Weighted Subspace Fitting for General Array Error Models

Magnus Jansson, *Member, IEEE*, A. Lee Swindlehurst, *Member, IEEE*, and Björn Ottersten, *Member, IEEE*

***Abstract*— Model error sensitivity is an issue common to all high-resolution direction-of-arrival estimators. Much attention has been directed to the design of algorithms for minimum variance estimation taking only finite sample errors into account. Approaches to reduce the sensitivity due to array calibration errors have also appeared in the literature. Herein, one such approach is adopted that assumes that the errors due to finite samples and model errors are of comparable size. A weighted subspace fitting method for very general array perturbation models is derived. This method provides minimum variance estimates under the assumption that the prior distribution of the perturbation model is known. Interestingly, the method reduces to the WSF (MODE) estimator if no model errors are present. Vice versa, assuming that model errors dominate, the method specializes to the corresponding "model-errors-only subspace fitting method." Unlike previous techniques for model errors, the estimator can be implemented using a two-step procedure if the nominal array is uniform and linear, and it is also consistent even if the signals are fully correlated. The paper also contains a large sample analysis of one of the alternative methods, namely, MAPprox. It is shown that MAPprox also provides minimum variance estimates under reasonable assumptions.**

***Index Terms*— Antenna arrays, array signal processing, direction-of-arrival estimation, error analysis, parameter estimation, robustness.**

I. Introduction

ALL SIGNAL parameter estimation methods in array signal processing rely on information about the array response and assume that the signal wavefronts impinging on the array have perfect spatial coherence (e.g., perfect plane waves). The array response may be determined by either empirical measurements, which is a process referred to as array *calibration*, or by making certain assumptions about the sensors in the array and their geometry (e.g., identical sensors in known locations). Unfortunately, an array cannot be perfectly calibrated, and analytically derived array responses relying on the array geometry and wave propagation are at best good approximations. Due to changes in antenna location, temperature, and the surrounding environment, the response of the array may be significantly different than when it was last calibrated. Furthermore, the calibration measurements themselves are subject to gain and phase errors, and they can only be obtained for discrete angles, thus necessitating interpolation techniques for uncalibrated directions.

Because of the many sources of error listed above, the limiting factor in the performance of array signal processing algorithms is most often not measurement noise but rather perturbations in the array response model. Depending on the size of such errors, estimates of the directions of arrival (DOA's) and the source signals may be significantly degraded. A number of studies have been conducted to quantify the performance degradation due to model errors for both DOA [1]–[12] and signal estimation [13]–[17]. These analyzes have assumed either a deterministic model for the errors, using bias as the performance metric, or a stochastic model, where variance is typically calculated.

A number of techniques have also been considered for improving the robustness of array processing algorithms. In one such approach, the array response is parameterized not only by the DOA's of the signals but by perturbation or "nuisance" parameters that describe deviations of the response from its nominal value as well. These parameters can include, for example, displacements of the antenna elements from their nominal positions, uncalibrated receiver gain, and phase offsets, etc. With such a model, a natural approach is to attempt to estimate the unknown nuisance parameters simultaneously with the signal parameters. Such methods are referred to as *auto-calibration* techniques and have been proposed by a number of authors, including [18]–[26]. When auto-calibration techniques are employed, it is critical to determine whether both the signal and nuisance parameters are identifiable. In certain cases, they are not; for example, we cannot uniquely estimate both DOA's and sensor positions unless, of course, additional information is available, such as sources in known locations [21], [27]–[29], cyclostationary signals with two or more known cycle frequencies [30], or partial information about the phase response of the array [31]. The identifiability problem can be alleviated if the perturbation parameters are assumed to be drawn from some known *a priori* distribution. Although this itself represents a form of additional information, it has the advantage of allowing an optimal maximum *a posteriori* (MAP) solution to the problem to be formulated [24]. In [24], it is shown that by using an asymptotically equivalent approximation to the resulting MAP criterion, the estimation of the signal and nuisance parameters can be decoupled, leading to a significant simplification of the problem.

Manuscript received August 4, 1997; revised January 29, 1998. This work was supported in part by the Swedish Research Council for Engineering Sciences (TFR) and the National Science Foundation under Grant MIP-9408154. The associate editor coordinating the review of this paper and approving it for publication was Dr. Yingbo Hua.

M. Jansson and B. Ottersten are with the Department of Signals, Sensors, and Systems, Royal Institute of Technology (KTH), Stockholm, Sweden (e-mail: magnusj@s3.kth.se).

A. L. Swindlehurst is with the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602 USA.

Publisher Item Identifier S 1053-587X(98)05960-1.

Another technique for reducing the sensitivity of array processing algorithms to model errors is through the use of statistically optimal weightings based on probabilistic models for the errors. Optimal weightings that ignore the finite sample effects of noise have been developed for MUSIC [6] and the general subspace fitting technique [7]. A more general analysis in [9] included the effects of noise and presented a weighted subspace fitting (WSF) method whose DOA estimates asymptotically achieve the Cramér–Rao bound (CRB) but only for a very special type of array error model. On the other hand, the MAP approach in [24] is asymptotically statistically efficient for very general error models. However, since it is implemented by means of *noise* subspace fitting [32], if the sources are highly correlated or closely spaced in angle, its finite sample performance may be poor. In fact, the method of [24] is not a consistent estimator of the DOA's if the signals are perfectly coherent.

In this paper, we develop a statistically efficient weighted *signal* subspace fitting (SSF) algorithm that holds for very general array perturbation models, that has much better finite sample performance than [24] when the signals are highly correlated or closely spaced, and that yields consistent estimates when coherent sources are present. An additional advantage of our SSF formulation is that if the array is nominally uniform and linear, a two-step procedure similar to that for MODE [33] can be used to eliminate the search for the DOA's. Furthermore, as part of our analysis, we show that the so-called MAPprox technique described in [22] also asymptotically achieves the CRB, and we draw some connections between it and the new SSF algorithm presented herein. The weighted SSF method presented here differs from other WSF techniques like [7], [9], and [34] in that all of the elements of the vectors spanning the signal subspace have their own individual weighting [35]. In these earlier algorithms, the form of the weighting forced certain subsets of the elements to share a particular weight. It is the generalization to individual weights that is the key to the algorithm's optimality for very general array response perturbation models. For this reason, we will refer to the algorithm as generalized weighted subspace fitting (GWSF).

The paper is organized as follows. In the next section, the data model is briefly introduced. In Section III, more insight into the studied problem is given by presenting previous approaches. In Section IV, the proposed GWSF estimator is introduced, and its statistical properties are analyzed in Section V. Details about the implementation of the algorithm are given in Section VI. In Section VII, we show that the MAPprox estimator has very close connections with GWSF and inherits the same asymptotic properties as GWSF. Finally, the theoretical observations are illustrated by numerical examples in Section VIII.

II. Problem Formulation

Assume that the output of an array of m sensors is given by the model

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho})\mathbf{s}(t) + \mathbf{n}(t)$$

where $\mathbf{s}(t)$ is a complex d vector containing the emitted signal waveforms, and $\mathbf{n}(t)$ is an additive noise vector. The array steering matrix is defined as

$$\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho}) = [\bar{\mathbf{a}}(\boldsymbol{\theta}_1, \boldsymbol{\rho}) \ \cdots \ \bar{\mathbf{a}}(\boldsymbol{\theta}_d, \boldsymbol{\rho})]$$

where $\bar{\mathbf{a}}(\boldsymbol{\theta}_i, \boldsymbol{\rho})$ denotes the array response to a unit waveform associated with the signal parameter $\boldsymbol{\theta}_i$ (possibly vector valued, although we will specialize to the scalar case). The parameters in the real n vector $\boldsymbol{\rho}$ are used to model the uncertainty in the array steering matrix. It is assumed that $\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho})$ is known for a nominal value $\boldsymbol{\rho} = \boldsymbol{\rho}_0$ and that the columns in $\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho}_0)$ are linearly independent as long as $\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_j$, $i \neq j$. The model for $\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho}_0)$ can be obtained, for example, by physical insight, or it could be the result of a calibration experiment. We may then have knowledge about the sensitivity of the nominal response to certain variations in $\boldsymbol{\rho}$, which can be modeled by considering $\boldsymbol{\rho}$ as a random vector. The *a priori* information is then in the form of a probability distribution that can be used in the design of an estimation algorithm to make it robust with regard to the model errors.

In this paper, we will use a stochastic model for the signals; more precisely, $\mathbf{s}(t)$ is considered to be a zero-mean Gaussian random vector with second moments

$$E\{\mathbf{s}(t)\mathbf{s}^*(s)\} = \mathbf{P}\delta_{t,s}, \quad E\{\mathbf{s}(t)\mathbf{s}^T(s)\} = \mathbf{0}$$

where

- $(\cdot)^*$ complex conjugate transpose;
- $(\cdot)^T$ transpose;
- $\delta_{t,s}$ Kronecker delta.

Let d' denote the rank of the signal covariance matrix $\mathbf{P}$. Special attention will be given to cases where $d' < d$, i.e., cases in which the signals may be fully correlated (coherent). In particular, we show that the proposed method is consistent even under such severe conditions. The noise is modeled as a zero-mean spatially and temporally white complex Gaussian vector with second-order moments

$$E\{\mathbf{n}(t)\mathbf{n}^*(s)\} = \sigma^2\mathbf{I}\delta_{t,s}, \quad E\{\mathbf{n}(t)\mathbf{n}^T(s)\} = \mathbf{0}.$$

The perturbation parameter vector $\boldsymbol{\rho}$ is modeled as a Gaussian random variable with mean $E\{\boldsymbol{\rho}\} = \boldsymbol{\rho}_0$ and covariance

$$E\{(\boldsymbol{\rho} - \boldsymbol{\rho}_0)(\boldsymbol{\rho} - \boldsymbol{\rho}_0)^T\} = \boldsymbol{\Omega}.$$

It is assumed that both $\boldsymbol{\rho}_0$ and $\boldsymbol{\Omega}$ are known. Similar to [9] and [24], we consider small perturbations in $\boldsymbol{\rho}$ and, more specifically, assume that the effect of the array errors on the estimates of $\boldsymbol{\theta}$ is of comparable size with those due to the finite sample effects of the noise. To model this, we assume that

$$\boldsymbol{\Omega} = \overline{\boldsymbol{\Omega}}/N$$

where N is the number of samples, and $\overline{\boldsymbol{\Omega}}$ is independent of N. This somewhat artificial assumption concerning the model errors is made only for convenience in showing the statistical optimality of the algorithm presented later. An identical result can be obtained for small $\boldsymbol{\rho} - \boldsymbol{\rho}_0$ if a first-order perturbation analysis is used, but this approach is somewhat less elegant. The assumption that $\boldsymbol{\rho} - \boldsymbol{\rho}_0 = O_p(1/\sqrt{N})$

implies that the calibration errors and finite sample effects are of comparable size. [Hereafter, we will use $O_p(\cdot)$ and $o_p(\cdot)$ to denote the probability version of the corresponding deterministic notation.] Despite this fact, we will see that the resulting algorithm is also optimal for the following limiting cases:

- $N \to \infty$ for fixed $(\boldsymbol{\rho} - \boldsymbol{\rho}_0)$—(model errors only);
- $(\boldsymbol{\rho} - \boldsymbol{\rho}_0) \to \mathbf{0}$ for fixed N—(finite sample errors only).

In other words, for each of the above cases, the proposed technique converges to an algorithm that has previously been shown to be statistically efficient for that case. We can thus conclude that the algorithm is globally optimal and not just under the assumption $\boldsymbol{\rho} - \boldsymbol{\rho}_0 = O_p(1/\sqrt{N})$.

III. Robust Estimation

This section introduces some of the previous approaches taken to solve the problem considered in this paper. The Cramér–Rao lower bound is also restated from [24].

A. MAP

As the problem has been formulated, it is natural to use a maximum *a posteriori* (MAP) estimator. Following [22], the cost function to be minimized for simultaneous estimation of signal and perturbation parameters is

$$V_{\mathrm{MAP}}(\boldsymbol{\theta}, \boldsymbol{\rho}) = V_{\mathrm{ML}}(\boldsymbol{\theta}, \boldsymbol{\rho}) + \tfrac{1}{2}(\boldsymbol{\rho} - \boldsymbol{\rho}_0)^T \boldsymbol{\Omega}^{-1}(\boldsymbol{\rho} - \boldsymbol{\rho}_0).$$

$V_{\mathrm{ML}}(\boldsymbol{\theta}, \boldsymbol{\rho})$ is the concentrated negative log-likelihood function and is given by (see, e.g., [36], [37])

$$V_{\mathrm{ML}}(\boldsymbol{\theta}, \boldsymbol{\rho}) = N \log \det\{\mathbf{A}\hat{\mathbf{P}}\mathbf{A}^* + \hat{\sigma}^2\mathbf{I}\}$$

where, for brevity, we have suppressed the arguments and where constant terms have been neglected. Here, $\hat{\mathbf{P}} = \hat{\mathbf{P}}(\boldsymbol{\theta}, \boldsymbol{\rho})$ and $\hat{\sigma}^2 = \hat{\sigma}^2(\boldsymbol{\theta}, \boldsymbol{\rho})$ denote the maximum likelihood estimates of $\mathbf{P}$ and σ^2, respectively. They are explicitly given by

$$\begin{aligned} \hat{\sigma}^2(\boldsymbol{\theta}, \boldsymbol{\rho}) &= \frac{1}{m-d} \operatorname{Tr}\{\boldsymbol{\Pi}^{\perp}_{\mathbf{A}}(\boldsymbol{\theta}, \boldsymbol{\rho})\hat{\mathbf{R}}\} \\ \hat{\mathbf{P}}(\boldsymbol{\theta}, \boldsymbol{\rho}) &= \mathbf{A}^{\dagger}(\boldsymbol{\theta}, \boldsymbol{\rho})(\hat{\mathbf{R}} - \hat{\sigma}^2(\boldsymbol{\theta}, \boldsymbol{\rho})\mathbf{I})\mathbf{A}^{\dagger^*}(\boldsymbol{\theta}, \boldsymbol{\rho}) \end{aligned} \tag{1}$$

where

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}(t)\mathbf{x}^*(t)$$

$\mathbf{A}^{\dagger} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$, and $\boldsymbol{\Pi}^{\perp}_{\mathbf{A}} = \mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}$.

For large N, it is further argued in [22] that $V_{\mathrm{ML}}(\boldsymbol{\theta}, \boldsymbol{\rho})$ can be replaced with any other cost function that has the same large sample behavior as ML. In particular, it was proposed that the weighted subspace fitting (WSF) cost function [34] be used. This results in

$$\begin{aligned} V_{\text{MAP-WSF}}(\boldsymbol{\theta}, \boldsymbol{\rho}) = \frac{N}{\hat{\sigma}^2} V_{\mathrm{WSF}}(\boldsymbol{\theta}, \boldsymbol{\rho}) \\ + \tfrac{1}{2}(\boldsymbol{\rho} - \boldsymbol{\rho}_0)^T \boldsymbol{\Omega}^{-1}(\boldsymbol{\rho} - \boldsymbol{\rho}_0) \end{aligned} \tag{2}$$

where

$$V_{\mathrm{WSF}}(\boldsymbol{\theta}, \boldsymbol{\rho}) = \operatorname{Tr}\{\boldsymbol{\Pi}^{\perp}_{\mathbf{A}}(\boldsymbol{\theta}, \boldsymbol{\rho})\hat{\mathbf{E}}_s \hat{\mathbf{W}}_{\mathrm{WSF}} \hat{\mathbf{E}}_s^*\}. \tag{3}$$

Here, $\hat{\mathbf{E}}_s$ and $\hat{\mathbf{W}}_{\mathrm{WSF}}$ are defined via the eigendecomposition of $\hat{\mathbf{R}}$

$$\hat{\mathbf{R}} = \hat{\mathbf{E}}_s \hat{\boldsymbol{\Lambda}}_s \hat{\mathbf{E}}_s^* + \hat{\mathbf{E}}_n \hat{\boldsymbol{\Lambda}}_n \hat{\mathbf{E}}_n^* \tag{4}$$

where $\hat{\boldsymbol{\Lambda}}_s$ is a diagonal matrix containing the d' largest eigenvalues, and $\hat{\mathbf{E}}_s$ is composed of the corresponding eigenvectors. The WSF weighting matrix is defined as $\hat{\mathbf{W}}_{\mathrm{WSF}} = \hat{\tilde{\boldsymbol{\Lambda}}}^2 \hat{\boldsymbol{\Lambda}}_s^{-1}$, where $\hat{\tilde{\boldsymbol{\Lambda}}} = \hat{\boldsymbol{\Lambda}}_s - \hat{\sigma}^2\mathbf{I}$, and $\hat{\sigma}^2$ is a consistent estimate of σ^2 (for example, the average of the $m - d'$ smallest eigenvalues of $\hat{\mathbf{R}}$). The minimization of either V_{MAP} or $V_{\text{MAP-WSF}}$ is over both $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$. The idea of the methods presented in the following two sections is to eliminate the need for an explicit minimization over $\boldsymbol{\rho}$.

