

# Analysis of the Combined Effects of Finite Samples and Model Errors on Array Processing Performance

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

**Abstract**—The principal sources of estimation error in sensor array signal processing applications are the finite sample effects of additive noise and imprecise models for the antenna array and spatial noise statistics. While the effects of these errors have been studied individually, their combined effect has not yet been rigorously analyzed. In this paper, we undertake such an analysis for the class of so-called *subspace fitting* algorithms. In addition to deriving first-order asymptotic expressions for the estimation error, we show that an overall optimal weighting exists for a particular array and noise covariance error model. In a companion paper, the optimally weighted subspace fitting method is shown to be asymptotically equivalent with the more complicated maximum *a posteriori* estimator. Thus, for the model in question, no other method can yield more accurate estimates for large samples and small model errors. Numerical examples and computer simulations are included to illustrate the obtained results and to verify the asymptotic analysis for realistic scenarios.

## I. INTRODUCTION

**D**URING the past decade, a number of so-called “high-resolution” subspace-based algorithms for array signal processing and parameter estimation have been introduced. Most of these techniques are presented in the context of estimating the directions-of-arrival (DOA’s) of multiple co-channel signals using an array of sensors. In recent years, attention has shifted from new algorithm development to algorithm performance analyses and comparisons. A number of research contributions have considered the asymptotic effects of additive noise on DOA estimation performance, assuming that the array response and noise model are perfectly known [1]–[4]. Conversely, others have investigated the effects of an imprecisely known array response and noise model, while ignoring the finite sample effects of noise [5]–[8]. The combined effects of additive noise and modelling errors are considered in [9] and [10] but only for the high signal-to-noise ratio (SNR) case.

In any practical situation, all of the above error sources will be present simultaneously. Not only is one forced to estimate the DOA’s using only a finite amount of noisy data, but the ar-

ray does not respond as expected, and the spatial characteristics of the noise field are not well understood. More precisely, the array element positions may not be accurately known, the gain and phase response of a given sensor may vary as a function of its surroundings, imprecise interpolation techniques must be used with arrays calibrated at discrete angles, the noise field may change with time, reverberation, channel cross-talk, or spatially distributed sources may be present, etc. Together with the noise itself, these sources of error will obviously degrade algorithm performance, although the amount of degradation will depend on the relative contribution of the noise and model errors.

Some approaches have been proposed in the literature to mitigate the effects of modeling errors. Optimal techniques assuming unstructured perturbations and an infinite number of snapshots are derived in [6] and [8]. In [11] and [12], a method that is robust against an unknown noise field is suggested, assuming that no information about the perturbation structure is available. Examples of techniques for structured perturbation models include the so-called auto-calibration techniques considered in [13]–[15] and the maximum *a posteriori* (MAP) approach [16], [17]. The methods considered herein belong to the class of *signal subspace fitting* (SSF) algorithms [4]. These methods have been shown to be first-order equivalent to many other parametric techniques for particular choices of a certain weighting matrix. Optimal choices of the weighting matrix have been derived in [4] and [8] for cases involving finite sample effects (FSE) or modeling errors (ME), respectively, and the resulting SSF techniques have been shown to give the best possible performance of *any* estimator under their respective assumptions.

The goal of this paper is to extend earlier results on algorithm performance analysis to the more realistic case where both modeling errors and finite sample effects are present. Not surprisingly, it is found that the estimation error covariance is, up to first order, a sum of the individual contributions from the two types of errors. It is shown that an overall optimal subspace weighting for the SSF class of methods exists for a particular class of modeling errors. This model assumes 1) additive random array response errors that are statistically uniform from sensor to sensor, but possibly correlated between different DOA’s (with known correlation), and 2) additive random perturbations to the noise covariance that are independent from element to element. The SSF method using the overall optimal weighting will outperform all other SSF techniques when considering both modeling errors and finite sample effects. Indeed, in the companion

Manuscript received December 3, 1992; revised May 10, 1994. The associate editor coordinating the review of this paper and approving it for publication was Prof. Daniel Fuhrmann. This work was supported by the National Science Foundation under grant MIP-9110112, by the SDIO/IST Program managed by the Army Research Office under Contract DAAL03-90-G-0108, and by the Advanced Research Projects Agency of the Department of Defense monitored by the Air Force Office of Scientific Research under Contract F49620-91-C-0086.

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

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

IEEE Log Number 9404786.

1053-587X/94\$04.00 © 1994 IEEE

paper [17], the optimally weighted SSF technique is found to be asymptotically equivalent with the more complicated MAP estimator [16] for the problem at hand. Thus, the covariance of the asymptotic distribution of the estimation error coincides with the Cramér-Rao bound (CRB) for Gaussian signals, perturbations, and noise, and hence, no other technique can give lower first-order error variance.

We begin in the next section by introducing the details of the problem considered, including the nominal data model as well as the models used for the array response and noise covariance errors. The class of SSF methods is also briefly reviewed. Section III contains the combined performance analysis, and Section IV presents the optimal SSF weightings. Some numerical examples and computer simulations are included in the final section. The performance of the overall optimal SSF method is compared with that resulting from the FSE- and ME-only optimal weightings. Comparisons with the appropriate CRB are also made for a particular case involving nonuniform sensor position errors (a simulated towed array).

## II. PROBLEM FORMULATION

### A. Nominal Data Model

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

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

The columns of the  $m \times d$  matrix  $\mathbf{A}$  are the so-called *array propagation vectors*, which are denoted  $\mathbf{a}(\theta_i)$ ,  $i = 1, \dots, d$ . These vectors are functions of the *signal parameters* and describe the array response to a unit waveform with parameter(s)  $\theta_i$ . Although not necessary, we shall assume that  $\theta_i$  is a real-valued scalar referred to as the  $i$ th DOA. The components of the  $d$ -vector  $\theta$  are the model DOA's, whereas the vector  $\theta_0$  contains their true values. It is assumed that  $\theta_0$  is an inner point of the parameter set of interest and that  $\mathbf{A}(\theta)$  is differentiable over this set.

The  $d$ -vector  $\mathbf{s}(t)$  is composed of the complex 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 the 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.

The signal waveforms are regarded as deterministic (i.e., fixed) sequences such that the following limit exists:

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

where  $\mathbf{X}^*$  denotes the complex conjugate transpose of  $\mathbf{X}$ . On the contrary, the noise term  $\mathbf{n}(t)$  is modeled as a stationary,

complex Gaussian random process that is uncorrelated with the signals. The noise has zero mean and is assumed to be both spatially and temporally white, i.e.

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

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

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

It should be noted that the analysis performed under the above model remains valid if the signal waveforms happen to be realizations of some stochastic process. The corresponding stochastic limit in (2) is then required to hold with probability one. See Chapter 2 of [18] for more details on this topic.

### B. Perturbation Models

The exact parametrization of the array propagation vectors is unknown in any practical situation. Thus, the available model  $\mathbf{a}(\theta)$  may differ from the "true" propagation vector. It will be assumed that the data have actually been generated by the equation

$$\begin{aligned} X(t) &= [\mathbf{a}(\theta_1) + \tilde{\mathbf{a}}(\theta_1)] \cdots [\mathbf{a}(\theta_d) + \tilde{\mathbf{a}}(\theta_d)] \begin{bmatrix} s_1(t) \\ \vdots \\ s_d(t) \end{bmatrix} + \mathbf{n}(t) \\ &= (\mathbf{A} + \tilde{\mathbf{A}})\mathbf{s}(t) + \mathbf{n}(t). \end{aligned} \quad (5) \quad (6)$$

In some applications, it is conceivable that the physical origin of the array model uncertainty could be precisely characterized. The perturbations may be due to sensor position errors, gain errors, phase errors, mutual coupling between sensors, receiver fluctuations due to temperature and humidity, quantization effects, etc. It is, in principle, possible to explain the effect of each of these error sources from physical insight, thus leading to a model where the propagation vectors are parametrized by the DOA's along with a set of extra "perturbation parameters." However, in a practical application, all of the above-mentioned phenomena (along with several others) are likely to be present simultaneously. Clearly, a model based on physical insight is impractical in such a case.

