SPICE: A Sparse Covariance-Based Estimation Method for Array Processing

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

***Abstract*—This paper presents a novel SParse Iterative Covariance-based Estimation approach, abbreviated as SPICE, to array processing. The proposed approach is obtained by the minimization of a covariance matrix fitting criterion and is particularly useful in many-snapshot cases but can be used even in single-snapshot situations. SPICE has several unique features not shared by other sparse estimation methods: it has a simple and sound statistical foundation, it takes account of the noise in the data in a natural manner, it does not require the user to make any difficult selection of hyperparameters, and yet it has global convergence properties.**

***Index Terms*—Array processing, covariance fitting, direction-of-arrival (DOA) estimation, sparse parameter estimation.**

I. Introduction and Preliminaries

CONSIDER an array processing scenario in which the main problem is to estimate the location parameters of a number of narrowband sources that are present in the array's viewing field. Let Ω denote the set of possible locations, and let θ be a generic location parameter. Also, let $\{\theta_k\}_{k=1}^K$ denote a grid that covers Ω. We assume that the grid is fine enough such that the true location parameters of the existing sources lie on (or, practically, close to) the grid. Under this reasonable assumption we can use the following *nonparametric* model for the output of the array (see, e.g., [1]):

$$\boldsymbol{y}(t) = \sum_{k=1}^{K} \boldsymbol{a}_k s_k(t) + \boldsymbol{\epsilon}(t) \quad t = 1, \ldots, M \quad (N \times 1) \tag{1}$$

where M is the total number of snapshots, N is the number of sensors in the array, $\boldsymbol{y}(t) \in \mathbb{C}^{N\times 1}$ is the tth observed snapshot, $\boldsymbol{a}_k \in \mathbb{C}^{N\times 1}$ denotes the array transfer vector (aka manifold or steering vector) corresponding to θ_k, $s_k(t) \in \mathbb{C}$ is the unknown signal impinging on the array from a possible source at θ_k, and $\boldsymbol{\epsilon}(t) \in \mathbb{C}^{N\times 1}$ is a noise term.

A *sparse* (or semi-parametric) estimation method makes the assumption, reminiscent of the parametric approach, that only a small number of sources exists and therefore that only a few rows of the signal matrix

$$\boldsymbol{S} = \begin{bmatrix} s_1(1) & \cdots & \cdots & s_1(M) \\ \vdots & \vdots & \vdots & \vdots \\ s_K(1) & \cdots & \cdots & s_K(M) \end{bmatrix} \tag{2}$$

are different from zero. The estimation problem is then to decide from the data $\{\boldsymbol{y}(t)\}$ which rows of the above matrix are nonzero. Indeed, once this is done, the solution to the location estimation problem, which is usually the main goal of array processing, is immediate : if the row $\tilde{k}$ (let us say) of (2) is deemed to be different from zero then we can infer that there is a corresponding source at an estimated location equal to $\theta_{\tilde{k}}$.

The previous formulation of the location problem begs for the use of basic ideas from the area of sparse parameter estimation, or rather simple extensions of those ideas to the present multi-snapshot case. To describe these ideas briefly, let

$$\boldsymbol{Y}^* = [\boldsymbol{y}(1), \ldots, \boldsymbol{y}(M)] \in \mathbb{C}^{N\times M} \tag{3}$$

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}_1^* \\ \vdots \\ \boldsymbol{s}_K^* \end{bmatrix} \in \mathbb{C}^{K\times M} \tag{4}$$

$$\boldsymbol{B}^* = [\boldsymbol{a}_1 \cdots \boldsymbol{a}_K] \in \mathbb{C}^{N\times K} \tag{5}$$

where the superscript $*$ denotes the conjugate transpose (we denote the matrix in (5) by $\boldsymbol{B}$ to reserve the notation $\boldsymbol{A}$ for an extended form of (5) that will be introduced in the next section). A direct application of the ℓ_1-norm minimization principle (see, e.g., [2]) to the present scenario described by (1)–(5) consists of estimating the matrix $\boldsymbol{S}$ as the solution to the following constrained minimization problem:

$$\min_{\boldsymbol{S}} \sum_{k=1}^{K} \|\boldsymbol{s}_k\| \quad \text{s.t. } \|\boldsymbol{Y}^* - \boldsymbol{B}^*\boldsymbol{S}\| \le \eta \tag{6}$$

where $\|\cdot\|$ denotes the Euclidean norm for vectors and the Frobenius norm for matrices, and η is a threshold that must be chosen by the user. Note that the objective in (6) is equal to the ℓ_1-norm of the vector $\{\|\boldsymbol{s}_k\|\}_{k=1}^K$, an observation that shows clearly that this approach is a direct extension of the standard single-snapshot approach of, e.g., [2].

A method for array processing based on (6) was pursued in [3]. Note that (6) is easily recognized to be an SOCP (second order cone program), see, e.g., [4], which can be efficiently solved provided that N, M, and K do not take on too large values. However, in the array processing application this is not necessarily the case : indeed, while $N \approx 10 - 10^2$ (tens to hundreds) is reasonably small, $M \approx 10^2 - 10^3$ and $K \approx 10^2 - 10^6$

Manuscript received June 11, 2010; revised August 09, 2010, October 22, 2010; accepted October 26, 2010. Date of publication November 01, 2010; date of current version January 12, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jean Pierre Delmas.

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

J. Li is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: li@dsp.ufl.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2090525

(depending on the desired resolution and the dimension of θ (1D, 2D, 3-D etc.)) can be rather large. For such large dimensional problems the currently available SOCP software (see, e.g., [5]) is too slow to use. In an attempt to overcome this computational problem, [3] suggested a way to reduce the number of columns of $\boldsymbol{Y}^*$ and $\boldsymbol{S}$ in (6) to manageable values by means of a singular value decomposition operation.

The approaches that rely on (6), such as the one in [3], suffer from a number of problems. First, the motivation of (6) is more clearly established in the noise-free case than in the more practical noisy data case. Second, and likely related to the first problem, there exist no clear-cut guidelines for the selection of η in (6) (see, e.g., [6] for a critical discussion on this aspect). Finally, solving (6) as an SOCP may be too time consuming for some applications in which N, M and especially K take on large values, and the available techniques for reducing the dimensions in (6) (such as that suggested in [3]) require the choice of further hyperparameters, besides η, and even so they are only approximate.

In this paper we present a new method of sparse parameter estimation in models of the type of (1). This method, which is called SPICE (SParse Iterative Covariance-based Estimation), is obtained using a novel covariance-based fitting approach that was recently introduced in [7] where the focus was on time-series data (the single-snapshot case) as opposed to the array data (the multi-snapshot scenario) considered here. SPICE has a number of useful features that are hardly shared by other sparse estimation methods. In particular, SPICE does not suffer from the problems described in the previous paragraph. Indeed, as will be shown in the next section, SPICE has a sound (covariance-based) statistical motivation which makes it possible to use the method in noisy data scenarios without the need for choosing any hyperparmeters. Additionally, the SPICE algorithm has a simple form, and yet it enjoys global convergence properties. We will show that the covariance fitting problem that SPICE solves can be reformulated as an SOCP of the form of (6) *but* with $\eta = 0$, with $\boldsymbol{Y}^*$ and $\boldsymbol{S}$ replaced by other matrices with only N columns (typically $M \gg N$), and with row-augmented matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ that, unlike $\boldsymbol{B}$ and $\boldsymbol{S}$ in (6), take account of the noise term in the data (1). This SOCP formulation of the SPICE estimation criterion shows that the ℓ_1-norm minimization problem in (6) can be given a simple statistical motivation based on covariance fitting, provided that the matrices in (6) are suitably defined, see the rest of this paper for details.