B. MAPprox

In [22], it is suggested that the minimization of $V_{\text{MAP-WSF}}$ over $\boldsymbol{\rho}$ can be done implicitly. Assuming that $\boldsymbol{\rho} - \boldsymbol{\rho}_0$ is small, a second-order Taylor expansion of $V_{\text{MAP-WSF}}$ around $\boldsymbol{\rho}_0$ is performed, and then, the result is analytically minimized with regard to $\boldsymbol{\rho}$. This leads to a concentrated form of the cost function

$$\begin{aligned} &V_{\mathrm{MAPprox}}(\boldsymbol{\theta}) = \\ &\frac{N}{\hat{\sigma}^2}\left\{V_{\mathrm{WSF}} - \frac{1}{2}\partial_{\boldsymbol{\rho}} V_{\mathrm{WSF}}^T \left(\partial_{\boldsymbol{\rho}\boldsymbol{\rho}} V_{\mathrm{WSF}} + \frac{\hat{\sigma}^2}{N}\boldsymbol{\Omega}^{-1}\right)^{-1} \partial_{\boldsymbol{\rho}} V_{\mathrm{WSF}}\right\} \end{aligned} \tag{5}$$

where $\partial_{\boldsymbol{\rho}} V_{\mathrm{WSF}}$ denotes the gradient of V_{WSF} with respect to $\boldsymbol{\rho}$, and $\partial_{\boldsymbol{\rho}\boldsymbol{\rho}} V_{\mathrm{WSF}}$ denotes the Hessian. The right-hand side of (5) is evaluated at $\boldsymbol{\rho}_0$ and $\boldsymbol{\theta}$. The method is referred to as *MAPprox*. In Section VII, we analyze the asymptotic properties of the MAPprox estimates and point out its relationship to the GWSF estimator derived in Section IV.

C. MAP-NSF

Another method for approximating the MAP estimator is the MAP-NSF approach of [24]. In MAP-NSF, we use the fact that the noise subspace fitting (NSF) method is also asymptotically equivalent to ML [32]. Thus, in lieu of using the WSF criterion in (2), the NSF cost function

$$V_{\mathrm{NSF}} = \operatorname{Tr}\{\mathbf{A}^*\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^*\mathbf{A}\hat{\mathbf{U}}\} \tag{6}$$

is employed. Here, $\hat{\mathbf{U}}$ is a consistent estimate of the matrix

$$\mathbf{U} = \mathbf{A}^{\dagger}\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^2\boldsymbol{\Lambda}_s^{-1}\mathbf{E}_s^*\mathbf{A}^{\dagger^*}.$$

Assuming that $\boldsymbol{\rho} - \boldsymbol{\rho}_0$ is small, we can perform a first-order Taylor expansion of $\mathbf{A}$ around $\boldsymbol{\rho}_0$

$$\mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho}) \triangleq \operatorname{vec}(\mathbf{A}) \approx \mathbf{a}_0 + \mathbf{D}_{\boldsymbol{\rho}}(\boldsymbol{\rho} - \boldsymbol{\rho}_0) \tag{7}$$

where $\operatorname{vec}(\cdot)$ is the vectorization operation, $\mathbf{a}_0 = \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho}_0)$ and

$$\mathbf{D}_{\boldsymbol{\rho}} = \left[\frac{\partial \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\rho}_1} \;\cdots\; \frac{\partial \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\rho}_n}\right]\Bigg|_{\boldsymbol{\theta}, \boldsymbol{\rho}_0}. \tag{8}$$

Inserting (7) in (6), the criterion becomes quadratic in $\boldsymbol{\rho}$ and can thus, for fixed $\boldsymbol{\theta}$, be explicitly minimized with respect to $\boldsymbol{\rho}$. This results in the MAP-NSF cost function (see [24])

$$V_{\text{MAP-NSF}}(\boldsymbol{\theta}) = \mathbf{a}_0^*\hat{\mathbf{M}}\mathbf{a}_0 - \hat{\mathbf{f}}^T\hat{\boldsymbol{\Gamma}}^{-1}\hat{\mathbf{f}} \tag{9}$$

where (below, $\otimes$ denotes the Kronecker product)

$$\hat{\mathbf{M}} = \hat{\mathbf{U}}^T \otimes (\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^*) \tag{10}$$

$$\hat{\mathbf{f}} = \text{Re}\{\mathbf{D}_{\boldsymbol{\rho}}^*\hat{\mathbf{M}}\mathbf{a}_0\}$$

$$\hat{\boldsymbol{\Gamma}} = \text{Re}\left\{\mathbf{D}_{\boldsymbol{\rho}}^*\hat{\mathbf{M}}\mathbf{D}_{\boldsymbol{\rho}} + \frac{\hat{\sigma}^2}{2}\,\overline{\boldsymbol{\Omega}}^{-1}\right\}. \tag{11}$$

In [24], the statistical properties of the MAP-NSF estimate of $\boldsymbol{\theta}$ are analyzed. It is shown that MAP-NSF is a consistent and asymptotically efficient estimator if $d' = d$. This means that the asymptotic covariance matrix of the estimation error in $\boldsymbol{\theta}$ is equal to the Cramér–Rao lower bound given in the next section.

D. Cramér–Rao Lower Bound

In [22] and [24], an asymptotically valid Cramér–Rao lower bound is derived for the problem of interest herein. Below, we give the lower bound on the signal parameters only.

Theorem 1: Assuming that $\hat{\boldsymbol{\theta}}$ is an asymptotically unbiased estimate of $\boldsymbol{\theta}_0$, then for large N

$$E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\} \geq \mathbf{CRB}_{\boldsymbol{\theta}} \triangleq \frac{\sigma^2}{2N}\left[\mathbf{C} - \mathbf{F}_{\boldsymbol{\theta}}^T\boldsymbol{\Gamma}^{-1}\mathbf{F}_{\boldsymbol{\theta}}\right]^{-1} \tag{12}$$

where[1]

$$\mathbf{C} = \text{Re}\{\mathbf{D}_{\boldsymbol{\theta}}^*\mathbf{M}\mathbf{D}_{\boldsymbol{\theta}}\}$$

$$\mathbf{M} = \mathbf{U}^T \otimes \boldsymbol{\Pi}_{\mathbf{A}}^{\perp} \tag{13}$$

$$\mathbf{D}_{\boldsymbol{\theta}} = \left[\frac{\partial \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\theta}_1} \quad \cdots \quad \frac{\partial \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\theta}_d}\right],$$

$$\mathbf{F}_{\boldsymbol{\theta}} = \text{Re}\{\mathbf{D}_{\boldsymbol{\rho}}^*\mathbf{M}\mathbf{D}_{\boldsymbol{\theta}}\} \tag{14}$$

$$\boldsymbol{\Gamma} = \text{Re}\left\{\mathbf{D}_{\boldsymbol{\rho}}^*\mathbf{M}\mathbf{D}_{\boldsymbol{\rho}} + \frac{\sigma^2}{2}\,\overline{\boldsymbol{\Omega}}^{-1}\right\}. \tag{15}$$

The above expressions are evaluated at $\boldsymbol{\theta}_0$ and $\boldsymbol{\rho}_0$.

Proof: See [22] and [24]. ∎

It is important to note that the CRB for the case with no calibration errors is $\sigma^2\mathbf{C}^{-1}/2N$ and clearly is a lower bound for the MAP-CRB.

IV. Generalized Weighted Subspace Fitting (GWSF)

It was shown in [24] that the MAP-NSF method is asymptotically statistically efficient. One drawback with MAP-NSF is that it is not consistent if the sources are fully correlated, that is, if $d' < d$. This problem can be overcome in the *signal* subspace formulation. In this section, we derive a new method in the signal subspace fitting (SSF) class of algorithms. The method will be shown to possess the same large sample performance as MAP-NSF and is thus also asymptotically efficient. A further advantage of the proposed SSF approach is that the cost function depends on the parameters in a relatively simple manner and, if the nominal array is uniform and linear, the nonlinear minimization can be replaced by a rooting technique (see Section VI).

According to the problem formulation, $\hat{\mathbf{R}} - \mathbf{R} = O_p(1/\sqrt{N})$, where

$$\mathbf{R} = E\{\mathbf{x}(t)\mathbf{x}^*(t)\} = \mathbf{A}_0\mathbf{P}\mathbf{A}_0^* + \sigma^2\mathbf{I}$$

and $\mathbf{A}_0 = \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)$. This implies that $\hat{\mathbf{E}}_s \to \mathbf{E}_s = \mathbf{A}_0\mathbf{T}$ at the same rate, where $\mathbf{T}$ is a $d \times d'$ full rank matrix. The idea of the proposed method is to minimize a suitable norm of the residuals

$$\boldsymbol{\varepsilon} = \text{vec}(\mathbf{B}^*(\boldsymbol{\theta})\hat{\mathbf{E}}_s)$$

where $\mathbf{B}(\boldsymbol{\theta})$ is an $m \times (m-d)$ full rank matrix whose columns span the null space of $\mathbf{A}^*(\boldsymbol{\theta})$. This implies that $\mathbf{B}^*(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) = \mathbf{0}$, and $\mathbf{B}^*(\boldsymbol{\theta}_0)\mathbf{E}_s = \mathbf{0}$.

Remark 1: For general arrays, there is no known closed-form parameterization of $\mathbf{B}$ in terms of $\boldsymbol{\theta}$. However, this will not introduce any problems since, as we show in Section VI, the final criterion can be written as an explicit function of $\boldsymbol{\theta}$.

For general array error models, we have to consider the real and imaginary parts of $\boldsymbol{\varepsilon}$ separately to get optimal (minimum variance) performance. Equivalently, we may study $\boldsymbol{\varepsilon}$ and its complex conjugate, which we denote by $\boldsymbol{\varepsilon}^c$. We believe that the analysis of the method is clearer if $\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}^c$ are used instead of the real and imaginary parts. However, from an implementational point of view, it may be more suitable to use real arithmetic. Before presenting the estimation criterion, we define

$$\hat{\mathbf{e}}_s = \begin{bmatrix}\text{vec}(\hat{\mathbf{E}}_s) \\ \text{vec}(\hat{\mathbf{E}}_s^c)\end{bmatrix}$$

$$\overline{\mathbf{A}} = \begin{bmatrix}(\mathbf{I}_{d'} \otimes \mathbf{A}) & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_{d'} \otimes \mathbf{A}^c)\end{bmatrix}$$

$$\overline{\mathbf{B}} = \begin{bmatrix}(\mathbf{I}_{d'} \otimes \mathbf{B}) & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_{d'} \otimes \mathbf{B}^c)\end{bmatrix}$$

$$\overline{\boldsymbol{\varepsilon}} = \begin{bmatrix}\boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}^c\end{bmatrix} = \overline{\mathbf{B}}^*\hat{\mathbf{e}}_s.$$

Notice that the columns of $\overline{\mathbf{B}}$ span the null space of $\overline{\mathbf{A}}^*$.

We propose to estimate the signal parameters $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} V_{\text{GWSF}}(\boldsymbol{\theta}) \tag{16}$$

$$V_{\text{GWSF}}(\boldsymbol{\theta}) = \overline{\boldsymbol{\varepsilon}}^*(\boldsymbol{\theta})\mathbf{W}\overline{\boldsymbol{\varepsilon}}(\boldsymbol{\theta}) \tag{17}$$

where $\mathbf{W}$ is a positive definite weighting matrix. The method will in the sequel be referred to as the *generalized weighted subspace fitting (GWSF)* method. To derive the weighting that leads to minimum variance estimates of $\boldsymbol{\theta}$, we need to compute the residual covariance matrix. As shown in Appendix A, the (asymptotic) second-order moment of the residual vector $\overline{\boldsymbol{\varepsilon}}$ at $\boldsymbol{\theta}_0$

$$\mathbf{C}_{\overline{\boldsymbol{\varepsilon}}} = \lim_{N\to\infty} NE\{\overline{\boldsymbol{\varepsilon}}\overline{\boldsymbol{\varepsilon}}^*\} \tag{18}$$

can be written as

$$\mathbf{C}_{\overline{\boldsymbol{\varepsilon}}} = \overline{\mathbf{L}} + \overline{\mathbf{G}\mathbf{G}}^* \tag{19}$$

[1] Notice that $\mathbf{M}$ is the limit of $\hat{\mathbf{M}}$ defined in (10) if $d' = d$.

where we have defined

$$\overline{\mathbf{L}} = \begin{bmatrix} \mathbf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}^c \end{bmatrix}, \qquad \mathbf{L} = (\sigma^2 \tilde{\mathbf{\Lambda}}^{-2} \mathbf{\Lambda}_s \otimes \mathbf{B}^* \mathbf{B}) \tag{20}$$

$$\overline{\mathbf{G}} = \begin{bmatrix} \mathbf{G} \\ \mathbf{G}^c \end{bmatrix}, \qquad \mathbf{G} = (\mathbf{T}^T \otimes \mathbf{B}^*) \mathbf{D}_{\boldsymbol{\rho}} \overline{\mathbf{\Omega}}^{1/2}. \tag{21}$$

Here, $\overline{\mathbf{\Omega}}^{1/2}$ is a (symmetric) square root of $\overline{\mathbf{\Omega}}$, and $\mathbf{T} = \mathbf{A}_0^{\dagger} \mathbf{E}_s$. It is easy to see that $\mathbf{C}_{\overline{\varepsilon}}$ is positive definite since $\mathbf{L}$ is. It is then well known that (in the class of estimators based on $\overline{\varepsilon}$) the optimal choice of the weighting in terms of minimizing the parameter estimation error variance is

$$\mathbf{W}_{\mathrm{GWSF}} = \mathbf{C}_{\overline{\varepsilon}}^{-1} \tag{22}$$

(see, e.g., [38]). In the next section, it is proven that this weighting in fact also gives asymptotically (in N) minimum variance estimates (in the class of unbiased estimators) for the estimation problem under consideration. The implementation of GWSF is discussed in Section VI. We conclude this section with two remarks regarding the GWSF formulation.

Remark 2: If there are no model errors, then GWSF reduces to the WSF estimator (3). This can easily be verified by setting $\overline{\mathbf{\Omega}} = \mathbf{0}$ and rewriting the GWSF criterion (17).

Remark 3: For the case of model errors only ($N \to \infty$ or $\sigma^2 \to 0$), GWSF becomes the "model errors only" algorithm [7]. The results obtained for GWSF are also consistent with weightings given in [9] for special array error models.