A pragmatic remedy to this situation is simply to assume that the array response is a random quantity, whose mean value is the known nominal model. Thus, we assume herein that the array propagation errors are random with zero mean and second-order moments

$$E[\tilde{\mathbf{a}}(\theta_i)\tilde{\mathbf{a}}^*(\theta_j)] = \nu_{ij}\mathbf{B} \quad (7)$$

$$E[\tilde{\mathbf{a}}(\theta_i)\tilde{\mathbf{a}}^T(\theta_j)] = 0. \quad (8)$$

The sensor-to-sensor covariances are collected in the matrix  $\mathbf{Y} = \{\nu_{ij}\}$ . Both  $\mathbf{B}$  and  $\mathbf{Y}$  are assumed to be available to the user, e.g., from system performance specifications. Note that the error model (7) allows for direction-dependent modeling errors, but the sensor-to-sensor correlation (if any) is independent of  $\theta$ .

Perturbation models similar to (7) and (8) have been used by a number of others, primarily in the analysis of adaptive beamforming algorithms [19]–[21]. The model presented here

is actually somewhat more general than what was assumed in these papers since we allow for some degree of angle-to-angle and sensor-to-sensor correlation. To connect the nonphysical model of (7) and (8) with the more intuitive idea of gain and phase perturbations to the array, let the nominal response of the  $k$ th sensor in the direction  $\theta$  be  $a_k(\theta) = e^{j\phi_k}$ , and let  $\tilde{g}_k$  and  $\tilde{\phi}_k$  represent the corresponding gain and phase perturbations. Then, to first order, we have

$$\tilde{a}_k(\theta) = (1 + \tilde{g}_k)e^{j(\phi_k + \tilde{\phi}_k)} \simeq a_k(\theta) + \tilde{a}_k(\theta)$$

where

$$\tilde{a}_k(\theta) = (\tilde{g}_k + j\tilde{\phi}_k)e^{j\phi_k}.$$

If we let  $\sigma_g^2$  and  $\sigma_\phi^2$  represent the variances of  $\tilde{g}_k$  and  $\tilde{\phi}_k$ , respectively, and if we assume that the gain and phase perturbations are zero-mean and independent, then

$$\begin{aligned} E[|\tilde{a}_k|^2] &= \sigma_g^2 + \sigma_\phi^2 \\ E[\tilde{a}_k^2] &= (\sigma_g^2 - \sigma_\phi^2)e^{j2\phi}. \end{aligned}$$

Thus, the model of (7) and (8) simply amounts to assuming that  $\nu_{ii} = \sigma_g^2 + \sigma_\phi^2$  for the  $i$ th DOA and that the gain and phase errors are roughly “of the same order” ( $\sigma_g^2$  and  $\sigma_\phi^2$  are equal). A similar result also holds for correlations between the errors on different sensors and in different directions.

The reason for a *random* perturbation model as opposed to a *deterministic* one lies in the consideration of how one chooses to quantify the effects of the perturbation. In a given fixed scenario, of course, the presence of array errors will introduce a bias in the DOA estimates. Presumably, if one wanted to measure the magnitude of this bias, it would simply be a matter of directly computing the limiting ( $N \rightarrow \infty$ ) estimate  $\hat{\theta}$  and then subtracting  $\theta_0$ . This procedure would obviously have to be repeated for each perturbation scenario considered since the bias would be different in each case. The advantage of using a random model is that one can obtain a measure of the *average* effect of the array errors on estimation performance, which is measured now in terms of *variance* rather than *bias*, without being forced to adopt a particular perturbation scenario (which may be no more representative than any other similar perturbation).

The covariance structure in (7) and (8) also allows one to optimize performance for the proposed SSF estimation method. However, as will be demonstrated in Section V, one can often use the resulting optimized SSF approach to also improve the performance for more complicated perturbation models. More precisely, Example 5.2 shows how the sensitivity to underestimation of the number of signals can be reduced, and Example 5.3 deals with the case of unknown sensor positions. We should also remark that (7) and (8) is indeed a reasonable model in the common case of an experimentally calibrated array, where the sources of error may be quantization errors in collecting the calibration data, interpolation errors in using a calibration grid, etc.

The effects of errors in the noise model on algorithm performance will also be studied. As with the array errors, we will model the perturbation to the noise covariance as a random variable with given moments. It is most convenient

to specify the conditional mean and covariances of the noise given  $\tilde{\Sigma}$ , the perturbation<sup>1</sup>

$$E[\mathbf{n}(t)|\tilde{\Sigma}] = 0 \quad (9)$$

$$E[\mathbf{n}(t)\mathbf{n}^T(s)|\tilde{\Sigma}] = 0 \quad (10)$$

$$E[\mathbf{n}(t)\mathbf{n}^*(s)|\tilde{\Sigma}] = (\sigma^2\mathbf{I} + \tilde{\Sigma})\delta_{t,s}. \quad (11)$$

Other than being Hermitian,  $\tilde{\Sigma}$  is treated as a random matrix with independent elements  $\tilde{\sigma}_{ij}$  of equal variance:

$$E[\tilde{\sigma}_{ij}] = 0 \quad (12)$$

$$E[\tilde{\sigma}_{ij}\tilde{\sigma}_{kl}] = \mu^2\delta_{i,l}\delta_{j,k}. \quad (13)$$

Thus, the real diagonal elements of  $\tilde{\Sigma}$  are independent of all other elements, whereas the off-diagonal terms are correlated only with their conjugate image. The off-diagonal elements are assumed to be circularly symmetric, i.e., the real and imaginary parts are uncorrelated and of equal variance  $\mu^2/2$ . It is assumed that  $\mu^2$  is known, whereas  $\sigma^2$  may be unknown. Note from (11) and (12) that

$$E[\mathbf{n}(t)\mathbf{n}^*(s)] = E\left[E\left\{\mathbf{n}(t)\mathbf{n}^*(s) \mid \tilde{\Sigma}\right\}\right] = \sigma^2\mathbf{I}\delta_{t,s} \quad (14)$$

i.e., the total covariance matrix of the noise is identical to the nominal spatially white covariance. However, the higher order moments of the noise are altered by the perturbation model, which will be seen to affect the behavior of the estimation techniques under consideration.

The perturbations of the nominal array propagation and noise models introduce a bias in the estimates. In other words, it is not possible to determine the DOA's exactly even from an infinite collection of observation vectors, i.e., as  $N \rightarrow \infty$ . However, the goal of our study is to make a mathematically consistent first-order analysis of the DOA estimation error in the presence of *both* finite sample and model errors. Thus, 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  $\nu_{ii} + \mu^2$  in the model-error-only case. To make the relative contribution of the two error sources of comparable magnitude, we introduce the artifice of expressing the array perturbation variances as

$$\Upsilon = \tilde{\Upsilon}/N \quad (15)$$

where  $\tilde{\Upsilon}$  is independent of  $N$ , and similarly

$$\mu^2 = \tilde{\mu}^2/N. \quad (16)$$

An asymptotic performance analysis is then carried out assuming  $N \rightarrow \infty$ . This leads to expressions for the asymptotic covariance matrix of the limiting distribution of the estimation error involving the combined effects of finite samples and modeling errors.

One could alternatively assume that  $N$  is “large” and that  $\nu_{ii} + \mu^2$  is “small” and perform a first-order analysis in  $1/N$

<sup>1</sup>The model used here is slightly different but no less general than that assumed in [6] and [8], where the conditional covariance of the noise in (11) was assumed to be  $\sigma^2(\mathbf{I} + \tilde{\Sigma})$ .

and  $\nu_{ii} + \mu^2$  simultaneously. This approach can be shown to yield results identical to those obtained with the more artificial assumption that  $\Upsilon$  and  $\mu$  are independent of  $N$ . However, such a first-order perturbation analysis is quite heuristic and does not allow for mathematically precise statements involving the limiting distribution (or second-order moments) of the estimation error since this distribution necessarily depends on the relative sizes of the different sources of errors. Indeed, if  $\Upsilon$  and  $\mu^2$  are both  $o(1/N)$ , the limiting distribution is independent of the modeling errors, and the analysis of [4] applies. On the other hand, if  $1/N \ll \nu_{ii}^2 + \mu^2$ , the finite sample errors are irrelevant, and the analysis of [8] is applicable. By adopting the assumptions in (15) and (16), we are simply restricting ourselves in this paper to the "in between" cases where neither finite sample effects or model errors dominate. Further comments on this issue are provided in Remark 3.1 of Section III-A.

