^{1/2} p_l(i) |\boldsymbol{a}_l^*\boldsymbol{R}^{-1}(i)\boldsymbol{y}|. \end{aligned} \tag{34}$$

Interestingly, the above equation has the same multiplicative form as the power update formula for SLIM, see (18), whereas the updating equation for IAA has a different form. Furthermore, the updating formulas for all three methods depend on $|\boldsymbol{a}_k^*\boldsymbol{R}^{-1}(i)\boldsymbol{y}|$, although the power to which this quantity is raised depends on the method (1st power for SPICE and 2nd power for IAA and SLIM).

With regard to implementation, SPICE and SLIM can be implemented quite efficiently by first computing $\boldsymbol{z}(i) = \boldsymbol{R}^{-1}(i)\boldsymbol{y}$ (possibly by means of a conjugate-gradient algorithm, see, e.g., [26]) and then evaluating the scalar products $\boldsymbol{a}_k^*\boldsymbol{z}(i)$ (for $k = 1, \ldots, K+N$). The implementation of IAA, on the other hand, requires comparatively more computations due to the need for evaluating the denominator in (13) for $k = 1, \ldots, K+N$ (the reader is reminded that $K \gg N$ usually). For easy reference, the SPICE algorithm is summarized in Table I.

C. Some Theoretical Properties

Because SPICE monotonically decreases the objective function (due to its cyclic operation) and since the minimization problem it solves is convex, we can expect that the algorithm has global convergence properties. In effect, it can be shown that under reasonably weak conditions (essentially requiring that $\{p_k(0) > 0\}$ and that the matrix $\boldsymbol{R}(i)$ remains positive definite as i increases), *the limit points of SPICE power sequence*, (34), *are global minimizers of* $\boldsymbol{y}^*\boldsymbol{R}^{-1}\boldsymbol{y}$ subject to the constraints in (24) [24]. In other words, the SPICE algorithm is globally convergent for any initial value that belongs to the interior of the set $\{p_k \geq 0\}$, which is usually the case [e.g., for (14)]. Despite the said initialization, the limit points of the algorithm tend often to be on the boundary of the above set, a fact in agreement with the previously made observation that SPICE is a sparse estimation method (see the comments following (24))[1].

The theoretical analysis of the global minimum points of $\boldsymbol{y}^*\boldsymbol{R}^{-1}\boldsymbol{y}$ appears to be difficult. Empirical experience with the SPICE algorithm suggests that typically the locations of the dominant peaks of the true power spectrum are well determined, but also that the corresponding power values require some form

[1]We believe that extensions of Elfving's theorem as well of Caratheodory representation result (see, e.g., [27]) to the present complex-valued data model, which is a topic left for future research, will shed more light on the sparse character of SPICE power estimates.

of correction. To understand this type of behavior, we consider the noise free case in which [see (1) with $\boldsymbol{\epsilon} = \mathbf{0}$]

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{b} \tag{35}$$

for some vector $\boldsymbol{b} \in \mathbb{C}^{C\times 1}$ and a matrix $\boldsymbol{B} \in \mathbb{C}^{N\times C}$ that is a block of $\boldsymbol{A}^*$. Furthermore, let $\sigma_1 = \cdots = \sigma_N \triangleq \sigma$ (the true value of which is $\sigma = 0$ in the present case), let all the other powers in $\boldsymbol{P}$ be zero except for those corresponding to the columns of $\boldsymbol{B}$, and let

$$\boldsymbol{D}^2 = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & d_C \end{bmatrix} \tag{36}$$

be the diagonal matrix corresponding to the nonzero powers. Then, by the matrix inversion lemma

$$\begin{aligned} \boldsymbol{R}^{-1} &= [(\boldsymbol{B}\boldsymbol{D})(\boldsymbol{D}\boldsymbol{B}^*) + \sigma\boldsymbol{I}]^{-1} \\ &= \frac{1}{\sigma}\boldsymbol{I} - \frac{1}{\sigma}\boldsymbol{B}\boldsymbol{D}(\sigma\boldsymbol{I} + \boldsymbol{D}\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{D})^{-1}\boldsymbol{D}\boldsymbol{B}^* \end{aligned} \tag{37}$$

which implies that

$$\begin{aligned} \boldsymbol{y}^*\boldsymbol{R}^{-1}\boldsymbol{y} &= \frac{1}{\sigma}\boldsymbol{b}^*[\boldsymbol{B}^*\boldsymbol{B} - \boldsymbol{B}^*\boldsymbol{B}\boldsymbol{D}(\sigma\boldsymbol{I} \\ &\quad + \boldsymbol{D}\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{D})^{-1}\boldsymbol{D}\boldsymbol{B}^*\boldsymbol{B}]\boldsymbol{b} \\ &= \frac{1}{\sigma}\boldsymbol{b}^*\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{D}(\sigma\boldsymbol{I} + \boldsymbol{D}\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{D})^{-1} \\ &\quad \times [(\sigma\boldsymbol{I} + \boldsymbol{D}\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{D})\boldsymbol{D}^{-1} - \boldsymbol{D}\boldsymbol{B}^*\boldsymbol{B}]\boldsymbol{b} \\ &= \boldsymbol{b}^*\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{D}(\sigma\boldsymbol{I} + \boldsymbol{D}\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{D})^{-1}\boldsymbol{D}^{-1}\boldsymbol{b} \\ &= \boldsymbol{b}^*[\boldsymbol{D}(\sigma\boldsymbol{I} + \boldsymbol{D}\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{D})(\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{D})^{-1}]^{-1}\boldsymbol{b} \\ &= \boldsymbol{b}^*[\sigma(\boldsymbol{B}^*\boldsymbol{B})^{-1} + \boldsymbol{D}^2]\boldsymbol{b}. \end{aligned} \tag{38}$$

In many cases of interest $(\boldsymbol{B}^*\boldsymbol{B})^{-1} \to \mathbf{0}$ as $N \to \infty$; for example this happens in the spectral analysis application (see Section IV), in which $\|\boldsymbol{a}_k\|^2 = N$—this is assumed in the following. So let us assume for simplicity that the term $\sigma(\boldsymbol{B}^*\boldsymbol{B})^{-1}$ in (38) can be neglected. Then the minimization problem in (24) becomes, approximately

$$\min_{\{d_k\},\sigma} \sum_{k=1}^{C} \frac{|b_k|^2}{d_k} \quad \text{s.t.} \sum_{k=1}^{C} d_k + \sigma = \frac{\|\boldsymbol{y}\|^2}{N} \tag{39}$$

the solution to which is (by a calculation similar to that in (32)):

$$\sigma = 0^+ \quad (\sigma \text{ must satisfy } \sigma > 0 \text{ for } \boldsymbol{R}^{-1} \text{ to exist}) \tag{40}$$

$$d_k = |b_k|\rho \tag{41}$$

where

$$\rho = \frac{\|\boldsymbol{y}\|^2}{N\left(\sum_{k=1}^{C}|b_k|\right)} \tag{42}$$

is a normalizing constant. It follows from (41) that the powers (at the true locations) which minimize the SPICE criterion are proportional to the square root of the true powers

$$d_k = \rho(|b_k|^2)^{1/2}. \tag{43}$$

Note, however, that for $C = 1$ we have $\|\boldsymbol{y}\|^2 = N|b_1|^2$ and hence the power is correctly determined in this case: $d_1 = |b_1|^2$.