Thus, the GWSF method can be considered to be optimal in general and not just when the model errors and the finite sample effects are of the same order.

V. Performance Analysis

The asymptotic properties of the GWSF estimates are analyzed in this section. It is shown that the estimates are consistent and have a limiting Gaussian distribution. It is also shown that the asymptotic covariance matrix of the estimation error is equal to the CRB.

A. Consistency

The cost function (17) converges uniformly in $\boldsymbol{\theta}$ with probability one (w.p. 1) to the limit

$$\overline{V}_{\mathrm{GWSF}}(\boldsymbol{\theta}) = \begin{bmatrix} \mathrm{vec}(\mathbf{B}^*(\boldsymbol{\theta})\mathbf{E}_s) \\ \mathrm{vec}^c(\mathbf{B}^*(\boldsymbol{\theta})\mathbf{E}_s) \end{bmatrix}^* \mathbf{W} \begin{bmatrix} \mathrm{vec}(\mathbf{B}^*(\boldsymbol{\theta})\mathbf{E}_s) \\ \mathrm{vec}^c(\mathbf{B}^*(\boldsymbol{\theta})\mathbf{E}_s) \end{bmatrix}. \tag{23}$$

This implies that $\hat{\boldsymbol{\theta}}$ converges to the minimizing argument of (23). Since $\mathbf{W}$ is positive definite, $\overline{V}_{\mathrm{GWSF}}(\boldsymbol{\theta}) \geq 0$ with equality if and only if $\mathbf{B}^*(\boldsymbol{\theta})\mathbf{E}_s = \mathbf{0}$. This implies that $\mathbf{E}_s = \mathbf{A}(\boldsymbol{\theta})\mathbf{T}_2$ for some full rank $d \times d'$ matrix $\mathbf{T}_2$. However, since $\mathbf{E}_s = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{T}$ by the definition of $\mathbf{E}_s$, it follows from [39] that if $d < (m + d')/2$, then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is the only solution to $\mathbf{B}^*(\boldsymbol{\theta})\mathbf{E}_s = \mathbf{0}$. We have thus proven that $\hat{\boldsymbol{\theta}}$ in (16) tends to $\boldsymbol{\theta}_0$ w.p. 1 as $N \to \infty$.

B. Asymptotic Distribution

As already mentioned, there exists a certain weighting matrix (22) that minimizes the estimation error variance. This matrix depends on the true parameters and, in practice, has to be replaced with a consistent estimate. However, this will not change the second-order asymptotic properties considered herein. In fact, it is easy to see below that the derivative of the cost function with respect to $\mathbf{W}$ only leads to higher order terms. Thus, whether $\mathbf{W}$ is considered to be a fixed matrix or a consistent estimate thereof does not matter for the analysis in this section.

The estimate that minimizes (17) satisfies by definition $V'(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, and since $\hat{\boldsymbol{\theta}}$ is consistent, a Taylor series expansion of $V'(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$ leads to

$$\tilde{\boldsymbol{\theta}} = -\mathbf{H}^{-1} V' + o_p(1/\sqrt{N}) \tag{24}$$

where $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$, $\mathbf{H}$ is the limiting Hessian, and V' is the gradient of $V(\boldsymbol{\theta})$. Both V' and $\mathbf{H}$ are evaluated at $\boldsymbol{\theta}_0$. In the following, let the index i denote the partial derivative with respect to $\boldsymbol{\theta}_i$. The derivative of $V(\boldsymbol{\theta})$ is given by

$$V_i' = \overline{\boldsymbol{\varepsilon}}_i^* \mathbf{W} \overline{\boldsymbol{\varepsilon}} + \overline{\boldsymbol{\varepsilon}}^* \mathbf{W} \overline{\boldsymbol{\varepsilon}}_i = 2\,\mathrm{Re}\{\overline{\boldsymbol{\varepsilon}}_i^* \mathbf{W} \overline{\boldsymbol{\varepsilon}}\}. \tag{25}$$

It is straightforward to verify that $\overline{\boldsymbol{\varepsilon}}_i^* \mathbf{W} \overline{\boldsymbol{\varepsilon}}$ is real for any weighting matrix having the form

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12}^c & \mathbf{W}_{11}^c \end{bmatrix}. \tag{26}$$

In particular, the matrix $\mathbf{W}_{\mathrm{GWSF}}$ in (22) has this structure. Next, notice that $\overline{\boldsymbol{\varepsilon}} = \overline{\mathbf{B}}^* \hat{\mathbf{e}}_s$ and

$$\overline{\boldsymbol{\varepsilon}}_i = \overline{\mathbf{B}}_i^* \hat{\mathbf{e}}_s = \overline{\mathbf{B}}_i^* \mathbf{e}_s + o_p(1) \tag{27}$$

where $\mathbf{e}_s$ denotes the limit of $\hat{\mathbf{e}}_s$. Define the vector $\mathbf{k}_i = \overline{\mathbf{B}}_i^* \mathbf{e}_s$ and insert (27) in (25) to show that

$$V_i' \simeq 2\mathbf{k}_i^* \mathbf{W} \overline{\boldsymbol{\varepsilon}} \tag{28}$$

for any $\mathbf{W}$ satisfying (26), where $\simeq$ represents an equality up to first order, that is, terms of order $o_p(1/\sqrt{N})$ are omitted. Recall (51) and that $\sqrt{N}(\hat{\mathbf{R}} - \mathbf{R})$ is asymptotically Gaussian by the central limit theorem. By (24) and (28), we thus conclude that the normalized estimation error has a limiting Gaussian distribution as well. In summary, we have the following result.

Theorem 2: If $d < (m + d')/2$, then the GWSF estimate (16) tends to $\boldsymbol{\theta}_0$ w.p. 1 as $N \to \infty$, and

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0},\, \mathbf{H}^{-1} \mathbf{Q} \mathbf{H}^{-1})$$

where

$$\mathbf{Q} = \lim_{N \to \infty} N E\{V' V'^T\} \tag{29}$$

and where $\xrightarrow{d} \mathcal{N}(\cdot)$ denotes convergence to a Gaussian distribution. If $\mathbf{W} = \mathbf{W}_{\mathrm{GWSF}}$ given in (22), then

$$\mathbf{H}^{-1} \mathbf{Q} \mathbf{H}^{-1} = N \mathbf{CRB}_{\boldsymbol{\theta}}.$$

Expressions for $\mathbf{H}$ and $\mathbf{Q}$ can be found in Appendix B, and $\mathbf{CRB}_{\boldsymbol{\theta}}$ is given in (12). All quantities are evaluated at $\boldsymbol{\theta}_0$.

Proof: The result follows from the analysis above and the calculations in Appendix B. ■

This result is similar to what was derived for the MAP-NSF method in [24]. It implies that GWSF and MAP-NSF are asymptotically equivalent and efficient for large N and small $\mathbf{\Omega}\ (=\overline{\mathbf{\Omega}}/N)$. An advantage of GWSF, as compared with MAP-NSF, is that GWSF is consistent even if $\mathbf{P}$ is rank deficient ($d' < d$). In short, GWSF is a generalization of MODE [33] and WSF [34] that allows us to include *a priori* knowledge of the array perturbations into the estimation criterion.

VI. Implementation

We will, in this section, discuss in some detail the implementation of the GWSF estimator. Recall that the GWSF criterion is written in terms of the matrix $\overline{\mathbf{B}}(\boldsymbol{\theta})$ and that the parameterization of $\overline{\mathbf{B}}$, in general, is not known. In this section, we first rewrite the GWSF criterion so that it becomes explicitly parameterized by $\boldsymbol{\theta}$. This function can then be minimized numerically to give the GWSF estimate of $\boldsymbol{\theta}$. Next, in Section VI-B, we present the implementation for the common special case when the nominal array is uniform and linear. For this case, it is possible to implement GWSF without the need for gradient search methods.

A. Implementation for General Arrays

The GWSF cost function in (17) reads

$$V_{\mathrm{GWSF}}(\boldsymbol{\theta}) = \bar{\boldsymbol{\varepsilon}}^*(\boldsymbol{\theta})\mathbf{W}\bar{\boldsymbol{\varepsilon}}(\boldsymbol{\theta}) = \hat{\mathbf{e}}_s{}^*\overline{\mathbf{B}}(\boldsymbol{\theta})\mathbf{W}\overline{\mathbf{B}}^*(\boldsymbol{\theta})\hat{\mathbf{e}}_s. \quad (30)$$

What complicates matters is that the functional form of $\overline{\mathbf{B}}(\theta)$ is not known for general array geometries. However, if the optimal weighting in (22) is used in (30), then it is possible to rewrite (30) so that it depends explicitly on $\boldsymbol{\theta}$. To see how this can be done, we will use the following lemma.

Lemma 1: Let $\mathbf{X}\ (m\times n)$ and $\mathbf{Y}\ (m\times(m-n))$ be two full column rank matrices. Furthermore, assume that $\mathbf{X}^*\mathbf{Y} = \mathbf{0}$, and let $\mathbf{C}$ and $\mathbf{D}$ be two nonsingular matrices. Then

$$\mathbf{X}\mathbf{C}^{-1}\mathbf{X}^* = \mathbf{\Pi}_{\mathbf{X}}(\mathbf{X}^{\dagger^*}\mathbf{C}\mathbf{X}^\dagger + \mathbf{Y}^{\dagger^*}\mathbf{D}\mathbf{Y}^\dagger)^{-1}\mathbf{\Pi}_{\mathbf{X}}$$

where $\mathbf{\Pi}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*$ is the orthogonal projection onto the range-space of $\mathbf{X}$.

Proof: Note that $\mathbf{\Pi}_{\mathbf{X}}\mathbf{Y} = \mathbf{0}$. The lemma is proven by

$$\begin{aligned}
\mathbf{X}\mathbf{C}^{-1}\mathbf{X}^* &= \mathbf{\Pi}_{\mathbf{X}}\mathbf{X}\mathbf{C}^{-1}\mathbf{X}^*\mathbf{\Pi}_{\mathbf{X}} \\
&= \mathbf{\Pi}_{\mathbf{X}}[\mathbf{X}\quad\mathbf{Y}]\begin{bmatrix}\mathbf{C}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{X}^*\\ \mathbf{Y}^*\end{bmatrix}\mathbf{\Pi}_{\mathbf{X}} \\
&= \mathbf{\Pi}_{\mathbf{X}}\left(\begin{bmatrix}\mathbf{X}^*\\ \mathbf{Y}^*\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{bmatrix}[\mathbf{X}\quad\mathbf{Y}]^{-1}\right)^{-1}\mathbf{\Pi}_{\mathbf{X}} \\
&= \mathbf{\Pi}_{\mathbf{X}}\left([\mathbf{X}^{\dagger^*}\mathbf{Y}^{\dagger^*}]\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{X}^\dagger\\ \mathbf{Y}^\dagger\end{bmatrix}\right)^{-1}\mathbf{\Pi}_{\mathbf{X}} \\
&= \mathbf{\Pi}_{\mathbf{X}}(\mathbf{X}^{\dagger^*}\mathbf{C}\mathbf{X}^\dagger + \mathbf{Y}^{\dagger^*}\mathbf{D}\mathbf{Y}^\dagger)^{-1}\mathbf{\Pi}_{\mathbf{X}}.
\end{aligned}$$ ■

Now apply Lemma 1 to the GWSF criterion (30). With reference to Lemma 1, take $\mathbf{X} = \overline{\mathbf{B}}$, $\mathbf{Y} = \overline{\mathbf{A}}$, $\mathbf{C} = \mathbf{W}^{-1}$ and, for example, $\mathbf{D} = (\overline{\mathbf{A}}^*\overline{\mathbf{A}})^2$ to show that

$$V_{\mathrm{GWSF}}(\boldsymbol{\theta}) = \hat{\mathbf{e}}_s^*\mathbf{\Pi}_{\overline{\mathbf{A}}}^{\perp}\tilde{\mathbf{W}}\mathbf{\Pi}_{\overline{\mathbf{A}}}^{\perp}\hat{\mathbf{e}}_s \quad (31)$$

where

$$\tilde{\mathbf{W}} = (\overline{\mathbf{B}}^{\dagger^*}\mathbf{W}^{-1}\overline{\mathbf{B}}^\dagger + \overline{\mathbf{A}\mathbf{A}}^*)^{-1}. \quad (32)$$

In (31), we used the fact that $\mathbf{\Pi}_{\overline{\mathbf{B}}} = \mathbf{\Pi}_{\overline{\mathbf{A}}}^{\perp}$. Let $\mathbf{W}$ be chosen according to (22). It is then readily verified that

$$\begin{aligned}
\overline{\mathbf{B}}^{\dagger^*}\mathbf{W}^{-1}\overline{\mathbf{B}}^\dagger &= \overline{\mathbf{B}}^{\dagger^*}\mathbf{C}_{\bar{\boldsymbol{\varepsilon}}}\overline{\mathbf{B}}^\dagger \\
&= \begin{bmatrix}\tilde{\mathbf{L}} & \mathbf{0}\\ \mathbf{0} & \tilde{\mathbf{L}}^c\end{bmatrix} + \begin{bmatrix}\tilde{\mathbf{G}}\\ \tilde{\mathbf{G}}^c\end{bmatrix}[\tilde{\mathbf{G}}^* \quad \tilde{\mathbf{G}}^T]
\end{aligned} \quad (33)$$

where

$$\tilde{\mathbf{L}} = (\sigma^2\tilde{\mathbf{\Lambda}}^{-2}\mathbf{\Lambda}_s \otimes \mathbf{\Pi}_{\mathbf{A}}^{\perp}), \quad \tilde{\mathbf{G}} = (\mathbf{T}^T \otimes \mathbf{\Pi}_{\mathbf{A}}^{\perp})\mathbf{D}_{\boldsymbol{\rho}}\overline{\mathbf{\Omega}}^{1/2}.$$

Combining (31)–(33), we see that the criterion is now explicitly parameterized by $\boldsymbol{\theta}$. The GWSF algorithm for general array geometries can be summarized as follows.

1) Based on $\hat{\mathbf{R}}$, compute $\hat{\mathbf{E}}_s$, $\hat{\mathbf{\Lambda}}_s$, and
$$\hat{\sigma}^2 = \frac{1}{m-d'}\{\mathrm{Tr}(\hat{\mathbf{R}}) - \mathrm{Tr}(\hat{\mathbf{\Lambda}}_s)\}, \qquad \hat{\tilde{\mathbf{\Lambda}}} = \hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2\mathbf{I}.$$
2) The GWSF estimate of $\boldsymbol{\theta}$ is then given by
$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \hat{\mathbf{e}}_s^*\mathbf{\Pi}_{\overline{\mathbf{A}}}^{\perp}\hat{\tilde{\mathbf{W}}}\mathbf{\Pi}_{\overline{\mathbf{A}}}^{\perp}\hat{\mathbf{e}}_s$$
where $\hat{\tilde{\mathbf{W}}}$ is given by (32) and (33) with the sample estimates from the first step inserted in place of the true quantities.

This results in a nonlinear optimization problem where $\boldsymbol{\theta}$ enters in a complicated way. From the analysis in Section V, it follows that the asymptotic second-order properties of $\hat{\boldsymbol{\theta}}$ are not affected if the weighting matrix $\tilde{\mathbf{W}}$ is replaced by a consistent estimate. Hence, if a consistent estimate of $\boldsymbol{\theta}_0$ is used to compute $\tilde{\mathbf{W}}$, then the criterion in the second step only depends on $\boldsymbol{\theta}$ via $\mathbf{\Pi}_{\overline{\mathbf{A}}}^{\perp}$. This may lead to some computational savings, but it still results in a nonlinear optimization problem. In the next section, we turn to the common special case of a uniform linear array, for which it is possible to avoid gradient search methods.