### C. Subspace Fitting Methods

Most parametric estimation methods depend on the measurements only through the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^N \mathbf{x}(t)\mathbf{x}^*(t). \quad (17)$$

Under the stated assumptions,  $\hat{\mathbf{R}}$  takes the form

$$\hat{\mathbf{R}} = (\mathbf{A} + \tilde{\mathbf{A}})\hat{\mathbf{P}}(\mathbf{A} + \tilde{\mathbf{A}})^* + (\mathbf{A} + \tilde{\mathbf{A}})\mathbf{Z} + \mathbf{Z}^*(\mathbf{A} + \tilde{\mathbf{A}})^* + \Sigma \quad (18)$$

where we have defined

$$\hat{\mathbf{P}} = \frac{1}{N} \sum_{t=1}^N \mathbf{s}(t)\mathbf{s}^*(t) \quad (19)$$

$$\mathbf{Z} = \frac{1}{N} \sum_{t=1}^N \mathbf{s}(t)\mathbf{n}^*(t) \quad (20)$$

$$\Sigma = \frac{1}{N} \sum_{t=1}^N \mathbf{n}(t)\mathbf{n}^*(t). \quad (21)$$

As  $N \rightarrow \infty$ ,  $\hat{\mathbf{R}}$  converges (with probability 1) in the absence of model errors to the limit

$$\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^* + \sigma^2\mathbf{I}. \quad (22)$$

The rank of the signal covariance matrix  $\mathbf{P}$  is denoted  $d'$ . Let the eigendecomposition of  $\mathbf{R}$  be given by

$$\mathbf{R} = \sum_{k=1}^m \lambda_k \mathbf{e}_k \mathbf{e}_k^* = \mathbf{E}_s \Lambda_s \mathbf{E}_s^* + \mathbf{E}_n \Lambda_n \mathbf{E}_n^* \quad (23)$$

where  $\Lambda_s$  is a diagonal matrix containing the  $d'$  largest eigenvalues, and the columns of  $\mathbf{E}_s$  are the corresponding eigenvectors. Similarly,  $\Lambda_n$  and  $\mathbf{E}_n$  are composed of the smallest eigenvalues and their corresponding eigenvectors. From (22), it is easy to see that the  $m-d'$  columns of  $\mathbf{E}_n$  are all orthogonal to the matrix  $\mathbf{A}\mathbf{P}\mathbf{A}^*$ , and they correspond to the multiple eigenvalue  $\sigma^2$ , i.e.,  $\Lambda_n = \sigma^2\mathbf{I}$ . Since  $\mathbf{E}_n$  and  $\mathbf{E}_s$

are orthogonal, it follows that there exists a full rank  $d \times d'$  matrix  $\mathbf{T}$  such that

$$\mathbf{E}_s = \mathbf{A}(\theta_0)\mathbf{T}. \quad (24)$$

Equation (24) is the key observation for all subspace-based techniques, including those that fall in the class of signal subspace fitting (SSF) methods [4]. In the SSF approach, an estimate of  $\mathbf{E}_s$  is obtained from the eigendecomposition of the sample covariance matrix

$$\hat{\mathbf{R}} = \hat{\mathbf{E}}_s \hat{\Lambda}_s \hat{\mathbf{E}}_s^* + \hat{\mathbf{E}}_n \hat{\Lambda}_n \hat{\mathbf{E}}_n^* \quad (25)$$

and the following weighted least-squares problem based on (24) is solved to estimate  $\theta$ :

$$\{\hat{\theta}, \mathbf{T}\} = \arg \min_{\theta, \mathbf{T}} \|\mathbf{W}_r [\hat{\mathbf{E}}_s - \mathbf{A}(\theta)\mathbf{T}] \mathbf{W}_c^{1/2}\|_F^2. \quad (26)$$

The weighting matrices  $\mathbf{W}_r$  and  $\mathbf{W}_c$  are assumed to be positive definite and Hermitian. Roughly speaking, the role of the weighting matrices is to whiten the equation error  $\hat{\mathbf{E}}_s - \mathbf{A}\mathbf{T}$ . The row weighting  $\mathbf{W}_r$  takes care of nonuniform modeling errors (it does not appear in the finite-sample-only case), whereas the column weighting  $\mathbf{W}_c$  essentially contains the inverse variances of the eigenvector estimates. The SSF minimization of (26) is often written in a more compact form by solving for the linear parameter  $\mathbf{T}$ :

$$\hat{\theta} = \arg \min_{\theta} V(\theta) \quad (27)$$

$$V(\theta) = \text{Tr}\{\mathbf{\Pi}^\perp \mathbf{W}_r \hat{\mathbf{E}}_s \mathbf{W}_c \hat{\mathbf{E}}_s^* \mathbf{W}_r\} \quad (28)$$

where  $\mathbf{\Pi}^\perp$  denotes the projector onto the nullspace of the matrix  $(\mathbf{W}_r \mathbf{A})^*$ , i.e.,

$$\mathbf{\Pi}^\perp = \mathbf{I} - \mathbf{\Pi} \mathbf{p} = \mathbf{I} - \mathbf{G}(\mathbf{G}^* \mathbf{G})^{-1} \mathbf{G}^* = \mathbf{I} - \mathbf{G} \mathbf{G}^\dagger \quad (29)$$

$$\mathbf{G} = \mathbf{W}_r \mathbf{A}. \quad (30)$$

As detailed in [4], a number of important algorithms can be cast in the SSF framework by making appropriate choices for  $\mathbf{W}_r$  and  $\mathbf{W}_c$ . In particular, it has been shown that for the finite-sample-only case, the weighting matrices

$$\mathbf{W}_r = \mathbf{I} \quad (31)$$

$$\mathbf{W}_c = \tilde{\Lambda}^2 \Lambda_s^{-1} \quad (32)$$

where

$$\tilde{\Lambda} = \Lambda_s - \sigma^2 \mathbf{I} \quad (33)$$

yield an algorithm that is asymptotically (large  $N$ ) equivalent to the maximum likelihood method, assuming stochastic Gaussian emitter signals. Hence, the DOA estimates obtained from (28) using the weighting matrices (31) and (32) asymptotically achieve the Cramér-Rao bound and thus have minimum variance. In a similar fashion, different sets of optimal weighting matrices have been derived assuming array and noise model errors are present and  $N \rightarrow \infty$  [8]. The development of an optimal SSF algorithm that takes both finite sample effects and model errors into account requires a combined performance analysis that assumes both error sources are present. Such an analysis is conducted in the next section. Optimal weighting matrices are then derived in Section IV for the combined case.

### III. PERFORMANCE ANALYSIS

It is well known that in the absence of modeling errors, the SSF estimate obtained from (28) converges to the true value (w.p.1) as  $N \rightarrow \infty$ . The convergence of the SSF estimate when both  $N \rightarrow \infty$  and  $\nu_{ii} + \mu^2 \rightarrow 0$  readily follows. Thus, we conclude that the estimate is strongly consistent under the mathematical assumptions (15) and (16). A useful measure for predicting the accuracy in “large enough” sample sizes (and, in this case, assuming “small enough” modeling errors) is the asymptotic covariance matrix of the parameter estimates. Under mild conditions, this is the same as the covariance of the limiting distribution of the estimation error. Note that this distribution is not necessarily Gaussian since no assumptions on the distributions of the perturbation terms are made.

#### A. Asymptotic Accuracy

Let  $\hat{\theta}$  minimize the criterion function (28). Then,  $V'(\hat{\theta}) = 0$ , where  $V'$  denotes the gradient. Since the estimate is consistent, a standard Taylor series expansion of this equation yields

$$0 = V'(\hat{\theta}) \simeq V'(\theta_0) + V''(\theta_0)(\hat{\theta} - \theta_0) \simeq V'(\theta_0) + \mathbf{H}\tilde{\theta} \quad (34)$$

where  $\mathbf{H}$  denotes the limiting ( $N \rightarrow \infty$ ) Hessian

$$\mathbf{H} = \lim_{N \rightarrow \infty} V''(\theta_0) \quad (\text{w.p.1}). \quad (35)$$

We use the symbol  $\simeq$  to specify that terms of order  $o_p(1/\sqrt{N})$  are neglected. The notation  $O_p(\cdot)$  and  $o_p(\cdot)$  represent “in probability” counterparts of the corresponding deterministic notation (see Section 2.9 of [22]). From (34), the estimation error can be expressed as  $\tilde{\theta} = \hat{\theta} - \theta_0 \simeq -\mathbf{H}^{-1}V'$ . Let  $\mathbf{Q}$  denote the limiting normalized covariance of the gradient

$$\mathbf{Q} = \lim_{N \rightarrow \infty} NE[V'(\theta_0)V'^T(\theta_0)]. \quad (36)$$