Another way to realize that (43) is true is based on (29) and (33). For given $\boldsymbol{P}$ (and hence $\boldsymbol{R}$), the $\{\beta_k\}$ in (29) (we omit the index i to simplify the notation) are readily recognized to be estimates of the signal amplitudes as well as of the noise term in (5); now the $\{p_k\}$ in (33) are proportional to $\{|\beta_k|\}$: therefore $\{p_k\}$ obtained from SPICE must be (scaled) estimates of the square root of the powers, as $\{|\beta_k|\}$ are. It also follows from this discussion that, whenever accurate determination of the heights of the power peaks (not only of their locations) is required, we can proceed in the following way. First we use the $\{|\beta_k|^2\}$ obtained from SPICE (at convergence) to reestimate the powers

$$\hat{p}_k = |\beta_k|^2 \quad k = 1, \ldots, K+N. \tag{44}$$

Note that the $\{\beta_k\}$ obtained from (29) would be the best linear unbiased estimates if $\boldsymbol{P}$ used in (29) were the true power matrix (or a scaled version thereof). However, as explained above, the powers produced directly by SPICE do not satisfy this condition, which means that the accuracy of (44) might not be satisfactory. To improve the estimation performance of (44) we use it three times to reestimate the powers : each time we use the most recent power estimates to build $\boldsymbol{P}$, compute $\{\beta_k\}$ with (29), and then obtain enhanced power estimates via (44).

Remark 1: Once the locations of the dominant power peaks are estimated with SPICE, their heights can be determined by means of several other methods (see, e.g., [28]), besides the one outlined above. Of these possible methods, we have tested the one based on multiple-frequency least squares (MFLS) and found out that its performance in the spectral analysis application presented in Section IV was quite satisfactory. However, we have chosen to omit the details on MFLS and to focus on the method based on (29) and (44), because the latter method is inherently intertwined with SPICE, unlike MFLS. ■

Before concluding this discussion, we remark on the fact that in the previous analysis we let $\boldsymbol{R} = \boldsymbol{B}\boldsymbol{D}^2\boldsymbol{B} + \sigma\boldsymbol{I}$ where $\boldsymbol{D}$ and σ were unknown, but $\boldsymbol{B}$ was the matrix made from the true signal vectors $\{\boldsymbol{a}(\tilde{\omega}_c)\}$ present in $\boldsymbol{y}$ [see (1)]. An interesting question is what happens if we replace $\boldsymbol{B}$ in $\boldsymbol{R}$ by a matrix that is also unknown, let us say $\boldsymbol{X} \in \mathbb{C}^{N\times C}$: will the minimization of the SPICE criterion yield the true solution $\boldsymbol{X} = \boldsymbol{B}$, or at least a close approximation of it? As shown in Appendix B the answer to this question is positive, therefore lending support to the previously asserted fact that the locations of the true power spectrum peaks are well determined by SPICE.

D. Some Extensions

In some applications the powers $\{p_k\}$ satisfy certain known linear relationships. For example, we might know that the *noise variance in the different measurements is constant*, i.e.

$$\sigma_1 = \cdots = \sigma_N \triangleq \sigma. \tag{45}$$

The extension of SPICE to include the above information is immediate. We only need to observe from (28) and (29) that, under

(45), the term that multiplies σ in (30) is

$$\sum_{k=K+1}^{K+N} |\beta_k(i)|^2 = \sigma^2(i)\|\boldsymbol{R}^{-1}(i)\boldsymbol{y}\|^2. \tag{46}$$

Consequently the updating formulas for SPICE become

$$p_k(i+1) = p_k(i)\frac{|\boldsymbol{a}_k^*\boldsymbol{R}^{-1}(i)\boldsymbol{y}|}{w_k^{1/2}\rho(i)} \tag{47}$$

$$\sigma(i+1) = \sigma(i)\frac{\|\boldsymbol{R}^{-1}(i)\boldsymbol{y}\|}{\nu^{1/2}\rho(i)} \tag{48}$$

where

$$\nu = \sum_{k=K+1}^{K+N} w_k \tag{49}$$

and

$$\rho(i) = \sum_{l=1}^{K} w_l^{1/2} p_l(i)|\boldsymbol{a}_l^*\boldsymbol{R}^{-1}(i)\boldsymbol{y}| + \nu^{1/2}\sigma(i)\|\boldsymbol{R}^{-1}(i)\boldsymbol{y}\|. \tag{50}$$

Similarly to the discussion in the previous subsection, whenever accurate estimation of the powers is necessary, we can use (29) and (44) to obtain refined estimates of $\{p_k\}_{k=1}^K$ and of σ

$$\hat{p}_k = |\beta_k|^2 \quad k = 1, \ldots, K \quad \hat{\sigma} = \frac{1}{N}\sum_{k=K+1}^{K+N} |\beta_k|^2. \tag{51}$$

As outlined before we repeat three times the calculation in (51), each time using the latest estimates of $\{p_k\}_{k=1}^K$ and σ to determine $\{\beta_k\}_{k=1}^{K+N}$ via (29). This is the SPICE power estimation formula that will be used in the numerical study in Section IV.

In other applications the power spectrum is known to possess certain symmetries. A prime example of this situation is the *processing of real-valued data*, in which case constraints of the type

$$p_k = p_{\bar{k}} \tag{52}$$

for certain values of $k \neq \bar{k}$ are known to hold. Modifying SPICE to take (52) into account can be done as indicated above for the similar constraint in (45), and therefore we omit the details of this extension.