II. SPICE Estimation Criterion

Let us assume that

$$E\left[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}^*(\bar{t})\right] = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_N \end{bmatrix} \delta_{t,\bar{t}} \tag{7}$$

where E stands for the expectation operator, and

$$\delta_{t,\bar{t}} = \begin{cases} 1, & \text{if } t = \bar{t} \\ 0, & \text{elsewhere.} \end{cases} \tag{8}$$

This assumption on the noise term in (1) is reasonable in most applications. Let us also assume that the signals $\{s_k(t)\}$ and the noise $\boldsymbol{\epsilon}(\bar{t})$ are uncorrelated with each other for any $(t, \bar{t})$ and that

$$E[s_k(t)s_{\bar{k}}^*(\bar{t})] = p_k\delta_{k,\bar{k}}\delta_{t,\bar{t}}. \tag{9}$$

Then the data snapshots $\{\boldsymbol{y}(1), \ldots, \boldsymbol{y}(M)\}$ are uncorrelated with one another and have the following covariance matrix:

$$\begin{aligned} \boldsymbol{R} &= E\left[\boldsymbol{y}(t)\boldsymbol{y}^*(t)\right] \\ &= \sum_{k=1}^{K} p_k\boldsymbol{a}_k\boldsymbol{a}_k^* + \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_N \end{bmatrix} \\ &= [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K\ \boldsymbol{I}] \begin{bmatrix} p_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & p_2 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & p_K & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \sigma_N \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1^* \\ \vdots \\ \boldsymbol{a}_K^* \\ \boldsymbol{I} \end{bmatrix} \\ &\triangleq \boldsymbol{A}^*\boldsymbol{P}\boldsymbol{A} \end{aligned} \tag{10}$$

where

$$\begin{aligned} \boldsymbol{A}^* &\triangleq [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K\ \boldsymbol{I}] \\ &\triangleq [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K\ \boldsymbol{a}_{K+1}, \ldots, \boldsymbol{a}_{K+N}] \end{aligned} \tag{11}$$

$$\begin{aligned} \boldsymbol{P} &\triangleq \begin{bmatrix} p_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & p_2 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & p_K & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \sigma_N \end{bmatrix} \\ &\triangleq \begin{bmatrix} p_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & p_2 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & p_K & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & p_{K+1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & p_{K+N} \end{bmatrix}. \end{aligned} \tag{12}$$

The assumption in (9) that the source signals are spatially uncorrelated, which led to the above expression for the covariance matrix $\boldsymbol{R}$, does not always hold: in some applications the signals can be correlated or even coherent. However the SPICE method proposed in this paper is *robust* to this assumption—we refer to [7] (see also [1]) for a theoretical explanation of this robustness property, which will be illustrated numerically in Section V.

We will consider the following covariance fitting criterion for the purpose of parameter estimation (see, e.g., [8] and the references therein):

$$f = \left\| \boldsymbol{R}^{-1/2}(\hat{\boldsymbol{R}} - \boldsymbol{R})\hat{\boldsymbol{R}}^{-1/2} \right\|^2 \tag{13}$$

where $\boldsymbol{R}^{-1/2}$ denotes the positive definite square-root of $\boldsymbol{R}^{-1}$

$$\hat{\boldsymbol{R}} = \frac{\boldsymbol{Y}^*\boldsymbol{Y}}{M} \tag{14}$$

and the inverses of $\hat{\boldsymbol{R}}$ and $\boldsymbol{R}$ are assumed to exist. Note that $\boldsymbol{R}^{-1}$ exists under weak conditions: indeed, while the "true" values of many $\{p_k\}$ in $\boldsymbol{R}$ may be zero, one typically has $\sigma_k > 0$ $(k = 1, \ldots, N)$ which renders $\boldsymbol{R}$ nonsingular. Regarding $\hat{\boldsymbol{R}}$, the inverse of this matrix exists with probability one as long as $M > N$. However, for $M < N$ the sample covariance matrix $\hat{\boldsymbol{R}}$ is singular and therefore (13) cannot be used. In the latter case, one can estimate the parameters $\{p_k\}$ by minimizing the following criterion, instead of (13):

$$\|\boldsymbol{R}^{-1/2}(\hat{\boldsymbol{R}} - \boldsymbol{R})\|^2. \tag{15}$$

The above criterion is the one used in [7] where the focus was on the time series case with $M = 1$. In the present array processing scenario, however, we prefer to consider (13) because usually the condition $M > N$ is satisfied *and* because (13) has a statistically stronger motivation than (15): indeed, under certain conditions specified in [8], it can be shown that the parameter estimates minimizing (13) are statistically asymptotically (in M) efficient, whereas the estimates obtained from (15) are suboptimal. Nevertheless, we should note that what we do in the following for (13) applies with minor modifications to (15) as well, see Remark 1 below—this observation can be useful in those array processing applications in which $M < N$ and therefore in which (15) should be used in lieu of (13).

A simple calculation shows that

$$\begin{aligned} f &= \operatorname{tr}\left[\boldsymbol{R}^{-1}(\hat{\boldsymbol{R}} - \boldsymbol{R})\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{R}} - \boldsymbol{R})\right] \\ &= \operatorname{tr}\left[(\boldsymbol{R}^{-1}\hat{\boldsymbol{R}} - \boldsymbol{I})(\boldsymbol{I} - \hat{\boldsymbol{R}}^{-1}\boldsymbol{R})\right] \\ &= \operatorname{tr}(\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}) + \operatorname{tr}(\hat{\boldsymbol{R}}^{-1}\boldsymbol{R}) - 2N \end{aligned} \tag{16}$$

where

$$\operatorname{tr}\left(\hat{\boldsymbol{R}}^{-1}\boldsymbol{R}\right) = \sum_{k=1}^{K+N} p_k \boldsymbol{a}_k^* \hat{\boldsymbol{R}}^{-1}\boldsymbol{a}_k. \tag{17}$$

It follows from (16) and (17) that the minimization of f is equivalent to the minimization of the function

$$g = \operatorname{tr}\left(\hat{\boldsymbol{R}}^{1/2}\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}^{1/2}\right) + \sum_{k=1}^{K+N}\left(\boldsymbol{a}_k^*\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}_k\right)p_k. \tag{18}$$

Remark 1: A similar calculation shows that in the case of (15) the function to be minimized with respect to the unknown powers $\{p_k\}$ is

$$\operatorname{tr}\left(\hat{\boldsymbol{R}}\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}\right) + \sum_{k=1}^{K+N}\|\boldsymbol{a}_k\|^2 p_k. \tag{19}$$

Owing to the analogy between (18) and (19), it should come as no surprise that what we do in the following for (18) can also be done for (19) if necessary (e.g., for $M < N$). ■

The problem of minimizing g in (18) with respect to $\{p_k\}$ can be easily shown to be an SDP (semi-definite program), see Appendix A, and is therefore *convex* (see, e.g., [9]). Solving (18) as an SDP, however, is not recommended due to the fact that the available SDP solvers are too computationally intensive for the values of N, M, and K encountered in many array processing applications. Consequently we adopt a different line of attack that consists of replacing the problem of minimizing (18) by a related problem, as described next.

It follows from (17) that a consistent (in M) estimate of the right-hand side of this equation is given by N. Therefore, we can think of reformulating the problem of minimizing g as the following constrained minimization:

$$\min_{\{p_k \geq 0\}} \operatorname{tr}\left(\hat{\boldsymbol{R}}^{1/2}\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}^{1/2}\right) \quad \text{s.t.} \sum_{k=1}^{K+N} w_k p_k = 1 \tag{20}$$

where

$$w_k = \frac{\boldsymbol{a}_k^*\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}_k}{N}. \tag{21}$$

Interestingly, the problems (18) and (20) are not only asymptotically equivalent (as M increases, and under the condition that $\boldsymbol{R}$ in (10) can represent the true covariance matrix) but they are *exactly equivalent* (in general) in the sense that their solutions are scaled versions of each other (note that a scaling of $\{p_k\}$ has no effect on source location estimation). This equivalence property is proved in Appendix B where we also show that the problem obtained from (18) by constraining the first term to N (a constraint suggested, once again, by asymptotic considerations), instead of constraining the second term as in (20), is equivalent to (18) and (20) as well.