Then, the covariance of the asymptotic distribution of the estimation error is

$$E[\tilde{\theta}\tilde{\theta}^T] = \mathbf{C}/N \quad (37)$$

$$\mathbf{C} = \mathbf{H}^{-1}\mathbf{Q}\mathbf{H}^{-1}. \quad (38)$$

The asymptotic Hessian depends only on the limiting sample covariance and its eigendecomposition and is obtained as in the separate error source cases [4], [8]:

$$\mathbf{H} = 2 \operatorname{Re}\{(\mathbf{D}^* \mathbf{W}_r \mathbf{\Pi}^\perp \mathbf{W}_r \mathbf{D}) \odot (\mathbf{T} \mathbf{W}_c \mathbf{T}^*)^T\} \quad (39)$$

where  $\operatorname{Re}\{\}$  denotes the real part,  $\odot$  means elementwise multiplication, and

$$\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_d] \quad (40)$$

$$\mathbf{d}_i = \frac{d\mathbf{a}(\theta_i)}{d\theta_i} \quad (41)$$

$$\mathbf{A}^\dagger = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* \quad (42)$$

$$\mathbf{T} = \mathbf{A}^\dagger \mathbf{E}_s. \quad (43)$$

To evaluate (36), the derivative of the criterion function is first calculated. Following [4], we have

$$\frac{dV}{d\theta_i} = V_i = -2 \operatorname{Re} \left[ \operatorname{Tr} \{ \mathbf{G}_i^* \mathbf{\Pi}^\perp \mathbf{W}_r \hat{\mathbf{E}}_s \mathbf{W}_c \hat{\mathbf{E}}_s^* \mathbf{W}_r \mathbf{G}_i^\dagger \} \right]. \quad (44)$$

where  $\mathbf{G}_i$  is the derivative of  $\mathbf{G}$  with respect to  $\theta_i$  and  $\mathbf{G}_i^\dagger = \mathbf{G}(\mathbf{G}^* \mathbf{G})^{-1}$ .

Since  $\mathbf{\Pi}^\perp \mathbf{W}_r \mathbf{E}_s = 0$ , we have  $\mathbf{\Pi}^\perp \mathbf{W}_r \hat{\mathbf{E}}_s \simeq O_p(1/\sqrt{N})$ , which gives

$$V_i \simeq -2 \operatorname{Re} \left[ \operatorname{Tr} \{ \mathbf{G}_i^* \mathbf{\Pi}^\perp \mathbf{W}_r \hat{\mathbf{E}}_s \mathbf{W}_c \mathbf{E}_s^* \mathbf{W}_r \mathbf{G}_i^\dagger \} \right]. \quad (45)$$

Next, note that only the  $i$ th column of  $\mathbf{G}_i$ , which is equal to  $\mathbf{W}_r \mathbf{d}_i$ , is nonzero. This observation gives

$$V_i \simeq -2 \operatorname{Re} \{ \mathbf{d}_i^* \mathbf{W}_r \mathbf{\Pi}^\perp \mathbf{W}_r \hat{\mathbf{E}}_s \mathbf{W}_c \mathbf{E}_s^* \mathbf{W}_r \{ \mathbf{G}_i^\dagger \}_{:,i} \} \quad (46)$$

where  $\{ \mathbf{G}_i^\dagger \}_{:,i} = \mathbf{G} \{ (\mathbf{G}^* \mathbf{G})^{-1} \}_{:,i}$  is the  $i$ th column of  $\mathbf{G}_i^\dagger$ . The first-order relation between the estimated signal eigenvectors and the sample covariance matrix is (e.g., see [8])

$$\mathbf{\Pi}^\perp \mathbf{W}_r \hat{\mathbf{E}}_s \simeq \mathbf{\Pi}^\perp \mathbf{W}_r \hat{\mathbf{R}} \mathbf{E}_s \tilde{\mathbf{\Lambda}}^{-1}. \quad (47)$$

Using (18) and noting that  $\mathbf{\Pi}^\perp \mathbf{W}_r \mathbf{A} = 0$  and  $\mathbf{Z} \simeq O_p(1/\sqrt{N})$ , (47) leads to

$$\mathbf{\Pi}^\perp \mathbf{W}_r \hat{\mathbf{E}}_s \simeq \mathbf{\Pi}^\perp \mathbf{W}_r (\tilde{\mathbf{A}} \mathbf{P} \mathbf{A}^* + \mathbf{Z}^* \mathbf{A}^* + \mathbf{\Sigma}) \mathbf{E}_s \tilde{\mathbf{\Lambda}}^{-1}. \quad (48)$$

Inserting (48) into (45) gives

$$V_i \simeq V_{i1} + V_{i2} + V_{i3} \quad (49)$$

where the approximate derivative of the criterion function is expressed as the sum of three terms, the first stemming from the array modeling errors and the last two from the effects of noise. These terms take the form

$$V_{i1} = -2 \operatorname{Re} \{ \alpha_i^* \tilde{\mathbf{A}} \mathbf{P} \mathbf{A}^* \beta_i \} \quad (50)$$

$$V_{i2} = -2 \operatorname{Re} \{ \alpha_i^* \mathbf{Z}^* \mathbf{A}^* \beta_i \} \quad (51)$$

$$V_{i3} = -2 \operatorname{Re} \{ \alpha_i^* \mathbf{\Sigma} \beta_i \} \quad (52)$$

where  $m\beta_i \alpha_i$  and  $\beta_i$  are the deterministic vectors

$$\alpha_i = \mathbf{W}_r \mathbf{\Pi}^\perp \mathbf{W}_r \mathbf{d}_i \quad (53)$$

$$\beta_i = \mathbf{E}_s \tilde{\mathbf{\Lambda}}^{-1} \mathbf{W}_c \mathbf{E}_s^* \mathbf{W}_r \{ \mathbf{G}_i^\dagger \}_{:,i}. \quad (54)$$

It is not difficult to see that the expected values of the crossterms ( $E[V_{i1}V_{j2}]$ , etc.) are zero, and hence, they do not contribute to  $\mathbf{Q}$ . The terms  $E[V_{ik}V_{jk}]$  are evaluated in Appendix A, leading to the following result.

*Theorem 1:* Let  $\hat{\theta}$  be obtained from (28), and let the perturbation models of Section II-B hold. The covariance of the asymptotic distribution of the estimation error is then the sum of the error covariances derived in [4] and [8] for each of the three sources of error considered separately. For large  $N$ , we have

$$E[\tilde{\theta}\tilde{\theta}^T] = \mathbf{C}/N \quad (55)$$

$$\mathbf{C} = \mathbf{H}^{-1}(\mathbf{Q}_{AP} + \mathbf{Q}_{NP} + \mathbf{Q}_{FS})\mathbf{H}^{-1} \quad (56)$$

$$= \mathbf{C}_{AP} + \mathbf{C}_{NP} + \mathbf{C}_{FS}. \quad (57)$$

An expression for the asymptotic Hessian matrix  $\mathbf{H}$  is given in (39). The effects of array perturbations (AP), noise covariance

perturbations (NP), and finite samples (FS) are expressed separately as

$$\mathbf{Q}_{AP} = 2 \operatorname{Re}\{(\mathbf{D}^* \mathbf{W}_r \mathbf{\Pi}^\perp \mathbf{W}_r \mathbf{B} \mathbf{W}_r \mathbf{\Pi}^\perp \mathbf{W}_r \mathbf{D}) \odot (\mathbf{T} \mathbf{W}_c \mathbf{T}^* \mathbf{\Upsilon}^T \mathbf{T} \mathbf{W}_c \mathbf{T}^*)^T\} \quad (58)$$

$$\mathbf{Q}_{NP} = 2\bar{\mu}^2 \operatorname{Re}\{(\mathbf{D}^* \mathbf{W}_r \mathbf{\Pi}^\perp \mathbf{W}_r^2 \mathbf{\Pi}^\perp \mathbf{W}_r \mathbf{D}) \odot (\mathbf{T} \mathbf{W}_c \tilde{\mathbf{\Lambda}}^{-2} \mathbf{W}_c \mathbf{T}^*)^T\} \quad (59)$$

$$\mathbf{Q}_{FS} = 2\sigma^2 \operatorname{Re}\{(\mathbf{D}^* \mathbf{W}_r \mathbf{\Pi}^\perp \mathbf{W}_r^2 \mathbf{\Pi}^\perp \mathbf{W}_r \mathbf{D}) \odot (\mathbf{T} \mathbf{W}_c \tilde{\mathbf{\Lambda}}^{-2} \mathbf{\Lambda}_s \mathbf{W}_c \mathbf{T}^*)^T\}. \quad (60)$$

*Proof:* The expressions (58)–(60) are derived in Appendix A.  $\square$

The fact that  $\mathbf{C}$  is written as a sum of the individual error covariances is independent of the specific models used in Section II-B; in fact, the above result would hold for any of the error models  $\mathbf{C}_{AP}$  and  $\mathbf{C}_{NP}$  described in [8]. In addition, even though our results have been derived specifically for certain multidimensional SSF algorithms, expressions identical to (56) and (57) hold for the MUSIC algorithm [23] if one replaces the individual error covariances with those derived in [2] and [6] and for the ESPRIT algorithm [24] if one uses the results of [25] and [26].