In the rest of this subsection we explain how to extend SPICE to the case of *multiple snapshots*. Let $\boldsymbol{Y} \in \mathbb{C}^{N\times M}$ denote the matrix whose columns are the M available data snapshots, and let

$$\hat{\boldsymbol{R}} = \frac{1}{M}\boldsymbol{Y}\boldsymbol{Y}^* \tag{53}$$

The weighted covariance fitting criterion associated with (53) is given by [compare to (19)]

$$\begin{aligned} f &= \|\boldsymbol{R}^{-1/2}(\hat{\boldsymbol{R}} - \boldsymbol{R})\|^2 \\ &= \mathrm{tr}(\hat{\boldsymbol{R}}\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}) + \mathrm{tr}(\boldsymbol{R}) - 2\mathrm{tr}(\hat{\boldsymbol{R}}) \end{aligned} \tag{54}$$

where

$$\mathrm{tr}(\boldsymbol{R}) = E(\mathrm{tr}(\hat{\boldsymbol{R}})) = \sum_{k=1}^{K+N} \|\boldsymbol{a}_k\|^2 p_k. \tag{55}$$

An extended version of the SPICE estimation criterion, see (24), follows easily from the above equation:

$$\min_{\{p_k\geq 0\}} \mathrm{tr}(\hat{\boldsymbol{R}}\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}) \quad \text{s.t.} \sum_{k=1}^{K+N} w_k p_k = 1 \tag{56}$$

where

$$w_k = \frac{\|\boldsymbol{a}_k\|^2}{\mathrm{tr}(\hat{\boldsymbol{R}})}. \tag{57}$$

The derivation of the SPICE algorithm for the extended problem in (56) parallels that in Section III-B for the case of $M = 1$. Consider the function

$$h = \mathrm{tr}(\hat{\boldsymbol{R}}\boldsymbol{Q}^*\boldsymbol{P}^{-1}\boldsymbol{Q}\hat{\boldsymbol{R}}). \tag{58}$$

For fixed $\boldsymbol{P}$, the matrix $\boldsymbol{Q}$ that minimizes (58) (s.t. $\boldsymbol{Q}^*\boldsymbol{A} = \boldsymbol{I}$) is still given by (26)

$$\boldsymbol{Q}_0 = \boldsymbol{P}\boldsymbol{A}\boldsymbol{R}^{-1} \tag{59}$$

and it is still true that

$$\begin{aligned} h|_{\boldsymbol{Q}=\boldsymbol{Q}_0} &= \mathrm{tr}(\hat{\boldsymbol{R}}\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}) \\ &= \text{the original objective in (56).} \end{aligned} \tag{60}$$

Next, observe that the function in (58) can be rewritten (for fixed $\boldsymbol{Q}$) as

$$h = \sum_{k=1}^{K+N} \frac{|\beta_k|^2}{p_k} \tag{61}$$

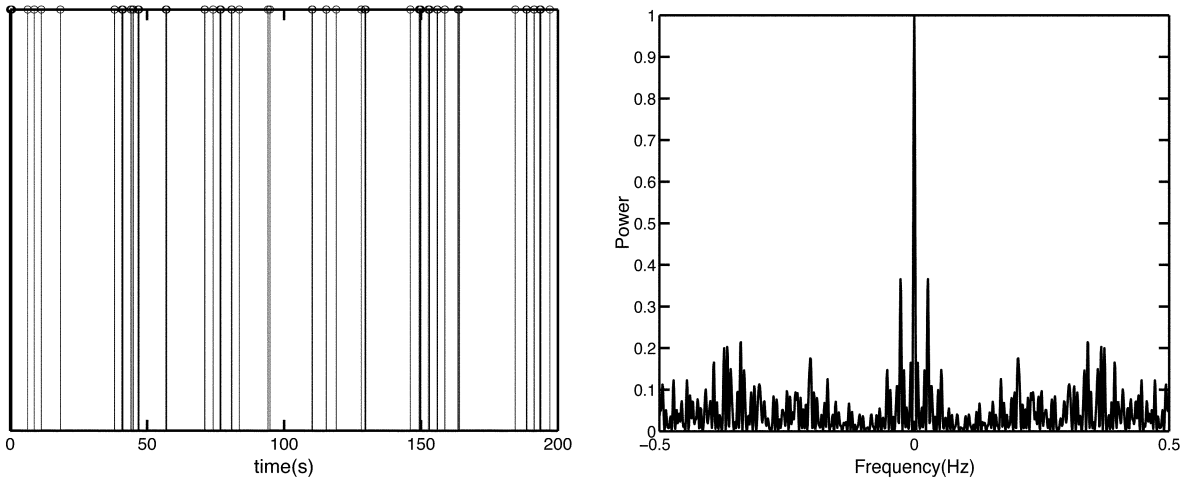
where now $\{|\beta_k|^2\}$ are the diagonal elements of the matrix $\boldsymbol{Q}\hat{\boldsymbol{R}}^2\boldsymbol{Q}^*$, i.e.

$$|\beta_k|^2 = [\boldsymbol{Q}\hat{\boldsymbol{R}}^2\boldsymbol{Q}^*]_{kk} = \frac{1}{M^2}[\boldsymbol{Q}\boldsymbol{Y}(\boldsymbol{Y}^*\boldsymbol{Y})\boldsymbol{Y}^*\boldsymbol{Q}^*]_{kk}. \tag{62}$$

Owing to the perfect analogy between (61) and the (30) corresponding to the case $M = 1$, we conclude that the extended algorithm has exactly the same form as the basic SPICE algorithm, the only modification being the different expressions for $\{w_k\}$ and for $\{|\beta_k|^2\}$ [see (57) and (62)]. Note also that whenever $M > N$ the $\{p_k\}$ obtained from SPICE are usually accurate estimates of the true powers and therefore, unlike in the case of $M = 1$, we can use them directly as power estimates—which is what we will do in the numerical study of the next section.

IV. Application to Spectral Analysis and Numerical Performance Study

Let $y_k = y(t_k), \quad k = 1, \ldots, N$, be the kth sample in a set of N measurements performed at possibly irregularly spaced



(a) (b)

Fig. 1. (a) The sampling pattern, for $N = 100$, mimicking a real-life case in astronomy [31]. (b) The corresponding spectral window.

times $\{t_k\}$. In spectral analysis applications the vectors $\{\boldsymbol{a}(\omega_k)\}$ correspond to sinusoidal components

$$\boldsymbol{a}(\omega_k) = \begin{bmatrix} e^{i\omega_k t_1} \\ \vdots \\ e^{i\omega_k t_N} \end{bmatrix} \tag{63}$$

[observe that in this case $\|\boldsymbol{a}(\omega_k)\|^2 = N$ is a constant; see (21), (22) for equations in which $\{\|\boldsymbol{a}(\omega_k)\|^2\}$ appear]. The interval for the angular frequency, *viz.* Ω, in which one can conduct spectral analysis without any aliasing problem, can be determined from the so-called spectral window, $|(1/N)\sum_{k=1}^{N} e^{i\omega t_k}|^2$, as the largest range for ω in which the only peak with height equal (or close) to 1 is at $\omega = 0$ [29], [30]. Let $\Omega = [-\omega_{\max}, \omega_{\max}]$ denote the so-obtained interval, or the subinterval of it that is of interest. We use a uniform grid $\{\omega_k\}_{k=1}^{K}$ to cover Ω, with a step equal to $(2\pi)/(5(t_N - t_1))$ (note that $(2\pi)/((t_N - t_1))$ is the best expected resolution in the class of nonparametric methods [1]; because we can hope for a better resolution in the sparse-estimation method class, we choose a five-times-smaller step for the frequency grid).

