

# A Bayesian Approach to Auto-Calibration for Parametric Array Signal Processing

Mats Viberg, *Member, IEEE*, and A. Lee Swindlehurst, *Member, IEEE*

**Abstract**— A number of techniques for parametric (high-resolution) array signal processing have been proposed in the last few decades. With few exceptions, these algorithms require an exact characterization of the array, including knowledge of the sensor positions, sensor gain/phase response, mutual coupling, and receiver equipment effects. Unless all sensors are identical, this information must typically be obtained by experimental measurements (calibration). In practice, of course, all such information is inevitably subject to errors. Recently, several different methods have been proposed for alleviating the inherent sensitivity of parametric methods to such modeling errors. The technique proposed herein is related to the class of so-called auto-calibration procedures, but it is assumed that certain prior knowledge of the array response errors is available. This is a reasonable assumption in most applications, and it allows for more general perturbation models than does pure auto-calibration. The optimal maximum *a posteriori* (MAP) estimator for the problem at hand is formulated, and a computationally more attractive large-sample approximation is derived. The proposed technique is shown to be statistically efficient, and the achievable performance is illustrated by numerical evaluation and computer simulation.

## I. INTRODUCTION

THE field of array signal processing is concerned with the problem of extracting information from measurements taken by an array of spatially distributed sensors. This research area has received much attention over the last few decades, and many relevant parameter estimation methods have been proposed. These techniques find application in such diverse disciplines as radar detection, radio and satellite communication, underwater source localization, and seismic exploration. As the terminology implies, parametric techniques require access to a parametrized model of the array measurements.

The assumed data model includes the response of the array to emitters as a function of their spatial and temporal parameters, such as bearing, range, polarization, frequency, etc. Hence, the array geometry and the gain/phase characteristics of the sensors and the associated receiver equipment must be

Manuscript received June 12, 1993; revised May 25, 1994. This work was supported by the Advanced Research Projects Agency of the Department of Defense, monitored by the Air Force Office of Scientific Research under Grant No. F49620-91-C-0086; by the SDIO/IST Program, managed by the Army Research Office under Grant DAAH04-93-G-0029; and by the National Science Foundation under Grant MIP-9110112. The associate editor coordinating the review of this paper and approving it for publication was Prof. Daniel Fuhrmann.

M. Viberg is with the Department of Applied Electronics, Chalmers University of Technology, S-412 96 Gothenburg, Sweden.

A. Swindlehurst is with the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602 USA.  
IEEE Log Number 9406019.

known to the user. If this is not the case, or if, for example, the sensors are subject to mutual coupling, the array must be calibrated. This is normally done by experimentally measuring the array response to sources at different locations. In either case, the assumed model is bound to differ from the actual response at the time of data collection. This may be due to time-varying environmental conditions, or simply to errors in the chosen model structure. The effects of such model errors have been studied by several authors, e.g., [1]–[6]. These effects can be contrasted with estimation errors induced by the necessity of collecting only a finite amount of data from the array, e.g., as studied in [7]–[13]. In many cases of practical interest, the effects of modeling errors have at least as great an influence on the total estimation error as do the finite sample effects of noise.

While most estimation algorithms have been developed with only finite sample effects in mind, alternative techniques that only take model errors into account have been proposed in [6] and [14]. An optimal approach to the problem must account for both sources of errors simultaneously. One idea would be to estimate the unknown parameters of the array response simultaneously with the signal parameters. Such techniques are referred to as *auto-calibration* methods, and have been proposed in, for example, [15]–[17]. An alternative is to view the modeling errors as random with known second-order statistics. Optimally weighted subspace fitting methods based on this model are derived in [6] and [14] (model errors only) and [18] and [19] (both model errors and finite sample effects).

One drawback associated with auto-calibration techniques is that, in many situations, the array response and signal location parameters are not independently identifiable. For instance, it is not possible to estimate both sensor phase characteristics and signal bearing angles. On the other hand, the techniques pursued in [6], [14], and [18] can only give optimal performance in special cases (as will be demonstrated in Section IV–C). Herein, we consider a combination of the two approaches. That is, the array perturbation parameters are assumed to be random with known *a priori* distribution, and are estimated in a Bayesian framework along with the signal parameters. Similar ideas have appeared, e.g., in [20]–[24], and the results of the latter reference are taken as a starting point for the present work.

One of the main contributions of this paper is the development of an algorithm that is computationally much simpler than the optimal maximum *a posteriori* (MAP) estimator, but is asymptotically equivalent in performance. The proposed

method combines the ideas of [24] and the noise subspace fitting (NSF) approach of [25], and hence is referred to as the MAP-NSF algorithm. Use of the NSF problem formulation decouples the estimation of the directions of arrival (DOA's) of the signals and the array perturbation parameters, and yields a criterion function that depends only on the DOA's. The array calibration parameters can then be solved for directly, if desired. In addition to developing the MAP-NSF algorithm and verifying its asymptotic efficiency, we also present a compact expression for the Cramér-Rao bound (CRB) for the case where both finite sample effects and model errors are present.

The remainder of the paper is organized as follows. In the next section, the data model is formulated along with some fundamental assumptions and results. Section III presents both the MAP and MAP-NSF estimators, and Section IV explores some connections between MAP-NSF and other existing approaches. The compact expression for the CRB is presented in Section V, and the statistical efficiency of MAP-NSF is established. The resulting expression for the asymptotic estimation error covariance is examined using some illustrative examples in Section VI, along with computer simulations to investigate the usefulness of the asymptotic expressions for performance prediction.

## II. PROBLEM FORMULATION

This section introduces notation and briefly presents some preliminary results that are necessary for the analysis that follows.

### A. Data Model

Consider an array of  $m$  sensors having arbitrary positions and characteristics. Impinging on the array are the waveforms of  $d$  narrowband point sources, where  $d < m$ . The vector of sensor outputs is denoted  $\mathbf{x}(t)$ , and is modeled by the following familiar equation:

$$\begin{aligned} \mathbf{x}(t) &= [a(\theta_1, \boldsymbol{\rho}) \cdots a(\theta_d, \boldsymbol{\rho})] \begin{bmatrix} s_1(t) \\ \vdots \\ s_d(t) \end{bmatrix} + \mathbf{n}(t) \\ &= \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho})\mathbf{s}(t) + \mathbf{n}(t). \end{aligned} \quad (1)$$

The columns of the  $m \times d$  matrix  $\mathbf{A}$  are the so-called *array propagation vectors*, denoted  $a(\theta_i, \boldsymbol{\rho})$ ,  $i = 1, \dots, d$ . These vectors describe the array response to a unit waveform with signal parameter(s)  $\theta_i$ . The above model also allows for unknown "perturbation parameters," collected in the real  $n$ -vector  $\boldsymbol{\rho} = [\rho_1, \dots, \rho_n]^T$ . This includes structured parameters, such as sensor gain, phase, position, and/or mutual coupling, as well as unstructured parameters (see Section IV-C). The nominal array response is assumed to be unambiguous; *i.e.*, the matrix  $[a(\theta_1, \boldsymbol{\rho}_0), \dots, a(\theta_m, \boldsymbol{\rho}_0)]$  has full rank for any collection of distinct parameters  $\theta_1, \dots, \theta_m$ .

Though not necessary, it is assumed in our discussion that  $\theta_i$  is a real scalar, referred to as the  $i$ th DOA. The components of the  $d$ -vector  $\boldsymbol{\theta}$  are the DOA's of the model, whereas the vector  $\boldsymbol{\theta}_0$  represents their true values. It is further assumed that the array response parameters in  $\boldsymbol{\rho}$  represent small deviations from their known nominal values, collected in the

vector  $\boldsymbol{\rho}_0$ . The *a priori* covariance matrix of the perturbation, denoted by  $\boldsymbol{\Omega}$ , is also assumed known. This matrix could be determined, for example, using sample statistics from a number of independent, identical calibration experiments, or using tolerance data specified by the manufacturer of the sensors. The complex  $d$ -vector  $\mathbf{s}(t)$  is composed of the emitter waveforms received at time  $t$ , and the  $m$ -vector  $\mathbf{n}(t)$  accounts for additive measurement noise. The array output is assumed to be sampled at  $N$  distinct time instants. Based on these measurements,  $\mathbf{x}(1), \dots, \mathbf{x}(N)$ , the problem of interest is to determine the DOA's of all emitters. The number of signals,  $d$ , is assumed to be known.

Although the signal waveforms will be assumed to be Gaussian random processes when deriving the exact MAP estimator and the CRB, the properties of the proposed method will be analyzed under a less restrictive assumption. The signal waveforms are then regarded as arbitrary deterministic (*i.e.*, fixed) sequences such that the following limit exists and is positive definite:

$$\mathbf{P} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \mathbf{s}(t)\mathbf{s}^*(t) > 0 \quad (2)$$

where  $\{\cdot\}^*$  denotes the complex conjugate transpose. Similarly, we assume that the perturbation parameters are drawn from a Gaussian distribution when deriving the MAP estimator and the CRB, although the Gaussian assumption is relaxed when analyzing the resulting estimator.

On the contrary, the noise term  $\mathbf{n}(t)$  is modeled as a stationary, complex Gaussian random process, uncorrelated with the signals. The noise has zero mean and is assumed to be circularly symmetric as well as spatially and temporally white:

$$E[\mathbf{n}(t)\mathbf{n}^*(s)] = \sigma^2 \mathbf{I} \delta_{t,s} \quad (3)$$

$$E[\mathbf{n}(t)\mathbf{n}^T(s)] = 0 \quad (4)$$

where  $\delta_{t,s}$  is the Kronecker delta.

Since we are interested in studying the combined effects of finite sample errors and modeling errors, the size of the perturbations relative to the number of available snapshots plays a crucial role. The variances of the estimated DOA's are known to be proportional to  $1/N$  in the finite-sample-only case, whereas they are proportional to  $\boldsymbol{\Omega}$  in the model-error-only case. In this analysis, the relative contribution of the two error sources will be assumed to be of comparable magnitude, and the covariance matrix of the perturbation parameters will be expressed as

$$E[(\boldsymbol{\rho} - \boldsymbol{\rho}_0)(\boldsymbol{\rho} - \boldsymbol{\rho}_0)^T] = \boldsymbol{\Omega} = \bar{\boldsymbol{\Omega}}/N \quad (5)$$

where  $\bar{\boldsymbol{\Omega}}$  is independent of  $N$ . In Section V, a performance analysis is carried out assuming  $N$  to be "large enough." An argument for the somewhat artificial assumption (5) is that if  $\boldsymbol{\Omega} = o(1/N)$ , then the effect of the modeling errors can be neglected and the methods designed for finite sample errors only are optimal. On the other hand, if  $\boldsymbol{\Omega}^{-1} = o(N)$ , the effect of the modeling errors dominates, thus rendering the methods designed solely for such errors optimal. Since the MAP approach presented herein is inherently more complicated than

either of the methods that take only finite samples or modeling errors into account, the former should be avoided when one type of error dominates the other.