The problem (20) is also an SDP, and therefore convex. Furthermore, note that the linear constraint in (20) is of the (weighted) ℓ_1-norm type, and thus it can be expected to be sparsity inducing for the solution to (20). Apparently, the criterion in (20) was never considered in the previous literature on sparse parameter estimation. On the other hand, interestingly enough, this *type* of criterion occurs frequently in the seemingly unrelated literature on optimal experiment design (see, e.g., [10] and the references of that paper). The latter literature (in particular [10]) has served as a source of inspiration for [7] and, to some extent, for this paper as well.

III. SPICE Updating Formulas

Let $\boldsymbol{C} \in \mathbb{C}^{(K+N)\times N}$ and consider the following problem:

$$\min_{\boldsymbol{C}} \operatorname{tr}\left(\boldsymbol{C}^*\boldsymbol{P}^{-1}\boldsymbol{C}\right) \quad \text{s.t. } \boldsymbol{A}^*\boldsymbol{C} = \hat{\boldsymbol{R}}^{1/2}. \tag{22}$$

The solution to (22) is given (for fixed $\boldsymbol{P}$) by

$$\boldsymbol{C}_0 = \boldsymbol{P}\boldsymbol{A}\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}^{1/2} \tag{23}$$

and the corresponding minimum value of the function in (22) is

$$\begin{aligned} \operatorname{tr}\left(\boldsymbol{C}_0^*\boldsymbol{P}^{-1}\boldsymbol{C}_0\right) &= \operatorname{tr}\left(\hat{\boldsymbol{R}}^{1/2}\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}^{1/2}\right) \\ &= \text{the original objective in (20).} \end{aligned} \tag{24}$$

To prove the above assertion, observe that (23) follows if we can show that (hereafter the notation $\boldsymbol{X} \geq \boldsymbol{Y}$, with $\boldsymbol{X}$ and $\boldsymbol{Y}$ being

Hermitian matrices of appropriate dimensions, means that the difference matrix $\boldsymbol{X}-\boldsymbol{Y}$ is positive semi-definite):

$$\begin{aligned}\boldsymbol{C}^*\boldsymbol{P}^{-1}\boldsymbol{C} &\geq \boldsymbol{C}_0^*\boldsymbol{P}^{-1}\boldsymbol{C}_0 \\ &= \hat{\boldsymbol{R}}^{1/2}\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}^{1/2} \quad \text{s.t. } \boldsymbol{A}^*\boldsymbol{C} = \hat{\boldsymbol{R}}^{1/2}. \end{aligned} \tag{25}$$

By standard properties of partitioned matrices (see, e.g., [11, Lemma A.3]) and the fact that $\boldsymbol{R} > \boldsymbol{0}$, (25) holds if and only if the following partitioned matrix is positive semi-definite:

$$\begin{aligned}\begin{bmatrix} \boldsymbol{C}^*\boldsymbol{P}^{-1}\boldsymbol{C} & \hat{\boldsymbol{R}}^{1/2} \\ \hat{\boldsymbol{R}}^{1/2} & \boldsymbol{R} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{C}^*\boldsymbol{P}^{-1}\boldsymbol{C} & \boldsymbol{C}^*\boldsymbol{A} \\ \boldsymbol{A}^*\boldsymbol{C} & \boldsymbol{A}^*\boldsymbol{P}\boldsymbol{A} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{C}^* & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{P}^{-1} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{P} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A} \end{bmatrix} \geq 0. \end{aligned} \tag{26}$$

The central matrix in (26) can be rewritten as

$$\begin{bmatrix} \boldsymbol{P}^{-1} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{P} \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}^{-1/2} \\ \boldsymbol{P}^{1/2} \end{bmatrix} \begin{bmatrix} \boldsymbol{P}^{-1/2} & \boldsymbol{P}^{1/2} \end{bmatrix} \tag{27}$$

and thus it is obviously positive semi-definite (because it has the form $\boldsymbol{X}^*\boldsymbol{X}$, with $\boldsymbol{X} = [\boldsymbol{P}^{-1/2}\boldsymbol{P}^{1/2}]$). Therefore, (25) (and hence (23)) is proved and then (24) follows via substitution.

To summarize, we have proved above (see (24)) that the minimization of the objective in (22) with respect to $\boldsymbol{C}$, for any fixed $\boldsymbol{P} \geq \boldsymbol{0}$, leads to the original function of $\boldsymbol{P}$ in (20). Then it follows that the minimization of (22) with respect to $\boldsymbol{C}$ and $\{p_k\}$ yields the same $\{p_k\}$ as the minimization of (20). The usefulness of this observation lies in the fact that the minimization of the augmented function in (22) can be conveniently done by means of a cyclic algorithm that minimizes (22) with respect to $\boldsymbol{C}$, for fixed $\boldsymbol{P}$, then minimizes (22) with respect to $\boldsymbol{P}$, for given $\boldsymbol{C}$, and so forth until convergence. The solution to the first step of this algorithm was already derived, see (23). The solution needed in the second step can also be obtained in closed form. To show how this can be done, let

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{c}_1^* \\ \vdots \\ \boldsymbol{c}_{K+N}^* \end{bmatrix} \tag{28}$$

and observe that

$$\operatorname{tr}\left(\boldsymbol{C}^*\boldsymbol{P}^{-1}\boldsymbol{C}\right) = \operatorname{tr}\left(\boldsymbol{P}^{-1}\boldsymbol{C}\boldsymbol{C}^*\right) = \sum_{k=1}^{K+N} \frac{\|\boldsymbol{c}_k\|^2}{p_k}. \tag{29}$$

A. The Case of Different $\{\sigma_k\}$

By the Cauchy–Schwarz inequality

$$\begin{aligned}\left[\sum_{k=1}^{K+N} w_k^{1/2}\|\boldsymbol{c}_k\|\right]^2 &\leq \left[\sum_{k=1}^{K+N} \frac{\|\boldsymbol{c}_k\|^2}{p_k}\right]\left[\sum_{k=1}^{K+N} w_k p_k\right] \\ &= \sum_{k=1}^{K+N} \frac{\|\boldsymbol{c}_k\|^2}{p_k}. \end{aligned} \tag{30}$$

From this, it follows that the minimization of the objective in (22) with respect to $\{p_k\}$ (s.t. $p_k \geq 0$ and $\sum_{k=1}^{K+N} w_k p_k = 1$) gives (for fixed $\boldsymbol{C}$):

$$p_k = \frac{\|\boldsymbol{c}_k\|}{w_k^{1/2}\rho}, \quad \rho = \sum_{m=1}^{K+N} w_m^{1/2}\|\boldsymbol{c}_m\| \tag{31}$$

and the corresponding minimum value of the objective is

$$\left(\sum_{k=1}^{K+N} w_k^{1/2}\|\boldsymbol{c}_k\|\right)^2. \tag{32}$$

Equation (31) above provides the solution to the second step of the cyclic algorithm, whereas the solution to the first step is given by (23). Combining (23) and (31) leads to the updating formulas of *the SPICE algorithm* in which only the powers $\{p_k\}$ (that are the quantities of main interest) appear explicitly:

$$p_k^{i+1} = p_k^i \frac{\|\boldsymbol{a}_k^*\boldsymbol{R}^{-1}(i)\hat{\boldsymbol{R}}^{1/2}\|}{w_k^{1/2}\rho(i)} \quad k = 1, \ldots, K+N \tag{33}$$

$$\rho(i) = \sum_{m=1}^{K+N} w_m^{1/2} p_m^i \left\|\boldsymbol{a}_m^*\boldsymbol{R}^{-1}(i)\hat{\boldsymbol{R}}^{1/2}\right\| \tag{34}$$

where the index i denotes the iteration number, and $\boldsymbol{R}(i)$ is the matrix $\boldsymbol{R}$ made from $\{p_k^i\}$. The algorithm can be initialized with the power estimates obtained by means of the periodogram method (see, e.g., [1]):

$$p_k^0 = \frac{\boldsymbol{a}_k^*\hat{\boldsymbol{R}}\boldsymbol{a}_k}{\|\boldsymbol{a}_k\|^4} \quad k = 1, \ldots, K+N. \tag{35}$$