We will consider $N = 100$ and the sampling pattern shown in Fig. 1(a). This sampling pattern, which will be fixed in the simulation runs, mimics the type of patterns encountered in certain applications of spectral analysis in astronomy (see [31], [7]) where data collection depends on many factors and therefore is typically performed at rather irregular time intervals. For the sampling pattern in question we selected $\omega_{\max} = \pi$. The spectral window is shown in Fig. 1(b) from which one can see that the only peak with height equal (or close) to 1 in the frequency range of $[-\pi, \pi]$ is at $\omega = 0$, as required. The grid size K is chosen as $K = 5(t_N - t_1) = 1000$.

The data samples were simulated using (1) and (63) with $C = 3, \tilde{\omega}_1 = 2\pi \times 0.3100, \tilde{\omega}_2 = 2\pi \times 0.3150, \tilde{\omega}_3 = 2\pi \times 0.1450, \tilde{s}_1 = 10e^{i\varphi_1}, \tilde{s}_2 = 10e^{i\varphi_2}$, and $\tilde{s}_3 = 3e^{i\varphi_3}$, where the phases $\{\varphi_k\}_{k=1}^{3}$ were independently and uniformly distributed in $[0, 2\pi]$. The disturbance term, $\boldsymbol{\epsilon}$, was normal white noise with mean zero and variance σ; the noise variance will be varied to control the signal-to-noise ratio defined as

$$\mathrm{SNR} = 10\log\left(\frac{100}{\sigma}\right) = 20 - 10\log\sigma \quad [\mathrm{dB}].$$

The following methods will be used for spectrum/parameter estimation:

M_1: SFLS.
M_2: IAA.
M_3: SLIM.
M_4: $\mathrm{SPICE_{SS}} = \mathrm{SPICE} - \mathrm{single\ snapshot\ with}\ \sigma_1 = \cdots = \sigma_N$.
M_5: $\mathrm{SPICE_{MS}} = \mathrm{SPICE} - \mathrm{multi{-}snapshot\ with}\ \sigma_1 = \cdots = \sigma_N$.

We will not consider SPICE with different $\{\sigma_k\}$ as it is unlikely to provide better performance in the present case in which the noise elements have the same variance. However, note that in some applications (such as astronomy) the noises affecting different measurements can indeed have different variances, in which case SPICE with unconstrained $\{\sigma_k\}$ should be used. Regarding M_5, we note that replicated measurements are rarely available in spectral analysis applications. However in other applications, such as in array processing, they are a common occurrence and this is the reason for including M_5 in the present comparison. As the number of snapshots increases, the accuracy of M_5 improves significantly; we use $M = 1000$ in the following. We note in this context that IAA, and probably SLIM too, can also be extended to the multisnapshot case [17]. However, the said extensions are not as well motivated as the proposed extension of SPICE, and on that basis they will not be considered in this numerical study. The stopping criterion for the iterative methods, namely, IAA, SLIM and SPICE, is $(\|\boldsymbol{p}^{i+1} - \boldsymbol{p}^{i}\|)/(\|\boldsymbol{p}^{i}\|) < 10^{-4}$, where $\boldsymbol{p}^{i}$ denotes the estimate of the power vector at the ith iteration. The average number of iterations and the average computation time (on a 2.26 GHz, 4 GB PC) for these methods in the present example (with $\mathrm{SNR} = 10$

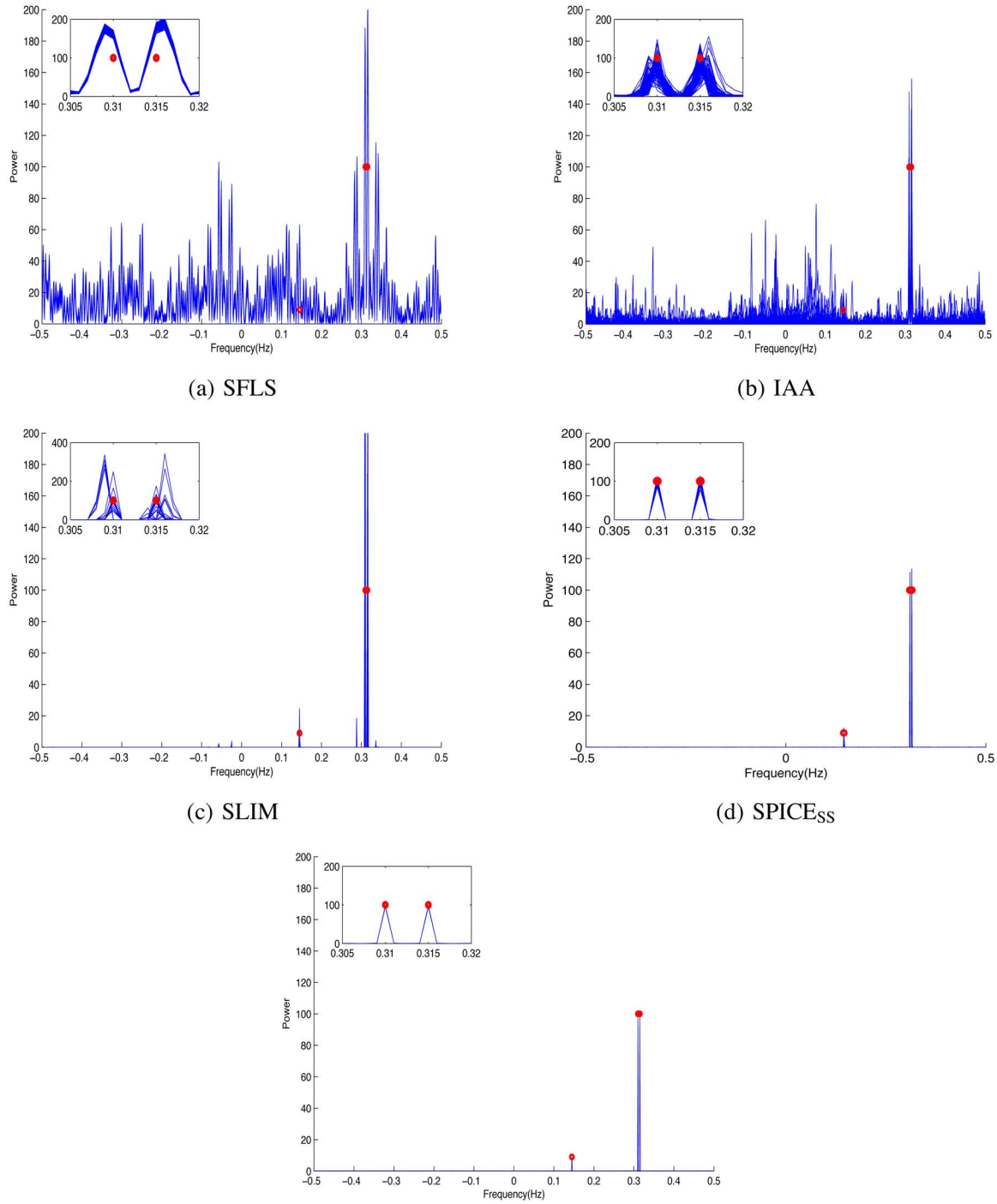


Fig. 2. Superimposed spectra obtained with the methods under consideration in 100 Monte Carlo runs, $\mathrm{SNR} = 10$ dB. The circles in the plots indicate the true value of the powers. (a) SFLS. (b) IAA. (c) SLIM. (d) $\mathrm{SPICE}_{\mathrm{SS}}$. (e) $\mathrm{SPICE}_{\mathrm{MS}}$.

dB) are as follows: IAA—940 iterations and 158 sec, SLIM—20 iterations and 3 sec and $\mathrm{SPICE}_{\mathrm{SS}}$—180 iterations and 20 sec.