### B. Preliminary Results

In the next section, the exact MAP estimator is presented along with a less computationally demanding approximation. The proposed approximate technique is a *subspace-based* method, in that it relies heavily on the properties of the eigendecomposition of the array covariance. Under the above assumptions, the covariance matrix of the array output takes the form

$$\begin{aligned} \mathbf{R} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N E[\mathbf{x}(t)\mathbf{x}^*(t)] \\ &= \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)\mathbf{P}\mathbf{A}^*(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) + \sigma^2 \mathbf{I}. \end{aligned} \quad (6)$$

Since the matrix  $\mathbf{A}\mathbf{P}\mathbf{A}^*$  has rank  $d$  by assumption (the arguments of  $\mathbf{A}$  will frequently be suppressed for notational convenience),  $\sigma^2$  is an eigenvalue of  $\mathbf{R}$  with multiplicity  $m-d$ , and the corresponding eigenvectors are all orthogonal to  $\mathbf{A}$ . The eigendecomposition of  $\mathbf{R}$  thus takes the form

$$\mathbf{R} = \sum_{i=1}^m \lambda_i \mathbf{e}_i \mathbf{e}_i^* = \mathbf{E}_s \boldsymbol{\Lambda}_s \mathbf{E}_s^* + \sigma^2 \mathbf{E}_n \mathbf{E}_n^* \quad (7)$$

where  $\boldsymbol{\Lambda}_s$  is a diagonal matrix containing the  $d$  largest eigenvalues, and the columns of the  $m \times d$  matrix  $\mathbf{E}_s$  are the corresponding unit-norm eigenvectors. Similarly, the columns of  $\mathbf{E}_n$  are the  $m-d$  eigenvectors corresponding to  $\sigma^2$ . Since  $\mathbf{E}_n$  is orthogonal to  $\mathbf{A}$ , it follows that the range space of  $\mathbf{E}_s$  coincides with that of  $\mathbf{A}$ . This observation forms the basis for all subspace-based estimation techniques, starting with the development of the popular MUSIC algorithm [26], [27]. Assuming orthonormal eigenvectors, the orthogonal projection onto the range space of  $\mathbf{A}$  is denoted

$$\boldsymbol{\Pi} = \mathbf{A}(\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* = \mathbf{E}_s \mathbf{E}_s^* \quad (8)$$

and its orthogonal complement is

$$\boldsymbol{\Pi}^\perp = \mathbf{I} - \mathbf{A}(\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* = \mathbf{E}_n \mathbf{E}_n^*. \quad (9)$$

We now derive some expressions for the signal covariance matrix that will be useful in the subsequent analysis. Combining (6) and (7) we get

$$\mathbf{A}\mathbf{P}\mathbf{A}^* = \mathbf{E}_s \tilde{\boldsymbol{\Lambda}} \mathbf{E}_s^* \quad (10)$$

where

$$\tilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}_s - \sigma^2 \mathbf{I}. \quad (11)$$

Denoting the pseudo-inverse of  $\mathbf{A}$  by

$$\mathbf{A}^\dagger = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* \quad (12)$$

(10) implies

$$\mathbf{P} = \mathbf{A}^\dagger \mathbf{E}_s \tilde{\boldsymbol{\Lambda}} \mathbf{E}_s^* \mathbf{A}^{\dagger*}. \quad (13)$$

Noting that  $(\mathbf{A}^\dagger \mathbf{E}_s)^{-1} = \mathbf{E}_s^* \mathbf{A}$ , the inverse of the above equation can be written as

$$\mathbf{P}^{-1} = \mathbf{A}^* \mathbf{E}_s \tilde{\boldsymbol{\Lambda}}^{-1} \mathbf{E}_s^* \mathbf{A}. \quad (14)$$

Under the stated assumptions, the eigendecomposition of  $\mathbf{R}$  can be consistently estimated by performing an eigendecomposition of the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^N \mathbf{x}(t)\mathbf{x}^*(t) = \hat{\mathbf{E}}_s \hat{\boldsymbol{\Lambda}}_s \hat{\mathbf{E}}_s^* + \hat{\mathbf{E}}_n \hat{\boldsymbol{\Lambda}}_n \hat{\mathbf{E}}_n^* \quad (15)$$

where the partitioning of the eigen-elements is similar to (7).

### III. ROBUST ESTIMATION

In this section, the exact MAP formulation of the problem is presented, along with a simplified but asymptotically equivalent approximation.

#### A. Exact MAP Estimation

When deriving the MAP estimator, it will be assumed that the *a priori* distribution of  $\boldsymbol{\rho}$  is Gaussian with known mean  $\boldsymbol{\rho}_0$  and covariance matrix  $\boldsymbol{\Omega}$ . Likewise, the emitter signals,  $\mathbf{s}(t)$  are modeled as zero-mean Gaussian with second-order moments

$$E[\mathbf{s}(t)\mathbf{s}^*(s)] = \mathbf{P}\delta_{t,s} \quad (16)$$

$$E[\mathbf{s}(t)\mathbf{s}^T(s)] = 0. \quad (17)$$

The signal parameters,  $\boldsymbol{\theta}$ , the emitter covariance,  $\mathbf{P}$ , and the noise variance,  $\sigma^2$ , are all regarded as unknown deterministic parameters (i.e., parameters with a non-informative *a priori* distribution). Following [24], the joint MAP estimate of  $\boldsymbol{\theta}$ ,  $\mathbf{P}$ ,  $\sigma^2$  and  $\boldsymbol{\rho}$  is then obtained as

$$\{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\rho}}, \hat{\mathbf{P}}, \hat{\sigma}^2\} = \arg \min_{\boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{P}, \sigma^2} V_{\text{MAP}}(\boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{P}, \sigma^2) \quad (18)$$

$$\begin{aligned} V_{\text{MAP}}(\boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{P}, \sigma^2) &= V_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{P}, \sigma^2) \\ &\quad + \frac{1}{2}(\boldsymbol{\rho} - \boldsymbol{\rho}_0)^T \boldsymbol{\Omega}^{-1}(\boldsymbol{\rho} - \boldsymbol{\rho}_0) \end{aligned} \quad (19)$$

when  $\boldsymbol{\Omega}$  is full rank. Here,  $V_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{P}, \sigma^2)$  is the negative log-likelihood function, which is given by [28]

$$\begin{aligned} V_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{P}, \sigma^2) &= N(\log |\mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{P}, \sigma^2)| \\ &\quad + \text{Tr}\{\mathbf{R}^{-1}(\boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{P}, \sigma^2)\hat{\mathbf{R}}\}) + \text{const}. \end{aligned} \quad (20)$$

where  $|\cdot|$  denotes the determinant. The ML criterion function is known to be separable in  $\mathbf{P}$  and  $\sigma^2$ . For fixed  $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho})$ , the minimizing signal covariance matrix and noise power are [28]

$$\hat{\mathbf{P}} = \mathbf{A}^\dagger (\hat{\mathbf{R}} - \hat{\sigma}^2 \mathbf{I}) \mathbf{A}^{\dagger*} \quad (21)$$

$$\hat{\sigma}^2 = \frac{1}{m-d} \text{Tr}\{\boldsymbol{\Pi}^\perp \hat{\mathbf{R}}\}. \quad (22)$$

Substituting (21)–(22) into (20) leads to [29]

$$V_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\rho}) = N \log \left| \mathbf{A} \hat{\mathbf{P}} \mathbf{A}^* + \hat{\sigma}^2 \mathbf{I} \right| + \text{const}. \quad (23)$$

Clearly,  $V_{\text{MAP}}(\boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{P}, \sigma^2)$  is also separable in  $\mathbf{P}$  and  $\sigma^2$ , and ignoring constant terms the concentrated MAP criterion function is

$$V_{\text{MAP}}(\boldsymbol{\theta}, \boldsymbol{\rho}) = V_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\rho}) + \frac{1}{2}(\boldsymbol{\rho} - \boldsymbol{\rho}_0)^T \boldsymbol{\Omega}^{-1}(\boldsymbol{\rho} - \boldsymbol{\rho}_0). \quad (24)$$

This can be interpreted as a *regularized* ML criterion. That is, the effect of the prior distribution is to force  $\hat{\boldsymbol{\rho}}$  to be close to the nominal value,  $\boldsymbol{\rho}_0$ . If the perturbation parameters are identifiable, this effect is diminished as the number of snapshots  $N$  increases. Thus, the MAP estimate has the same asymptotic properties as the ML estimate (*i.e.*, the pure auto-calibration technique). However, in many applications of interest,  $\boldsymbol{\rho}$  cannot be consistently estimated along with the signal parameters. In such cases the prior distribution has a crucial influence on the asymptotic properties of the estimates of both  $\boldsymbol{\theta}$  and  $\boldsymbol{\rho}$ .

### B. The MAP-NSF Method

It has been assumed that the signal covariance matrix has full rank; *i.e.*, the signals are noncoherent. Then it is known [29] that in the absence of model errors, the ML criterion is asymptotically equivalent to the following noise subspace fitting (NSF) criterion

$$V_{\text{NSF}} = N \text{Tr}\{\mathbf{A}^* \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A} \hat{\mathbf{U}}\} \quad (25)$$

where  $\hat{\mathbf{U}}$  denotes a consistent estimate of the matrix

$$\mathbf{U} = \sigma^{-2} \mathbf{A}^\dagger \mathbf{E}_s \tilde{\Lambda}^2 \Lambda_s^{-1} \mathbf{E}_s^* \mathbf{A}^{\dagger*} = \sigma^{-2} \mathbf{P} \mathbf{A}^* \mathbf{R}^{-1} \mathbf{A} \mathbf{P}. \quad (26)$$

In the present case, the term ‘‘asymptotic equivalence’’ is to be interpreted as

$$\left. \frac{\partial V_{\text{MAP}}}{\partial \eta} \right|_{\boldsymbol{\theta}_0, \boldsymbol{\rho}_0} = \left. \frac{\partial V_{\text{NSF}}}{\partial \eta} \right|_{\boldsymbol{\theta}_0, \boldsymbol{\rho}_0} + o_p(1/\sqrt{N}). \quad (27)$$

Here  $\eta$  refers to any component of  $\boldsymbol{\theta}$  or  $\boldsymbol{\rho}$ , and the symbol  $o_p(1/\sqrt{N})$  represents a term that tends to zero faster than  $1/\sqrt{N}$  in probability. The extension of this result to the case where model errors are also present is immediate since the proof only depends on the fact that  $\hat{\mathbf{R}} = \mathbf{R} + O_p(1/\sqrt{N})$ . By standard first-order arguments, this implies that the MAP estimate is asymptotically (for large  $N$ ) equivalent to the minimizing arguments of the criterion function

$$V_{\text{NSF}}(\boldsymbol{\theta}, \boldsymbol{\rho}) + \frac{1}{2}(\boldsymbol{\rho} - \boldsymbol{\rho}_0)^T \boldsymbol{\Omega}^{-1}(\boldsymbol{\rho} - \boldsymbol{\rho}_0). \quad (28)$$

This criterion depends on its parameters in a simpler way than the exact MAP criterion (24). However, it still requires a nonlinear minimization over both  $\boldsymbol{\theta}$  and  $\boldsymbol{\rho}$ .

Making use of the assumption that  $\boldsymbol{\Omega}$  decreases as  $1/N$ , a further simplification of the criterion is possible that enables separation with respect to  $\boldsymbol{\rho}$ . Recall the following formulas for the  $\text{vec}(\cdot)$  operator (vectorization of a matrix by stacking its columns) and the Kronecker product  $\otimes$  (see [30], [31])

$$\text{Tr}\{\mathbf{ABCD}\} = \text{vec}(\mathbf{D}^T)^T (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B}) \quad (29)$$

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B}) \quad (30)$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = ((\mathbf{AC}) \otimes (\mathbf{BD})) \quad (31)$$

$$(\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T. \quad (32)$$

Using (29), the NSF criterion can be rewritten as

$$N \text{Tr}\{\mathbf{A}^* \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A} \hat{\mathbf{U}}\} = N \mathbf{a}^* \hat{\mathbf{M}} \mathbf{a} \quad (33)$$

where

$$\mathbf{a} = \text{vec}(\mathbf{A}) \quad (34)$$

$$\hat{\mathbf{M}} = \hat{\mathbf{U}}^T \otimes (\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^*). \quad (35)$$

Next, the vectorized steering matrix is approximated locally around  $\boldsymbol{\rho}_0$  as

$$\mathbf{a} = \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho}) \approx \mathbf{a}_0 + \mathbf{D}_\rho \tilde{\boldsymbol{\rho}} \quad (36)$$

where

$$\mathbf{a}_0 = \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho}_0) \quad (37)$$

$$\mathbf{D}_\rho = \left[ \frac{\partial \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \rho_1}, \dots, \frac{\partial \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \rho_n} \right] \Big|_{\boldsymbol{\theta}, \boldsymbol{\rho}_0} \quad (38)$$

$$\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho} - \boldsymbol{\rho}_0. \quad (39)$$

Note that, when evaluated at  $\boldsymbol{\rho}_0$ , the derivative of  $\mathbf{a}$  with respect to  $\boldsymbol{\theta}$  or  $\boldsymbol{\rho}$  is identical to that of  $\mathbf{a}_0 + \mathbf{D}_\rho \tilde{\boldsymbol{\rho}}$ . It follows that the minimizing arguments of (33) are asymptotically identical to the estimates obtained by minimizing the following approximate MAP-NSF criterion with respect to  $\boldsymbol{\theta}$  and  $\boldsymbol{\rho}$ :

$$(\mathbf{a}_0 + \mathbf{D}_\rho \tilde{\boldsymbol{\rho}})^* \hat{\mathbf{M}} (\mathbf{a}_0 + \mathbf{D}_\rho \tilde{\boldsymbol{\rho}}) + \frac{1}{2} \tilde{\boldsymbol{\rho}}^T \tilde{\boldsymbol{\Omega}}^{-1} \tilde{\boldsymbol{\rho}} \quad (40)$$

where we have normalized by  $N$  and used (5).