B. Implementation for Uniform Linear Arrays

In this section, we present a way to significantly simplify the implementation of GWSF if the nominal array is linear and uniform. The form of the estimator is in the spirit of IQML [40] and MODE [33], [41]. Define the following polynomial with d roots on the unit circle corresponding to the DOAs

$$b_0z^d + b_1z^{d-1} + \cdots + b_d = b_0\prod_{k=1}^{d}(z - e^{j\omega_k}), \qquad b_0 \neq 0 \quad (34)$$

where $\omega_k = 2\pi\delta\sin(\boldsymbol{\theta}_k)$ and where δ is the separation between the sensors measured in wavelengths. The idea is to reparameterize the minimization problem in terms of the polynomial coefficients instead of $\boldsymbol{\theta}$. As will be seen below, this leads to a considerable computational simplification. Observe that the roots of (34) lie on the unit circle. A common way to impose this constraint on the coefficients in the estimation procedure is to enforce the conjugate symmetry property

$$b_k = b_{d-k}^c; \qquad k = 0, \cdots, d. \quad (35)$$

This only approximately ensures that the roots are on the unit circle. However, notice that if z_i is a root of (34) with conjugate symmetric coefficients, then so is z_i^{-c}. Assume now that we have a set of perturbed polynomial coefficients satisfying (35). As long as the perturbation is small enough, the d roots of (34) will be on the unit circle since the true DOA's are distinct. That is, for large enough N, the conjugate symmetry property in fact guarantees that the roots are on the unit circle and that the asymptotic properties of the estimates will not be affected.

Next, define the $m \times (m-d)$ Toeplitz matrix $\mathbf{B}$ as

$$\mathbf{B}^* = \begin{bmatrix} b_d & \cdots & b_1 & b_0 & 0 & \cdots & 0 \\ 0 & b_d & \cdots & b_1 & b_0 & \ddots & 0 \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & b_d & \cdots & b_1 & b_0 \end{bmatrix}. \tag{36}$$

It is readily verified that $\mathbf{B}^*\mathbf{A}(\boldsymbol{\theta}) = \mathbf{0}$, and since the rank of $\mathbf{B}$ is $m-d$ (by construction, since $b_0 \neq 0$), it follows that the columns of $\mathbf{B}$ span the null space of $\mathbf{A}^*$. This implies that we can use $\mathbf{B}$ in (36) in the GWSF criterion, minimize with respect to the polynomial coefficients, and then compute $\boldsymbol{\theta}$ from the roots of the polynomial (34). In the following, we give the details on how this can be performed.

It is useful to rewrite the GWSF criterion as [cf. (17) and (22)]

$$\begin{aligned} V_{\mathrm{GWSF}} &= \bar{\boldsymbol{\varepsilon}}^* \mathbf{W}_{\mathrm{GWSF}} \bar{\boldsymbol{\varepsilon}} \\ &= \begin{bmatrix} \mathrm{vec}(\mathbf{B}^*\hat{\mathbf{E}}_W) \\ \mathrm{vec}(\mathbf{B}^T\hat{\mathbf{E}}_W^c) \end{bmatrix}^* \overline{\mathbf{W}} \begin{bmatrix} \mathrm{vec}(\mathbf{B}^*\hat{\mathbf{E}}_W) \\ \mathrm{vec}(\mathbf{B}^T\hat{\mathbf{E}}_W^c) \end{bmatrix} \end{aligned} \tag{37}$$

where

$$\hat{\mathbf{E}}_W = \frac{1}{\sqrt{\hat{\sigma}^2}} \hat{\mathbf{E}}_s \hat{\tilde{\mathbf{\Lambda}}} \hat{\mathbf{\Lambda}}_s^{-1/2} \tag{38}$$

$$\overline{\mathbf{W}}^{-1} = \begin{bmatrix} (\overline{\mathbf{W}}^{-1})_{11} & (\overline{\mathbf{W}}^{-1})_{12} \\ (\overline{\mathbf{W}}^{-1})_{12}^c & (\overline{\mathbf{W}}^{-1})_{11}^c \end{bmatrix}. \tag{39}$$

The blocks in $\overline{\mathbf{W}}^{-1}$ are given by

$$\begin{aligned} (\overline{\mathbf{W}}^{-1})_{11} &= (\mathbf{I}_{d'} \otimes \mathbf{B}^*\mathbf{B}) + (\overline{\mathbf{T}}^T \otimes \mathbf{B}^*)\mathbf{D}_{\boldsymbol{\rho}} \overline{\mathbf{\Omega}} \mathbf{D}_{\boldsymbol{\rho}}^* (\overline{\mathbf{T}}^c \otimes \mathbf{B}) \\ (\overline{\mathbf{W}}^{-1})_{12} &= (\overline{\mathbf{T}}^T \otimes \mathbf{B}^*)\mathbf{D}_{\boldsymbol{\rho}} \overline{\mathbf{\Omega}} \mathbf{D}_{\boldsymbol{\rho}}^T (\overline{\mathbf{T}} \otimes \mathbf{B}^c) \end{aligned}$$

where $\overline{\mathbf{T}} = \mathbf{T}\hat{\tilde{\mathbf{\Lambda}}}\hat{\mathbf{\Lambda}}_s^{-1/2}/\sqrt{\hat{\sigma}^2} = \mathbf{A}^\dagger \hat{\mathbf{E}}_W$. Let $\hat{\mathbf{E}}_{W,k}$ denote column k and $\hat{\mathbf{E}}_{W_{i,j}}$ the i,jth element of $\hat{\mathbf{E}}_W$ and define the vector $\mathbf{b} = [b_0\ b_1\ \cdots\ b_d]^T$. Observe that

$$\begin{aligned} \mathbf{B}^*\hat{\mathbf{E}}_{W,k} &= \begin{bmatrix} \hat{\mathbf{E}}_{W_{d+1,k}} & \hat{\mathbf{E}}_{W_{d,k}} & \cdots & \hat{\mathbf{E}}_{W_{1,k}} \\ \hat{\mathbf{E}}_{W_{d+2,k}} & \hat{\mathbf{E}}_{W_{d+1,k}} & \cdots & \hat{\mathbf{E}}_{W_{2,k}} \\ \vdots & & & \vdots \\ \hat{\mathbf{E}}_{W_{m,k}} & \hat{\mathbf{E}}_{W_{m-1,k}} & \cdots & \hat{\mathbf{E}}_{W_{m-d,k}} \end{bmatrix} \mathbf{b} \\ &\triangleq \tilde{\mathbf{S}}_k \mathbf{b} \end{aligned}$$

for $k = 1, \cdots, d'$, and define

$$\mathbf{S} = [\tilde{\mathbf{S}}_1^T\ \cdots\ \tilde{\mathbf{S}}_{d'}^T]^T.$$

Then, it follows that

$$\mathrm{vec}(\mathbf{B}^*\hat{\mathbf{E}}_W) = \mathbf{S}\mathbf{b}.$$

Let $\boldsymbol{\beta} = [b_0\ \cdots\ b_{\lfloor (d-1)/2 \rfloor}]^T$, where $\lfloor(\cdot)\rfloor$ denotes the largest integer less or equal to $(\cdot)$. In view of the conjugate symmetry constraint (35)

$$\mathbf{b} = \begin{bmatrix} \boldsymbol{\beta} \\ (b_{d/2}) \\ \tilde{\mathbf{I}}\boldsymbol{\beta}^c \end{bmatrix}$$

where the real-valued parameter $b_{d/2}$ only appears if d is even. Here, $\tilde{\mathbf{I}}$ is the exchange matrix with ones along the antidiagonal and zeros elsewhere. Clearly, we have

$$\begin{aligned} \mathbf{S}\mathbf{b} &= [\mathbf{S}_1 \quad (\mathbf{s}) \quad \mathbf{S}_2] \begin{bmatrix} \boldsymbol{\beta}_r + j\boldsymbol{\beta}_i \\ (b_{d/2}) \\ \tilde{\mathbf{I}}(\boldsymbol{\beta}_r - j\boldsymbol{\beta}_i) \end{bmatrix} \\ &= [\mathbf{S}_1 + \mathbf{S}_2\tilde{\mathbf{I}} \quad (\mathbf{s}) \quad j(\mathbf{S}_1 - \mathbf{S}_2\tilde{\mathbf{I}})] \begin{bmatrix} \boldsymbol{\beta}_r \\ (b_{d/2}) \\ \boldsymbol{\beta}_i \end{bmatrix} \triangleq \mathbf{F}\boldsymbol{\mu} \end{aligned}$$

where the subscripts r and i denote real and imaginary parts, respectively. With the notation introduced above, the criterion (37) can be written as

$$V_{\mathrm{GWSF}} = \boldsymbol{\mu}^T [\mathbf{F}^* \quad \mathbf{F}^T] \overline{\mathbf{W}} [\mathbf{F}^* \quad \mathbf{F}^T]^* \boldsymbol{\mu}.$$

To obtain a proper parameterization, we should also include a nontriviality constraint to avoid the solution $\mathbf{b} = \mathbf{0}$ ($\boldsymbol{\mu} = \mathbf{0}$). In [41], it is proposed that either b_{0r} or b_{0i} can be set to one. In what follows, we assume, for simplicity, that it is possible to put $b_{0r} = \boldsymbol{\mu}_1 = 1$. For particular cases, this may not be possible. In such cases, we should instead use the constraint $b_{0i} = 1$. As an alternative to these linear constraints, we can use a norm constraint on $\boldsymbol{\mu}$ (see, e.g., [42]). The key advantage of the reformulation of the problem in terms of $\boldsymbol{\mu}$ is that V_{GWSF} is quadratic in $\boldsymbol{\mu}$, and thus, it can be solved explicitly.

Since $\mathbf{B}^*\hat{\mathbf{E}}_W = O_p(1/\sqrt{N})$ and since $\mathbf{B}^*\mathbf{E}_s$ uniquely determines $\boldsymbol{\theta}_0$, a consistent estimate of $\boldsymbol{\theta}_0$ can be obtained by minimizing $||\mathrm{vec}(\mathbf{B}^*\hat{\mathbf{E}}_W)||_2^2$. Furthermore, this also implies that replacing the weighting matrix $\overline{\mathbf{W}}$ in (37) by a consistent estimate does not change the asymptotic distribution of the estimates. In summary, we have the following GWSF algorithm for uniform linear arrays.

1) Based on $\hat{\mathbf{R}}$, compute $\hat{\mathbf{E}}_s$, $\hat{\mathbf{\Lambda}}_s$, and

$$\hat{\sigma}^2 = \frac{1}{m-d'} \{\mathrm{Tr}(\hat{\mathbf{R}}) - \mathrm{Tr}(\hat{\mathbf{\Lambda}}_s)\}, \qquad \hat{\tilde{\mathbf{\Lambda}}} = \hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2\mathbf{I}.$$

2) Using the notation introduced above, an initial estimate can be obtained by minimizing

$$||\mathrm{vec}(\mathbf{B}^*\hat{\mathbf{E}}_W)||_2^2 = ||\mathbf{S}\mathbf{b}||_2^2 = ||\mathbf{F}\boldsymbol{\mu}||_2^2 = ||\mathbf{J}\boldsymbol{\eta} + \mathbf{c}||_2^2 \tag{40}$$

over $\boldsymbol{\eta} \in \mathbb{R}^d$, where $\boldsymbol{\eta}$, $\mathbf{J}$, and $\mathbf{c}$ are defined via

$$\mathbf{F}\boldsymbol{\mu} = [\mathbf{c} \quad \mathbf{J}][1 \quad \boldsymbol{\eta}^T]^T = \mathbf{J}\boldsymbol{\eta} + \mathbf{c}.$$

The norm should be interpreted as $||\mathbf{x}||_2^2 = \mathbf{x}^*\mathbf{x}$. Using the minimizer $\hat{\boldsymbol{\eta}}$ of (40), an estimate of $\mathbf{b}$ can be constructed. The initial estimate of $\boldsymbol{\theta}_0$ is then found via the roots of (34). Notice that $\hat{\boldsymbol{\eta}}$ is the least-squares solution to the overdetermined set of equations given in (40).

3) Given $\overline{\mathbf{\Omega}}$ and an initial estimate $\hat{\boldsymbol{\theta}}$ (and $\hat{\mathrm{b}}$), compute $\hat{\overline{\mathbf{T}}}\mathbf{A}^{\dagger}(\hat{\boldsymbol{\theta}})\hat{\mathbf{E}}_W$, $\mathbf{B}(\hat{\mathrm{b}})$, and $\mathbf{D}_{\boldsymbol{\rho}}(\hat{\boldsymbol{\theta}})$. Use these quantities to obtain a consistent estimate of $\overline{\mathbf{W}}^{-1}$ in (39). Let $\overline{\mathbf{W}}^{-1/2}$ denote the Cholesky factor of $\overline{\mathbf{W}}^{-1}$ such that $\overline{\mathbf{W}}^{-1} = \overline{\mathbf{W}}^{-1/2}\overline{\mathbf{W}}^{-*/2}$. Solve the following linear system of equations for $\mathbf{F}_w$

$$\overline{\mathbf{W}}^{-1/2}\mathbf{F}_w = [\mathbf{F}^* \quad \mathbf{F}^T]^*$$

and define $\mathbf{J}_w$ and $\mathbf{c}_w$ via

$$\mathbf{F}_w\boldsymbol{\mu} = [\mathbf{c}_w \quad \mathbf{J}_w][1 \quad \boldsymbol{\eta}^T]^T = \mathbf{J}_w\boldsymbol{\eta} + \mathbf{c}_w.$$

The solution of this step is given by

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}\in\mathbb{R}^d} \|\mathbf{J}_w\boldsymbol{\eta} + \mathbf{c}_w\|_2^2 = -\mathbf{J}_w^{\dagger}\mathbf{c}_w.$$

The GWSF estimate $\hat{\boldsymbol{\theta}}$ is then constructed from $\hat{\boldsymbol{\eta}}$ as described before.

In finite samples and difficult scenarios, it may be useful to reiterate the third step several times; however, this does not improve the asymptotic statistical properties of the estimate.

In this section, we have shown that GWSF can be implemented in a very attractive manner if the nominal array is uniform and linear. In particular, the solution can be obtained in a "closed form" by solving a set of linear systems of equations and rooting the polynomial in (34). Thus, there is no need for an iterative optimization procedure like that necessary in, for example, MAP-NSF.

VII. Performance Analysis of MAPprox

The MAPprox cost function (5) was derived by a second-order approximation of the MAP-WSF cost function (2) around $\boldsymbol{\rho}_0$. The MAPprox estimates are thus expected to have a performance similar to ML. This has also been observed in simulations [22], [24]. However, a formal analysis of the asymptotic properties of MAPprox has not been conducted. In this section, we show that using asymptotic approximations, the MAPprox cost function in fact can be rewritten to coincide with the GWSF cost function. This establishes the observed asymptotic efficiency of the MAPprox method.