Remark 2: The SPICE algorithm for the alternative covariance fitting criterion in (15) (or (19)) can be readily derived by paralleling the above calculations. The result is an updating formula similar to (33) above with the only difference that $\hat{\boldsymbol{R}}^{1/2}$ in (33) and (34) should be replaced by $\hat{\boldsymbol{R}}$ and $\{w_k\}$ in (21) by $w_k = \|\boldsymbol{a}_k\|^2/\operatorname{tr}(\hat{\boldsymbol{R}})$. ■

B. The Case of Identical $\{\sigma_k\}$

In some applications, it is known that the noise components in the different elements of the array output vector have the same variance:

$$\sigma_1 = \cdots = \sigma_N \triangleq \sigma. \tag{36}$$

Using this information is important since, based on it, we can reduce the number of powers that need to be estimated. To derive the necessary modification of SPICE, that takes (36) into account, first observe that the minimization of (22) with respect to $\boldsymbol{C}$ is not affected by (36). However the minimization with respect to $\{p_k\}_{k=1}^K$ and σ, for fixed $\boldsymbol{C}$, is slightly different. Under the above constraint on $\{\sigma_k\}$, (29) becomes

$$\operatorname{tr}(\boldsymbol{C}^*\boldsymbol{P}^{-1}\boldsymbol{C}) = \sum_{k=1}^{K} \frac{\|\boldsymbol{c}_k\|^2}{p_k} + \sum_{k=K+1}^{K+N} \frac{\|\boldsymbol{c}_k\|^2}{\sigma}. \tag{37}$$

This function is to be minimized with respect to $\{p_k \geq 0\}$ and $\sigma \geq 0$, subject to

$$\sum_{k=1}^{K} w_k p_k + \gamma\sigma = 1 \tag{38}$$

where

$$\gamma = \sum_{k=K+1}^{K+N} w_k. \tag{39}$$

By the same argument as above, see (30)–(32), the solution to this optimization problem is

$$p_k = \frac{\|\boldsymbol{c}_k\|}{w_k^{1/2}\rho} \quad k = 1, \ldots K \tag{40}$$

$$\sigma = \frac{\left[\sum\limits_{k=K+1}^{K+N} \|\boldsymbol{c}_k\|^2\right]^{1/2}}{\gamma^{1/2}\rho} \tag{41}$$

where

$$\rho = \sum_{k=1}^{K} w_k^{1/2}\|\boldsymbol{c}_k\| + \gamma^{1/2}\left[\sum_{k=K+1}^{K+N} \|\boldsymbol{c}_k\|^2\right]^{1/2} \tag{42}$$

and the corresponding minimum function (for fixed $\boldsymbol{C}$) is

$$\left(\sum_{k=1}^{K} w_k^{1/2}\|\boldsymbol{c}_k\| + \gamma^{1/2}\left[\sum_{k=K+1}^{K+N} \|\boldsymbol{c}_k\|^2\right]^{1/2}\right)^2. \tag{43}$$

Inserting the expression (23) for $\boldsymbol{C}$ in (40) and (41), we obtain the following *modified SPICE algorithm* for the case of (36):

$$p_k^{i+1} = p_k^i \frac{\|\boldsymbol{a}_k^* \boldsymbol{R}^{-1}(i)\hat{\boldsymbol{R}}^{1/2}\|}{w_k^{1/2}\rho(i)} \quad k = 1, \ldots, K \tag{44}$$

$$\sigma^{i+1} = \sigma^i \frac{\|\boldsymbol{R}^{-1}(i)\hat{\boldsymbol{R}}^{1/2}\|}{\gamma^{1/2}\rho(i)} \tag{45}$$

$$\rho(i) = \sum_{k=1}^{K} w_k^{1/2} p_k^i \|\boldsymbol{a}_k^* \boldsymbol{R}^{-1}(i)\hat{\boldsymbol{R}}^{1/2}\| + \gamma^{1/2}\sigma^i \|\boldsymbol{R}^{-1}(i)\hat{\boldsymbol{R}}^{1/2}\|. \tag{46}$$

Initial estimates of the powers, for $k = 1, \ldots, K$, can still be obtained with the periodogram (see (35)), and σ can be initialized for instance as the average of the N smallest values of $\{p_k^0\}_{k=1}^K$ each multiplied by $\|\boldsymbol{a}_k\|^2$. To motivate this choice for σ^0 note that, at least for M sufficiently large, we have $p_k^0 = \boldsymbol{a}_k^* \hat{\boldsymbol{R}} \boldsymbol{a}_k / \|\boldsymbol{a}_k\|^4 \geq \sigma / \|\boldsymbol{a}_k\|^2$ with the equality being likely to hold for the smallest values of p_k^0, e.g., for the N smallest such values. This observation implies that (below $\{\tilde{p}_k^0\}_{k=1}^N$ are the N smallest values of the set $\{p_k^0\}$ and $\{\tilde{\boldsymbol{a}}_k\}$ the corresponding manifold vectors)

$$\sigma^0 = \sum_{k=1}^{N} \frac{\tilde{p}_k^0 \|\tilde{\boldsymbol{a}}_k\|^2}{N} \tag{47}$$

can be expected to be a reasonable estimate of σ.

Because SPICE monotonically decreases the objective function (due to its cyclic operation) and as the minimization problem it solves is convex, we can expect that the algorithm has global convergence properties. That this is indeed the case follows from the general analysis in [10] where it was proved that, under weak conditions (essentially requiring that $p_k^0 > 0$ and that the matrix $\boldsymbol{R}(i)$ stays positive definite as the iteration proceeds), *the limit points of the SPICE iterative process are global solutions to* (20).

IV. SOCP Formulation of SPICE

SPICE takes account of the noise without introducing any hyperparameters to be selected by the user. Indeed, SPICE is a *fully data adaptive* approach that estimates both $\{p_k\}$ and $\{\sigma_k\}$ from the available data snapshots. As already mentioned, to our knowledge, the SPICE criterion was never used in the previous literature on sparse parameter estimation. It is therefore somewhat unexpected that SPICE can be related to (an extended version of) the classical ℓ_1-norm minimization approach in (6), as we explain in the rest of this section.

As indicated in Section II the SPICE estimation problem, (20), can be cast and solved as an SDP. However it can also be reformulated as an SOCP which is a special case of SDP. It is this reformulation of (20) as an SOCP that reveals the connection between SPICE and a certain extended ℓ_1-norm minimization problem. To reformulate SPICE as an SOCP we proceed in the following way. Instead of minimizing (22) cyclically (or alternately) with respect to $\boldsymbol{C}$ and $\boldsymbol{P}$, we minimize this function with respect to $\{p_k\}$ for arbitrary $\boldsymbol{C}$. The result of this minimization operation is already available from the calculations in the previous section where it was shown that the minimizing powers are given by (31) and the corresponding minimum function by (32). It remains to minimize (32) with respect to $\boldsymbol{C}$ (under the constraint in (22)):

$$\min_{\boldsymbol{C}} \sum_{k=1}^{K+N} w_k^{1/2}\|\boldsymbol{c}_k\| \quad \text{s.t. } \boldsymbol{A}^*\boldsymbol{C} = \hat{\boldsymbol{R}}^{1/2}. \tag{48}$$

Once $\boldsymbol{C}$ is obtained from (48), $\{p_k\}$ can be calculated using (31).