Since the criterion function in (40) is quadratic in  $\tilde{\boldsymbol{\rho}}$ , we easily obtain the minimum with respect to  $\tilde{\boldsymbol{\rho}}$  (for fixed  $\boldsymbol{\theta}$ ) as

$$\hat{\boldsymbol{\rho}}_{\text{MAP-NSF}} = \boldsymbol{\rho}_0 - \boldsymbol{\Gamma}^{-1} \mathbf{f} \quad (41)$$

where

$$\boldsymbol{\Gamma} = \text{Re} \left\{ \mathbf{D}_\rho^* \hat{\mathbf{M}} \mathbf{D}_\rho + \frac{1}{2} \tilde{\boldsymbol{\Omega}}^{-1} \right\} \quad (42)$$

$$\mathbf{f} = \text{Re} \{ \mathbf{D}_\rho^* \hat{\mathbf{M}} \mathbf{a}_0 \}. \quad (43)$$

Substituting (41) into (40) leads to the following separated criterion function

$$\mathbf{a}_0^* \hat{\mathbf{M}} \mathbf{a}_0 - \mathbf{f}^T \boldsymbol{\Gamma}^{-1} \mathbf{f}. \quad (44)$$

Note that  $\boldsymbol{\Gamma}$  and  $\mathbf{f}$  depend on  $\boldsymbol{\theta}$  through  $\mathbf{D}_\rho$ , and in principle  $\hat{\mathbf{M}}$  also depends on  $\boldsymbol{\theta}$  through  $\hat{\mathbf{U}}$ . However, it will be assumed that a consistent estimate of  $\boldsymbol{\theta}$  is available to form the estimates  $\hat{\mathbf{M}}$ ,  $\hat{\mathbf{f}}$  and  $\hat{\boldsymbol{\Gamma}}$ . Under the stated assumptions, such an estimate can be obtained, for instance, by letting  $\boldsymbol{\rho} = \boldsymbol{\rho}_0$  and  $\hat{\mathbf{U}} = \mathbf{I}$  in (25), which leads to the well known MUSIC algorithm (see [25] for details). As will be seen later, the approximations made in  $\hat{\boldsymbol{\Gamma}}$ ,  $\hat{\mathbf{f}}$  and  $\hat{\mathbf{M}}$  do not change the asymptotic properties of the final estimate. The definitions of the quantities in the MAP-NSF cost function are repeated below for easy reference, followed by a summary of the proposed algorithm:

$$\mathbf{a}_0 = \text{vec}(\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho}_0)) \quad (45)$$

$$\hat{\mathbf{M}} = \hat{\sigma}^{-2} (\hat{\mathbf{A}}^\dagger \hat{\mathbf{E}}_s (\hat{\Lambda}_s - \hat{\sigma}^2 \mathbf{I})^2 \hat{\Lambda}_s^{-1} \hat{\mathbf{E}}_s^* \hat{\mathbf{A}}^{\dagger*})^T \otimes (\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^*) \quad (46)$$

$$\hat{\sigma}^2 = \frac{1}{m-d} \text{Tr}\{(\mathbf{I} - \hat{\mathbf{A}} \hat{\mathbf{A}}^\dagger) \hat{\mathbf{R}}\} \quad (47)$$

$$\hat{\Gamma} = \text{Re} \left\{ \hat{\mathbf{D}}_\rho^* \hat{\mathbf{M}} \hat{\mathbf{D}}_\rho + \frac{1}{2} \hat{\boldsymbol{\Omega}}^{-1} \right\} \quad (48)$$

$$\hat{\mathbf{f}} = \text{Re} \{ \hat{\mathbf{D}}_\rho^* \hat{\mathbf{M}} \mathbf{a}_0 \}. \quad (49)$$

The MAP-NSF Algorithm: Given the sample covariance  $\hat{\mathbf{R}}$  and an initial estimate  $\hat{\boldsymbol{\theta}}$  of the DOA's:

- a) Compute the eigendecomposition  $\hat{\mathbf{R}} = \hat{\mathbf{E}}_s \hat{\boldsymbol{\Lambda}}_s \hat{\mathbf{E}}_s^* + \hat{\mathbf{E}}_n \hat{\boldsymbol{\Lambda}}_n \hat{\mathbf{E}}_n^*$
- b) Compute the quantities (46)–(48)
- c) Using  $\hat{\boldsymbol{\theta}}$  as an initial guess, use a numerical method to solve the following optimization problem

$$\hat{\boldsymbol{\theta}}_{\text{MAP-NSF}} = \arg \min_{\boldsymbol{\theta}} V(\boldsymbol{\theta}) \quad (50)$$

$$V(\boldsymbol{\theta}) = \mathbf{a}_0^* \hat{\mathbf{M}} \mathbf{a}_0 - \hat{\mathbf{f}}^T \hat{\Gamma}^{-1} \hat{\mathbf{f}}. \quad (51)$$

#### IV. RELATIONSHIP TO OTHER TECHNIQUES

This section presents some connections between the proposed MAP-NSF method and existing techniques for DOA estimation under model uncertainties.

##### A. Autocalibration

Autocalibration usually refers to techniques that simultaneously estimate the DOA's and sensor positions. For most geometries, this is possible if the location of one sensor and the direction to another is known [20]. Herein, we consider a more general model where the array response depends on both the DOA's as well as an arbitrary set of perturbation parameters. By assuming an *a priori* distribution for the perturbation parameters, these need not be identifiable from the data only, hence allowing for more flexible models. If the perturbation parameters are indeed identifiable, one may delete the influence of the prior by letting  $\boldsymbol{\Omega}^{-1} = 0$  in (48). Thus, maximum likelihood autocalibration is a special case of the MAP approach considered herein. However, it should be noted that [17] uses a deterministic signal model as opposed to the stochastic Gaussian signal model assumed here. It is known that the stochastic signal model leads to superior performance regardless of the actual signal distribution, at least in the absence of model errors [11], [13].

##### B. The MAPprox Approach

The MAPprox (MAP approximation) method is proposed in [24] as an alternative approximation of the exact MAP estimator. The method is derived in two steps. First,  $V_{\text{ML}}(\boldsymbol{\theta}, \boldsymbol{\rho})$  in (24) is replaced by the asymptotically equivalent signal subspace fitting criterion,

$$V_{\text{WSF}} = \frac{N}{\sigma^2} \text{Tr} \{ \boldsymbol{\Pi}^\perp \hat{\mathbf{E}}_s \mathbf{W}_{\text{WSF}} \hat{\mathbf{E}}_s^* \} \quad (52)$$

where

$$\mathbf{W}_{\text{WSF}} = (\hat{\boldsymbol{\Lambda}}_s - \sigma^2 \mathbf{I})^2 \hat{\boldsymbol{\Lambda}}_s^{-1}. \quad (53)$$

Second, an implicit Newton-step over  $\boldsymbol{\rho}$  is performed, starting at the nominal value,  $\boldsymbol{\rho}_0$ . The DOA parameter is kept fixed at

an arbitrary value. This results in the concentrated approximate criterion function

$$V_{\text{MAPprox}}(\boldsymbol{\theta}) = V_{\text{WSF}} - \frac{1}{2} \partial_\rho V_{\text{WSF}}^T \times [\partial_{\rho\rho} V_{\text{WSF}} + \boldsymbol{\Omega}^{-1}]^{-1} \partial_\rho V_{\text{WSF}} \quad (54)$$

where  $\partial_\rho V_{\text{WSF}}$  denotes the gradient of  $V_{\text{WSF}}$  with respect to  $\boldsymbol{\rho}$ , evaluated at  $\boldsymbol{\rho}_0$  and  $\boldsymbol{\theta}$ . Similarly,  $\partial_{\rho\rho} V_{\text{WSF}}$  denotes the ‘‘ $\boldsymbol{\rho}$ -corner’’ of the Hessian matrix. In practice, the Gauss-Newton Hessian (see [32]) is used to avoid taking second derivatives. The MAP-NSF technique can also be derived by essentially following the above steps, but using the NSF criterion (25) in lieu of (52). It is conceivable that the two approaches are asymptotically equivalent when  $\mathbf{P} > 0$  (although this remains to be formally proved). The asymptotic equivalence of the methods is also indicated by the simulation examples in Section VI, where it is also observed that the MAPprox method appears to have better finite sample properties. The following important differences between the approaches should be noted:

- The MAP-NSF method is only applicable if the signal covariance matrix is nonsingular.
- The MAP-NSF method allows ‘‘more approximations’’ of the criterion function without losing asymptotic efficiency. In particular,  $\partial_{\rho\rho} V_{\text{WSF}}$  in (54) cannot be replaced by a consistent estimate (see Section V-B).

This latter fact is the reason why MAP-NSF is preferred herein. It implies that the MAP-NSF criterion depends on  $\boldsymbol{\theta}$  in a simpler way than does the MAPprox criterion, which considerably simplifies both the analysis and the implementation.

##### C. Signal Subspace Fitting

For the moment, consider the following simple unstructured model for the perturbed array response:

$$\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho}) = \mathbf{A}(\boldsymbol{\theta}) + \tilde{\mathbf{A}} \quad (55)$$

where we define

$$\boldsymbol{\rho} = \begin{bmatrix} \text{Re}\{\text{vec}(\tilde{\mathbf{A}})\} \\ \text{Im}\{\text{vec}(\tilde{\mathbf{A}})\} \end{bmatrix} \quad (56)$$

and where the columns of  $\tilde{\mathbf{A}}$ , denoted  $\tilde{a}_i$ , are modeled as random variables with moments

$$E[\tilde{a}_i] = 0 \quad i = 1, \dots, d \quad (57)$$

$$E[\tilde{a}_i \tilde{a}_k^*] = \nu_{ik} \mathbf{I} \quad i, k = 1, \dots, d \quad (58)$$

$$E[\tilde{a}_i \tilde{a}_k^T] = 0 \quad i, k = 1, \dots, d. \quad (59)$$

This model corresponds to an additive, circularly symmetric complex array perturbation that is uncorrelated from sensor to sensor, but possibly  $\boldsymbol{\theta}$ -dependent. It is easy to verify that under these assumptions, the covariance of  $\boldsymbol{\rho}$  is given by

$$\boldsymbol{\Omega} = \frac{1}{2} \begin{bmatrix} \text{Re}\{\boldsymbol{\Upsilon}\} \otimes \mathbf{I} & -\text{Im}\{\boldsymbol{\Upsilon}\} \otimes \mathbf{I} \\ \text{Im}\{\boldsymbol{\Upsilon}\} \otimes \mathbf{I} & \text{Re}\{\boldsymbol{\Upsilon}\} \otimes \mathbf{I} \end{bmatrix} \quad (60)$$

where the  $i, k$ th element of the matrix  $\boldsymbol{\Upsilon}$  is  $\nu_{ik}$ .