Consider the normalized MAPprox cost function [cf., (5)]

$$\frac{1}{N}V_{\mathrm{MAPprox}}(\boldsymbol{\theta}) = V(\boldsymbol{\theta}) - \tfrac{1}{2}\partial_{\boldsymbol{\rho}}V^T(\boldsymbol{\theta})[\partial_{\boldsymbol{\rho}\boldsymbol{\rho}}V(\boldsymbol{\theta}) + \overline{\mathbf{\Omega}}^{-1}]^{-1}\partial_{\boldsymbol{\rho}}V(\boldsymbol{\theta}) \quad (41)$$

where

$$V(\boldsymbol{\theta}) = V(\boldsymbol{\theta}, \boldsymbol{\rho}_0) \quad (42)$$

$$V(\boldsymbol{\theta}, \boldsymbol{\rho}) = \frac{1}{\hat{\sigma}^2}\mathrm{Tr}\{\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}, \boldsymbol{\rho})\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^*\}. \quad (43)$$

The gradient and the Hessian in (41) are evaluated at $\boldsymbol{\rho}_0$. One problem with $V_{\mathrm{MAPprox}}(\boldsymbol{\theta})$ is that the inverse $(\partial_{\boldsymbol{\rho}\boldsymbol{\rho}}V(\boldsymbol{\theta}) + \overline{\mathbf{\Omega}}^{-1})^{-1}$ may not exist for all $\boldsymbol{\theta}$. This implies that the consistency of the MAPprox estimate cannot be guaranteed. However, locally around $\boldsymbol{\theta}_0$, the function is well behaved. In the following, we will approximate (41) around $\boldsymbol{\theta}_0$. The approximation retains the second-order behavior of $V_{\mathrm{MAPprox}}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$ and is well defined for all $\boldsymbol{\theta}$, implying that the consistency problem is eliminated. More precisely, we will study (41) for $\boldsymbol{\theta}$ such that $\boldsymbol{\theta} - \boldsymbol{\theta}_0 = O_p(1/\sqrt{N})$.

Let us start by deriving the gradient $\partial_{\boldsymbol{\rho}}V$. In what follows, we write $\mathbf{\Pi}^{\perp}$ in place of $\mathbf{\Pi}_{\mathbf{A}}^{\perp}$ for simplicity of notation. The derivative of (43) with respect to $\boldsymbol{\rho}_i$ is

$$\partial_{\boldsymbol{\rho}_i}V = \frac{1}{\hat{\sigma}^2}\mathrm{Tr}\{\mathbf{\Pi}_i^{\perp}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^*\} = -\frac{2}{\hat{\sigma}^2}\mathrm{Re}[\mathrm{Tr}\{\mathbf{\Pi}^{\perp}\mathbf{A}_i\mathbf{A}^{\dagger}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^*\}]$$

where we have used

$$\mathbf{\Pi}_i^{\perp} = -\mathbf{\Pi}^{\perp}\mathbf{A}_i\mathbf{A}^{\dagger} - \mathbf{A}^{\dagger^*}\mathbf{A}_i^*\mathbf{\Pi}^{\perp}.$$

Using the definition (8) and the general identity $\mathrm{Tr}\{\mathbf{AB}\} \mathrm{vec}^T(\mathbf{A}^T)\mathrm{vec}(\mathbf{B})$, we can write

$$\begin{aligned}\partial_{\boldsymbol{\rho}}V(\boldsymbol{\theta}) &= \partial_{\boldsymbol{\rho}}V|_{\boldsymbol{\theta},\boldsymbol{\rho}_0}\\ &= -\frac{2}{\hat{\sigma}^2}\mathrm{Re}[\mathbf{D}_{\boldsymbol{\rho}}^*\mathrm{vec}(\mathbf{\Pi}^{\perp}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^*\mathbf{A}^{\dagger^*})]\\ &= O_p(1/\sqrt{N}).\end{aligned} \quad (44)$$

Consider next the i, jth element of the Hessian matrix

$$\begin{aligned}\partial_{\boldsymbol{\rho}_i\boldsymbol{\rho}_j}V = -\frac{2}{\hat{\sigma}^2}\mathrm{Re}[\mathrm{Tr}\{&\underbrace{\mathbf{\Pi}_j^{\perp}\mathbf{A}_i\mathbf{A}^{\dagger}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^*}_{O_p(1)}\\ &+ \underbrace{\mathbf{\Pi}^{\perp}\mathbf{A}_{ij}\mathbf{A}^{\dagger}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^*}_{O_p(1/\sqrt{N})}\\ &+ \underbrace{\mathbf{\Pi}^{\perp}\mathbf{A}_i\mathbf{A}_j^{\dagger}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^*}_{O_p(1/\sqrt{N})}\}].\end{aligned} \quad (45)$$

Notice that

$$\begin{aligned}\frac{1}{N}V_{\mathrm{MAPprox}} &= \underbrace{V}_{O_p(1/N)} - \frac{1}{2}\underbrace{\partial_{\boldsymbol{\rho}}V^T}_{O_p(1/\sqrt{N})}(\underbrace{\partial_{\boldsymbol{\rho}\boldsymbol{\rho}}V}_{O_p(1)} + \underbrace{\overline{\mathbf{\Omega}}^{-1}}_{O(1)})^{-1}\underbrace{\partial_{\boldsymbol{\rho}}V}_{O_p(1/\sqrt{N})}\\ &= O_p(1/N).\end{aligned}$$

All of the order relations above are valid locally around $\boldsymbol{\theta}_0$. Next, study the derivative of $(1/N)V_{MAPprox}$ with respect to $\boldsymbol{\theta}_i$ (here, an index i denotes a partial derivative with respect to $\boldsymbol{\theta}_i$)

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\theta}_i}\frac{1}{N}V_{\mathrm{MAPprox}} =\ & \underbrace{V_i}_{O_p(1/\sqrt{N})} - \frac{1}{2}\underbrace{\partial_{\boldsymbol{\rho}}V_i^T}_{O_p(1)}(\underbrace{\partial_{\boldsymbol{\rho}\boldsymbol{\rho}}V}_{O_p(1)} + \underbrace{\overline{\mathbf{\Omega}}^{-1}}_{O(1)})^{-1}\underbrace{\partial_{\boldsymbol{\rho}}V}_{O_p(1/\sqrt{N})}\\ &+ \frac{1}{2}\underbrace{\partial_{\boldsymbol{\rho}}V^T}_{O_p(1/\sqrt{N})}(\underbrace{\partial_{\boldsymbol{\rho}\boldsymbol{\rho}}V}_{O_p(1)} + \underbrace{\overline{\mathbf{\Omega}}^{-1}}_{O(1)})^{-1}\\ &\times\underbrace{\partial_{\boldsymbol{\rho}\boldsymbol{\rho}}V_i}_{O_p(1)}(\underbrace{\partial_{\boldsymbol{\rho}\boldsymbol{\rho}}V}_{O_p(1)} + \underbrace{\overline{\mathbf{\Omega}}^{-1}}_{O(1)})^{-1}\underbrace{\partial_{\boldsymbol{\rho}}V}_{O_p(1/\sqrt{N})}\\ &- \frac{1}{2}\underbrace{\partial_{\boldsymbol{\rho}}V^T}_{O_p(1/\sqrt{N})}(\underbrace{\partial_{\boldsymbol{\rho}\boldsymbol{\rho}}V}_{O_p(1)} + \underbrace{\overline{\mathbf{\Omega}}^{-1}}_{O(1)})^{-1}\underbrace{\partial_{\boldsymbol{\rho}}V_i}_{O_p(1)}.\end{aligned} \quad (46)$$

Since the term containing $\partial_{\boldsymbol{\rho}\boldsymbol{\rho}} V_i$ is of order $O_p(1/N)$, we conclude that $\partial_{\boldsymbol{\rho}\boldsymbol{\rho}} V$ can be replaced by a consistent estimate without affecting the asymptotic second-order properties. This implies, in particular, that we can neglect the last two terms in (45) and only retain the first. [Notice that all three terms in $\partial_{\boldsymbol{\rho}\boldsymbol{\rho}} V_i$ are of order $O_p(1)$. If the second term in (46) were $O_p(1/\sqrt{N})$, then all three terms in (45) would have contributed.] Similarly, it can be shown that asymptotically, the two last terms in (45) do not affect the second-order derivatives of V_{MAPprox}/N with respect to $\boldsymbol{\theta}$. Recall (45) and rewrite it as

$$\begin{aligned}\partial_{\boldsymbol{\rho}_i\boldsymbol{\rho}_j} V &= -\frac{2}{\hat{\sigma}^2}\,\mathrm{Re}[\mathrm{Tr}\{\boldsymbol{\Pi}_j^{\perp}\mathbf{A}_i\mathbf{A}^{\dagger}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^{*}\}] + o_p(1)\\ &= -\frac{2}{\hat{\sigma}^2}\,\mathrm{Re}[\mathrm{Tr}\{(-\boldsymbol{\Pi}^{\perp}\mathbf{A}_j\mathbf{A}^{\dagger} - \mathbf{A}^{\dagger^*}\mathbf{A}_j^{*}\boldsymbol{\Pi}^{\perp})\\ &\quad\times\mathbf{A}_i\mathbf{A}^{\dagger}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^{*}\}] + o_p(1)\\ &= \frac{2}{\hat{\sigma}^2}\,\mathrm{Re}[\mathrm{Tr}\{\mathbf{A}^{\dagger^*}\mathbf{A}_j^{*}\boldsymbol{\Pi}^{\perp}\mathbf{A}_i\mathbf{A}^{\dagger}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^{*}\}]\\ &\quad + o_p(1)\\ &= \frac{2}{\hat{\sigma}^2}\,\mathrm{Re}[\mathrm{vec}^{*}(\mathbf{A}_i)\\ &\quad\times((\mathbf{A}^{\dagger}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^{*}\mathbf{A}^{\dagger^*})^{T}\otimes\boldsymbol{\Pi}^{\perp})\mathrm{vec}(\mathbf{A}_j)]\\ &\quad + o_p(1). \end{aligned} \tag{47}$$

The dominating term of (47) can be written in matrix form as

$$\begin{aligned}\tilde{V}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta}) &\triangleq \frac{2}{\hat{\sigma}^2}\mathrm{Re}[\mathbf{D}_{\boldsymbol{\rho}}^{*}((\mathbf{A}^{\dagger}\hat{\mathbf{E}}_s\hat{\mathbf{W}}_{\mathrm{WSF}}\hat{\mathbf{E}}_s^{*}\mathbf{A}^{\dagger^*})^{T}\otimes\boldsymbol{\Pi}^{\perp})\mathbf{D}_{\boldsymbol{\rho}}]\\ &= \frac{2}{\hat{\sigma}^2}\,\mathrm{Re}[\mathbf{D}_{\boldsymbol{\rho}}^{*}\hat{\mathbf{M}}\mathbf{D}_{\boldsymbol{\rho}}]. \end{aligned} \tag{48}$$

If $\tilde{V}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta})$ is used in lieu of $\partial_{\boldsymbol{\rho}\boldsymbol{\rho}} V(\boldsymbol{\theta})$ in (41), we get the alternate cost function

$$V_{\mathrm{MAPprox2}}(\boldsymbol{\theta}) = V(\boldsymbol{\theta}) - \tfrac{1}{2}\partial_{\boldsymbol{\rho}} V^{T}(\boldsymbol{\theta})[\tilde{V}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta}) + \overline{\boldsymbol{\Omega}}^{-1}]^{-1}\partial_{\boldsymbol{\rho}} V(\boldsymbol{\theta}).$$

It can be seen that (48) is non-negative definite, which ensures that the inverse $(\tilde{V}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta}) + \overline{\boldsymbol{\Omega}}^{-1})^{-1}$ always exists. Furthermore, we have [cf. (11)]

$$\tilde{V}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta}) + \overline{\boldsymbol{\Omega}}^{-1} = \frac{2}{\hat{\sigma}^2}\,\hat{\boldsymbol{\Gamma}}.$$

It can also be verified that $V(\boldsymbol{\theta})$ and $\partial_{\boldsymbol{\rho}} V(\boldsymbol{\theta})$, as defined in (42) and (44), can be written as

$$V(\boldsymbol{\theta}) = \tfrac{1}{2}\bar{\boldsymbol{\varepsilon}}^{*}\hat{\overline{\mathbf{L}}}^{-1}\bar{\boldsymbol{\varepsilon}}, \quad \partial_{\boldsymbol{\rho}} V(\boldsymbol{\theta}) = -\overline{\boldsymbol{\Omega}}^{-1/2}\hat{\overline{\mathbf{G}}}^{*}\hat{\overline{\mathbf{L}}}^{-1}\bar{\boldsymbol{\varepsilon}}$$

where $\hat{\overline{\mathbf{L}}}$ and $\hat{\overline{\mathbf{G}}}$ are defined similarly to (20) and (21) but computed with sample data quantities and in $\boldsymbol{\theta}$ (not in $\boldsymbol{\theta}_0$). This implies that

$$V_{\mathrm{MAPprox2}}(\boldsymbol{\theta}) = \tfrac{1}{2}\bar{\boldsymbol{\varepsilon}}^{*}\hat{\mathbf{W}}_{\mathrm{M2}}\bar{\boldsymbol{\varepsilon}}$$

where

$$\hat{\mathbf{W}}_{\mathrm{M2}} = \left(\hat{\overline{\mathbf{L}}}^{-1} - \hat{\overline{\mathbf{L}}}^{-1}\hat{\overline{\mathbf{G}}}\,\overline{\boldsymbol{\Omega}}^{-1/2}\,\frac{\hat{\sigma}^2}{2}\,\hat{\boldsymbol{\Gamma}}^{-1}\overline{\boldsymbol{\Omega}}^{-1/2}\hat{\overline{\mathbf{G}}}^{*}\hat{\overline{\mathbf{L}}}^{-1}\right). \tag{49}$$

The MAPprox2 cost function can thus be written in the same form as the GWSF criterion [see (17)]. The weighting matrix in (49) depends on $\boldsymbol{\theta}$ and on sample data quantities. However, in Section V, it was shown that any two weighting matrices related as $\mathbf{W}_1 = \mathbf{W}_2 + o_p(1)$ give the same asymptotic performance. This implies that in the following, we can consider the "limit" of (49) in lieu of $\hat{\mathbf{W}}_{\mathrm{M2}}$. In what follows, let $\mathbf{W}_{\mathrm{M2}}$ denote $\hat{\mathbf{W}}_{\mathrm{M2}}$ evaluated in $\boldsymbol{\theta}_0$ and for infinite data. Thus, if we can show that $\mathbf{W}_{\mathrm{M2}}$ is proportional to $\mathbf{W}_{\mathrm{GWSF}}$, then we have proven that MAPprox2 has the same asymptotic performance as GWSF and, hence, that MAPprox2 is asymptotically efficient. Combining (22), (64), and (66), it is immediately clear that $\mathbf{W}_{\mathrm{M2}} = \mathbf{W}_{\mathrm{GWSF}}$.

In this section, the MAPprox cost function was approximated to retain the local properties that affect the estimation error variance. The approximation (MAPprox2) was then shown to coincide with the GWSF cost function (with a "consistent" estimate of the weighting matrix), and hence, it can be concluded that MAPprox(2) provides minimum variance estimates of the DOA's. However, the MAPprox cost function depends on $\boldsymbol{\theta}$ in quite a complicated manner. In Section VIII we compare MAPprox2 and GWSF by means of several simulations, and the strong similarities between these two approaches are clearly visible. Since the implementation of GWSF is much easier, its use is preferred.

VIII. Simulation Examples

In this section, we illustrate the findings of this paper by means of simulation examples. The MAP-NSF, MAPprox2, and GWSF methods are compared with one another, as well as with methods like WSF (MODE) that do not take the array perturbations into account. In [24], the asymptotic performance of WSF was analyzed for the case under study. Unfortunately, the result given in that paper contains a printing error and, therefore, we provide the following result.