The above problem, which is an SOCP (see Appendix A), is similar to the ℓ_1-norm minimization problem (6). However, there are also significant differences between these two problems, as well as between the ways in which they have been obtained, as follows.

i) The ℓ_1-norm in (48) has a weighted form, with the weights determined by the data. It is well known that $w_k = \boldsymbol{a}_k^* \hat{\boldsymbol{R}}^{-1} \boldsymbol{a}_k / N$ is an estimate of the inverse power corresponding to the kth point of the location grid (see, e.g., [1]). Therefore, the smaller the said power the larger the $w_k^{1/2}$ and consequently more weight is put on minimizing the associated $\|\boldsymbol{c}_k\|$ in (48), which intuitively is the way it should be. Furthermore, the ℓ_1-norm objective in (48) follows naturally from the statistically sound covariance fitting criterion in (13), in contrast with the similar objective in (6) whose motivation is essentially heuristical.

ii) The matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ in (48) comprise extra rows that account for the noise in the data, unlike the similar quantities $\boldsymbol{B}$ and $\boldsymbol{S}$ in (6).

iii) The inequality constraint in (6) is replaced in (48) by a hyperparameter-free equality constraint where $\hat{\boldsymbol{R}}^{1/2}$ is used in lieu of $\boldsymbol{Y}^*$ in (6). Note that the matrix $\boldsymbol{C}$ in (48) has N columns instead of M columns as $\boldsymbol{S}$ in (6), which evidently makes (48) more convenient computationally (typically $M \gg N$). This dimensional reduction also follows naturally from the employed covariance fitting criterion, whereas the idea suggested in [3] for reducing the number of columns of $\boldsymbol{Y}^*$ and $\boldsymbol{S}$ in (6) was motivated on more or less heuristical grounds and its use depends on the selection of an additional hyperparameter, besides η.

The reader might wonder what is the relationship between $\boldsymbol{C}$ in (48) and $\boldsymbol{S}$ in (6). To clarify this aspect, observe that we can write the data (1) in the following matrix form:

$$\boldsymbol{Y}^* = \boldsymbol{B}^*\boldsymbol{S} + \boldsymbol{\Delta} \tag{49}$$

where

$$\boldsymbol{\Delta} = [\boldsymbol{\epsilon}(1), \ldots, \boldsymbol{\epsilon}(M)]. \tag{50}$$

It follows from (49) that we can assume, as in (6), that $\|\boldsymbol{Y}^* - \boldsymbol{B}^*\boldsymbol{S}\| \leq \eta$ provided that η is "sufficiently large"; furthermore, a proper selection of η would require information about the noise variance(s) and the noise distribution, which is usually unavailable, and even then the inequality constraint in (6) could not in general be made to hold surely for any finite practically convenient value of η. On the other hand, it also follows from (49) that

$$\boldsymbol{Y}^* = \boldsymbol{A}^* \begin{bmatrix} \boldsymbol{S} \\ \boldsymbol{\Delta} \end{bmatrix} \tag{51}$$

and therefore that

$$\boldsymbol{A}^* \underbrace{\begin{bmatrix} \boldsymbol{S} \\ \boldsymbol{\Delta} \end{bmatrix} \frac{\boldsymbol{Y}\hat{\boldsymbol{R}}^{-1/2}}{M}}_{\triangleq \boldsymbol{C}} = \hat{\boldsymbol{R}}^{1/2}. \tag{52}$$

The above equation clarifies the relationship between $\boldsymbol{C}$ and $\boldsymbol{S}$. Unlike (6), the equality constraint (52) holds deterministically and it does not involve any hyperparameter. Note also that if $\boldsymbol{S}$ is row-sparse then so is the matrix $\boldsymbol{S}\boldsymbol{Y}\hat{\boldsymbol{R}}^{-1/2}/M$ in (52). Regarding the other matrix appearing in (52), viz. $\boldsymbol{\Delta}\boldsymbol{Y}\hat{\boldsymbol{R}}^{-1/2}/M$, it tends to zero as the signal-to-noise ratio (SNR) increases; however, for small or medium SNR values this matrix is not row-sparse, which may be perceived as a downside of (52), although a relatively minor one as the matrix $\boldsymbol{C}$ can still be considered to be row-sparse because $K \gg N$.

Whenever *the constraint* $\sigma_1 = \cdots = \sigma_N \triangleq \sigma$ *is enforced*, the above reformulation of the SPICE estimation problem should be changed as follows. First, $\boldsymbol{C}$ is obtained as the solution of the following SOCP (see (43)):

$$\begin{aligned} \min_{\boldsymbol{C}} \quad & \sum_{k=1}^{K} w_k^{1/2} \|\boldsymbol{c}_k\| + \gamma^{1/2} \left[\sum_{k=K+1}^{K+N} \|\boldsymbol{c}_k\|^2 \right]^{1/2} \\ \text{s.t.} \quad & \boldsymbol{A}^*\boldsymbol{C} = \hat{\boldsymbol{R}}^{1/2} \end{aligned} \tag{53}$$

(see Appendix A for the formulation of the above problem as an SOCP). Then $\{p_k\}$ and σ are determined using (40)–(42).

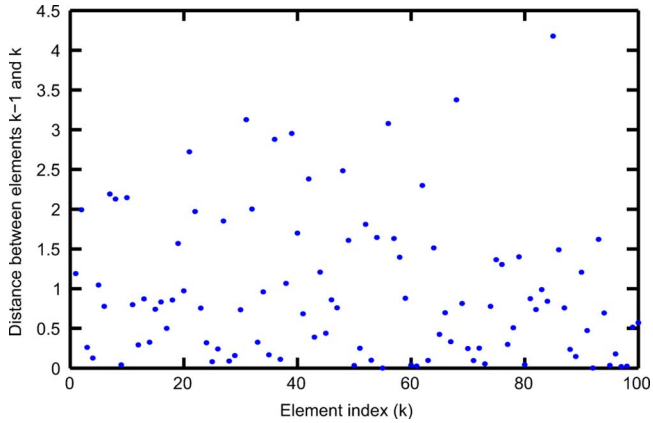

Fig. 1. Inter-element spacings (in units of $\lambda/2$) for the NULA with $N = 100$.

According to our albeit limited experience, the SPICE multiplicative algorithm in (33) and (34) (or (44)–(46)) is usually visibly faster than the SOCP-based algorithm that was outlined in this section. However, the SOCP formulation of the SPICE estimation problem remains important for understanding the connection between the proposed approach and the standard one based on ℓ_1-norm minimization.

V. Numerical Illustrations and Concluding Remarks

In this section, we illustrate numerically the performance of the proposed methods and compare it with the performance of some standard methods from the literature. We will consider both a uniform linear array (ULA) and a nonuniform linear array (NULA). In the ULA case, the sensors are uniformly placed with a spacing of $\lambda/2$, where λ denotes the wavelength of the sources. The inter-element spacings in the case of NULA are as shown in Fig. 1. The number of sensors in the array is $N = 10$ for ULA and $N = 100$ for NULA. In the ULA case the number of snapshots is $M = 200$ whereas it is $M = 100$ in the NULA case. The steering vector $\boldsymbol{a}_k$ for the ULA, corresponding to a direction of arrival (DOA) equal to θ_k, is given by

$$\boldsymbol{a}_k = \begin{bmatrix} e^{i\pi \sin(\theta_k)} \\ \vdots \\ e^{i\pi N \sin(\theta_k)} \end{bmatrix} \tag{54}$$

(observe that $\|\boldsymbol{a}_k\|^2 = N$ is a constant). The steering vector for the NULA can be similarly defined. The interval for the DOA is $\Omega = (-90°, 90°]$. We use a uniform grid $\{\theta_k\}_{k=1}^K$ to cover Ω, with a step of $0.1°$, which means that $K = 1800$.