In [18] and [19], an ‘‘optimal’’ signal subspace fitting algorithm is derived that takes into account both finite sample effects and the above array perturbation model. The algorithm

involves minimizing a criterion function identical to (52), but using the weighting matrix

$$\mathbf{W}_{\text{OSF}} = \left( \hat{\mathbf{E}}_s^* \hat{\mathbf{A}}^* \mathbf{\Upsilon}^T \hat{\mathbf{A}} \hat{\mathbf{E}}_s + \frac{\hat{\sigma}^2}{N} (\hat{\mathbf{A}}_s - \hat{\sigma}^2 \mathbf{I})^{-2} \hat{\mathbf{A}}_s \right)^{-1} \quad (61)$$

in place of  $\mathbf{W}_{\text{WSF}}$ . Note that  $\mathbf{W}_{\text{OSF}}$  is just a weighted combination of  $\mathbf{W}_{\text{WSF}}$ , the optimal weighting when finite sample effects alone are considered, and the optimal weighting derived in [6] for the unstructured error model above when  $N \rightarrow \infty$ . The use of the term "optimal" here means that asymptotically, no other weighting matrix in (52) will yield DOA estimates with lower variance. We will now show that using  $\mathbf{W}_{\text{OSF}}$  in (52) yields an algorithm that is in fact asymptotically equivalent to MAP-NSF for the array perturbation model described above. Since the MAP-NSF method is asymptotically efficient (see Section V-C), no algorithm, whether or not it is in the form of (52), will yield asymptotically unbiased estimates with lower variance than the minimizing arguments of

$$V_{\text{OSF}} = \frac{N}{\hat{\sigma}^2} \text{Tr} \{ \mathbf{\Pi}^\perp \hat{\mathbf{E}}_s \mathbf{W}_{\text{OSF}} \hat{\mathbf{E}}_s^* \}. \quad (62)$$

To begin with, note that in this case,  $\hat{\mathbf{D}}_\rho = [\mathbf{I} \ j\mathbf{I}]$ . Some simple algebra then yields (63), at the bottom of this page, where we have used the fact that, for any invertible matrix  $\mathbf{X}$ ,

$$\begin{bmatrix} \text{Re}\{\mathbf{X}\} & -\text{Im}\{\mathbf{X}\} \\ \text{Im}\{\mathbf{X}\} & \text{Re}\{\mathbf{X}\} \end{bmatrix}^{-1} = \begin{bmatrix} \text{Re}\{\mathbf{X}^{-1}\} & -\text{Im}\{\mathbf{X}^{-1}\} \\ \text{Im}\{\mathbf{X}^{-1}\} & \text{Re}\{\mathbf{X}^{-1}\} \end{bmatrix}. \quad (64)$$

Multiplying out all terms in (63) and collecting real and imaginary parts leads to the simpler expression

$$\hat{\mathbf{f}}^T \hat{\mathbf{F}}^{-1} \hat{\mathbf{f}} = \mathbf{a}_0^* \hat{\mathbf{M}} \left( \hat{\mathbf{M}} + \frac{1}{N} \mathbf{\Upsilon}^{-1} \otimes \mathbf{I} \right)^{-1} \hat{\mathbf{M}} \mathbf{a}_0. \quad (65)$$

Using the definition of  $\hat{\mathbf{M}}$  and an obvious notation, write  $\hat{\mathbf{M}} = \hat{\mathbf{U}}^T \otimes (\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^*)$ , and note that

$$\begin{aligned} \left( \hat{\mathbf{M}} + \frac{1}{N} \mathbf{\Upsilon}^{-1} \otimes \mathbf{I} \right)^{-1} &= \left( \left( \hat{\mathbf{U}}^T + \frac{1}{N} \mathbf{\Upsilon}^{-1} \right) \otimes (\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^*) \right. \\ &\quad \left. + \frac{1}{N} \mathbf{\Upsilon}^{-1} \otimes (\hat{\mathbf{E}}_s \hat{\mathbf{E}}_s^*) \right)^{-1} \\ &= \left( \hat{\mathbf{U}}^T + \frac{1}{N} \mathbf{\Upsilon}^{-1} \right)^{-1} \otimes (\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^*) \\ &\quad + N \mathbf{\Upsilon} \otimes (\hat{\mathbf{E}}_s \hat{\mathbf{E}}_s^*). \end{aligned} \quad (66)$$

Finally, we have

$$\begin{aligned} V_{\text{MAP-NSF}} &= \mathbf{a}_0^* \hat{\mathbf{M}} \mathbf{a}_0 - \hat{\mathbf{f}}^T \hat{\mathbf{F}}^{-1} \hat{\mathbf{f}} \\ &= \mathbf{a}_0^* \left( \left[ \hat{\mathbf{U}}^T - \hat{\mathbf{U}}^T (\hat{\mathbf{U}}^T + \frac{1}{N} \mathbf{\Upsilon}^{-1})^{-1} \hat{\mathbf{U}}^T \right] \otimes (\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^*) \right) \mathbf{a}_0 \\ &= \mathbf{a}_0^* \left( (\hat{\mathbf{U}}^{-T} + N \mathbf{\Upsilon})^{-1} \otimes (\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^*) \right) \mathbf{a}_0 \\ &= \text{Tr} \left\{ \mathbf{A}^* \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A} (\hat{\mathbf{U}}^{-1} + N \mathbf{\Upsilon}^T)^{-1} \right\} \\ &= \frac{1}{N} \text{Tr} \left\{ \mathbf{A}^* \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A} \hat{\mathbf{U}}_{\text{OSF}} \right\} \end{aligned} \quad (67)$$

where

$$\hat{\mathbf{U}}_{\text{OSF}} = \hat{\mathbf{A}}^\dagger \hat{\mathbf{E}}_s \mathbf{W}_{\text{OSF}} \hat{\mathbf{E}}_s^* \hat{\mathbf{A}}^{\dagger*}. \quad (68)$$

Using arguments similar to those in [29], it can be shown that the two cost functions

$$V_{\text{NSF}}(\boldsymbol{\theta}) = \text{Tr} \left\{ \mathbf{A}^* \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A} (\hat{\mathbf{A}}^\dagger \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^* \hat{\mathbf{A}}^{\dagger*}) \right\} \quad (69)$$

$$V_{\text{SSF}}(\boldsymbol{\theta}) = \text{Tr} \left\{ \mathbf{\Pi}^\perp \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^* \right\} \quad (70)$$

are asymptotically equivalent for any weighting matrix  $\mathbf{W}$ , as long as the emitter covariance  $\mathbf{P}$  has full rank. Thus, for the simple perturbation model described by (55)–(59), we can conclude that MAP-NSF and OSF are asymptotically equivalent as well. When both algorithms are applicable, it is preferable to use the OSF rather than the MAP-NSF criterion, since OSF is computationally less demanding and can handle coherent emitters.

## V. ASYMPTOTIC PERFORMANCE ANALYSIS

In this section, the asymptotic properties of the proposed approximate MAP-NSF method are investigated. However, let us start the discussion by presenting an approximate bound on the achievable performance.

### A. Cramér-Rao Bound

The Cramér-Rao bound (CRB) gives a lower bound on the (asymptotic) covariance matrix of any (asymptotically) unbiased estimator. In the present case, the full parameter set contains both the deterministic parameters  $\boldsymbol{\theta}$ ,  $\mathbf{P}$ ,  $\sigma^2$ , and the stochastic parameter  $\boldsymbol{\rho}$ . We will collectively denote the deterministic parameters by the vector  $\boldsymbol{\eta}$ . The CRB on  $\boldsymbol{\eta}$  and  $\boldsymbol{\rho}$  is the inverse of the Fisher information matrix (FIM)  $\mathbf{J}$ . Assuming Gaussian signals and perturbation parameters, we have from [33] that  $\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2$ , where the  $ij$ th element of

$$\begin{aligned} \hat{\mathbf{f}}^T \hat{\mathbf{F}}^{-1} \hat{\mathbf{f}} &= \left[ \text{Re}\{\hat{\mathbf{M}} \mathbf{a}_0\}^T \quad \text{Im}\{\hat{\mathbf{M}} \mathbf{a}_0\}^T \right] \\ &\quad \times \begin{bmatrix} \text{Re} \left\{ \left( \hat{\mathbf{M}} + \frac{1}{N} \mathbf{\Upsilon}^{-1} \otimes \mathbf{I} \right)^{-1} \right\} & -\text{Im} \left\{ \left( \hat{\mathbf{M}} + \frac{1}{N} \mathbf{\Upsilon}^{-1} \otimes \mathbf{I} \right)^{-1} \right\} \\ \text{Im} \left\{ \left( \hat{\mathbf{M}} + \frac{1}{N} \mathbf{\Upsilon}^{-1} \otimes \mathbf{I} \right)^{-1} \right\} & \text{Re} \left\{ \left( \hat{\mathbf{M}} + \frac{1}{N} \mathbf{\Upsilon}^{-1} \otimes \mathbf{I} \right)^{-1} \right\} \end{bmatrix} \times \begin{bmatrix} \text{Re}\{\hat{\mathbf{M}} \mathbf{a}_0\} \\ \text{Im}\{\hat{\mathbf{M}} \mathbf{a}_0\} \end{bmatrix} \end{aligned} \quad (63)$$

each term is given by

$$\mathbf{J}_{1,ij} = E_{\mathbf{x},\rho} \left[ \frac{\partial^2}{\partial \Theta_i \partial \Theta_j} V_{\text{ML}} \right] \quad (71)$$

$$\mathbf{J}_{2,ij} = \frac{1}{2} E_{\rho} \left[ \frac{\partial^2}{\partial \Theta_i \partial \Theta_j} (\boldsymbol{\rho} - \boldsymbol{\rho}_0)^T \boldsymbol{\Omega}^{-1} (\boldsymbol{\rho} - \boldsymbol{\rho}_0) \right] \quad (72)$$

where  $\Theta_i$  denotes the  $i$ th element of the compound parameter vector  $[\boldsymbol{\eta}^T \boldsymbol{\rho}^T]^T$ . Unfortunately,  $\mathbf{J}_1$  is not easily evaluated. The  $\boldsymbol{\rho}$ -parameter generally enters in a nonlinear fashion, rendering the expectation with respect to  $\boldsymbol{\rho}$  difficult to compute. The exact expression has been computed in [20] for some special cases involving only a single emitter. However, a more common approach is to ignore the expectation with respect to  $\boldsymbol{\rho}$ , and evaluate  $\mathbf{J}_1$  at  $\boldsymbol{\rho}_0$  [22]. Following [24] and using (5), it can be shown that this approximation is of order  $O(1)$ . Since  $J_1 = O(N)$ , we will thus obtain an asymptotically valid CRB. The following ‘‘compact CRB’’ formula involving  $\boldsymbol{\theta}$  and  $\boldsymbol{\rho}$  only is given in [24] and is restated herein for completeness:

*Theorem 1:* Let  $\hat{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\rho}}$  be asymptotically unbiased estimates of  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\rho}_0$ , and assume that  $\mathbf{s}(t)$  and  $\boldsymbol{\rho}$  are Gaussian distributed. Then for large  $N$  we have

$$E \begin{bmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0 \end{bmatrix}^T \geq \mathbf{CRB} \quad (73)$$

where the  $ij$ th element of  $\mathbf{CRB}^{-1}$  is given by

$$\mathbf{CRB}_{ij}^{-1} = \frac{2N}{\sigma^2} \text{Re} \left\{ \text{Tr} \left\{ \mathbf{A}_i^* \boldsymbol{\Pi}^\perp \mathbf{A}_j \right. \right. \\ \left. \left. \times \mathbf{P} \mathbf{A}^* \mathbf{R}^{-1} \mathbf{A} \mathbf{P} \right\} \right\} + \boldsymbol{\Omega}_{ij}^{-1} u(i, j). \quad (74)$$

Here,  $\mathbf{A}_i$  denotes the derivative of  $\mathbf{A}$  with respect to the  $i$ th component of  $[\boldsymbol{\theta}^T \boldsymbol{\rho}^T]^T$ , and  $u(\cdot, \cdot)$  is an indicator function such that  $u(i, j) = 0$  if either  $i \leq d$  or  $j \leq d$ , and  $u(i, j) = 1$  otherwise. All expressions are evaluated at the true parameter values  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\rho}_0$ .