Theorem 3: Let $\hat{\boldsymbol{\theta}}_{\mathrm{WSF}}$ be the minimizing argument of $V_{\mathrm{WSF}}(\boldsymbol{\theta}) = V_{\mathrm{WSF}}(\boldsymbol{\theta}, \boldsymbol{\rho}_0)$, where $V_{\mathrm{WSF}}(\boldsymbol{\theta}, \boldsymbol{\rho})$ is given in (3). Then

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\mathrm{WSF}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathrm{WSF}})$$

where

$$\mathbf{C}_{\mathrm{WSF}} = \frac{\sigma^2}{2}\,\mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{F}_{\boldsymbol{\theta}}^{T}\overline{\boldsymbol{\Omega}}\mathbf{F}_{\boldsymbol{\theta}}\mathbf{C}^{-1}. \tag{50}$$

Proof: Since WSF is a special case of GWSF, the result follows from the analysis in Section V. Indeed, if in GWSF we use the weighting $\mathbf{W} = \overline{\mathbf{L}}^{-1}$ (which makes V_{GWSF} proportional to V_{WSF}), then $\mathbf{C}_{\mathrm{WSF}} = \mathbf{H}^{-1}\mathbf{Q}\mathbf{H}^{-1}$, where $\mathbf{H}$ and $\mathbf{Q}$ are given in (61) and (62), respectively. ■

It can be seen from (50) that the first term corresponds to the CRB for the case with no model errors. The second term in (50) reflects the performance degradation of WSF due to model errors.

Example 1: Consider a uniform linear array consisting of $m = 10$ sensors separated by a half wavelength. Two signals impinge from the directions $\boldsymbol{\theta}_0 = [0^{\circ}\ 5^{\circ}]^{T}$ relative to broadside. The signals are uncorrelated, and the SNR is 5 dB. $\mathbf{P}/\sigma^2 = 10^{5/10}\mathbf{I}_2$. The nominal unit gain sensors are perturbed by additive Gaussian random variables with variance 0.01. This corresponds to an uncertainty in the gain with a standard

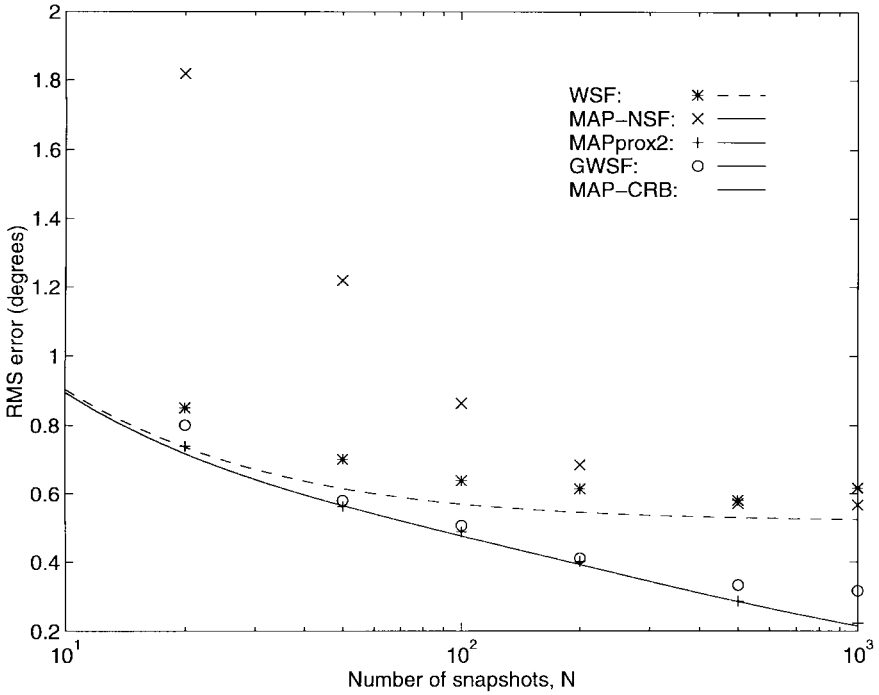

Fig. 1. RMS errors for $\boldsymbol{\theta}_1$ versus N.

deviation of 10%. Fig. 1 depicts the root-mean-square (RMS) error versus the sample size for different methods. (Only the RMS values for $\boldsymbol{\theta}_1$ are displayed; the results corresponding to $\boldsymbol{\theta}_2$ are similar.) In Fig. 1 and in the graphs that follow, we show simulation results using different symbols, whereas the theoretical results are displayed by lines. The empirical RMS values are computed from 1000 independent trials in all simulations. The MAP-NSF and MAPprox2 cost functions are minimized with Newton-type methods initialized by the WSF estimate. WSF (or, equivalently, MODE [33]) is implemented in the "rooting form" described in [41]. The GWSF method is implemented as described in Section VI-B. In the examples presented here, we reiterated the third step of the algorithm three times. The poor performance of MAP-NSF in Fig. 1 is due to two main reasons. First, MAP-NSF fails to resolve the two signals in many cases. Second, the numerical search is sometimes trapped in a local minima. The performance of GWSF and MAPprox2 is excellent and close to the accuracy predicted by the asymptotic analysis. However, the numerical minimization of the MAPprox2 cost function is complicated. In this example, MAPprox2 produced "outliers" in about 5% of the 1000 trials. These outliers were removed before the RMS value for MAPprox2 was calculated. For small N, GWSF and WSF have a similar performance since the finite sample effects dominate. However, for larger N (when calibration errors affect the performance), it can be seen that GWSF and MAPprox2 are significantly better than the standard WSF method.

Example 2: Here, we have the same set-up as in the previous example, but the sample size is fixed to $N = 200$, and the variance of the gain uncertainty is varied. In Fig. 2, the RMS errors for $\boldsymbol{\theta}_1$ are plotted versus the variance of the sensor gains. The curve denoted by CRB in Fig. 2 is the Cramér–Rao lower bound for the case with no model errors, that is, CRB $= \sigma^2\mathbf{C}^{-1}/2N$. Notice that GWSF and MAPprox2 begin to differ for large model errors. Indeed, the equivalence of GWSF and MAPprox shown in Section VII holds for large N and small model errors. The three methods (GWSF, MAPprox2, and MAP-NSF) are known to be asymptotically efficient estimators for this case only; the performance for large model errors is not predicted by the theory in this paper.

Example 3: This example involves uncertainty in the phase response of the sensors and is taken from [22]. Two plane waves are incident on a uniform linear array of $m = 6$ sensors with half wavelength interelement spacing. The signals are uncorrelated and 10 dB above the noise, and have DOA's $\boldsymbol{\theta}_0 = [0^\circ \ \ 10^\circ]^T$ relative to broadside. The uncertainty in the phase of the sensors is modeled by adding uncorrelated Gaussian random variables with covariance $\boldsymbol{\Omega} = 10^{-3}\,\mathrm{diag}[10,\ 1,\ 1,\ 1,\ 1,\ 10]$ to the nominal phase, which is zero for all sensors. RMS errors for $\boldsymbol{\theta}_1$ are shown in Fig. 3. It appears that MAP-NSF has a higher small sample threshold than GWSF and MAPprox2. For large samples, the performance of the methods is similar. Again, it is clearly useful to take the array uncertainty into account when designing the estimator.

Example 4: Here, we consider the same setup as in the previous example, but the signals are correlated. More precisely, the signal covariance matrix is

$$\mathbf{P} = 10\begin{bmatrix} 1 & -0.99 \\ -0.99 & 1 \end{bmatrix}$$

and the noise variance is $\sigma^2 = 1$. The RMS errors for $\boldsymbol{\theta}_1$ are shown in Fig. 4. It can be seen that the two *signal* subspace fitting methods GWSF and WSF can resolve the two signals,

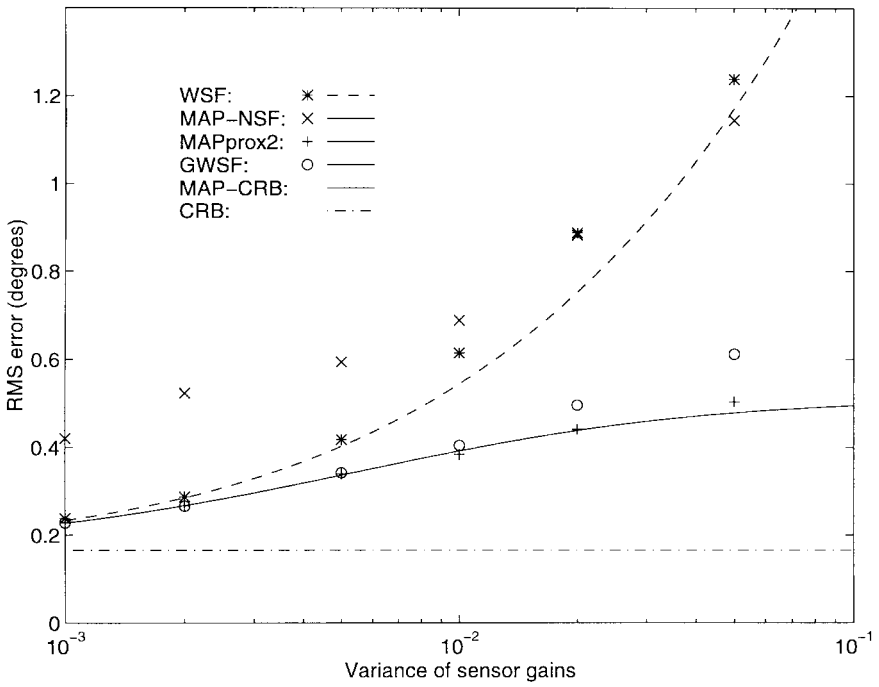

Fig. 2. RMS errors for $\boldsymbol{\theta}_1$ versus the variance of the uncertainty in the gain of the sensors.

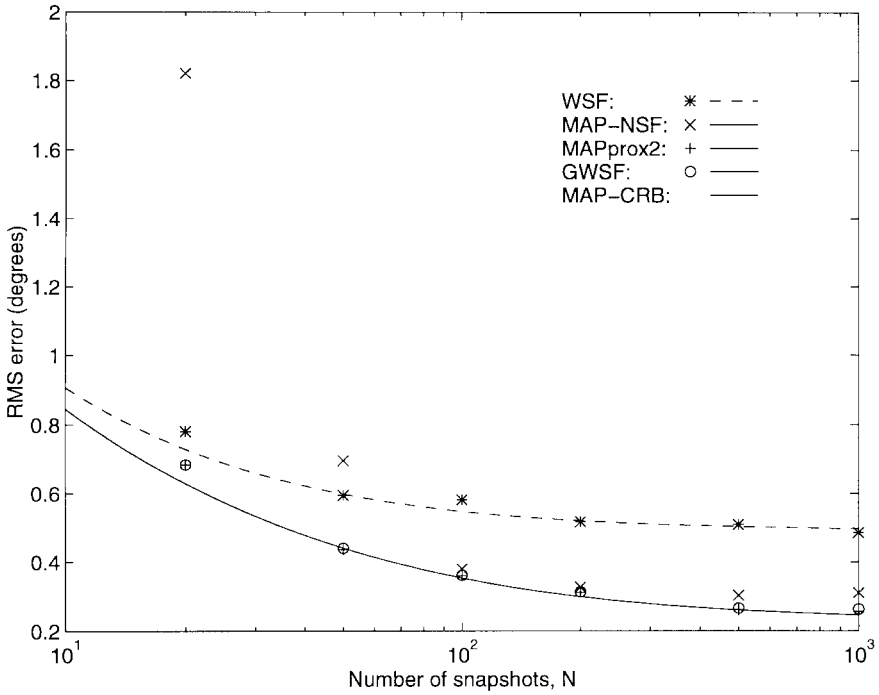

Fig. 3. RMS errors for $\boldsymbol{\theta}_1$ versus the number of snapshots N.

whereas the *noise* subspace fitting method MAP-NSF has a much higher resolution threshold.

IX. Conclusions

This paper has studied the problem of developing robust weighted signal subspace fitting algorithms for arrays with calibration errors. It was shown that if the second-order statistics of the calibration errors are known, then asymptotically statistically efficient weightings can be derived for very general perturbation models. Earlier research had resulted in optimal weightings that were applicable only for very special cases. The GWSF technique derived herein unifies earlier work and enjoys several advantages over the MAP-NSF approach, which is another statistically efficient technique developed for calibration errors. In particular, since GWSF is based on the signal rather than noise subspace, it is a consistent estimator, even when the signals are perfectly coherent, and has better finite sample performance than MAP-NSF when the signals

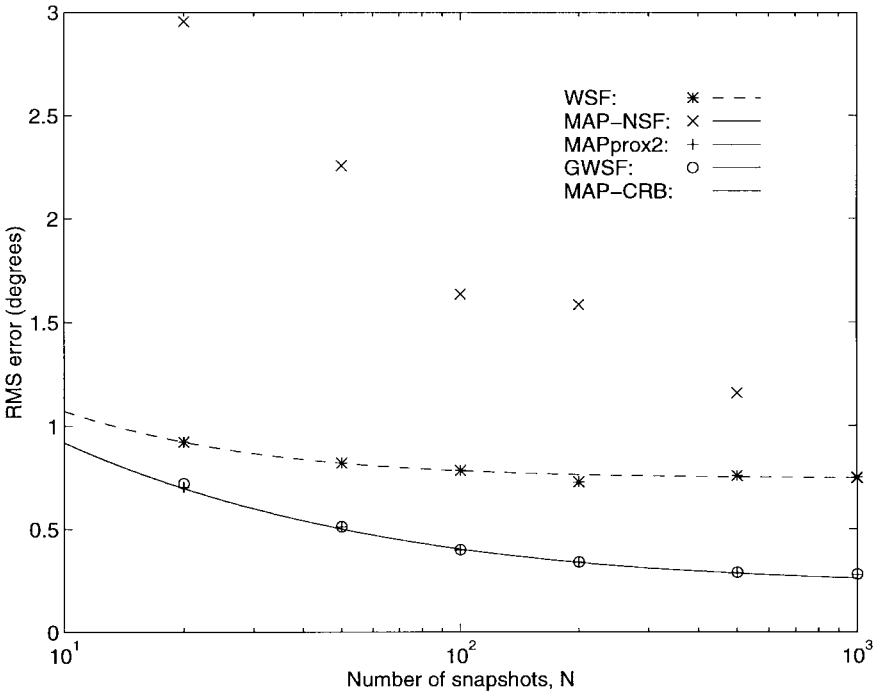

Fig. 4. RMS errors for $\boldsymbol{\theta}_1$ versus the number of snapshots N.

are highly correlated or closely spaced. In addition, unlike MAP-NSF, when the array is nominally uniform and linear, the GWSF criterion can be reparameterized in such a way that the directions of arrival may be solved for by rooting a polynomial rather than via a gradient search. As a byproduct of our analysis, we also demonstrated the asymptotic statistical efficiency of the MAPprox algorithm, which is another robust DOA estimator for perturbed arrays. Whereas the MAPprox and GWSF estimators have a very similar performance, GWSF depends on the parameters in a simpler way than MAPprox, and this facilitates the implementation. Our simulations indicate that in the presence of calibration errors with known statistics, both algorithms can yield a significant performance improvement over techniques that ignore such errors.