In Fig. 2 we display, in a superimposed manner, the power spectrum estimates obtained with M_1-M_5 in 100 Monte Carlo (MC) runs, for $\mathrm{SNR} = 10$ dB. It can be inferred from the plots in this figure that SFLS suffers from heavy leakage problems; as a result the peak at $\tilde{\omega}_3$ is buried completely in "clutter," and the frequencies $\tilde{\omega}_1$ and $\tilde{\omega}_2$ are estimated with a bias that can be seen clearly from the insert. IAA, on the other hand, resolves the two closely spaced peaks without any bias but misses the weaker component at $\tilde{\omega}_3$ possibly due to ill-conditioning of the matrix $\boldsymbol{R}$ caused by irregular sampling (note that IAA, in general, works reasonably well for uniform sampling schemes). The semiparametric method SLIM yields a sparse spectrum but misses the true peaks in some MC runs, and also overestimates the powers in some realizations (note that some of the peaks at $\tilde{\omega}_1$ and $\tilde{\omega}_2$

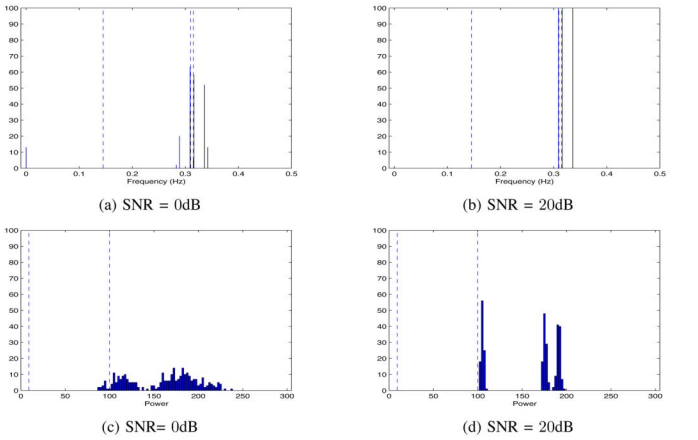


Fig. 3. Histograms of (a)–(b) frequency estimates and (c)–(d) power estimates for SFLS at two SNR values, obtained from 100 Monte Carlo runs. The dashed lines show the true frequencies and the true powers. (a) $\mathrm{SNR} = 0$ dB. (b) $\mathrm{SNR} = 20$ dB. (c) $\mathrm{SNR} = 0$ dB. (d) $\mathrm{SNR} = 20$ dB.

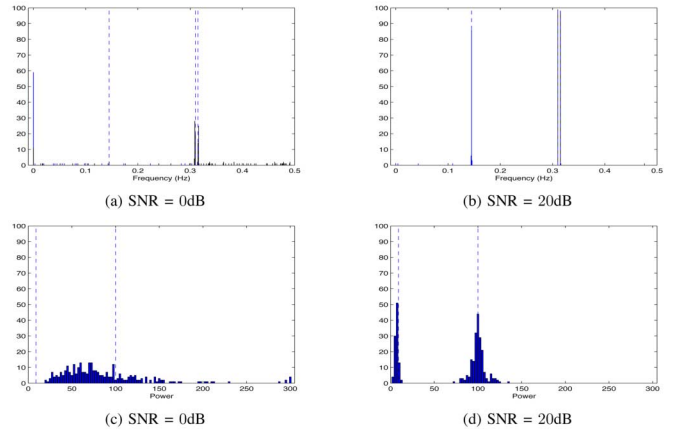


Fig. 4. Histograms of (a)–(b) frequency estimates and (c)–(d) power estimates for IAA at two SNR values, obtained from 100 Monte Carlo runs. The dashed lines show the true frequencies and the true powers. (a) $\mathrm{SNR} = 0$ dB. (b) $\mathrm{SNR} = 20$ dB. (c) $\mathrm{SNR} = 0$ dB. (d) $\mathrm{SNR} = 20$ dB.

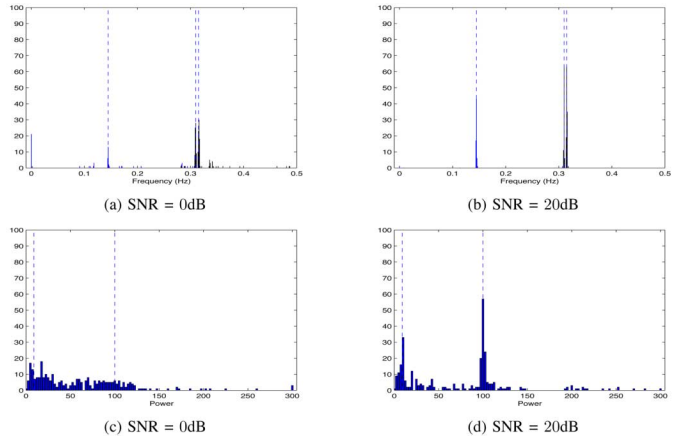


Fig. 5. Histograms of (a)–(b) frequency estimates and (c)–(d) power estimates for SLIM at two SNR values, obtained from 100 Monte Carlo runs. The dashed lines show the true frequencies and the true powers. (a) $\mathrm{SNR} = 0$ dB. (b) $\mathrm{SNR} = 20$ dB. (c) $\mathrm{SNR} = 0$ dB. (d) $\mathrm{SNR} = 20$ dB.