The source signals $\{s_k(t)\}$, see (1), have constant modulus, which is usually the situation in communications applications. We will consider cases with both uncorrelated and coherent sources. The noise term in (1) is chosen to be white, both temporally and spatially, and Gaussian distributed with zero mean and variance σ.

A. Fixed Sources : DOA Estimation

In this subsection we consider DOA estimation of fixed sources using the ULA. The data samples were simulated using (1) with three sources at $\theta_1 = 10°$, $\theta_2 = 40°$, $\theta_3 = 55°$ and the following signals $s_1(t) = 3e^{i\varphi_1(t)}$, $s_2(t) = 10e^{i\varphi_2(t)}$ and $s_3(t) = 10e^{i\varphi_3(t)}$, where the phases $\{\varphi_k(t)\}_{k=1}^3$ were

independently and uniformly distributed in $[0, 2\pi]$. In the simulation with coherent sources, the sources at θ_1 and θ_3 were the coherent ones (i.e., they had identical phases), whereas the source at θ_2 was uncorrelated to them. The noise variance was varied to control the signal-to-noise ratio defined as

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

The following methods were used for DOA estimation :

- M_1: the periodogram (PER), see (35);
- M_2: the iterative adaptive approach (IAA), see [12];
- M_3: the multiple signal classification method (MUSIC), see e.g., [1];
- M_4: SPICE, see (15), (19) as well as Remark 2 and [7];
- M_5: the enhanced SPICE or, for short, SPICE-plus (SPICE+), see (44)–(46).

Note that IAA was shown in [12] to outperform the method of [3] and its variations in the literature, which is why we selected IAA for the present comparison. As a cross-check and for comparison's sake we have also computed the SPICE+ estimates by means of an SDP solver (see [5] for the latter). The so-obtained estimates were identical (within numerical accuracy) to those provided by the SPICE+ algorithm; however, as expected, the SDP solver was either much slower or, worse, it could not be executed for instance due to memory problems.

In Fig. 2, we show the average root mean-square errors (RMSEs) of the DOA estimates obtained with $M_1 - M_5$ in 1000 Monte Carlo runs, for several SNR values:

$$\text{RMSE} = \left[\frac{1}{3000} \sum_{k=1}^{3} \sum_{m=1}^{1000} \left(\hat{\theta}_k^m - \theta_k \right)^2 \right]^{1/2}$$

where $\hat{\theta}_k^m$ denotes the estimate of θ_k in the mth Monte Carlo run. In this figure, the top plot shows the RMSEs for uncorrelated sources and the bottom plot shows the RMSEs for coherent sources. As can be seen from this figure, the PER estimates of the DOAs have poor accuracy in all cases under study, due to a significant bias (which does not decrease as the SNR increases). Regarding IAA, this method provides competitive estimates only for SNR $\geq$ 0 dB. The parametric method of MUSIC, which requires knowledge on the number of sources in the data, yields reasonably accurate DOA estimates in the uncorrelated source case (at least for SNR $\geq$ 0 dB) but completely fails in the coherent source case (as expected). The proposed methods of SPICE and SPICE+ give the best performance in the cases considered: in particular, their threshold SNR appears to be lower by some 10 dB than the threshold SNR of IAA and of MUSIC. For uncorrelated sources, SPICE+ is more accurate than SPICE, whereas SPICE outperforms SPICE+ in the coherent source case. From a computational standpoint SPICE+ converged faster than SPICE presumably due to the data-dependent weights used by SPICE+.

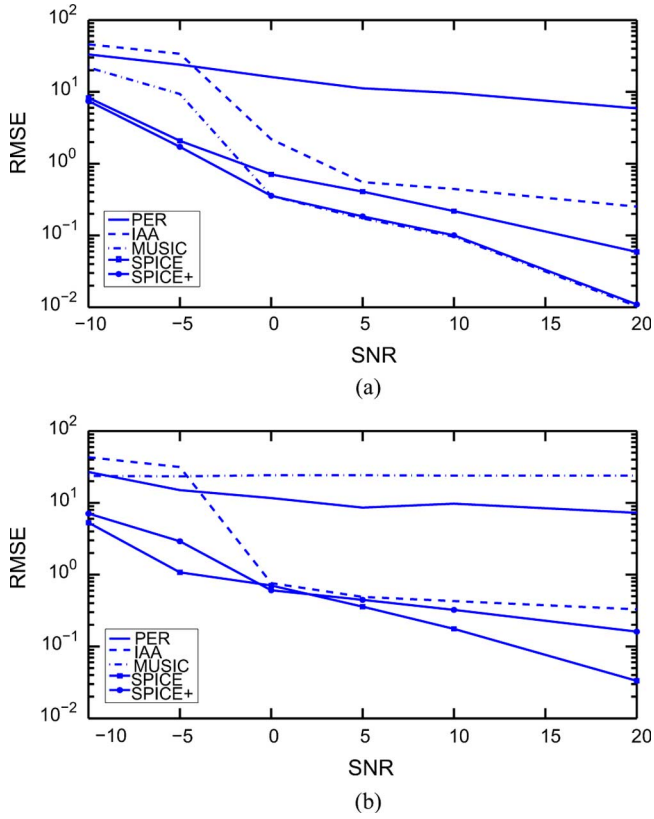

Fig. 2. Estimation performance of PER, IAA, MUSIC, SPICE and SPICE+ (the ULA case): RMSE versus SNR for the DOA estimates obtained in 1000 Monte Carlo runs. (a) Uncorrelated sources. (b) Coherent sources.

B. Mobile Sources : DOA Tracking

In this subsection, we consider DOA estimation of mobile sources using the NULA. The data samples were generated assuming two mobile sources, one moving linearly from 30° to 60° and the other from 60° to 30°, in steps of 0.03°, over a course of 1000 data snapshots. The uncorrelated signals of these two sources were given by $s_1(t) = 10e^{i\varphi_1(t)}$ and $s_2(t) = 10e^{i\varphi_2(t)}$, where the phases $\{\varphi_k(t)\}$ were independently and uniformly distributed in $[0, 2\pi]$.

In this example, only the methods M_1, M_4, and M_5 are considered. At any given time instant t, the sample covariance matrix $\hat{\boldsymbol{R}}$ is formed from the most recent 100 data snapshots and the DOA estimates are recomputed by each of the three methods. As before, the SPICE methods, M_4 and M_5, are initialized with the PER estimate. Alternatively, one may think of initializing the SPICE methods with their respective estimates obtained from the previous data window; however, as discussed in Section III-B, for global convergence SPICE should be initialized with a dense estimate rather than a sparse one. The SPICE methods were iterated only five times except at $t = 100$ where they were applied for the first time and were iterated till convergence. We have also tried a larger number of iterations, such as 20, but noticed no significant improvement in the DOA estimates [compare Fig. 3(c) and (d)]. The SNR in this simulation was 20 dB (however, very similar results were obtained for SNR $=$ 10 dB and 0 dB).