*Remark 1:* The precise mathematical meaning of Theorem 1 is that the covariance matrix of the estimation error in the limiting distribution is given by (55). However, a pragmatic use of the result for finite  $N$  is, of course, to substitute  $\tilde{\mathbf{Y}} = N \mathbf{Y}$  in (58) and  $\bar{\mu}^2 = N \mu^2$  in (59) and to use the resulting matrix  $\mathbf{C}/N$  to predict the covariance matrix of the estimation error itself (rather than the covariance of the distribution). If  $N$  is “large enough” and  $\nu_{ii}^2, i = 1, \dots, d$  and  $\mu^2$  are “small enough,” the theoretical results do indeed agree well with empirical observations obtained via simulations, as demonstrated in Section V.

We also note that with the substitutions  $\tilde{\mathbf{Y}} = N \mathbf{Y}$  and  $\bar{\mu}^2 = N \mu^2$ , the covariance  $\mathbf{C}/N$  does not go to zero as  $N \rightarrow \infty$  but rather converges to the expression obtained in [8] for model errors only. Similarly, for fixed  $N$ , letting  $\mathbf{Y} \rightarrow 0$  and  $\mu \rightarrow 0$  in  $\mathbf{C}/N$  yields the covariance for the finite sample only case considered in [4]. Thus, even though our analysis was conducted assuming  $\nu_{ii}^2 + \mu^2 \simeq O(1/N)$ , the results of Theorem 1 can be applied to the limiting cases as well.  $\square$

#### IV. OPTIMALLY WEIGHTED SUBSPACE FITTING

The expression for the asymptotic variance of the DOA estimates derived in the previous section is useful for predicting algorithm performance. This can be done by simply substituting into the error expressions the appropriate error model, error variances, and weighting matrices corresponding to the algorithm of interest. Another important problem is selecting the weighting matrices  $\mathbf{W}_r$  and  $\mathbf{W}_c$  so that the resulting estimation error is minimized. Because of the nature of the expressions (58)–(60), it is convenient to first consider the case of array perturbations only and then noise covariance perturbations plus finite sample effects separately. The weight optimization closely follows the corresponding results in [4] and [8], and the proofs are therefore omitted. The expressions

will be given using the original perturbation variances  $\mathbf{Y}$  and  $\mu^2$  rather than the  $N$ -normalized quantities.

##### A. Array Perturbations

This case is treated in [8], where the result is repeated below for reference.

*Theorem 2:* Assume that  $\mu^2 = 0$  and  $|\nu_{ij}| \gg \sigma^2/(N\|\mathbf{\Lambda}_s\|)$  so that only array perturbations contribute to the estimation error. Write the asymptotic covariance matrix of the estimation error as  $\mathbf{C} = \mathbf{C}(\mathbf{W}_r, \mathbf{W}_c)$  to emphasize the dependence on the weighting matrices. Then

$$\mathbf{W}_r = \mathbf{B}^{-1/2} \quad (61)$$

$$\mathbf{W}_c = (\mathbf{T}^* \mathbf{Y}^T \mathbf{T})^{-1} \quad (62)$$

yield minimum variance estimates in the sense that

$$\mathbf{C}(\mathbf{B}^{-1/2}, (\mathbf{T}^* \mathbf{Y}^T \mathbf{T})^{-1}) \leq \mathbf{C}(\mathbf{W}_r, \mathbf{W}_c) \quad (63)$$

for all possible  $\mathbf{W}_r > 0$  and  $\mathbf{W}_c > 0$ . Here, the matrix inequality  $\mathbf{A} \geq \mathbf{B}$  means that the difference  $\mathbf{A} - \mathbf{B}$  is positive semidefinite.

##### B. Noise Covariance and Finite Sample Effects

In this case, we have the following result:

*Theorem 3:* Assume that  $\mathbf{Y} = 0$  so that only the effects of noise (including possibly an imprecisely known covariance matrix) contribute to the estimation error. Then, it holds that

$$\mathbf{C}(\mathbf{I}, \tilde{\mathbf{\Lambda}}^2(N\mu^2\mathbf{I} + \sigma^2\mathbf{\Lambda}_s)^{-1}) \leq \mathbf{C}(\mathbf{W}_r, \mathbf{W}_c). \quad (64)$$

Since a scaling does not affect the result, the optimal weighting matrices are

$$\mathbf{W}_r = \mathbf{I} \quad (65)$$

$$\mathbf{W}_c = \tilde{\mathbf{\Lambda}}^2 \left( \frac{\mu^2}{\sigma^2/N} \mathbf{I} + \mathbf{\Lambda}_s \right)^{-1}. \quad (66)$$

*Proof:* This result follows by applying Lemma A.2 of [27] to the expressions (39), (56), (59), and (60).

Note that the interesting quantity is the ratio of the variance of the noise covariance perturbation and the variance of the noise itself, i.e., the relative perturbation on each element of the noise covariance.

##### C. Combined Errors

For the special case of uniform array perturbations, the following result provides the overall optimal weighting matrices when all error sources are present.

*Theorem 4:* Assume that  $\mathbf{B} = \mathbf{I}$ . Then, the choices

$$\mathbf{W}_r = \mathbf{I} \quad (67)$$

$$\mathbf{W}_c = \left( \mathbf{T}^* \mathbf{Y}^T \mathbf{T} + \mu^2 \tilde{\mathbf{\Lambda}}^{-2} + \frac{\sigma^2}{N} \tilde{\mathbf{\Lambda}}^{-2} \mathbf{\Lambda}_s \right)^{-1} \quad (68)$$

provide minimum variance DOA estimates.

*Proof:* This result is also a straightforward application of Lemma A.2 of [27].  $\square$

It is interesting to see how (68) “interpolates” between the optimal weighting matrices for the individual AP, NP, and FS cases. Unfortunately, no overall optimal choice of weighting matrices has been derived in the general case for arbitrary array and noise model errors. However, one may still suggest reasonable choices that normally lead to improved performance, even if it is not proven to be optimal. A possible extension of Theorem 4 when (7) and (8) holds is simply to use (67) and (68) as is, even if  $\mathbf{B} \neq \mathbf{I}$ . Another interesting possibility, which has been motivated by the form of (68), is to modify the row weighting as

$$\mathbf{W}_r = (\mathbf{B} + \alpha \mathbf{I})^{-1/2} \quad (69)$$

where  $\alpha$  is a scaling that controls the relative contribution of the AP optimal row weighting  $\mathbf{W}_r = \mathbf{B}^{-1/2}$  and the NP + FS choice  $\mathbf{W}_r = \mathbf{I}$ . Briefly,  $\alpha$  should be small if array perturbations are the major source of error (high SNR and/or large  $N$ ), whereas it should be chosen large if noise modeling errors and/or finite sample effects dominate. In more general cases, where (7) and (8) and/or (13) do not hold as stated, one can still use the above results to generate weighting matrices that improve performance, even if not proven to be optimal. This possibility is further investigated in Examples 5.2 and 5.3.

#### D. Using Estimated Weighting Matrices

A difficulty with the implementation of the optimal SSF technique is that the weighting matrices depend, in general, on unknown quantities. A natural approach is to replace these by consistent estimates, but one may wonder if this approach leads to any performance loss. To verify that estimated weighting matrices do not deteriorate the asymptotic estimation accuracy, we shall first show the following more general result.