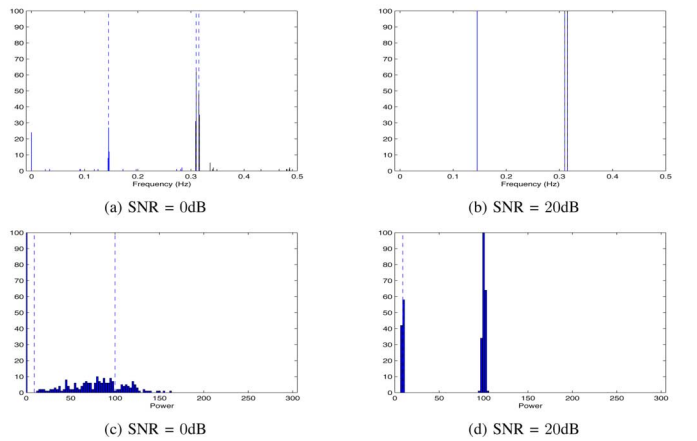


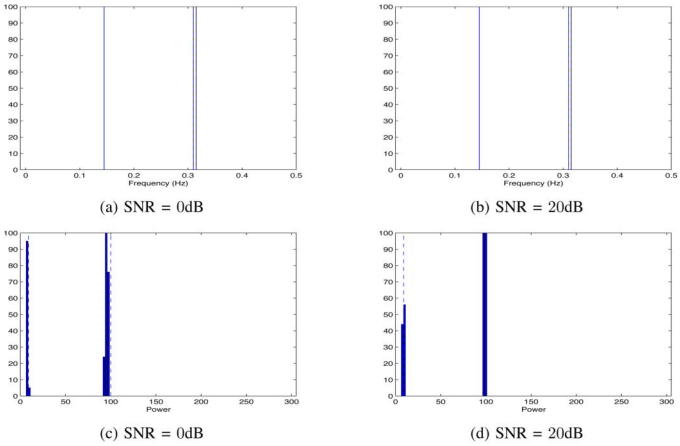
Fig. 6. Histograms of (a)–(b) frequency estimates and (c)–(d) power estimates for $\mathrm{SPICE}_{\mathrm{SS}}$ at two SNR values, obtained from 100 Monte Carlo runs. The dashed lines show the true frequencies and the true powers. (a) $\mathrm{SNR} = 0$ dB. (b) $\mathrm{SNR} = 20$ dB. (c) $\mathrm{SNR} = 0$ dB. (d) $\mathrm{SNR} = 20$ dB.

in Fig. 2(c) are larger than 200 (see the insert); however we cut them off at 200 in order to be able to use the same scale for all plots in Fig. 2). $\mathrm{SPICE}_{\mathrm{SS}}$, on the other hand, locates all three peaks in most MC runs and also gives more accurate power estimates. Finally $\mathrm{SPICE}_{\mathrm{MS}}$, yields a nearly ideal spectrum.

In Figs. 3–7 we show the histograms, obtained from 100 MC runs, of the frequency and power estimates obtained with M_1-M_5 for two different values of SNR. For all methods the frequency estimates are obtained as the locations of the three largest peaks of the corresponding estimated spectrum and the power estimates are computed at the estimated frequencies. In the histogram for power estimates, the values were saturated at 300 (i.e., larger values are not shown to focus on the power range of interest). Similarly, in the histogram for frequency estimates, the values were saturated at 0 (i.e., negative values are not shown). As expected, SFLS works poorly at low SNR and even at $\mathrm{SNR} = 20$ dB it fails to locate the component at $\tilde{\omega}_3$ (instead, in all 100 MC runs, it picked a spurious peak to the right of $\tilde{\omega}_2$, which is an artifact due to the sampling scheme). Note also that the SFLS power estimates at $\tilde{\omega}_1$ and $\tilde{\omega}_2$ are significantly overestimated. IAA gives poor frequency and power estimates at the low SNR, which is primarily due to the ill-conditioning of $\boldsymbol{R}$ caused by the irregular sampling scheme; however, it gives satisfactory frequency and power estimates at the SNR value of 20 dB. The accuracy of the frequency and power estimates obtained with SLIM is relatively poor at both SNR values considered; this is mainly due to the fact that SLIM gives a too sparse spectrum and hence fails to detect some of the components present in the data. The frequency estimation performance of $\mathrm{SPICE}_{\mathrm{SS}}$ is superior to that of SFLS, IAA, and SLIM, at both low and high SNR values. The $\mathrm{SPICE}_{\mathrm{SS}}$ power estimates are also reasonable, despite the fact that they are somewhat biased downwards at $\tilde{\omega}_1$ and $\tilde{\omega}_2$. Finally $\mathrm{SPICE}_{\mathrm{MS}}$ gives very accurate frequency estimates and precise power estimates for both SNR values.

V. Concluding Remarks

The SPICE (**s**emi**p**arametric/**sp**arse **i**terative **c**ovariance-based **e**stimation) method introduced in this paper enjoys global convergence properties, is user parameter free, can be easily used in the multisnapshot case, and has a small computational complexity. There are very few (if any) other sparse



(a) SNR = 0dB (b) SNR = 20dB

(c) SNR = 0dB (d) SNR = 20dB

Fig. 7. Histograms of (a)–(b) frequency estimates and (c)–(d) power estimates for SPICE_{MS} at two SNR values, obtained from 100 Monte Carlo runs. The dashed lines show the true frequencies and the true powers. (a) $\text{SNR} = 0$ dB. (b) $\text{SNR} = 20$ dB. (c) $\text{SNR} = 0$ dB. (d) $\text{SNR} = 20$ dB.

estimation methods that share all these useful features. In our opinion the capability of SPICE to operate in noisy scenarios without requiring prior information about the noise variance or the sparsity index of the parameter vector as well as its simple extension to multisnapshot situations are particularly useful characteristics. Regarding the statistical properties of the parameter estimates provided by SPICE, the proposed method was shown to outperform two competing methods in a numerical spectral analysis application.

Separable models appear in many branches of science and engineering, and we are planning to investigate the use of SPICE in several other applications, besides spectral analysis. Fully understanding the theoretical properties of SPICE also requires some additional research work. For example, as mentioned briefly in the footnote to the discussion in Section III-C, extensions of certain representation results for solutions to SPICE-type optimization problems are likely to shed further light on the computational and statistical properties of SPICE as well as on its relationship to the group of sparse estimation methods based on ℓ_1-norm minimization principles. We leave working out such theoretical extensions and providing a more detailed analysis of SPICE to a possible future publication.

Appendix A
Proof of (26) and (27)

To prove (26) we need to show that

$$\boldsymbol{Q}_0^*\boldsymbol{P}^{-1}\boldsymbol{Q}_0 = \boldsymbol{R}^{-1} \le \boldsymbol{Q}^*\boldsymbol{P}^{-1}\boldsymbol{Q} \tag{64}$$

for any $\boldsymbol{Q}$ that satisfies $\boldsymbol{Q}^*\boldsymbol{A} = \boldsymbol{I}$. However (64) is equivalent to

$$\begin{bmatrix} \boldsymbol{Q}^*\boldsymbol{P}^{-1}\boldsymbol{Q} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{A}^*\boldsymbol{P}\boldsymbol{A} \end{bmatrix} \ge 0 \Leftrightarrow$$
$$\begin{bmatrix} \boldsymbol{Q}^*\boldsymbol{P}^{-1}\boldsymbol{Q} & \boldsymbol{Q}^*\boldsymbol{A} \\ \boldsymbol{A}^*\boldsymbol{Q} & \boldsymbol{A}^*\boldsymbol{P}\boldsymbol{A} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q}^* & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{P}^{-1/2} \\ \boldsymbol{P}^{1/2} \end{bmatrix}$$
$$\times \begin{bmatrix} \boldsymbol{P}^{-1/2} & \boldsymbol{P}^{1/2} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A} \end{bmatrix} \ge 0 \tag{65}$$

and (65) evidently holds true. Therefore (26) is proved, and (27) follows by substitution.