To obtain a compact expression for the CRB on  $\boldsymbol{\theta}$  only, introduce the notation

$$\mathbf{D}_\theta = \left[ \frac{\partial \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \theta_1}, \dots, \frac{\partial \mathbf{a}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \theta_d} \right] \quad (75)$$

$$\mathbf{F}_\theta = \text{Re} \{ \mathbf{D}_\rho^* \mathbf{M} \mathbf{D}_\theta \} \quad (76)$$

$$\mathbf{C} = \text{Re} \{ \mathbf{D}_\theta^* \mathbf{M} \mathbf{D}_\theta \} \quad (77)$$

where the matrices are evaluated at the ‘‘true’’ values,  $\boldsymbol{\theta}_0, \boldsymbol{\rho}_0$ . Then we have the following result:

*Corollary 1:* Let  $\hat{\boldsymbol{\theta}}$  be an asymptotically unbiased estimate of  $\boldsymbol{\theta}_0$ , and assume that  $\mathbf{s}(t)$  and  $\boldsymbol{\rho}$  are Gaussian distributed. Then, for large  $N$  the following inequality holds:

$$E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T] \geq \mathbf{CRB}_\theta = \frac{1}{2N} [\mathbf{C} - \mathbf{F}_\theta^T \boldsymbol{\Gamma}^{-1} \mathbf{F}_\theta]^{-1}. \quad (78)$$

*Proof:* Straightforward from (74).  $\square$

Note that, as derived in [11] and [13], the asymptotic CRB on  $\boldsymbol{\theta}$  in the absence of modeling errors is given by  $\mathbf{C}^{-1}/(2N)$ . We thus confirm the intuitively clear result that modeling errors can only deteriorate the attainable estimation accuracy. As will be seen later in this section, the proposed MAP-NSF

method attains the bound given in (78) *regardless* of the actual distribution of  $\mathbf{s}(t)$  and  $\boldsymbol{\rho}$  (although the right hand side of (78) is a lower bound on the attainable estimation error variance only for Gaussian signals and perturbations).

### B. Consistency

Let us first verify that the MAP-NSF estimates converge with probability 1 (w.p.1) to the true values as  $N \rightarrow \infty$ . This holds true for any choice of weighting matrix  $\hat{\mathbf{U}} > 0$ , and for any value of  $\hat{\boldsymbol{\theta}}$  used in forming  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{D}}_\rho$ . To see this, note that the criterion function (51) is the Schur complement of the following matrix

$$\begin{bmatrix} \mathbf{a}_0^* \hat{\mathbf{M}} \mathbf{a}_0 & \mathbf{f}^T \\ \mathbf{f} & \boldsymbol{\Gamma} \end{bmatrix} = \text{Re} \left\{ \begin{bmatrix} \mathbf{a}_0^* \\ \mathbf{D}_\rho^* \end{bmatrix} \hat{\mathbf{M}} \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{D}_\rho \end{bmatrix}^* \right\} + \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2} \boldsymbol{\Omega}^{-1} \end{bmatrix}. \quad (79)$$

Clearly, this matrix is positive definite if  $\mathbf{a}_0^* \hat{\mathbf{M}} \mathbf{a}_0 > 0$ , and singular otherwise. Thus, we have  $V(\boldsymbol{\theta}) \geq 0$ , with equality iff  $\hat{\mathbf{M}} \mathbf{a}_0 = 0$ . Let  $\mathbf{M}$  denote the limiting (as  $N \rightarrow \infty$ ) value of  $\hat{\mathbf{M}}$ . Then

$$\mathbf{M} \mathbf{a}_0 = (\mathbf{U}^T \otimes (\mathbf{E}_n \mathbf{E}_n^*)) \text{vec}(\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho}_0)) \\ = \text{vec}(\boldsymbol{\Pi}^\perp \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho}_0) \mathbf{U}) \quad (80)$$

where we have used (30) and the fact that  $\mathbf{E}_n \mathbf{E}_n^* = \boldsymbol{\Pi}^\perp$ . Since the array is assumed to be unambiguous,  $\mathbf{M} \mathbf{a}_0 = 0$  iff  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ . The convergence of the criterion function is uniform in  $\boldsymbol{\theta}$  if the derivative of  $\mathbf{a}_0$  is bounded, and it follows that the minimizing argument of (51) converges to  $\boldsymbol{\theta}_0$  w.p.1 as  $N \rightarrow \infty$ .

It is interesting to note that any consistent estimate of  $\boldsymbol{\theta}$  can be used in forming  $\hat{\mathbf{M}}$  in (79), without affecting the consistency of the final MAP-NSF estimate. As we will see later, this replacement does not even affect the asymptotic distribution of the estimate. One might be tempted to guess that it would similarly be possible to insert a consistent estimate of  $\boldsymbol{\theta}$  into the Hessian matrix  $\partial_{\rho\rho} V_{\text{WSF}}$ , appearing in the MAPprox criterion function (54). However, the resulting criterion function then generally becomes unbounded from below, and consistency of the estimates cannot be guaranteed. Since  $\partial_{\rho\rho} V_{\text{WSF}}$  is usually a complicated function of  $\boldsymbol{\theta}$ , so is (54); therefore (51) results in a computationally simpler method.

### C. Asymptotic Distribution

Although the estimates were shown to be consistent for arbitrary weighting matrices, their asymptotic properties are certainly affected by the weighting used. In the following, it is assumed that the optimal weightings (46)–(48) are used, and that the estimates of these quantities are consistent, i.e.,  $\hat{\mathbf{M}} \rightarrow \mathbf{M}$ ,  $\hat{\boldsymbol{\Gamma}} \rightarrow \boldsymbol{\Gamma}$  and  $\hat{\mathbf{f}} \rightarrow \mathbf{f}$  (in probability) as  $N \rightarrow \infty$ . Since  $\hat{\boldsymbol{\theta}}_{\text{MAP-NSF}}$  is consistent, a standard Taylor series expansion of the gradient of the criterion function around the estimate can be performed. This yields

$$\tilde{\boldsymbol{\theta}} = -\mathbf{H}^{-1} \mathbf{V}' + o_p(1/\sqrt{N}) \quad (81)$$

where  $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\text{MAP-NSF}} - \boldsymbol{\theta}_0$  denotes the estimation error,  $\mathbf{V}'$  is the gradient of  $V(\boldsymbol{\theta})$ , and  $\mathbf{H}$  is the asymptotic Hessian. Both  $\mathbf{V}'$  and  $\mathbf{H}$  are evaluated at the true value  $\boldsymbol{\theta}_0$ .

Consider first the derivative of (51) with respect to  $\theta_i$ , denoted  $V_i$ :

$$\begin{aligned} V_i &= 2\text{Re}\{\mathbf{a}_0^* \hat{\mathbf{M}} \mathbf{a}_i\} - 2\hat{\mathbf{f}}^T \hat{\mathbf{\Gamma}}^{-1} \hat{\mathbf{f}}_i \\ &= 2\text{Re}\{\mathbf{a}_0^* \hat{\mathbf{M}} \hat{\mathbf{y}}_i\} \end{aligned} \quad (82)$$

where  $\hat{\mathbf{y}}$  is defined by

$$\hat{\mathbf{y}} = \mathbf{a}_0 - \hat{\mathbf{D}}_\rho \hat{\mathbf{\Gamma}}^{-1} \hat{\mathbf{f}} \quad (83)$$

and its derivative with respect to  $\theta_i$  is

$$\hat{\mathbf{y}}_i = \mathbf{a}_i - \hat{\mathbf{D}}_\rho \hat{\mathbf{\Gamma}}^{-1} \hat{\mathbf{f}}_i \quad (84)$$

where  $\hat{\mathbf{f}}_i$  denotes the derivative of  $\hat{\mathbf{f}}$  with respect to  $\theta_i$ . Observe that  $\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* = \mathbf{\Pi}^\perp + O_p(1/\sqrt{N})$ , so that  $\mathbf{a}_0^* \hat{\mathbf{M}} = O_p(1/\sqrt{N})$  for any value of  $\hat{\mathbf{U}}$  (see (80)). Therefore,  $\hat{\mathbf{U}}$ ,  $\hat{\mathbf{\Gamma}}$  and  $\hat{\mathbf{f}}$  can be replaced by their limiting values, denoted  $\mathbf{U}$ ,  $\mathbf{\Gamma}$ , and  $\mathbf{f}$  respectively, without affecting the asymptotic (second-order) properties of  $V_i$ . The corresponding limiting  $\hat{\mathbf{y}}_i$  is denoted  $\mathbf{y}_i$ . Thus,

$$V_i = 2\text{Re}\{\mathbf{a}_0^* \hat{\mathbf{M}} \mathbf{y}_i\} + o_p(1/\sqrt{N}). \quad (85)$$

Note that the only stochastic term in (85) is the  $\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^*$  term of  $\hat{\mathbf{M}}$ . To approximate this term we write

$$\hat{\mathbf{M}} \mathbf{a}_0 \simeq \text{vec}(\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A} \mathbf{U}) \quad (86)$$

where the arguments of  $\mathbf{a} = \mathbf{a}(\theta_0, \rho_0)$  have been suppressed for notational convenience. The calculations in Appendix A show that

$$\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A} \simeq -\mathbf{\Pi}^\perp \hat{\mathbf{R}} \mathbf{E}_s \tilde{\mathbf{A}}^{-1} \mathbf{E}_s^* \mathbf{A}. \quad (87)$$

By the central limit theorem (see e.g., Lemma 9.A.2 of [34]),  $\hat{\mathbf{R}} = \mathbf{R} + O_p(1/\sqrt{N})$  is asymptotically (for large  $N$ ) Gaussian distributed, and it follows from (81) and (85)–(87) that also the normalized estimation error has a limiting Gaussian distribution

$$\sqrt{N} \tilde{\boldsymbol{\theta}} \in \text{As}\mathcal{N}(0, \mathbf{H}^{-1} \mathbf{Q} \mathbf{H}^{-1}) \quad (88)$$

where  $\mathbf{Q}$  is the asymptotic covariance matrix of the normalized gradient

$$\mathbf{Q} = \lim_{N \rightarrow \infty} N E[V' V'^T]. \quad (89)$$

We thus have the following result:

*Theorem 2:* Let  $\hat{\boldsymbol{\theta}}$  be the minimizing argument of the MAP-NSF cost function (51). Then  $\hat{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}_0$  as  $N \rightarrow \infty$ , and the limiting distribution of the normalized estimation error is

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \in \text{As}\mathcal{N}(0, N \mathbf{CRB}_\theta) \quad (90)$$

where  $\mathbf{CRB}_\theta$  is the asymptotic CRB given by (78).

*Proof:* We need to verify that  $\mathbf{H} = \mathbf{Q} = (N \mathbf{CRB}_\theta)^{-1}$ . However, these calculations are deferred to Appendix B.  $\square$

As expected from the fact that the MAP-NSF method is derived by first-order approximations to the optimal MAP estimator, the former (and hence also the latter) yields asymptotically efficient estimates. Recall that the interpretation of this result is that  $\hat{\boldsymbol{\theta}}$  is a minimum variance estimate when both  $N$  is “large enough” and  $\boldsymbol{\Omega}$  is “small enough”. The requirements on  $N$  and  $\boldsymbol{\Omega}$  for the estimates to be “practically efficient” may be quite demanding in difficult scenarios, as will be seen in the next section.

#### D. Performance Without Auto-Calibration

It is of course of interest to assess the relative improvement offered by the auto-calibration technique as compared to techniques that do not exploit the perturbation model. We choose to compare with the WSF estimates [12], obtained by minimizing (52) with respect to  $\boldsymbol{\theta}$ . This method is known to be asymptotically equivalent with the stochastic maximum likelihood technique [28] and also with the NSF method (25)–(26) for the case  $\mathbf{P} > 0$ . Thus, the following result applies to these other techniques as well.