*Lemma 1:* Let  $\hat{\theta}$  be obtained by minimizing the criterion function  $V(\theta, \eta)$ , where the dependence on the parameter of interest,  $\theta$  and a nuisance parameter  $\eta$  has been stressed. The true values of  $\theta$  and  $\eta$  are assumed to be inner points of their respective definition sets, and the criterion function is assumed to be differentiable with respect to both arguments. Express the estimate of  $\theta$  as a function of  $\eta$ ,  $\hat{\theta} = \hat{\theta}(\eta)$ , and assume that the estimate is root- $N$  consistent for all choices of  $\eta$ , i.e.,

$$\hat{\theta}(\eta) - \theta_0 = O_p(1/\sqrt{N}) \quad (70)$$

for all  $\eta$ . Then,  $\eta$  can be replaced by any (weakly) consistent estimate without affecting the asymptotic properties of  $\hat{\theta}$ , i.e.

$$\hat{\theta}(\eta) = \hat{\theta}(\hat{\eta}) + o_p(1/\sqrt{N}). \quad (71)$$

*Proof:* The mean-value theorem implies that

$$\hat{\theta}(\eta) = \hat{\theta}(\bar{\eta}) + \frac{\partial}{\partial \eta} \hat{\theta}(\bar{\eta})(\eta - \bar{\eta}) \quad (72)$$

for some  $\bar{\eta}$ . However, the assumption (70) readily implies that  $\frac{\partial}{\partial \eta} \hat{\theta} = O_p(1/\sqrt{N})$  for all  $\eta$ , and the result follows.  $\square$

The weighting matrices of interest herein are restricted to be Hermitian and positive definite. To apply Lemma 1, we can, for example, parametrize  $\mathbf{W}_c$  using its Cholesky decomposition

$$\mathbf{W}_c = \mathbf{W}_c^{1/2} \mathbf{W}_c^{*/2} \quad (73)$$

where  $\mathbf{W}_c^{1/2}$  is lower triangular. The parameters  $\eta_i$  of Lemma 1 can then be chosen as the real and imaginary parts of the elements of the nonzero part of  $\mathbf{W}_c^{1/2}$ . Since this parametrization clearly satisfies the requirements of the lemma, the following results.

*Corollary 1:* The weighting matrices of Theorems 2–4 can be replaced by consistent estimates without affecting the asymptotic optimality.

## V. EXAMPLES

In this section, some numerical examples are presented to illustrate our results. Computer simulations are also included to investigate the applicability of the first-order expressions to realistic scenarios.

### A. Example 5.1: Unstructured Array and Noise Model Errors

In this scenario, the wavefield of two Gaussian signal sources is recorded using a perturbed uniform linear array (ULA) of  $m = 6$  sensors with half-wavelength interelement spacing. The emitters are located at  $\theta = [0^\circ, 5^\circ]$  relative to array broadside, which corresponds to an angle separation of a quarter of the Rayleigh beamwidth. The array response is perturbed according to the model (7) and (8), with  $\mathbf{B} = \mathbf{I}$  and

$$\mathbf{\Upsilon} = 0.001 \begin{bmatrix} 1.0 & 0.7 \\ 0.7 & 1.0 \end{bmatrix}.$$

A nondiagonal  $\mathbf{\Upsilon}$  is assumed here since the DOA's are closely spaced so that some correlation between the perturbations is expected (it should be mentioned that this does not drastically affect the performance of any of the methods). The covariance of the additive noise is also assumed to be slightly perturbed from its nominal value  $\sigma^2 \mathbf{I}$ . The noise model errors are generated according to the model (12) and (13), with  $\mu^2 = 0.001$ . The baseband signals are zero-mean Gaussian, 90% correlated, and of equal power, and the noise is assumed to have unit power ( $\sigma^2 = 1$ ).

A batch of  $N = 100$  snapshots is generated for a variety of SNR's, and the SSF technique of Section II-C is applied to each data set using  $\mathbf{W}_r = \mathbf{I}$  and three different column weighting matrices:

- $\mathbf{W}_c = \tilde{\Lambda}^2 \Lambda_s^{-1}$  (finite sample errors only), referred to as *weighted subspace fitting* (WSF) [4]
- $\mathbf{W}_c = (\mathbf{T}^* \mathbf{\Upsilon}^T \mathbf{T})^{-1}$  (array perturbations only), referred to as *robust subspace fitting* (RSF) [8]
- $\mathbf{W}_c = (\mathbf{T}^* \mathbf{\Upsilon}^T \mathbf{T} + \mu^2 \tilde{\Lambda}^{-2} + \frac{\sigma^2}{N} \tilde{\Lambda}^{-2} \Lambda_s)^{-1}$  (overall optimal weighting), referred to as *optimal subspace fitting* (OSF).

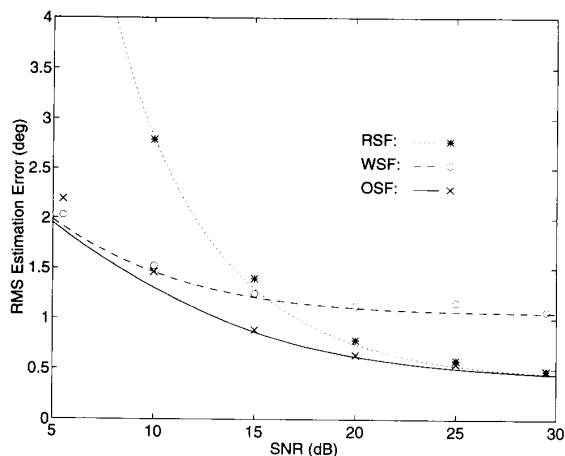


Fig. 1. Theoretical and empirical RMS error versus SNR for the SSF technique using different column weighting matrices, and perturbations in both the array and noise models.

Fig. 1 depicts the results of a Monte-Carlo simulation involving 256 independent trials for each SNR. The lines represent the error predicted by the theoretical expressions of Section III, whereas the symbols indicate the empirically calculated RMS errors. Only the results for the estimation of  $\theta_1 = 0^\circ$  are shown, where the results for  $\theta_2$  are similar.

The empirical RMS errors agree well with the theoretically predicted values for this case. Note also that, as expected, WSF is optimal for low SNR, RSF is optimal for high SNR, whereas OSF provides overall optimal estimates.

The previous example serves as an illustration of the performance improvement offered by the proposed method when the fairly restrictive perturbation model of Section II-B holds exactly. A perhaps more interesting question in practice is whether the OSF weightings also can be used to reduce the sensitivity under more physically motivated error models. The next two examples address this point.

#### B. Example 5.2: A Weak Unmodeled Source

An important source of errors in the covariance matrix of the noise is the presence of (weak) undetected signal sources. This type of error gives rise to a highly directional perturbation that does not fit the model of (12) and (13). Nevertheless, as illustrated in this example, the optimal weighting as derived in the previous section can be used to potentially reduce the sensitivity to such a perturbation. The scenario used here is as in Example 5.1 but without the unstructured array and noise perturbations. The SNR of the two signals is fixed to 10 dB, and an additional emitter of -10 dB SNR, uncorrelated with the signals of interest, is located at a varying DOA, ranging from 10 to  $50^\circ$ . The probability of detecting the weak source using the technique proposed in [28] is less than 30% for all cases. The empirical RMS error of the OSF (using (66) with  $\mu^2/\sigma^2 = 0.1$ ) and WSF estimates is shown in Fig. 2. Both methods were applied assuming  $d = 2$  signals. In this case,

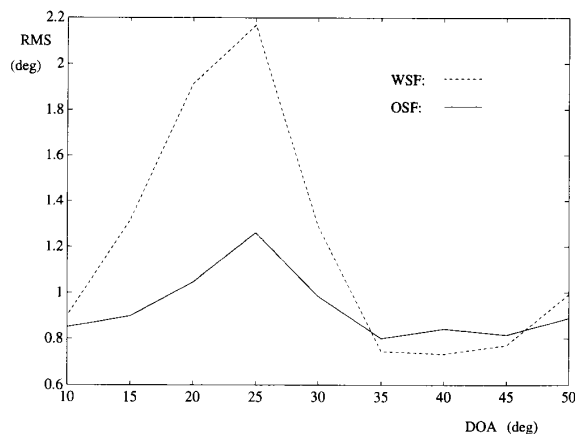


Fig. 2. Empirical RMS error versus the location of a weak (-10 dB) undetected signal source.

only the results for  $\hat{\theta}_2$  are shown; the performance difference for  $\theta_1$  is smaller in this scenario.

This example suggests that by taking noise model errors into account, one can obtain an estimator that is less sensitive to the presence of weak undetected sources.

#### Example 5.3: Antenna Position Errors

This example involves a ten-element, nominally uniform linear array with perturbed sensor positions. The standard deviation of the position errors in the direction perpendicular to the array is assumed to vary quadratically from  $\lambda/3000$  at one end of the array to  $\lambda/30$  at the other end, whereas the errors in the direction parallel to the array are assumed to be independent and a factor of  $\sqrt{10}$  smaller. This is similar to an example in [6] that attempts to model the nonlinearity of an underwater towed array.