Appendix B
On the Minimizers of the SPICE Criterion in the Noise-Free Case

Let $\boldsymbol{y}$ be given by (35), and let

$$\boldsymbol{R} = \boldsymbol{X}\boldsymbol{D}^2\boldsymbol{X}^* + \sigma\boldsymbol{I} \tag{66}$$

where $\boldsymbol{D}^2$ is as in (36), and $\boldsymbol{X} \in \mathbb{C}^{N\times C}$ is a matrix, made from columns of $\boldsymbol{A}^*$, that has full column rank. A straightforward calculation yields

$$\begin{aligned}\boldsymbol{R}^{-1} &= \frac{1}{\sigma}\boldsymbol{I} - \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{D}(\sigma\boldsymbol{I} + \boldsymbol{D}\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{D})^{-1}\boldsymbol{D}\boldsymbol{X}^* \\ &= \frac{1}{\sigma}[\boldsymbol{I} - \boldsymbol{X}\boldsymbol{D}(\boldsymbol{D}\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{D})^{-1}\boldsymbol{D}\boldsymbol{X}^*] \\ &\quad + \frac{1}{\sigma}[\boldsymbol{X}\boldsymbol{D}(\boldsymbol{D}\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{D})^{-1}\boldsymbol{D}\boldsymbol{X}^* \\ &\quad - \boldsymbol{X}\boldsymbol{D}(\sigma\boldsymbol{I} + \boldsymbol{D}\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{D})^{-1}\boldsymbol{D}\boldsymbol{X}^*] \\ &= \frac{1}{\sigma}[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^*\boldsymbol{X})^{-1}\boldsymbol{X}^*] \\ &\quad + \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{D}(\boldsymbol{D}\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{D})^{-1}[(\sigma\boldsymbol{I} + \boldsymbol{D}\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{D}) \\ &\quad - \boldsymbol{D}\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{D}](\sigma\boldsymbol{I} + \boldsymbol{D}\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{D})^{-1}\boldsymbol{D}\boldsymbol{X}^* \\ &= \frac{1}{\sigma}\boldsymbol{\Pi} + \boldsymbol{X}(\boldsymbol{X}^*\boldsymbol{X})^{-1}\boldsymbol{D}^{-2}(\sigma\boldsymbol{D}^{-2} + \boldsymbol{X}^*\boldsymbol{X})^{-1}\boldsymbol{X}^*\end{aligned} \tag{67}$$

where $\boldsymbol{\Pi}$ is the orthogonal projector matrix onto the null space of $\boldsymbol{X}^*$. Similarly to what we have done in Section III.C we can neglect the term $\sigma\boldsymbol{D}^{-2}$ in (68), as it is typically much smaller than $\boldsymbol{X}^*\boldsymbol{X}$. Therefore we have (approximately)

$$\begin{aligned}\boldsymbol{y}^*\boldsymbol{R}^{-1}\boldsymbol{y} &= \frac{1}{\sigma}\boldsymbol{b}^*\boldsymbol{B}^*\boldsymbol{\Pi}\boldsymbol{B}\boldsymbol{b} \\ &\quad + \boldsymbol{b}^*\boldsymbol{B}^*\boldsymbol{X}(\boldsymbol{X}^*\boldsymbol{X})^{-1}\boldsymbol{D}^{-2}(\boldsymbol{X}^*\boldsymbol{X})^{-1}\boldsymbol{X}^*\boldsymbol{B}\boldsymbol{b}.\end{aligned} \tag{68}$$

The minimization of (68) with respect to $\{d_k\}$, s.t. $\sum_{k=1}^{C} d_k = (\|\boldsymbol{y}\|^2/N) - \sigma$, can be done as in (30)–(33). The result of this minimization is the following function, which is to be minimized with respect to σ and $\boldsymbol{X}$

$$F = \frac{\alpha^2}{\sigma} + \frac{\beta^2}{(\gamma - \sigma)} \tag{69}$$

with $\alpha^2 = \boldsymbol{b}^*\boldsymbol{B}^*\boldsymbol{\Pi}\boldsymbol{B}\boldsymbol{b}$, $\gamma = (\|\boldsymbol{y}\|^2/N)$ and

$$\beta^2 = \left(\sum_{k=1}^{C} |\lambda_k|\right)^2 \tag{70}$$

where λ_k is the kth element of the vector $\boldsymbol{\lambda} = (\boldsymbol{X}^*\boldsymbol{X})^{-1}\boldsymbol{X}^*\boldsymbol{B}\boldsymbol{b}$.

Next, we consider the minimization of (69) with respect to $\sigma \in (0, \gamma)$, for fixed $\boldsymbol{X}$. The corresponding equation for the stationary points of F is

$$\begin{aligned}F' &= -\frac{\alpha^2}{\sigma^2} + \frac{\beta^2}{(\gamma - \sigma)^2} = 0 \Leftrightarrow \alpha(\gamma - \sigma) = \pm\beta\sigma \Leftrightarrow \\ \sigma &= \frac{\alpha\gamma}{\alpha \pm \beta}.\end{aligned} \tag{71}$$

Because $\sigma = (\alpha\gamma)/(\alpha - \beta)$ does not lie in $(0, \gamma)$, the only possible stationary point is

$$\sigma = \frac{\alpha\gamma}{\alpha + \beta}. \tag{72}$$

The second-order derivative of F

$$\frac{1}{2}F'' = \frac{\alpha^2}{\sigma^3} + \frac{\beta^2}{(\gamma - \sigma)^3} \tag{73}$$

is positive for any $\sigma \in (0, \gamma)$, and therefore (72) is a minimum point for F. The corresponding minimum value of F, as a function of $\boldsymbol{X}$, is given by

$$\tilde{F} = \frac{(\alpha + \beta)^2}{\gamma}. \tag{74}$$

The exact minimization of the above function with respect to $\boldsymbol{X}$ does not appear to lead to a simple closed-form solution. However, an approximate solution can be obtained as follows. Under quite general conditions $\beta^2 = \mathcal{O}(1)$ as N increases, whereas $\alpha^2 = \mathcal{O}(N)$ (unless $\boldsymbol{X} = \boldsymbol{B}$). It follows from this observation that, for a reasonably large value of N, the minimization of α in (74) is much more important than the minimization of β (note that γ in (74) is just a constant). Because $\alpha = 0$ for $\boldsymbol{X} = \boldsymbol{B}$, this means that the true locations of the power peaks are well determined via the minimization of the SPICE criterion with respect to $\boldsymbol{X}$, which was the fact to be shown.