*Theorem 3:* Let  $\hat{\boldsymbol{\theta}}_{\text{WSF}}$  be the minimizer of (52). Then

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{WSF}} - \boldsymbol{\theta}_0) \in \text{As}\mathcal{N}(0, \mathbf{C}_{\text{WSF}}) \quad (91)$$

where

$$\mathbf{C}_{\text{WSF}} = \frac{1}{2}(\mathbf{C}^{-1} + \mathbf{C}^{-1} \mathbf{F}_\theta^T \boldsymbol{\Omega} \mathbf{F}_\theta \mathbf{C}^{-1}) \quad (92)$$

and where  $\mathbf{F}_\theta$  and  $\mathbf{C}$  are defined in (76) and (77).

*Proof:* This result is a straightforward extension of Theorem 1 in [19], and the proof is therefore omitted.  $\square$

Since the WSF method gives asymptotically unbiased estimates under the model considered herein, the Cramér-Rao inequality implies that  $\mathbf{C}_{\text{WSF}} \geq N \mathbf{CRB}_\theta$ . The difference between the two depends, of course, on the scenario. However, note that for identifiable perturbation models (i.e., when  $\mathbf{D}_\rho^* \mathbf{M} \mathbf{D}_\rho > 0$ ), the “efficiency ratio”  $\mathbf{C}_{\text{WSF},ii}/(N \mathbf{CRB}_{\theta,ii})$  may be arbitrarily large for large  $N$ , since we then have  $\mathbf{CRB}_\theta \rightarrow 0$  as  $N \rightarrow \infty$  for a fixed value of  $\boldsymbol{\Omega}$ , but  $\mathbf{C}_{\text{WSF}}/N$  does not tend to zero unless  $\boldsymbol{\Omega} \rightarrow 0$  as  $N \rightarrow \infty$ . In the more interesting case of unidentifiable perturbation parameters ( $\mathbf{D}_\rho^* \mathbf{M} \mathbf{D}_\rho$  singular), the performance improvement offered by the MAP approach over WSF and other “conventional” DOA estimators may be less dramatic. An example is provided in the next section.

## VI. SIMULATION EXAMPLES

In this section we study the performance of the MAP-NSF method in finite samples and under “moderately sized” perturbations. A comparison with the MAPprox and WSF methods is also included. As an application of the general perturbation model considered herein, we study a model suitable for arrays mounted on a flexible structure. The array is assumed to be planar, although an extension to the 3-D case is straightforward. It is assumed that the distances between the sensors is known and fixed. For simplicity, we approximate the flexible structure using a piecewise linear model, as illustrated in Fig. 1. This approximation is reasonable for small perturbations, which is the case of interest herein. The nominal array is assumed to be a uniform linear array (ULA) of  $m = 10$  sensors, oriented along the  $x$ -axis of a coordinate system with its origin at the first sensor. The interelement spacing ( $r$  in the figure) is fixed at a half wavelength. In many cases of practical interest, there may be physical constraints that relate the incremental angles, thus leading to a perturbation model with fewer parameters (for example a polynomial model). However, here we assume that  $\rho_i, i = 1, \dots, m-1$ , are independent zero-mean Gaussian random variables with variance  $\nu^2$ , so



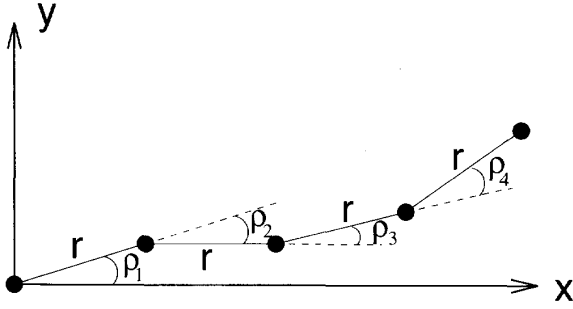


Fig. 1. An array mounted on a flexible structure.

that  $\Omega = \nu^2 \mathbf{I}$ . This is a reasonable model, e.g., for a towed array.

Assume that the sensors have unit gain and zero phase delay over the frequency band and DOA's of interest. Let the DOA's  $\theta_k$  be measured counter-clockwise relative to the  $x$ -axis. The response of the  $i$ th sensor in the direction  $\theta$  is then

$$\{a(\theta, \boldsymbol{\rho})\}_i = e^{j2\pi(r_{x,i} \cos \theta + r_{y,i} \sin \theta)} \quad (93)$$

where  $(r_{x,i}, r_{y,i})$  are the coordinates of the  $i$ th sensor location, measured in fractions of the wavelength. From Fig. 1, the sensor coordinates can be recursively calculated as

$$\phi_{i+1} = \phi_i + \rho_i \quad (94)$$

$$r_{x,i+1} = r_{x,i} + r \cos \phi_{i+1} \quad (95)$$

$$r_{y,i+1} = r_{y,i} + r \sin \phi_{i+1} \quad (96)$$

with initial conditions  $r_{x,1} = r_{y,1} = \phi_1 = 0$ . To implement the MAP-NSF and MAPprox methods, the derivative of  $\mathbf{a}(\theta, \boldsymbol{\rho})$  with respect to  $\boldsymbol{\rho}$ , evaluated at  $\boldsymbol{\rho}_0 = \mathbf{0}$ , must be calculated. After some straightforward algebra, we obtain

$$\begin{aligned} \left. \frac{\partial a(\theta, \boldsymbol{\rho})}{\partial \boldsymbol{\rho}} \right|_{\boldsymbol{\rho}=\mathbf{0}} &= j\pi \sin \theta \text{diag}[a(\theta, 0)] \\ &\times \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 2 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ m-1 & m-2 & \dots & 1 \end{bmatrix} \quad (97) \end{aligned}$$

where  $\text{diag}[a(\theta, 0)]$  is a diagonal matrix with the elements of  $a(\theta, 0)$  on the diagonal.

The perturbed array receives the waveforms of two uncorrelated signal sources located at  $\theta_1 = 85^\circ$  and  $\theta_2 = 90^\circ$ . The MAP-NSF and MAPprox methods estimate the DOA's and simultaneously calibrate the  $\boldsymbol{\rho}$ -parameters. For comparison, we also include the WSF estimates using the nominal ULA parametrization. The signal-to-noise ratio (SNR) is fixed at 10 dB for both signals. In the first experiment, the perturbation variance  $\nu^2$  is varied, whereas the number of snapshots in each batch is fixed at  $N = 1000$ . Fig. 2 displays the theoretical and empirical RMS errors of the estimate of  $\theta_1$  versus  $\nu$  (both axes in degrees). The theoretical results are obtained by evaluating the expressions (78) (MAP-NSF) and (92) (WSF). The empirical results are calculated from 512 independent Monte Carlo runs. The methods are implemented

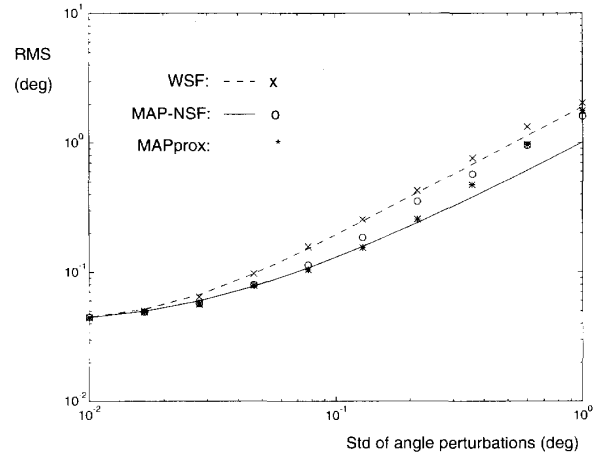


Fig. 2. Theoretical and empirical RMS error of the DOA estimates versus the standard deviation of the incremental sensor angle perturbations.

using a Newton-type search. The WSF search is initialized at  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ , whereas the MAP-NSF and MAPprox searches are initialized using the WSF estimates. As seen from the figure, the theoretical RMS error of the MAP-NSF estimate (i.e., the CRB) is slightly but clearly smaller than that of the WSF estimate, the difference being more pronounced for large  $\nu$ . However, the CRB appears to be difficult to reach in practice. The empirical results agree well with the theory only for  $\nu \leq 0.1^\circ$  (MAP-NSF) and  $\nu \leq 0.2^\circ$  (MAPprox). It is interesting to see that the empirical performance of the methods is similar for both large and small  $\nu$ , but there is a certain range of "moderate values" where MAP-NSF and MAPprox outperform WSF. Notice also that the first-order approximation is better for the MAPprox method than for MAP-NSF in this scenario. We also tried the MAP-NSF method initialized at  $\boldsymbol{\theta}_0$ , but this modification did not result in any significant improvement. Hence, the MAPprox method appears to have the best second-order properties.

As suggested by the analysis, the range of  $\nu$ -values for which the Bayesian auto-calibration approaches outperform WSF depends on  $N$ . To theoretically achieve the CRB we should have  $\nu^2 = O(1/N)$ . Thus, in the next experiment the number of snapshots is varied from  $N = 100$  to  $N = 10000$ , whereas  $\nu^2$  is simultaneously decreased as  $\nu^2(N) = c/N$ , where  $c = 10$  is determined such that  $\nu(1000) = 0.1^\circ$ . Fig. 3 shows the theoretical and empirical RMS errors for  $\hat{\theta}_1$  versus the number of snapshots. In this case, the empirical RMS errors of the MAP-NSF estimates approach the CRB for  $N > 1000$  and  $\nu < 0.1^\circ$ , whereas the MAPprox estimates are efficient already at  $0.3N = 100$  and  $\nu = 0^\circ$ . The RMS error of the WSF estimates is about 50% (3.5 dB) higher.

## VII. CONCLUSION

A method for DOA estimation in the presence of structured uncertainty in the array parametrization has been proposed. The array model was assumed to be a function of not only the DOA's, but also a set of random perturbation parameters with known first and second moments. The exact maximum *a poste-*

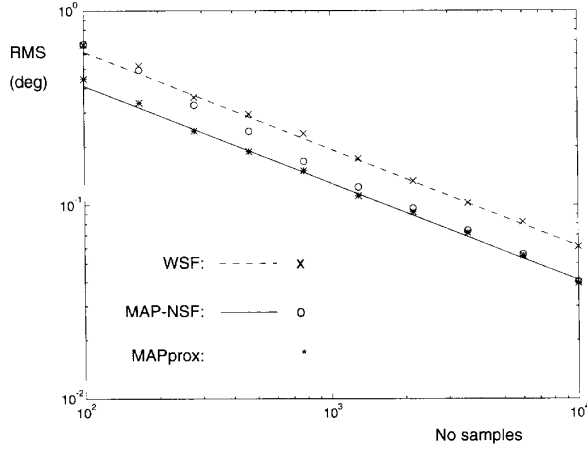


Fig. 3. Theoretical and empirical RMS error of the DOA estimates versus the number of snapshots.

*riori* (MAP) method for simultaneous estimation of the DOA's and the perturbations was first derived, assuming Gaussian perturbations and noise. A computationally simpler method, termed MAP-NSF (noise subspace fitting) that only involves a search over the DOA's was then proposed and analyzed, assuming a large number of samples and small perturbations. A compact expression for the approximate Cramér-Rao bound (CRB) was also derived, and the asymptotic covariance matrix of the MAP-NSF estimation error was found to coincide with the CRB. As a by-product, we also verified that the optimal signal subspace fitting method proposed in [18] is also statistically efficient for the unstructured perturbation model considered therein.