This type of array perturbation leads to values for  $\Upsilon$  and  $\mathbf{B}$  that are angle-dependent (see equation (37) of [6]), and hence, no combined optimal weighting is possible. However, in an attempt to find a weighting that, at least heuristically, balances the two types of errors, we choose to use the weighting scheme (cf. (68) and (69) and note that  $\mu = 0$  in this example):

$$\mathbf{W}_c = \left( \mathbf{T}^* \mathbf{\Upsilon}^T \mathbf{T} + \frac{\sigma^2}{N} \tilde{\mathbf{A}}^{-2} \mathbf{\Lambda}_s \right)^{-1} \quad (74)$$

$$\mathbf{W}_r = \left( \mathbf{B} + \frac{\sigma^2}{N} \mathbf{I} \right)^{-1/2} \quad (75)$$

where

$$\mathbf{\Upsilon} = 2 * \mathbf{\Pi} * \beta \mathbf{I} \quad (76)$$

$$\mathbf{B} = \frac{1}{\beta} \hat{\mathbf{B}}, \quad (77)$$

and where

$$\beta = \sqrt{\text{Tr}(\hat{\mathbf{B}})} \quad (78)$$

$$\hat{\mathbf{B}} = \frac{2\mathbf{\Pi}}{\lambda} (\mathbf{B}_x + \mathbf{B}_y). \quad (79)$$



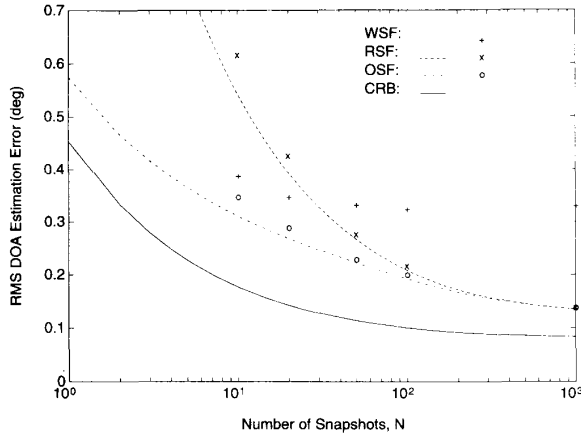


Fig. 3. Theoretical and empirical RMS error versus number of snapshots for a nominal linear array with sensor position perturbations.

Here,  $\mathbf{B}_x$  and  $\mathbf{B}_y$  are, respectively, the covariances of the position errors in directions parallel and perpendicular to the array. As alluded to above, in this example, both  $\mathbf{B}_x$  and  $\mathbf{B}_y$  are diagonal,  $\mathbf{B}_x = \mathbf{B}_y/10$ , and the  $k$ th diagonal element of  $\mathbf{B}_y$  is given by

$$\{\mathbf{B}_y\}_{kk} = \left( \frac{k^2 \lambda}{3000} \right)^2. \quad (80)$$

The structure assumed for  $\hat{\mathbf{B}}$  in (79) is motivated by the covariance expressions given in equation (37) of [6] for antenna position errors. The scaling term  $\beta$  is used to guarantee that our heuristic weighting matrices converge to the WSF weightings when  $\|\hat{\mathbf{B}}\| \ll \sigma^2/(N\|\mathbf{A}_s\|)$ .

The performance of the SSF approach using these weighting matrices is compared with WSF and RSF for a scenario involving two uncorrelated, 20-dB emitters located at 10 and 15° relative to array broadside. A total of 1000 Monte-Carlo trials are conducted for various values of  $N$ , and the number of snapshots and the results are plotted in Fig. 3 along with the corresponding CRB [16]. Here, OSF refers to the combined weighting matrices above. The lines in the figure correspond to our theoretical analysis, whereas the symbols represent the empirical results. The advantage of the combined weighting is clearly evident in these results, and the agreement with the theoretical curves is reasonably good. The slight discrepancy for  $N = 10, 20$  is due to the fact that our analysis is asymptotic in  $N$ . Despite this, the empirical results are within 5% of the theoretical predictions for  $N$  as small as 20.

## VI. CONCLUSION

We have shown how to conduct a combined performance analysis of subspace fitting algorithms that accounts for errors introduced by additive noise and modeling errors simultaneously. For the special case of random unstructured perturbations to the array and noise models, we have derived row and column subspace weighting matrices that yield minimum variance DOA estimates. The theoretical error variances for the optimally weighted subspace fitting estimates were

found to accurately predict empirical mean-square errors down to the threshold region. The derivation of optimal weighting matrices required a fairly simplistic model of the array response and noise covariance perturbations. However, for more complicated models, heuristic weighting matrices based on our analysis will often lead to significantly improved performance as compared with not taking the model errors into account. This is illustrated by two examples involving underestimation of the number of signals and nonuniform sensor position errors.

Implementation of the overall optimal technique requires knowledge of the covariances of the perturbation terms. An interesting question not considered herein is the sensitivity to misspecifications of these quantities. However, we should point out that the first-order effect of such errors vanish (by Corollary 1 in Section IV). Moreover, if the misspecification is a small scaling error, the proposed optimal weighting matrix is still a reasonable interpolation between the optimal weightings designed for one error source (array errors, noise covariance errors, or finite sample errors) only.

## APPENDIX A PROOF OF THEOREM 1

In this Appendix, the terms  $E[V_{ik}V_{jk}]$  are evaluated, where  $V_{ik}$  is defined in (50)–(52) for  $k = 1, 2, 3$ . Towards this end, the following formulas are first presented. If  $\mathbf{x}_1, \dots, \mathbf{x}_4$  denote arbitrary deterministic vectors, we have

$$\begin{aligned} E[\mathbf{x}_1^* \hat{\mathbf{A}} \mathbf{x}_2 \mathbf{x}_3^* \hat{\mathbf{A}} \mathbf{x}_4] &= 0, \\ E[\mathbf{x}_1^* \hat{\mathbf{A}}^* \mathbf{x}_2 \mathbf{x}_3^* \hat{\mathbf{A}} \mathbf{x}_4] &= \mathbf{x}_1^* \mathbf{Y}^T \mathbf{x}_4 \mathbf{x}_3^* \mathbf{B} \mathbf{x}_2 \end{aligned} \quad (\text{A.1})$$

$$\begin{aligned} E[\mathbf{x}_1^* \mathbf{Z} \mathbf{x}_2 \mathbf{x}_3^* \mathbf{Z} \mathbf{x}_4] &= 0, \\ E[\mathbf{x}_1^* \mathbf{Z}^* \mathbf{x}_2 \mathbf{x}_3^* \mathbf{Z} \mathbf{x}_4] &= \frac{\sigma^2}{N} \mathbf{x}_1^* \mathbf{x}_4 \mathbf{x}_3^* \hat{\mathbf{P}} \mathbf{x}_2 \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned} E[\mathbf{x}_1^* \Sigma \mathbf{x}_2 \mathbf{x}_3^* \Sigma \mathbf{x}_4] &= \left( \sigma^4 + \frac{\mu^2}{N} \right) \mathbf{x}_1^* \mathbf{x}_2 \mathbf{x}_3^* \mathbf{x}_4 \\ &\quad + \left( \frac{\sigma^4}{N} + \mu^2 \right) \mathbf{x}_3^* \mathbf{x}_2 \mathbf{x}_1^* \mathbf{x}_4. \end{aligned} \quad (\text{A.3})$$

We shall prove the most difficult of these, namely, (A.3):

$$\begin{aligned} E[\mathbf{x}_1^* \Sigma \mathbf{x}_2 \mathbf{x}_3^* \Sigma \mathbf{x}_4] &= \frac{1}{N^2} \sum_{t,s} \sum_{i,j,k,l} E \\ &\quad \times \left[ E \left\{ x_{1i}^* n_i(t) n_j^*(t) x_{2j} x_{3k} n_k(s) n_l^*(s) x_{4l} \mid \tilde{\Sigma} \right\} \right] \\ &= \frac{1}{N^2} \sum_{t,s} \sum_{i,j,k,l} x_{1i}^* x_{2j} x_{3k} x_{4l} E \left[ (\sigma^2 \delta_{ij} + \tilde{\sigma}_{ij}) \right. \\ &\quad \times (\sigma^2 \delta_{kl} + \tilde{\sigma}_{kl}) \\ &\quad \left. + (\sigma^2 \delta_{il} + \tilde{\sigma}_{il})(\sigma^2 \delta_{kj} + \tilde{\sigma}_{kj}) \delta_{ts} \right] \\ &= \frac{1}{N^2} \sum_{t,s} \sum_{i,j} (\sigma^4 + \mu^2 \delta_{ts}) x_{1i}^* x_{2i} x_{3j}^* x_{4j} \\ &\quad + (\sigma^4 \delta_{ts} + \mu^2) x_{1i}^* x_{4i} x_{3j}^* x_{2j} \\ &= (\sigma^4 + \mu^2/N) \mathbf{x}_1^* \mathbf{x}_2 \mathbf{x}_3^* \mathbf{x}_4 + (\sigma^4/N + \mu^2) \mathbf{x}_3^* \mathbf{x}_2 \mathbf{x}_1^* \mathbf{x}_4. \end{aligned}$$