APPENDIX A
ASYMPTOTIC COVARIANCE OF THE RESIDUAL

Since $\tilde{\mathbf{R}} = \hat{\mathbf{R}} - \mathbf{R} = O_p(1/\sqrt{N})$, it is straightforward to obtain the first-order relation from (4)

$$\mathbf{B}^*\hat{\mathbf{E}}_s = \mathbf{B}^*\tilde{\mathbf{R}}\mathbf{E}_s\tilde{\mathbf{\Lambda}}^{-1} + o_p(1/\sqrt{N}) \tag{51}$$

where $\mathbf{B} = \mathbf{B}(\boldsymbol{\theta}_0)$ (see, e.g., [43]). Next, notice that the observations are given by the model

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho})\mathbf{s}(t) + \mathbf{n}(t) = \mathbf{A}_0\mathbf{s}(t) + \mathbf{n}(t) + \tilde{\mathbf{A}}\mathbf{s}(t) \tag{52}$$

where $\tilde{\mathbf{A}} = \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}) - \mathbf{A}_0$ denotes the error in the steering matrix due to $\boldsymbol{\rho}$. It is important to observe that the first two terms in (52) represent the observation from the nominal array, whereas the third term is due to model errors. The sample covariance matrix can be partitioned accordingly as

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{x}(t)\mathbf{x}^*(t) = \hat{\mathbf{R}}_{\mathrm{FS}} + \hat{\mathbf{R}}_{\mathrm{ME}} \tag{53}$$

where

$$\hat{\mathbf{R}}_{\mathrm{FS}} = \frac{1}{N}\sum_{t=1}^{N}(\mathbf{A}_0\mathbf{s}(t) + \mathbf{n}(t))(\mathbf{A}_0\mathbf{s}(t) + \mathbf{n}(t))^* \tag{54}$$

$$\hat{\mathbf{R}}_{\mathrm{ME}} = \frac{1}{N}\sum_{t=1}^{N}[(\mathbf{A}_0\mathbf{s}(t) + \mathbf{n}(t))\mathbf{s}^*(t)\tilde{\mathbf{A}}^* + \tilde{\mathbf{A}}\mathbf{s}(t)(\mathbf{A}_0\mathbf{s}(t) + \mathbf{n}(t))^* + \tilde{\mathbf{A}}\mathbf{s}(t)\mathbf{s}^*(t)\tilde{\mathbf{A}}^*] \tag{55}$$

and the subscripts "FS" and "ME" stand for "finite sample" and "model error," respectively. A first-order approximation of $\mathbf{B}^*\hat{\mathbf{R}}$ is then

$$\mathbf{B}^*\hat{\mathbf{R}} = \mathbf{B}^*\hat{\mathbf{R}}_{\mathrm{FS}} + \mathbf{B}^*\tilde{\mathbf{A}}\mathbf{P}\mathbf{A}_0^* + o_p(1/\sqrt{N}). \tag{56}$$

It is interesting to note that the residuals are, to first order, due to two independent terms. The first is the finite sample errors from the nominal array, and the second is the model error contribution. Since the terms are independent, they can be considered one at a time. In view of (53), we denote the error in the principal eigenvectors due to finite samples and model errors by $\tilde{\mathbf{E}}_{s_{\mathrm{FS}}}$ and $\tilde{\mathbf{E}}_{s_{\mathrm{ME}}}$, respectively, that is, $\hat{\mathbf{E}}_s = \mathbf{E}_s + \tilde{\mathbf{E}}_{s_{\mathrm{FS}}} + \tilde{\mathbf{E}}_{s_{\mathrm{ME}}}$.

Let us start by deriving the part of the residual covariance matrix due to finite samples. Define $\tilde{\mathbf{R}}_{\mathrm{FS}} = \hat{\mathbf{R}}_{\mathrm{FS}} - \mathbf{R}$. By straightforward calculations, it is possible to show that

$$\mathrm{E}\{\mathrm{vec}(\tilde{\mathbf{R}}_{\mathrm{FS}})\,\mathrm{vec}^*(\tilde{\mathbf{R}}_{\mathrm{FS}})\} = \frac{1}{N}(\mathbf{R}^T \otimes \mathbf{R}) \tag{57}$$

since $\{\mathbf{x}(t)\}_{t=1}^N$ is a sequence of independent circular Gaussian random vectors. Recalling the first-order equality (51) and

making use of (57), we get

$$\begin{aligned}\mathbf{C}_{\boldsymbol{\varepsilon}_{\mathrm{FS}}} &\triangleq \lim_{N\to\infty} N\mathrm{E}\{\mathrm{vec}(\mathbf{B}^*\tilde{\mathbf{E}}_{s_{\mathrm{FS}}})\,\mathrm{vec}^*(\mathbf{B}^*\tilde{\mathbf{E}}_{s_{\mathrm{FS}}})\}\\ &= \lim_{N\to\infty} N\mathrm{E}\{\mathrm{vec}(\mathbf{B}^*\tilde{\mathbf{R}}_{\mathrm{FS}}\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1})\\ &\quad\times \mathrm{vec}^*(\mathbf{B}^*\tilde{\mathbf{R}}_{\mathrm{FS}}\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1})\}\\ &= \lim_{N\to\infty} N((\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1})^T\otimes\mathbf{B}^*)\\ &\quad\times \mathrm{E}\{\mathrm{vec}(\tilde{\mathbf{R}}_{\mathrm{FS}})\,\mathrm{vec}^*(\tilde{\mathbf{R}}_{\mathrm{FS}})\}((\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1})^c\otimes\mathbf{B})\\ &= ((\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1})^T\otimes\mathbf{B}^*)(\mathbf{R}^T\otimes\mathbf{R})((\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1})^c\otimes\mathbf{B})\\ &= (\tilde{\boldsymbol{\Lambda}}^{-1}\mathbf{E}_s^T\mathbf{R}^T\mathbf{E}_s^c\tilde{\boldsymbol{\Lambda}}^{-1}\otimes\mathbf{B}^*\mathbf{R}\mathbf{B})\\ &= (\tilde{\boldsymbol{\Lambda}}^{-2}\boldsymbol{\Lambda}_s\otimes\sigma^2\mathbf{B}^*\mathbf{B}). \end{aligned}\tag{58}$$

In the third equality, we used the general identity $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T\otimes\mathbf{A})\,\mathrm{vec}(\mathbf{B})$. Next, let $\mathbf{Z}$ be the permutation matrix such that $\mathrm{vec}(\mathbf{B}^*\tilde{\mathbf{E}}_{s_{\mathrm{FS}}}) = \mathbf{Z}\,\mathrm{vec}([\mathbf{B}^*\tilde{\mathbf{E}}_{s_{\mathrm{FS}}}]^T)$ [44]. By making use of $\mathbf{Z}$, (51), and (57), we have

$$\begin{aligned}\overline{\mathbf{C}}_{\boldsymbol{\varepsilon}_{\mathrm{FS}}} &\triangleq \lim_{N\to\infty} N\mathrm{E}\{\mathrm{vec}(\mathbf{B}^*\tilde{\mathbf{E}}_{s_{\mathrm{FS}}})\,\mathrm{vec}^T(\mathbf{B}^*\tilde{\mathbf{E}}_{s_{\mathrm{FS}}})\}\\ &= \lim_{N\to\infty} N((\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1})^T\otimes\mathbf{B}^*)\\ &\quad\times \mathrm{E}\{\mathrm{vec}(\tilde{\mathbf{R}}_{\mathrm{FS}})\,\mathrm{vec}^*(\tilde{\mathbf{R}}_{\mathrm{FS}})\}(\mathbf{B}^c\otimes(\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1}))\mathbf{Z}^T\\ &= ((\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1})^T\otimes\mathbf{B}^*)(\mathbf{R}^T\otimes\mathbf{R})\\ &\quad\times(\mathbf{B}^c\otimes(\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1}))\mathbf{Z}^T = \mathbf{0}. \end{aligned}$$

Next, we proceed and study the model error contribution to the residual covariance matrix. Recall (51), (56), and that $\mathrm{vec}(\tilde{\mathbf{A}}) = \mathbf{D}_{\boldsymbol{\rho}}\tilde{\boldsymbol{\rho}} + o_p(1/\sqrt{N})$. We have

$$\begin{aligned}\mathbf{C}_{\boldsymbol{\varepsilon}_{\mathrm{ME}}} &\triangleq \lim_{N\to\infty} N\mathrm{E}\{\mathrm{vec}(\mathbf{B}^*\tilde{\mathbf{E}}_{s_{\mathrm{ME}}})\mathrm{vec}^*(\mathbf{B}^*\tilde{\mathbf{E}}_{s_{\mathrm{ME}}})\}\\ &= \lim_{N\to\infty} N\mathrm{E}\{\mathrm{vec}(\mathbf{B}^*\tilde{\mathbf{A}}\mathbf{P}\mathbf{A}_0^*\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1})\\ &\quad\times \mathrm{vec}^*(\mathbf{B}^*\tilde{\mathbf{A}}\mathbf{P}\mathbf{A}_0^*\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}^{-1})\}\\ &= \lim_{N\to\infty} N(\mathbf{T}^T\otimes\mathbf{B}^*)\mathrm{E}\{\mathrm{vec}(\tilde{\mathbf{A}})\,\mathrm{vec}^*(\tilde{\mathbf{A}})\}(\mathbf{T}^c\otimes\mathbf{B})\\ &= (\mathbf{T}^T\otimes\mathbf{B}^*)\mathbf{D}_{\boldsymbol{\rho}}\overline{\boldsymbol{\Omega}}\mathbf{D}_{\boldsymbol{\rho}}^*(\mathbf{T}^c\otimes\mathbf{B}) \end{aligned}\tag{59}$$

where $\mathbf{T} = \mathbf{A}_0^\dagger\mathbf{E}_s$, and the relation [cf. (1)] $\mathbf{P} = \mathbf{A}_0^\dagger\mathbf{E}_s\tilde{\boldsymbol{\Lambda}}\mathbf{E}_s^*\mathbf{A}_0^{\dagger *}$ is used. Similarly

$$\begin{aligned}\overline{\mathbf{C}}_{\boldsymbol{\varepsilon}_{\mathrm{ME}}} &\triangleq \lim_{N\to\infty} N\mathrm{E}\{\mathrm{vec}(\mathbf{B}^*\tilde{\mathbf{E}}_{s_{\mathrm{ME}}})\,\mathrm{vec}^T(\mathbf{B}^*\tilde{\mathbf{E}}_{s_{\mathrm{ME}}})\}\\ &= (\mathbf{T}^T\otimes\mathbf{B}^*)\mathbf{D}_{\boldsymbol{\rho}}\overline{\boldsymbol{\Omega}}\mathbf{D}_{\boldsymbol{\rho}}^T(\mathbf{T}\otimes\mathbf{B}^c). \end{aligned}\tag{60}$$

Hence, the asymptotic covariance matrix of the residuals in (18) is given by

$$\mathbf{C}_{\bar{\boldsymbol{\varepsilon}}} = \begin{bmatrix}\mathbf{C}_{\boldsymbol{\varepsilon}} & \overline{\mathbf{C}}_{\boldsymbol{\varepsilon}}\\ \overline{\mathbf{C}}_{\boldsymbol{\varepsilon}}^c & \mathbf{C}_{\boldsymbol{\varepsilon}}^c\end{bmatrix}$$

with $\mathbf{C}_{\boldsymbol{\varepsilon}} = \mathbf{C}_{\boldsymbol{\varepsilon}_{\mathrm{FS}}} + \mathbf{C}_{\boldsymbol{\varepsilon}_{\mathrm{ME}}}$, and $\overline{\mathbf{C}}_{\boldsymbol{\varepsilon}} = \overline{\mathbf{C}}_{\boldsymbol{\varepsilon}_{\mathrm{ME}}}$, where $\mathbf{C}_{\boldsymbol{\varepsilon}_{\mathrm{FS}}}$, $\mathbf{C}_{\boldsymbol{\varepsilon}_{\mathrm{ME}}}$, and $\overline{\mathbf{C}}_{\boldsymbol{\varepsilon}_{\mathrm{ME}}}$ are given by (58)–(60), respectively.

APPENDIX B
PROOF OF THEOREM 2

Recall (28). Differentiating with respect to $\boldsymbol{\theta}_j$ and letting N tend to infinity leads to

$$\mathbf{H}_{ij} = 2\mathbf{k}_i^*\mathbf{W}\mathbf{k}_j.$$

Next, study the ijth element of $\mathbf{Q}$ in (29)

$$\begin{aligned}\mathbf{Q}_{ij} &= \lim_{N\to\infty} N\mathrm{E}\{V_i'V_j'\} = \lim_{N\to\infty} N\mathrm{E}\{4\mathbf{k}_i^*\mathbf{W}\overline{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^*\mathbf{W}\mathbf{k}_j\}\\ &= 4\mathbf{k}_i^*\mathbf{W}\mathbf{C}_{\bar{\boldsymbol{\varepsilon}}}\mathbf{W}\mathbf{k}_j. \end{aligned}$$

Defining the matrix $\mathbf{K} = [\mathbf{k}_1\ \cdots\ \mathbf{k}_d]$, the matrices $\mathbf{H}$ and $\mathbf{Q}$ can be written in compact form as

$$\mathbf{H} = 2\mathbf{K}^*\mathbf{W}\mathbf{K}, \tag{61}$$

$$\mathbf{Q} = 4\mathbf{K}^*\mathbf{W}\mathbf{C}_{\bar{\boldsymbol{\varepsilon}}}\mathbf{W}\mathbf{K}. \tag{62}$$

These expressions are valid for any $\mathbf{W}$ satisfying (26). It remains to be proven that $\mathbf{H}^{-1}\mathbf{Q}\mathbf{H}^{-1}$ equals $N\mathbf{CRB}_{\boldsymbol{\theta}}$ for the choice $\mathbf{W} = \mathbf{W}_{\mathrm{GWSF}}$ given in (22). Inserting $\mathbf{W} = \mathbf{C}_{\bar{\boldsymbol{\varepsilon}}}^{-1}$ in (61) and (62) gives

$$\mathbf{H}^{-1}\mathbf{Q}\mathbf{H}^{-1} = (\mathbf{K}^*\mathbf{C}_{\bar{\boldsymbol{\varepsilon}}}^{-1}\mathbf{K})^{-1}.$$

Next, we rewrite $\mathbf{k}_i$ as

$$\begin{aligned}\mathbf{k}_i &= \begin{bmatrix}\mathrm{vec}(\mathbf{B}_i^*\mathbf{E}_s)\\ \mathrm{vec}(\mathbf{B}_i^T\mathbf{E}_s^c)\end{bmatrix}\\ &= \begin{bmatrix}\mathrm{vec}(\mathbf{B}_i^*\mathbf{A}\mathbf{T})\\ \mathrm{vec}(\mathbf{B}_i^T\mathbf{A}^c\mathbf{T}^c)\end{bmatrix}\\ &= -\begin{bmatrix}\mathrm{vec}(\mathbf{B}^*\mathbf{A}_i\mathbf{T})\\ \mathrm{vec}(\mathbf{B}^T\mathbf{A}_i^c\mathbf{T}^c)\end{bmatrix}\\ &= -\begin{bmatrix}(\mathbf{T}^T\otimes\mathbf{B}^*)\,\mathrm{vec}(\mathbf{A}_i)\\ (\mathbf{T}^*\otimes\mathbf{B}^T)\,\mathrm{vec}(\mathbf{A}_i^c)\end{bmatrix} \end{aligned}\tag{63}$$

and hence

$$\mathbf{K} = -\begin{bmatrix}(\mathbf{T}^T\otimes\mathbf{B}^*)\mathbf{D}_{\boldsymbol{\theta}}\\ (\mathbf{T}^*\otimes\mathbf{B}^T)\mathbf{D}_{\boldsymbol{\theta}}^c\end{bmatrix}.$$

In (63), we made use of the fact that $\mathbf{B}_i^*\mathbf{A} = -\mathbf{B}^*\mathbf{A}_i$, which follows from $\mathbf{B}^*\mathbf{A} = \mathbf{0}$. We also rewrite the inverse of $\mathbf{C}_{\bar{\boldsymbol{\varepsilon}}}$ using the "matrix inversion lemma"

$$\begin{aligned}\mathbf{C}_{\bar{\boldsymbol{\varepsilon}}}^{-1} &= (\overline{\mathbf{L}} + \overline{\mathbf{G}}\,\overline{\mathbf{G}}^*)^{-1}\\ &= \overline{\mathbf{L}}^{-1} - \overline{\mathbf{L}}^{-1}\overline{\mathbf{G}}(\overline{\mathbf{G}}^*\overline{\mathbf{L}}^{-1}\overline{\mathbf{G}} + \mathbf{I})^{-1}\overline{\mathbf{G}}^*\overline{\mathbf{L}}^{-1}. \end{aligned}\tag{64}$$

In view of this, the product $\mathbf{K}^*\mathbf{C}_{\bar{\boldsymbol{\varepsilon}}}^{-1}\mathbf{K}$ contains two terms. In the sequel, we consider these two separately and show that they correspond to the two terms in the expression for $\mathbf{CRB}_{\boldsymbol{\theta}}^{-1}/N$ in (12).