If the DOA's and perturbation parameters are simultaneously identifiable from the data only, the improvement of autocalibration with or without a prior as compared to "traditional" DOA estimation techniques can be quite dramatic (see the examples in [24]), at least for large data records. However, the case of most interest herein is the unidentifiable case for which the prior is necessary for enabling simultaneous estimation of all parameters. To investigate the performance of the methods under such perturbations, a model of an array mounted on a flexible structure was presented and studied. This example indicates that our approach is particularly suited for the case where the effect of the modeling errors is of the same magnitude as that due to having only a finite number of noisy observations, i.e. when the variance of the perturbation parameters is  $O(1/N)$ , where  $N$  is the number of samples. We also verified empirically that the MAP-NSF and MAPprox methods do achieve the CRB in the case studied, but the MAPprox method performs better in cases involving small samples and/or "moderately sized" array perturbations. If finite sample effects dominate errors due to model mismatch, the more computationally efficient techniques that only use the nominal array model should be preferred over the approach presented herein. If, on the other hand, the modeling errors are the major source of estimation error, the CRB appears difficult to reach using a MAP approach. For this latter case,

it seems more natural to derive optimal estimators that ignore the finite sample effects, such as those proposed in [4] and [6] for an unstructured perturbation model. Some interesting areas for future research would be to extend the methods of [4] and [6] to the more general perturbation model considered herein, and to theoretically verify the empirical evidence presented in Section VI that suggests the MAPprox and MAP-NSF methods are asymptotically equivalent.

#### APPENDIX A

##### PERTURBATION OF THE NOISE EIGENVECTORS

In this appendix, the calculations leading to (87) are presented. Since the noise eigenvectors (the columns of  $\mathbf{E}_n$ ) correspond to a multiple eigenvalue, they are not uniquely defined. Dubious statements involving consistency of the estimated noise eigenvectors, or functions thereof, will therefore be avoided. However, the projection  $\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^*$  is well-defined, and converges to  $\mathbf{\Pi}^\perp$ , with an error of order  $O_p(1/\sqrt{N})$ . Consider an approximation of the quantity

$$\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A} = (\mathbf{I} - \hat{\mathbf{E}}_s \hat{\mathbf{E}}_s^*) \mathbf{A}. \quad (\text{A.1})$$

Expressing  $\hat{\mathbf{E}}_s = \mathbf{E}_s + \tilde{\mathbf{E}}_s$  and neglecting terms of order  $o_p(1/\sqrt{N})$  gives

$$\begin{aligned} \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A} &\simeq (\mathbf{I} - \tilde{\mathbf{E}}_s \mathbf{E}_s^* - \mathbf{E}_s \tilde{\mathbf{E}}_s^* - \mathbf{E}_s \mathbf{E}_s^*) \mathbf{A} \\ &= -(\tilde{\mathbf{E}}_s \mathbf{E}_s^* + \mathbf{E}_s \tilde{\mathbf{E}}_s^*) \mathbf{A}. \end{aligned} \quad (\text{A.2})$$

Next we make use of the relation

$$\mathbf{I} = \mathbf{E}_s^* \mathbf{E}_s = \hat{\mathbf{E}}_s^* \hat{\mathbf{E}}_s \quad (\text{A.3})$$

which gives

$$\tilde{\mathbf{E}}_s^* \mathbf{E}_s \simeq -\mathbf{E}_s^* \tilde{\mathbf{E}}_s. \quad (\text{A.4})$$

Noting that  $\mathbf{A} = \mathbf{E}_s \mathbf{E}_s^* \mathbf{A}$  then yields

$$\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A} \simeq -(\tilde{\mathbf{E}}_s - \mathbf{E}_s \mathbf{E}_s^* \tilde{\mathbf{E}}_s) \mathbf{E}_s^* \mathbf{A} = -\mathbf{\Pi}^\perp \tilde{\mathbf{E}}_s \mathbf{E}_s^* \mathbf{A}. \quad (\text{A.5})$$

The perturbation of the noise-projected signal eigenvectors is related to the sample covariance in, e.g., [10], and is obtained as

$$\mathbf{\Pi}^\perp \tilde{\mathbf{E}}_s \simeq \mathbf{\Pi}^\perp \hat{\mathbf{R}} \mathbf{E}_s \hat{\mathbf{A}}^{-1}. \quad (\text{A.6})$$

Clearly, (87) follows from (A.5)–(A.6).

#### APPENDIX B

##### ASYMPTOTIC EFFICIENCY OF MAP-NSF ESTIMATE

This appendix contains a proof of Theorem 2. The asymptotic Hessian is readily obtained by differentiating (85) with respect to  $\theta_j$  and letting  $N$  tend to infinity:

$$\mathbf{H}_{ij} = 2\text{Re}\{\mathbf{a}_j^* \mathbf{M} \mathbf{y}_i\}. \quad (\text{B.1})$$

From this expression, it is a simple exercise to verify that  $\mathbf{H} = (N\mathbf{C}\mathbf{R}\mathbf{B}_\theta)^{-1} = \mathbf{C} - \mathbf{F}_\theta^T \mathbf{\Gamma}^{-1} \mathbf{F}_\theta$ .

To calculate  $\mathbf{Q}$ , we first express the sample covariance matrix as

$$\hat{\mathbf{R}} = \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}) \hat{\mathbf{P}} \mathbf{A}^*(\boldsymbol{\theta}_0, \boldsymbol{\rho}) + \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}) \mathbf{Z} + \mathbf{Z}^* \mathbf{A}^*(\boldsymbol{\theta}_0, \boldsymbol{\rho}) + \boldsymbol{\Sigma} \quad (\text{B.2})$$

where we have defined

$$\hat{\mathbf{P}} = \frac{1}{N} \sum_{t=1}^N \mathbf{s}(t) \mathbf{s}^*(t) \quad (\text{B.3})$$

$$\mathbf{Z} = \frac{1}{N} \sum_{t=1}^N \mathbf{s}(t) \mathbf{n}^*(t) \quad (\text{B.4})$$

$$\boldsymbol{\Sigma} = \frac{1}{N} \sum_{t=1}^N \mathbf{n}(t) \mathbf{n}^*(t). \quad (\text{B.5})$$

Note that  $\boldsymbol{\rho}$  in (B.2) is regarded as a stochastic variable with  $\boldsymbol{\rho} - \boldsymbol{\rho}_0 = O_p(1/\sqrt{N})$ . This gives

$$\boldsymbol{\Pi}^\perp \hat{\mathbf{R}} = \boldsymbol{\Pi}^\perp (\tilde{\mathbf{A}} \mathbf{P} \mathbf{A}^* + \mathbf{Z}^* \mathbf{A}^* + \boldsymbol{\Sigma}) + o_p(1/\sqrt{N}) \quad (\text{B.6})$$

where

$$\tilde{\mathbf{A}} = \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}) - \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) \quad (\text{B.7})$$

i.e.,  $\text{vec}(\tilde{\mathbf{A}}) = \mathbf{D}_\rho(\boldsymbol{\rho} - \boldsymbol{\rho}_0) + o_p(1/\sqrt{N})$ . Inserting (87) and (B.6) into (86) gives

$$\hat{\mathbf{M}} \mathbf{a}_0 \simeq -\text{vec}\{\boldsymbol{\Pi}^\perp (\tilde{\mathbf{A}} \mathbf{P} \mathbf{A}^* + \mathbf{Z}^* \mathbf{A}^* + \boldsymbol{\Sigma}) \boldsymbol{\Phi}\} \quad (\text{B.8})$$

where we have defined

$$\boldsymbol{\Phi} = \mathbf{E}_s \tilde{\mathbf{\Lambda}}^{-1} \mathbf{E}_s^* \mathbf{A} \mathbf{U} = \sigma^{-2} \mathbf{E}_s \tilde{\mathbf{\Lambda}} \mathbf{\Lambda}_s^{-1} \mathbf{E}_s^* \mathbf{A}^\dagger. \quad (\text{B.9})$$

Note further from (13) and (26) that  $\mathbf{P} \mathbf{A}^* \boldsymbol{\Phi} = \mathbf{U}$ , which gives  $\mathbf{A}^* \boldsymbol{\Phi} = \mathbf{P}^{-1} \mathbf{U}$ . Using this and (B.8) in (82), we arrive at

$$V_i = V_{i1} + V_{i2} + V_{i3} + o_p(1/\sqrt{N}) \quad (\text{B.10})$$

where

$$V_{i1} = -2\text{Re}\{\text{vec}(\boldsymbol{\Pi}^\perp \tilde{\mathbf{A}} \mathbf{U})^* \mathbf{y}_i\} \quad (\text{B.11})$$

$$V_{i2} = -2\text{Re}\{\text{vec}(\boldsymbol{\Pi}^\perp \mathbf{Z}^* \mathbf{P}^{-1} \mathbf{U})^* \mathbf{y}_i\} \quad (\text{B.12})$$

$$V_{i3} = -2\text{Re}\{\text{vec}(\boldsymbol{\Pi}^\perp \boldsymbol{\Sigma} \boldsymbol{\Phi})^* \mathbf{y}_i\}. \quad (\text{B.13})$$

Note that the ‘‘cross-terms’’ do not contribute to (89), since for  $k \neq l$

$$\lim_{N \rightarrow \infty} NE[V_{ik} V_{il}] = 0. \quad (\text{B.14})$$

Hence, only the ‘‘diagonal terms’’ need to be evaluated.

Let us first rewrite (B.11) using (30) and (46):

$$V_{i1} = -2\text{Re}\{\text{vec}(\tilde{\mathbf{A}})^* \mathbf{M} \mathbf{y}_i\} \simeq -2\tilde{\boldsymbol{\rho}}^T \text{Re}\{\mathbf{D}_\rho^* \mathbf{M} \mathbf{y}_i\}. \quad (\text{B.15})$$

This gives immediately

$$E[V_{i1} V_{j1}] = 4\text{Re}\{\mathbf{y}_j^* \mathbf{M} \mathbf{D}_\rho\} \boldsymbol{\Omega} \text{Re}\{\mathbf{D}_\rho^* \mathbf{M} \mathbf{y}_i\} + o(1/N). \quad (\text{B.16})$$

To evaluate  $E[V_{i2} V_{j2}]$  and  $E[V_{i3} V_{j3}]$ , some useful covariance formulas are stated. Let  $\mathbf{x}$  and  $\mathbf{y}$  be deterministic vectors of appropriate dimensions. Then,

$$E[\text{vec}(\mathbf{Z}^*)^* \mathbf{x} \text{vec}(\mathbf{Z}^*)^* \mathbf{y}] = 0 \quad (\text{B.17})$$

$$E[\mathbf{x}^* \text{vec}(\mathbf{Z}^*) \text{vec}(\mathbf{Z}^*)^* \mathbf{y}] = \frac{\sigma^2}{N} \mathbf{x}^* (\hat{\mathbf{P}}^T \otimes \mathbf{I}) \mathbf{y} \quad (\text{B.18})$$

$$E[\text{vec}(\boldsymbol{\Sigma})^* \mathbf{x} \text{vec}(\boldsymbol{\Sigma})^* \mathbf{y}] = \sigma^4 \mathbf{x}^T \text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})^T \mathbf{y} + \frac{\sigma^4}{N} \mathbf{x}^T \text{btr}\{\text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})^T\} \mathbf{y} \quad (\text{B.19})$$

$$E[\mathbf{x}^* \text{vec}(\boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma})^* \mathbf{y}] = \sigma^4 \mathbf{x}^* \text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})^T \mathbf{y} + \frac{\sigma^4}{N} \mathbf{x}^* \mathbf{y} \quad (\text{B.20})$$

where  $\text{btr}(\mathbf{A})$  refers to the matrix  $\mathbf{A}$  with all blocks transposed. Next, note that for arbitrary scalars  $x$  and  $y$ , we have

$$2\text{Re}(x)\text{Re}(y) = \text{Re}(\bar{x}y + xy). \quad (\text{B.21})$$

Using (30) and (46), (B.12) can be expressed in the form

$$V_{i2} = -2\text{Re}\{\text{vec}(\mathbf{Z}^*)^* \mathbf{M} (\mathbf{P}^{-T} \otimes \boldsymbol{\Pi}^\perp) \mathbf{y}_i\}. \quad (\text{B.22})$$

Applying (B.17)–(B.18) and (B.21) now gives

$$\begin{aligned} E[V_{i2} V_{j2}] &= \frac{2\sigma^2}{N} \text{Re}\{\mathbf{y}_j^* \mathbf{M} (\mathbf{P}^{-T} \otimes \boldsymbol{\Pi}^\perp) \\ &\quad \times (\hat{\mathbf{P}}^T \otimes \mathbf{I}) (\mathbf{P}^{-T} \otimes \boldsymbol{\Pi}^\perp) \mathbf{M} \mathbf{y}_i\} \\ &= \frac{2\sigma^2}{N} \text{Re}\{\mathbf{y}_j^* \mathbf{M} (\mathbf{P}^{-T} \otimes \boldsymbol{\Pi}^\perp) \mathbf{M} \mathbf{y}_i\} + o(1/N). \end{aligned} \quad (\text{B.23})$$