Fig. 3 shows the plots of DOA estimates versus t for the three considered methods. In each plot, the estimates shown at any time t were obtained as the locations of the two largest peaks in the spatial spectrum provided by the method in question. As can be seen from the figure, PER performs poorly as it

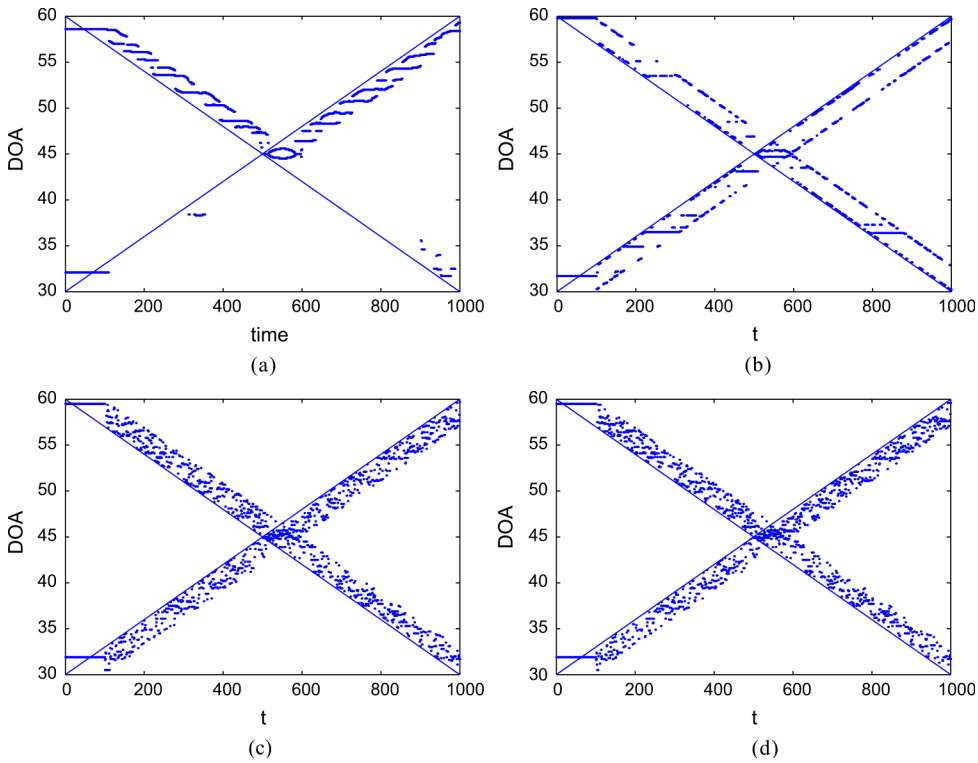

Fig. 3. Tracking performance of PER, SPICE and SPICE+ (the NULA case). The solid lines denote the trajectories of the true DOAs: (a) PER; (b) SPICE with five iterations; (c) SPICE+ with five iterations; and (d) SPICE+ with 20 iterations.

cannot track either the source moving from 30° to 45° or the one moving from 45° to 30°. Note that at each time instant, the data window (made of the most recent 100 data snapshots) comprises two sets of DOAs each with a width of 3°. For instance, at $t = 400$, the data window will contain signals with DOAs lying in the interval $[39°, 42°]$ and, respectively, in $[48°, 51°]$. PER often picks wrongly the two peaks only from the set with larger DOAs. On the other hand, the SPICE methods correctly pick the two peaks from both sets which leads to the band-like appearance of the corresponding plots in Fig. 3(b)–(d); observe that, as expected, the width of the bands in these figures (measured along the DOA axis) is equal to 3°. Note also that the true DOA trajectories are well approximated by the upper edge of the bands in Fig. 3(b)–(d) corresponding to the increasing DOA and by the lower edge of the bands associated with the decreasing DOA—this behavior, which was expected in view of the above discussion, is in fact quite stable with respect to the SNR [in simulations not shown here, we have observed that decreasing the SNR from 20 to 0 dB caused only 1 DOA estimate, out of 2000, to lie outside bands similar to those in Fig. 3(b)–(d)]. To provide further insight into this type of behavior, a single source moving along a sinusoidal trajectory was considered. Fig. 4 shows the DOA estimates for this source obtained with SPICE+. It is clear from this plot that the width of the band decreases till $t = 500$, where it is nearly zero, and then starts increasing again; hence, as expected, the width of the band is proportional to the slope of the DOA variation. Furthermore in this case, too, the true DOA trajectory is well approximated by the upper edge of the band (for increasing DOA) and the lower edge (for decreasing DOA).

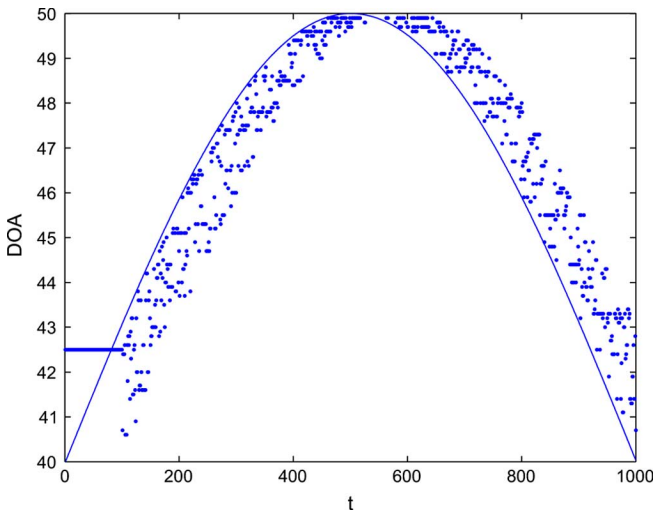

Fig. 4. Tracking performance of SPICE+ for a sinusoidal trajectory (the NULA case). The solid line denotes the true DOA trajectory.

APPENDIX A
SOME SDP AND SOCP FORMULATIONS

First, we show that the problem of minimizing (18), s.t. $p_k \geq 0$ $(k = 1, \ldots, K+N)$, can be formulated as an SDP. The proof that the same is true for (20) is similar and therefore its details are omitted. Let

$$\hat{\boldsymbol{R}}^{1/2} = \begin{bmatrix} \boldsymbol{r}_1^* \\ \vdots \\ \boldsymbol{r}_N^* \end{bmatrix} \tag{55}$$

and let

$$v_k = \left(\boldsymbol{a}_k^* \hat{\boldsymbol{R}}^{-1} \boldsymbol{a}_k \right). \tag{56}$$

Using this notation we can rewrite the function g in (18) as

$$g = \sum_{k=1}^{N} \boldsymbol{r}_k^* \boldsymbol{R}^{-1} \boldsymbol{r}_k + \sum_{k=1}^{K+N} v_k p_k. \tag{57}$$

Let $\{\alpha_k\}_{k=1}^N$ be auxiliary variables satisfying $\alpha_k \geq \boldsymbol{r}_k^* \boldsymbol{R}^{-1} \boldsymbol{r}_k$, or equivalently

$$\begin{bmatrix} \alpha_k & \boldsymbol{r}_k^* \\ \boldsymbol{r}_k & \boldsymbol{R} \end{bmatrix} \geq \boldsymbol{0}. \tag{58}$$

Then the minimization problem under discussion can be stated as

$$\begin{aligned} \min_{\{\alpha_k\},\{p_k\}} \quad & \sum_{k=1}^{N} \alpha_k + \sum_{k=1}^{K+N} v_k p_k \\ \text{s.t.} \quad & p_k \geq 0 \quad k = 1, \ldots, K+N \\ & \begin{bmatrix} \alpha_k & \boldsymbol{r}_k^* \\ \boldsymbol{r}_k & \boldsymbol{R} \end{bmatrix} \geq \boldsymbol{0} \quad k = 1, \ldots, N \end{aligned} \tag{59}$$

which is an SDP (as $\boldsymbol{R}$ is a linear function of $\{p_k\}$) [4], [9].

Next, we show that (48) can be cast as an SOCP. Let $\{\beta_k\}$ denote $(K+N)$ auxiliary variables that satisfy

$$\beta_k \geq w_k^{1/2} \|\boldsymbol{c}_k\| \quad k = 1, \ldots, K+N. \tag{60}$$

Then simply observe that (48) can be rewritten as

$$\begin{aligned} \min_{\{\beta_k\},\boldsymbol{C}} \quad & \sum_{k=1}^{K+N} \beta_k \\ \text{s.t.} \quad & \|\boldsymbol{c}_k\| \leq \frac{\beta_k}{w_k^{1/2}} \quad k = 1, \ldots, K+N \\ & \boldsymbol{A}^* \boldsymbol{C} = \hat{\boldsymbol{R}}^{1/2} \end{aligned} \tag{61}$$

which is an SOCP [4].