REFERENCES

[1] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice-Hall, 2005.

[2] S. Kay and S. Marple, Jr., "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.

[3] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 37, no. 5, pp. 720–741, 1989.

[4] J. Guerci, *Space-Time Adaptive Processing for Radar*. Norwood, MA: Artech House, 2003.

[5] *MIMO Radar Signal Processing*, J. Li and P. Stoica, Eds. New York: Wiley , 2008.

[6] J. D. Scargle, "Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data," *Astrophys. J.*, vol. 263, pp. 835–853, Dec. 1982.

[7] P. Babu, P. Stoica, J. Li, Z. Chen, and J. Ge, "Analysis of radial velocity data by a novel adaptive approach," *Astronom. J.*, vol. 139, pp. 783–793, Feb. 2010.

[8] G. Golub and V. Pereyra, "Separable nonlinear least squares: The variable projection method and its applications," *Inverse Problems*, vol. 19, pp. 1–26, 2003.

[9] G. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numer. Anal.*, vol. 10, no. 2, pp. 413–432, 1973.

[10] 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 Trans. Aerosp. Electron. Syst.*, vol. 46, pp. 425–443, 2010.

[11] P. Stoica, J. Li, and H. He, "Spectral analysis of nonuniformly sampled data: A new approach versus the periodogram," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 843–858, 2009.

[12] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Statist. Soc. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[13] A. Maleki and D. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *IEEE J. Sel. Topics in Signal Process.*, vol. 4, no. 2, pp. 330–341, 2010.

[14] X. Tan, W. Roberts, J. Li, and P. Stoica, "MIMO radar imaging via SLIM," *IEEE Trans. Signal Process.* 2009 [Online]. Available: http://www.it.uu.se/katalog/praba420/SLIM.pdf, submitted for publication

[15] Z. Chen, J. Li, X. Tan, H. He, B. Guo, P. Stoica, and M. Datum, "On probing waveforms and adaptive receivers for active sonar," *IEEE J. Ocean. Eng.*, 2010, submitted for publication.

[16] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, 1997.

[17] W. Roberts, P. Stoica, J. Li, T. Yardibi, and F. Sadjadi, "Iterative adaptive approaches to MIMO radar imaging," *IEEE J. Sel. Topics in Signal Process.*, vol. 4, no. 1, pp. 5–20, 2010.

[18] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digit. Signal Process.*, vol. 8, no. 3, pp. 185–210, 1998.

[19] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1314–1323, 1999.

[20] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Transactions on Signal Processing* 2010 [Online]. Available: http://www.it.uu.se/katalog/praba420/SPICEarray.pdf, submitted for publication

[21] O. Besson and P. Stoica, "Decoupled estimation of DOA and angular spread for a spatially distributed source," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 1872–1882, 2000.

[22] Y. Nesterov and A. Nemirovskii, "Interior-point polynomial algorithms in convex programming," *Soc. Indust. Math.*, 1994.

[23] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimiz. Methods and Software*, vol. 11, no. 1, pp. 625–653, 1999.

[24] Y. Yu, "Monotonic convergence of a general algorithm for computing optimal designs," *Ann. Statist.*, vol. 38, no. 3, pp. 1593–1606, 2010.

[25] L. Pronzato, "Optimal experimental design and some related control problems," *Automatica*, vol. 44, no. 2, pp. 303–325, 2008.

[26] J. Daniel, "The conjugate gradient method for linear and nonlinear operator equations," *SIAM J. Numer. Anal.*, pp. 10–26, 1967.

[27] G. Elfving, "Optimum allocation in linear regression theory," *Ann. Math. Statist.*, pp. 255–262, 1952.

[28] P. Stoica, H. Li, and J. Li, "Amplitude estimation of sinusoidal signals: Survey, new results, and an application," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 338–352, 2000.

[29] L. Eyer and P. Bartholdi, "Variable stars: Which Nyquist frequency?," *Astron. Astrophys., Suppl. Ser.*, vol. 135, pp. 1–3, Feb. 1999.

[30] P. Babu and P. Stoica, "Spectral analysis of nonuniformly sampled data—A review," *Digit. Signal Process.*, vol. 20, no. 2, pp. 359–378, 2010.

[31] G. Marcy, R. Butler, D. Fischer, S. Vogt, J. Lissauer, and E. Rivera, "A pair of resonant planets orbiting GJ 876," *Astrophys. J.*, vol. 556, no. 1, pp. 296–301, 2001.

Petre Stoica (F'94) received the D.Sc. degree in automatic control from the Polytechnic Institute of Bucharest (BPI), Bucharest, Romania, in 1979 and an honorary doctorate degree in science from Uppsala University, Uppsala, Sweden, in 1993.

He is a professor of systems modeling with the Department of Information Technology, Uppsala University, Sweden. More details about him can be found at http://www.it.uu.se/katalog/ps.

Prabhu Babu received the M.Sc. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, in 2007.

He is pursuing the Ph.D. degree in electrical engineering with applications in signal processing at the Department of Information Technology, Uppsala University, Sweden.

Jian Li (F'05) received the M.Sc. and Ph.D. degrees in electrical engineering from The Ohio State University (OSU), Columbus, in 1987 and 1991, respectively.

From April to June 1991, she was an Adjunct Assistant Professor with the Department of Electrical Engineering, OSU. From July 1991 to June 1993, she was an Assistant Professor with the Department of Electrical Engineering, University of Kentucky, Lexington. Since August 1993, she has been with the Department of Electrical and Computer Engineering, University of Florida,

Gainesville, where she is currently a Professor. In fall 2007, she was on sabbatical leave at the Massachusetts Institute of Technology, Cambridge. Her current research interests include spectral estimation, statistical and array signal processing, and their applications.

Dr. Li is a Fellow of IET and a member of Sigma Xi and Phi Kappa Phi. She received the 1994 National Science Foundation Young Investigator Award and the 1996 Office of Naval Research Young Investigator Award. She was an Executive Committee Member of the 2002 International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, in May 2002. She was a member of the Editorial Board of Signal Processing from 2005 to 2007. She has been a member of the Editorial Board of *Digital Signal Processing—A Review Journal* since 2006. She was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2005 and an Associate Editor of the IEEE SIGNAL PROCESSING MAGAZINE from 2003 to 2005. She is presently a member of the Sensor Array and Multichannel Technical Committee, IEEE Signal Processing Society. She is a coauthor of papers that received the First and Second Place Best Student Paper Awards, respectively, at the 2005 and 2007 Annual Asilomar Conferences on Signals, Systems, and Computers, Pacific Grove, CA. She is also a coauthor of the paper that received the M. Barry Carlton Award for the best paper published in IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS in 2005.