For the last diagonal term we note that

$$(\boldsymbol{\Phi}^T \otimes \boldsymbol{\Pi}^\perp)^T \text{vec}(\mathbf{I}) = \text{vec}\{(\boldsymbol{\Phi} \boldsymbol{\Pi}^\perp)^T\} = 0 \quad (\text{B.24})$$

$$(\boldsymbol{\Phi}^T \otimes \boldsymbol{\Pi}^\perp)^T \text{btr}\{\text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})^T\} = 0 \quad (\text{B.25})$$

which leads to

$$E[V_{i3} V_{j3}] = \frac{2\sigma^4}{N} \text{Re}\{\mathbf{y}_j^* (\boldsymbol{\Phi}^T \otimes \boldsymbol{\Pi}^\perp) (\boldsymbol{\Phi}^{T*} \otimes \boldsymbol{\Pi}^\perp) \mathbf{y}_i\}. \quad (\text{B.26})$$

Applying (B.9) yields

$$\boldsymbol{\Phi}^T \boldsymbol{\Phi}^{T*} = (\boldsymbol{\Phi}^* \boldsymbol{\Phi})^T = (\mathbf{U} \mathbf{A}^* \mathbf{E}_s \tilde{\mathbf{\Lambda}}^{-2} \mathbf{E}_s^* \mathbf{A} \mathbf{U})^T \quad (\text{B.27})$$

which gives

$$\begin{aligned} E[V_{i3} V_{j3}] &= \frac{2\sigma^4}{N} \text{Re}\{\mathbf{y}_j^* \mathbf{M} \left[ (\mathbf{A}^* \mathbf{E}_s \tilde{\mathbf{\Lambda}}^{-2} \mathbf{E}_s^* \mathbf{A})^T \right. \\ &\quad \left. \otimes \boldsymbol{\Pi}^\perp \right] \mathbf{M} \mathbf{y}_i\}. \end{aligned} \quad (\text{B.28})$$

Equation (14) implies

$$\begin{aligned} \mathbf{P}^{-1} + \sigma^2 \mathbf{A}^* \mathbf{E}_s \tilde{\mathbf{\Lambda}}^{-2} \mathbf{E}_s^* \mathbf{A} &= \mathbf{A}^* \mathbf{E}_s (\tilde{\mathbf{\Lambda}}^{-1} + \sigma^2 \tilde{\mathbf{\Lambda}}^{-2}) \mathbf{E}_s^* \mathbf{A} \\ &= \mathbf{A}^* \mathbf{E}_s \tilde{\mathbf{\Lambda}}^{-2} \mathbf{\Lambda}_s \mathbf{E}_s^* \mathbf{A} = \sigma^{-2} \mathbf{U}^{-1}. \end{aligned} \quad (\text{B.29})$$

Combining (B.23) and (B.28)–(B.29) now yields

$$E[V_{i2}V_{j2}] + E[V_{i3}V_{j3}] = \frac{2}{N} \{ \mathbf{y}_j^* \mathbf{M} \mathbf{y}_i \} + o(1/N) \quad (\text{B.30})$$

and thus from (B.16) and (B.30) the asymptotic covariance of the gradient is obtained as

$$\mathbf{Q}_{ij} = \lim_{N \rightarrow \infty} NE[V_i V_j] = 2\text{Re}\{ \mathbf{y}_j^* \mathbf{M} \mathbf{y}_i \} + 4\text{Re}\{ \mathbf{y}_j^* \mathbf{M} \mathbf{D}_\rho \} \bar{\Omega} \text{Re}\{ \mathbf{D}_\rho^* \mathbf{M} \mathbf{y}_i \}. \quad (\text{B.31})$$

Next, we go through some calculations to verify that  $\mathbf{Q}_{ij} = \mathbf{H}_{ij}$ . Using (83) we have

$$\begin{aligned} \text{Re}\{ \mathbf{D}_\rho^* \mathbf{M} \mathbf{y}_j \} &= \text{Re}\{ \mathbf{D}_\rho^* \mathbf{M} (\mathbf{a}_i - \mathbf{D}_\rho \Gamma^{-1} \mathbf{f}_j) \} \\ &= \mathbf{f}_j - \text{Re}\{ \mathbf{D}_\rho^* \mathbf{M} \mathbf{D}_\rho \} \Gamma^{-1} \mathbf{f}_j \\ &= (\mathbf{I} - \text{Re}\{ \mathbf{D}_\rho^* \mathbf{M} \mathbf{D}_\rho \}) \Gamma^{-1} \mathbf{f}_j \\ &= \frac{1}{2} \bar{\Omega}^{-1} \Gamma^{-1} \mathbf{f}_j \end{aligned} \quad (\text{B.32})$$

where in the last step equation (48) is used. The above algebra implies

$$4\text{Re}\{ \mathbf{y}_j^* \mathbf{M} \mathbf{D}_\rho \} \bar{\Omega} \text{Re}\{ \mathbf{D}_\rho^* \mathbf{M} \mathbf{y}_i \} = 2\mathbf{f}_j^T \Gamma^{-1} \text{Re}\{ \mathbf{D}_\rho^* \mathbf{M} \mathbf{y}_i \}. \quad (\text{B.33})$$

Next, note from (B.1) that

$$\begin{aligned} 2\text{Re}\{ \mathbf{y}_j^* \mathbf{M} \mathbf{y}_i \} &= \mathbf{H}_{ij} - 2\text{Re}\{ \mathbf{f}_i^T \Gamma^{-1} \mathbf{D}_\rho^* \mathbf{M} \mathbf{y}_i \} \\ &= \mathbf{H}_{ij} - 2\mathbf{f}_i^T \Gamma^{-1} \text{Re}\{ \mathbf{D}_\rho^* \mathbf{M} \mathbf{y}_i \}. \end{aligned} \quad (\text{B.34})$$

Inserting (B.33)–(B.34) into (B.31) shows that

$$\mathbf{Q}_{ij} = \mathbf{H}_{ij} \quad (\text{B.35})$$

which was the original goal of the proof.

#### REFERENCES

- [1] B. Friedlander, "A sensitivity analysis of the MUSIC algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 10, pp. 1740–1751, Oct. 1990.
- [2] ———, "Sensitivity analysis of the maximum likelihood direction-finding algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 6, pp. 953–968, Nov. 1990.
- [3] F. Li and R. Vaccaro, "Unified analysis for DOA estimation algorithms in array signal processing," *Signal Processing*, vol. 25, pp. 147–169, 1991.
- [4] 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, no. 7, pp. 1758–1774, July 1992.
- [5] F. Li and R. J. Vaccaro, "Sensitivity analysis of DOA estimation algorithms to sensor errors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 708–717, July 1992.
- [6] A. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors—Part II: Multidimensional algorithms," *IEEE Trans. Signal Processing*, vol. 41, no. 9, pp. 1277–1308, Sept. 1993.
- [7] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 331–341, Apr. 1986.
- [8] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, p. 532–544, Apr. 1988.
- [9] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.
- [10] 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, no. 11, pp. 1703–1720, Nov. 1989.
- [11] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.
- [12] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1110–1121, May 1991.
- [13] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp. 590–600, Mar. 1992.
- [14] A. Swindlehurst, "Robust algorithms for direction-finding in the presence of model errors," in *Proc. 5th ASSP Wkshp. Spectral Estimation, Modeling* (Rochester, NY), Oct. 1990, pp. 362–366.
- [15] 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.
- [16] B. Friedlander and A. J. Weiss, "Eigenstructure methods for direction finding with sensor gain and phase uncertainties," in *Proc. IEEE ICASSP*, 1988, pp. 2681–2684.
- [17] 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, no. 12, pp. 1958–1966, Dec. 1989.
- [18] M. Viberg and A. Swindlehurst, "Analysis of the combined effects of finite samples and model errors on array processing performance," in *Proc. IEEE ICASSP* (Minneapolis, MN), Apr. 1993.
- [19] ———, "Analysis of the combined effects of finite samples and model errors on array processing performance," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 3073–3083, Nov. 1994.
- [20] 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.
- [21] ———, "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, no. 6, pp. 724–735, June 1987.
- [22] 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 Wkshp. Spectral Estimation, Modeling* (Minneapolis, MN), Aug. 1988, pp. 98–102.
- [23] J. T.-H. Lo and S. L. Marple, "Observability conditions for multiple signal direction finding and array sensor location," *IEEE Trans. Signal Processing*, vol. 40, no. 11, pp. 2641–2650, Nov. 1992.
- [24] B. Wahlberg, B. Ottersten, and M. Viberg, "Robust signal parameter estimation in the presence of array perturbations," in *Proc. IEEE ICASSP* (Toronto), May 1991.
- [25] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 7, pp. 1132–1143, July 1990.
- [26] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Wkshp.* (Rome, NY), 1979, pp. 243–258.
- [27] G. Bienvenu and L. Kopp, "Principe de la goniometrie passive adaptive," in *Proc. 7eme Colloque GRESIT* (Nice, France), 1979, pp. 106/1–106/10.
- [28] J. F. Böhrne, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Processing*, vol. 10, pp. 329–337, 1986.
- [29] 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: Springer-Verlag, 1993, pp. 99–151.
- [30] A. Graham, *Kronecker Products and Matrix Calculus with Applications*. Chichester, UK: Ellis Horwood, 1981.
- [31] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester, UK: Wiley, 1988.
- [32] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. London: Academic, 1981.
- [33] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Vol. I. New York: Wiley, 1968.
- [34] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.



**Mats Viberg** (S'87-M'90) was born in Linköping, Sweden, on December 21, 1961. He received the M.S. degree in applied mathematics in 1985, the Lic. Eng. degree in 1987, and the Ph.D. degree in electrical engineering in 1989, all from Linköping University, Sweden.

He joined the Division of Automatic Control at the Department of Electrical Engineering, Linköping University, in 1984, and from November 1989 until August 1993, he was a Research Associate. From October 1988 to March 1989, he was on leave

at the Information Systems Laboratory, Stanford University, as a Visiting Scholar. From August 1992 until August 1993, he held a Fulbright-Hayes grant scholarship as a Visiting Researcher at the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT, and at the Information Systems Laboratory, Stanford University. Since September 1993, he has been a Professor of Signal Processing at the Department of Applied Electronics, Chalmers University of Technology, Sweden. His research interests are in statistical signal processing and its application to sensor array signal processing, system identification, and communication.

Dr. Viberg received the IEEE Signal Processing Society's 1993 Paper Award (Statistical Signal and Array Processing Area), for the paper "Sensor Array Processing Based on Subspace Fitting," which he coauthored with B. Ottersten.



**A. Lee Swindlehurst** was born in Boulder City, NV, USA, in 1960. He received the B.S. degree, summa cum laude, and the M.S. degree in electrical engineering from Brigham Young University, Provo, UT, in 1985 and 1986, respectively, and the Ph.D. degree in electrical engineering from Stanford University in 1991.

From 1983 to 1984, he was employed with Eyring Research Institute, Provo, UT, as a scientific programmer, writing software for a Minuteman missile simulation system. During 1984-1986, he was a

Research Assistant in the Department of Electrical Engineering at Brigham Young University. He was awarded an Office of Naval Research Graduate Fellowship for 1985-1988, and during that time, he was affiliated with the Information Systems Laboratory at Stanford University. From 1986-1990, he was also employed at ESL, Inc., of Sunnyvale, CA, where he was involved in the design of algorithms and architectures for a variety of radar and sonar signal processing systems. He joined the faculty of the Department of Electrical and Computer Engineering at Brigham Young University in 1990, where he is currently an Assistant Professor. His research interests include sensor array signal processing, detection and estimation theory, system identification, and control.