Finally, we show how to reformulate (53) as an SOCP. Let $\{\beta_k\}$ still denote some auxiliary variables (but now only $K+1$ of them) and constrain them to satisfy

$$\begin{aligned} \beta_k &\geq w_k^{1/2} \|\boldsymbol{c}_k\| \quad k = 1, \ldots, K \\ \beta_{K+1} &\geq \gamma^{1/2} \left[\sum_{k=1}^{K+N} \|\boldsymbol{c}_k\|^2 \right]^{1/2}. \end{aligned} \tag{62}$$

Using $\{\beta_k\}$ we can restate (53) in the following form:

$$\begin{aligned} \min_{\{\beta_k\},\boldsymbol{C}} \quad & \sum_{k=1}^{K+1} \beta_k \\ \text{s.t.} \quad & \|\boldsymbol{c}_k\| \leq \frac{\beta_k}{w_k^{1/2}} \quad k = 1, \ldots, K \\ & \left\| \begin{bmatrix} \boldsymbol{c}_{K+1} \\ \vdots \\ \boldsymbol{c}_{K+N} \end{bmatrix} \right\| \leq \frac{\beta_{K+1}}{\gamma^{1/2}} \\ & \boldsymbol{A}^* \boldsymbol{C} = \hat{\boldsymbol{R}}^{1/2} \end{aligned} \tag{63}$$

which, once again, is an SOCP.

APPENDIX B
EQUIVALENCE OF (18) AND (20)

Let P_0 and P_1 denote the minimization problems corresponding to (18) and (20). Therefore

$$\mathrm{P}_0: \min_{\{p_k \geq 0\}} \operatorname{tr}(\boldsymbol{R}^{-1} \hat{\boldsymbol{R}}) + \operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \boldsymbol{R}) \tag{64}$$

and

$$\mathrm{P}_1: \min_{\{p_k \geq 0\}} \operatorname{tr}(\boldsymbol{R}^{-1} \hat{\boldsymbol{R}}) \quad \text{s.t.} \operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \boldsymbol{R}) = 1. \tag{65}$$

Note that we have replaced N by 1 in the right-hand side of the constraint in (65) for notational convenience (this replacement has only a scaling effect on the solution of P_1). For completeness, we will also consider the following problem obtained by constraining the first term in (64) to one:

$$\mathrm{P}_2: \min_{\{p_k \geq 0\}} \operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \boldsymbol{R}) \quad \text{s.t.} \operatorname{tr}(\boldsymbol{R}^{-1} \hat{\boldsymbol{R}}) = 1. \tag{66}$$

First, we prove the equivalence of P_1 and P_2. By making use of an auxiliary variable α we can rewrite (65) and (66) as follows:

$$\mathrm{P}_1: \min_{\alpha,\{p_k \geq 0\}} \alpha \quad \text{s.t.} \operatorname{tr}(\boldsymbol{R}^{-1} \hat{\boldsymbol{R}}) = \alpha \text{ and } \operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \boldsymbol{R}) = 1 \tag{67}$$

$$\mathrm{P}_2: \min_{\alpha,\{p_k \geq 0\}} \alpha \quad \text{s.t.} \operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \boldsymbol{R}) = \alpha \text{ and } \operatorname{tr}(\boldsymbol{R}^{-1} \hat{\boldsymbol{R}}) = 1. \tag{68}$$

Let $\{\tilde{p}_k = \alpha p_k\}$ (note that $\alpha > 0$) and reformulate P_1 as

$$\mathrm{P}_1: \min_{\alpha,\{\tilde{p}_k \geq 0\}} \alpha \quad \text{s.t.} \operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{R}}) = \alpha \text{ and } \operatorname{tr}(\tilde{\boldsymbol{R}}^{-1} \hat{\boldsymbol{R}}) = 1 \tag{69}$$

where $\tilde{\boldsymbol{R}}$ is made from $\{\tilde{p}_k\}$. It follows from (67) and (69) that if $\{p_k\}$ is the solution to P_2 then $\{p_k/\alpha\}$ is the solution to P_1, and thus the proof of the equivalence between P_1 and P_2 is concluded.

Next, we consider P_0 and P_2. By writing down the Karush–Kuhn–Tucker conditions for P_0 it can be shown that the two terms of the objective in (64) must be equal to one another at the minimizing $\{p_k\}$. A more direct proof of this fact runs as follows. Assume that $\{p_k\}$ is the optimal solution of P_0 and that at $\{p_k\}$ we have

$$\operatorname{tr}(\boldsymbol{R}^{-1} \hat{\boldsymbol{R}}) = \rho^2 \operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \boldsymbol{R}) \tag{70}$$

for some $\rho^2 \neq 1$; note that the scaling factor in (70) must necessarily be positive, which is why we wrote it as ρ^2 ; below we will let $\rho > 0$ denote the square root of ρ^2. We will show in the following that the previous assumption leads to a contradiction : $\{p_k\}$ cannot be the solution to P_0 if $\rho \neq 1$. To do so let $\{\tilde{p}_k = \rho p_k\}$ and observe that, at $\{\tilde{p}_k\}$, the two terms in the objective of P_0 are identical:

$$\operatorname{tr}(\tilde{\boldsymbol{R}}^{-1} \hat{\boldsymbol{R}}) = \operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{R}}). \tag{71}$$

Making use of (70) and (71) along with the assumption that $\{p_k\}$ is the solution to P_0, we obtain the following inequality:

$$\operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \boldsymbol{R})(1 + \rho^2) < 2 \operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{R}}) = 2\rho \operatorname{tr}(\hat{\boldsymbol{R}}^{-1} \boldsymbol{R}) \tag{72}$$

or, equivalently,

$$(1 - \rho)^2 < 0 \tag{73}$$

which cannot hold, and hence the solution to P_0 must satisfy (70) with $\rho = 1$. Using this fact we can reformulate P_0 as follows:

$$\mathrm{P}_0\colon \min_{\alpha,\{p_k\geq 0\}} \alpha \quad \text{s.t. } \mathrm{tr}(\boldsymbol{R}^{-1}\hat{\boldsymbol{R}}) = \alpha \text{ and } \mathrm{tr}(\hat{\boldsymbol{R}}^{-1}\boldsymbol{R}) = \alpha \tag{74}$$

or, equivalently (as $\alpha > 0$),

$$\mathrm{P}_0\colon \min_{\beta,\{\tilde{p}_k\geq 0\}} \beta \quad \text{s.t. } \mathrm{tr}(\hat{\boldsymbol{R}}^{-1}\tilde{\boldsymbol{R}}) = \beta \text{ and } \mathrm{tr}(\tilde{\boldsymbol{R}}^{-1}\hat{\boldsymbol{R}}) = 1 \tag{75}$$

where $\{\tilde{p}_k = \alpha p_k\}$, as before, and $\beta = \alpha^2$ is a new auxiliary variable. Comparing (75) and (67) concludes the proof of the equivalence of P_0 and P_2.

REFERENCES

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

[2] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. .B, Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[3] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[4] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, no. 1-3, pp. 193–228, 1998.

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

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

[7] 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 Trans. Signal Process.*, vol. 59, no. 1, pp. 35–47, Jan. 2011.

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

[9] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.

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

[11] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[12] 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, no. , pp. 425–443, Jan. 2010.

Petre Stoica (SM'91–F'94) is a Professor of systems modeling in the Department of Information Technology, Uppsala University, Uppsala, 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 currently working towards the Ph.D. degree in electrical engineering with applications in signal processing at the Department of Information Technology, Uppsala University, Uppsala, Sweden.

Jian Li (S'88–M'90–SM'97–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 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 Executive Committee Member of the 2002 International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, in May 2002. 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 received the 1994 National Science Foundation Young Investigator Award and the 1996 Office of Naval Research Young Investigator Award. 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 the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS in 2005. She is a Fellow of IET and a member of Sigma Xi and Phi Kappa Phi.