Let us first study

$$\begin{aligned}&\mathbf{K}^*\overline{\mathbf{L}}^{-1}\mathbf{K}\\ &= \begin{bmatrix}(\mathbf{T}^T\otimes\mathbf{B}^*)\mathbf{D}_{\boldsymbol{\theta}}\\ (\mathbf{T}^*\otimes\mathbf{B}^T)\mathbf{D}_{\boldsymbol{\theta}}^c\end{bmatrix}^*\begin{bmatrix}\mathbf{L}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{L}^{-c}\end{bmatrix}\begin{bmatrix}(\mathbf{T}^T\otimes\mathbf{B}^*)\mathbf{D}_{\boldsymbol{\theta}}\\ (\mathbf{T}^*\otimes\mathbf{B}^T)\mathbf{D}_{\boldsymbol{\theta}}^c\end{bmatrix}\\ &= 2\,\mathrm{Re}\{\mathbf{D}_{\boldsymbol{\theta}}^*(\mathbf{T}^c\sigma^{-2}\tilde{\boldsymbol{\Lambda}}^2\boldsymbol{\Lambda}_s^{-1}\mathbf{T}^T\otimes\mathbf{B}(\mathbf{B}^*\mathbf{B})^{-1}\mathbf{B}^*)\mathbf{D}_{\boldsymbol{\theta}}\}\\ &= 2\sigma^{-2}\,\mathrm{Re}\{\mathbf{D}_{\boldsymbol{\theta}}^*\mathbf{M}\mathbf{D}_{\boldsymbol{\theta}}\} = 2\sigma^{-2}\mathbf{C}. \end{aligned}\tag{65}$$

Here, $\mathbf{M}$ is defined in (13), and we used the fact that $\mathbf{B}(\mathbf{B}^*\mathbf{B})^{-1}\mathbf{B}^* = \boldsymbol{\Pi}_{\mathbf{A}}^{\perp}$. Thus, we have shown that (65) equals the first term in $\mathbf{CRB}_{\boldsymbol{\theta}}^{-1}/N$.

To simplify the second term, we first need some preliminary calculations. Consider the inverse appearing in (64)

$$\begin{aligned}
&(\overline{\mathbf{G}}^*\overline{\mathbf{L}}^{-1}\overline{\mathbf{G}}+\mathbf{I})^{-1}\\
&\quad=\left([\mathbf{G}^* \quad \mathbf{G}^T]\begin{bmatrix}\mathbf{L}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{L}^{-c}\end{bmatrix}\begin{bmatrix}\mathbf{G}\\ \mathbf{G}^c\end{bmatrix}+\mathbf{I}\right)^{-1}\\
&\quad=(2\,\mathrm{Re}\{\mathbf{G}^*\mathbf{L}^{-1}\mathbf{G}\}+\mathbf{I})^{-1}\\
&\quad=(2\,\mathrm{Re}\{\overline{\mathbf{\Omega}}^{1/2}\mathbf{D}_{\boldsymbol{\rho}}^*(\mathbf{T}^c\otimes\mathbf{B})(\sigma^{-2}\tilde{\mathbf{\Lambda}}^2\mathbf{\Lambda}_s^{-1}\otimes(\mathbf{B}^*\mathbf{B})^{-1})\\
&\qquad\times(\mathbf{T}^T\otimes\mathbf{B}^*)\mathbf{D}_{\boldsymbol{\rho}}\overline{\mathbf{\Omega}}^{1/2}\}+\mathbf{I})^{-1}\\
&\quad=\frac{\sigma^2}{2}\,\overline{\mathbf{\Omega}}^{-1/2}\left(\mathrm{Re}\{\mathbf{D}_{\boldsymbol{\rho}}^*\mathbf{M}\mathbf{D}_{\boldsymbol{\rho}}\}+\frac{\sigma^2}{2}\,\overline{\mathbf{\Omega}}^{-1}\right)^{-1}\overline{\mathbf{\Omega}}^{-1/2}\\
&\quad=\frac{\sigma^2}{2}\,\overline{\mathbf{\Omega}}^{-1/2}\mathbf{\Gamma}^{-1}\overline{\mathbf{\Omega}}^{-1/2} \qquad (66)
\end{aligned}$$

where $\mathbf{\Gamma}$ is defined in (15). We also need to compute

$$\begin{aligned}
\mathbf{K}^*\overline{\mathbf{L}}^{-1}\overline{\mathbf{G}}=&-2\,\mathrm{Re}\{\mathbf{D}_{\boldsymbol{\theta}}^*(\mathbf{T}^c\sigma^{-2}\tilde{\mathbf{\Lambda}}^2\mathbf{\Lambda}_s^{-1}\mathbf{T}^T\\
&\otimes\mathbf{B}(\mathbf{B}^*\mathbf{B})^{-1}\mathbf{B}^*)\mathbf{D}_{\boldsymbol{\rho}}\overline{\mathbf{\Omega}}^{1/2}\}\\
=&-2\sigma^{-2}\mathbf{F}_{\boldsymbol{\theta}}^T\overline{\mathbf{\Omega}}^{1/2} \qquad (67)
\end{aligned}$$

where $\mathbf{F}_{\boldsymbol{\theta}}$ is defined in (14). Combining (64), (66), and (67), we see that the second term in the product $\mathbf{K}^*\mathbf{C}_{\bar{\boldsymbol{\varepsilon}}}^{-1}\mathbf{K}$ contributes with

$$\begin{aligned}
&-2\sigma^{-2}\mathbf{F}_{\boldsymbol{\theta}}^T\overline{\mathbf{\Omega}}^{1/2}\,\frac{\sigma^2}{2}\,\overline{\mathbf{\Omega}}^{-1/2}\mathbf{\Gamma}^{-1}\overline{\mathbf{\Omega}}^{-1/2}\overline{\mathbf{\Omega}}^{1/2}\mathbf{F}_{\boldsymbol{\theta}}2\sigma^{-2}\\
&\quad=-2\sigma^{-2}\mathbf{F}_{\boldsymbol{\theta}}^T\mathbf{\Gamma}^{-1}\mathbf{F}_{\boldsymbol{\theta}}
\end{aligned}$$

which is exactly equal to the second term in $\mathbf{CRB}_{\boldsymbol{\theta}}^{-1}/N$ [see (12)].

REFERENCES

[1] J. X. Zhu and H. Wang, "Effects of sensor position and pattern perturbations on CRLB for direction finding of multiple narrowband sources," in *Proc. 4th ASSP Workshop Spectral Estimation Modeling*, Minneapolis, MN, Aug. 1988, pp. 98–102.

[2] K. M. Wong, R. S. Walker, and G. Niezgoda, "Effects of random sensor motion on bearing estimation by the MUSIC algorithm," *Proc. Inst. Elect. Eng.*, vol. 135, pt. F, no. 3, pp. 233–250, June 1988.

[3] A. Swindlehurst and T. Kailath, "On the sensitivity of the ESPRIT algorithm to nonidentical subarrays," *Sādhanā, Academy Proc. Eng. Sci.,*, vol. 15, no. 3, pp. 197–212, Nov. 1990.

[4] B. Friedlander, "A sensitivity analysis of the MUSIC algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1740–1751, Oct. 1990.

[5] ———, "Sensitivity of the maximum likelihood direction finding algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, pp. 953–968, Nov. 1990.

[6] A. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors—Part I: The MUSIC algorithm," *IEEE Trans. Signal Processing*, vol. 40, pp. 1758–1774, July 1992.

[7] ———, "A performance analysis of subspace-based methods in the presence of model errors—Part II: Multidimensional algorithms," *IEEE Trans. Signal Processing*, vol. 41, pp. 2882–2890, Sept. 1993.

[8] F. Li and R. Vaccaro, "Sensitivity analysis of DOA estimation algorithms to sensor errors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, pp. 708–717, July 1992.

[9] M. Viberg and A. L. Swindlehurst, "Analysis of the combined effects of finite samples and model errors on array processing performance," *IEEE Trans. Signal Processing*, vol. 42, pp. 3073–3083, Nov. 1994.

[10] V. C. Soon and Y. F. Huang, "An analysis of ESPRIT under random sensor uncertainties," *IEEE Trans. Signal Processing*, vol. 40, pp. 2353–2358, Sept. 1992.

[11] A. Kangas, P. Stoica, and T. Söderström, "Finite sample and modeling error effects on ESPRIT and MUSIC direction estimators," *Proc. Inst. Elect. Eng. Radar, Sonar, Navig.*, vol. 141, pp. 249–255, 1994.

[12] ———, "Large sample analysis of MUSIC and min-norm direction estimators in the presence of model errors," *Circ., Syst., Signal Process.*, vol. 15, pp. 377–393, 1996.

[13] D. J. Ramsdale and R. A. Howerton, "Effect of element failure and random errors in amplitude and phase on the sidelobe level attainable with a linear array," *J. Acoust. Soc. Amer.*, vol. 68, no. 3, pp. 901–906, Sept. 1980.

[14] R. T. Compton, "The effect of random steering vector errors in the Applebaum adaptive array," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-18, pp. 392–400, Sept. 1982.

[15] A. H. Quazi, "Array beam response in the presence of amplitude and phase fluctuations," *J. Acoust. Soc. Amer.*, vol. 72, no. 1, pp. 171–180, July 1982.

[16] B. Friedlander and A. J. Weiss, "Effects of model errors on waveform estimation using the MUSIC algorithm," *IEEE Trans. Signal Processing*, vol. 42, pp. 147–155, Jan. 1994.

[17] J. Yang and A. Swindlehurst, "The effects of array calibration errors on DF-based signal copy performance," *IEEE Trans. Signal Processing*, vol. 43, pp. 2724–2732, Nov. 1995.

[18] A. Paulraj and T. Kailath, "Direction-of-arrival estimation by eigenstructure methods with unknown sensor gain and phase," in *Proc. IEEE ICASSP*, Tampa, FL, Mar. 1985, pp. 17.7.1–17.7.4.

[19] Y. Rockah and P. M. Schultheiss, "Array shape calibration using sources in unknown locations—Part I: Far-field sources," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 286–299, Mar. 1987.

[20] ———, "Array shape calibration using sources in unknown locations—Part II: Near-field sources and estimator implementation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 724–735, June 1987.

[21] A. J. Weiss and B. Friedlander, "Array shape calibration using sources in unknown locations—A maximum likelihood approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1958–1966, Dec. 1989.

[22] B. Wahlberg, B. Ottersten, and M. Viberg, "Robust signal parameter estimation in the presence of array perturbations," in *Proc. IEEE ICASSP*, Toronto, Ont., Canada, 1991, pp. 3277–3280.

[23] M. Wylie, S. Roy, and H. Messer, "Joint DOA estimation and phase calibration of linear equispaced (LES) arrays," *IEEE Trans. Signal Processing*, vol. 42, pp. 3449–3459, Dec. 1994.

[24] M. Viberg and A. L. Swindlehurst, "A Bayesian approach to auto-calibration for parametric array signal processing," *IEEE Trans. Signal Processing*, vol. 42, pp. 3495–3507, Dec. 1994.

[25] A. Flieller, A. Ferreol, P. Larzabal, and H. Clergeot, "Robust bearing estimation in the presence of direction-dependent modeling errors: Identifiability and treatment," in *Proc. ICASSP*, Detroit, MI, 1995, pp. 1884–1887.

[26] A. Swindlehurst, "A maximum *a posteriori* approach to beamforming in the presence of calibration errors," in *Proc. 8th Workshop Stat. Signal Array Process.*, Corfu, Greece, June 1996.

[27] J. Lo and S. Marple, "Eigenstructure methods for array sensor localization," in *Proc. ICASSP*, Dallas, TX, 1987, pp. 2260–2263.

[28] B. Ng and A. Nehorai, "Active array sensor location calibration," in *Proc. ICASSP*, Minneapolis, MN, 1993, pp. IV21–IV24.

[29] M. Koerber and D. Fuhrmann, "Array calibration by Fourier series parameterization: Scaled principle components method," in *Proc. ICASSP*, Minneapolis, MN, 1993, pp. IV340–IV343.

[30] P. Yip and Y. Zhou, "A self-calibration algorithm for cyclostationary signals and its uniqueness analysis," in *Proc. ICASSP*, Detroit, MI, 1995, pp. 1892–1895.

[31] D. McArthur and J. Reilly, "A computationally efficient self-calibrating direction of arrival estimator," in *Proc. ICASSP*, Adelaide, Australia, 1994, pp. IV-201–IV-205.

[32] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample ML techniques for parameter estimation and detection in array processing," in *Radar Array Processing*, Haykin, Litva, and Shepherd, Eds. Berlin, Germany: Springer-Verlag, 1993, pp. 99–151.

[33] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1132–1143, July 1990.

[34] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 1110–1121, May 1991.

[35] B. Ottersten, D. Asztély, M. Kristensson, and S. Parkvall, "A statistical approach to subspace based estimation with applications in telecommunications," in *Proc. 2nd Int. Workshop TLS Errors-in-Variables Modeling*, S. Van Huffel, Ed. Leuven, Belgium: SIAM, 1997, pp. 285–294.

[36] J. F. Böhme, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Process.*, vol. 10, pp. 329–337, 1986.

[37] A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources: A separable solution," in *Proc. ICASSP'88*, New York, Apr. 1988, vol. 5, pp. 2893–2896.
[38] T. Söderström and P. Stoica, *System Identification.* Hemel Hempstead, U.K.: Prentice-Hall, 1989.
[39] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 996–1000, July 1989.
[40] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 1081–1089, Oct. 1986.
[41] P. Stoica and K. Sharman, "Novel eigenanalysis method for direction estimation," *Proc. Inst. Elect. Eng.*, vol. 137, pt. F, no. 1, pp. 19–26, Feb. 1990.
[42] V. Nagesha and S. Kay, "On frequency estimation with the IQML algorithm," *IEEE Trans. Signal Processing*, vol. 42, pp. 2509–2513, Sept. 1994.
[43] H. Clergeot, S. Tressens, and A. Ouamri, "Performance of high resolution frequencies estimation methods compared to the Cramér–Rao bounds," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1703–1720, Nov. 1989.
[44] A. Graham, *Kronecker Products and Matrix Calculus with Applications.* Chichester, U.K.: Ellis Horwood, 1981.

Magnus Jansson (S'93–M'98) was born in Enköping, Sweden, in 1968. He received the M.S., Technical Licentiate, and the Ph.D. degrees in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 1992, 1995, and 1997, respectively.

He is currently a Research Associate in the Automatic Control Group at the Department of Signals, Sensors, and Systems, Royal Institute of Technology. His research interests include sensor array processing, time series analysis, and system identification.

A. Lee Swindlehurst (M'90), for photograph and biography, see this issue, p. 2483.

Björn Ottersten (S'87–M'89) was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986. In 1989, he received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

He has held research positions at the Department of Electrical Engineering, Linköping University, the Royal Institute of Technology, Stockholm, and the Information Systems Laboratory, Stanford University. From 1996 to 1997, he was Director of Research at ArrayComm Inc., San Jose, CA. In 1991, he was appointed Professor of Signal Processing at the Royal Institute of Technology (KTH), Stockholm, and he is currently head of the Department of Signals, Sensors, and Systems there. His research interests include wireless communications, stochastic signal processing, sensor array processing, and time series analysis.

Dr. Ottersten received the IEEE Signal Processing Society Paper Award in 1993.