For two arbitrary scalars  $x_1$  and  $x_2$ , it is easily seen that

$$2 \operatorname{Re}(x_1)\operatorname{Re}(x_2) = \operatorname{Re}(x_1^*x_2 + x_1x_2).$$

Using this fact and applying (A.1) and (50) leads to

$$\begin{aligned} E[V_{i1}V_{j1}] &= 2 \operatorname{Re}\left\{E[\beta_i^* \mathbf{A} \mathbf{P} \tilde{\mathbf{A}}^* \alpha_i \alpha_j^* \tilde{\mathbf{A}} \mathbf{P} \mathbf{A}^* \beta_j\right. \\ &\quad \left.+ \alpha_i^* \tilde{\mathbf{A}} \mathbf{P} \mathbf{A}^* \beta_i \alpha_j \tilde{\mathbf{A}} \mathbf{P} \mathbf{A}^* \beta_j\right\} \\ &= 2 \operatorname{Re}\{\alpha_i^* \mathbf{B} \alpha_j \beta_j^* \mathbf{A} \mathbf{P} \tilde{\mathbf{A}}^T \mathbf{P} \mathbf{A}^* \beta_i\}. \end{aligned} \quad (\text{A.4})$$

Similarly, by observing that  $\alpha_i^* \beta_j = 0$  (which follows directly from (53) and (54)) and using (A.2) and (A.3), the noise-dependent terms are obtained as

$$E[V_{i2}V_{j2}] = \frac{2\sigma^2}{N} \operatorname{Re}\{\alpha_i^* \alpha_j \beta_j^* \mathbf{A} \hat{\mathbf{P}} \mathbf{A}^* \beta_i\} \quad (\text{A.5})$$

$$E[V_{i3}V_{j3}] = 2 \left( \frac{\sigma^4}{N} + \mu^2 \right) \operatorname{Re}\{\alpha_i^* \alpha_j \beta_j^* \beta_i\}. \quad (\text{A.6})$$

Combining (36), (A.4)–(A.6), and using (15), (16), and (22) gives

$$\begin{aligned} \mathbf{Q}_{ij} &= \lim_{N \rightarrow \infty} NE[V_i V_j] = 2 \operatorname{Re}\{\alpha_i^* \mathbf{B} \alpha_j \beta_j^* \mathbf{A} \mathbf{P} \tilde{\mathbf{A}}^T \mathbf{P} \mathbf{A}^* \beta_i \\ &\quad + (\tilde{\mu}^2 + \sigma^4) \alpha_i^* \alpha_j \beta_j^* \beta_i \\ &\quad + \sigma^2 \alpha_i^* \alpha_j \beta_j^* \mathbf{A} \mathbf{P} \mathbf{A}^* \beta_i\} \\ &= 2 \operatorname{Re}\{\alpha_i^* \mathbf{B} \alpha_j \beta_j^* \mathbf{A} \mathbf{P} \tilde{\mathbf{A}}^T \mathbf{P} \mathbf{A}^* \beta_i \\ &\quad + \tilde{\mu}^2 \alpha_i^* \alpha_j \beta_j^* \beta_i + \sigma^2 \alpha_i^* \alpha_j \beta_j^* \mathbf{R} \beta_i\}. \end{aligned} \quad (\text{A.7})$$

To obtain a more compact matrix formulation for  $\mathbf{Q}$ , we use some identities that are derived from (22)–(24):

$$\mathbf{G}^\dagger \mathbf{W}_r \mathbf{E}_s = \mathbf{G}^\dagger \mathbf{W}_r \mathbf{A} \mathbf{A}^\dagger \mathbf{E}_s = \mathbf{A}^\dagger \mathbf{E}_s = \mathbf{T} \quad (\text{A.8})$$

$$\mathbf{P} = \mathbf{A}^\dagger \mathbf{E}_s \tilde{\mathbf{A}} \mathbf{E}_s^* \mathbf{A}^\dagger = \mathbf{T} \tilde{\mathbf{A}} \mathbf{T}^* \quad (\text{A.9})$$

$$\mathbf{E}_s^* \mathbf{R} \mathbf{E}_s = \mathbf{E}_s^* (\mathbf{E}_s \mathbf{A}_s \mathbf{E}_s^* + \sigma^2 \mathbf{E}_n \mathbf{E}_n^*) \mathbf{E}_s = \mathbf{A}_s. \quad (\text{A.10})$$

Inserting (53) and (54) and (A.8)–(A.10) into (A.7) results in

$$\mathbf{Q} = \mathbf{Q}_{\text{AP}} + \mathbf{Q}_{\text{NP}} + \mathbf{Q}_{\text{FS}} \quad (\text{A.11})$$

where the expressions for the matrices involved are as given in (58)–(60).

#### REFERENCES

- [1] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 36, no. 4, pp. 532–544, Apr. 1988.
- [2] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramer-Rao bound," *IEEE Trans. IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 720–741, May 1989.
- [3] H. Clergeot, S. Tressens, and A. Ouamri, "Performance of high resolution frequencies estimation methods compared to the Cramér-Rao bounds," *IEEE Trans. IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, no. 11, pp. 1703–1720, Nov. 1989.
- [4] 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.
- [5] B. Friedlander, "Sensitivity analysis of the maximum likelihood direction-finding algorithm," *IEEE Trans. Aerospace Electron. Syst.*, vol. 26, no. 6, 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, no. 7, pp. 1758–1774, July 1992.
- [7] F. Li and R. Vaccaro, "Sensitivity analysis of DOA estimation algorithms to sensor errors," *IEEE Trans. Aerospace Electron. Syst.*, vol. 28, no. 3, pp. 708–717, July 1992.
- [8] 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.
- [9] F. Li and R. J. Vaccaro, "Performance degradation of DOA estimators due to unknown noise fields," In *Proc. ICASSP 91* (Toronto, Canada), May 1991, pp. 1413–1416, vol. 2.
- [10] ———, "Performance degradation of DOA estimators due to unknown noise fields," *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp. 686–689, Mar. 1992.
- [11] K. M. Wong, J. P. Reilly, Q. Wu, and S. Qiao, "Estimation of the directions of arrival of signals in unknown correlated noise, Parts I and II," *IEEE Trans. Signal Processing*, vol. 40, no. 8, pp. 2007–2028, Aug. 1992.
- [12] M. Wax, "Detection and localization of multiple sources in noise with unknown covariance," *IEEE Trans. Signal Processing*, vol. 40, no. 1, pp. 245–249, Jan. 1992.
- [13] 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. 35, pp. 286–299, Mar. 1987.
- [14] 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.
- [15] 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, pp. 2641–2650, 1992.
- [16] B. Wahlberg, B. Ottersten, and M. Viberg, "Robust signal parameter estimation in the presence of array perturbations," In *Proc. ICASSP* (Toronto, Canada), May 1991.
- [17] M. Viberg and A. Swindlehurst, "A Bayesian approach to auto-calibration for parametric array signal processing," to be published in *IEEE Trans. Signal Processing*.
- [18] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [19] R. T. Compton, "The effect of random steering vector errors in the Applebaum adaptive array," *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-18, no. 5, pp. 392–400, Sept. 1982.
- [20] L. C. Godara, "Error analysis of the optimal antenna array processors," *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-22, no. 4, pp. 395–409, July 1986.
- [21] N. K. Jablon, "Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 996–1012, Aug. 1986.
- [22] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. London: Academic, 1979.
- [23] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. thesis, Stanford Univ., Stanford, CA, Nov. 1981.
- [24] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.
- [25] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1122–1135, May 1991.
- [26] A. Swindlehurst and T. Kailath, "On the sensitivity of the ESPRIT algorithm to non-identical subarrays," *Sādhanā, Acad. Proc. Eng. Sci.*, vol. 15, no. 3, pp. 197–212, Nov. 1990.
- [27] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér-Rao bound: Further results and comparisons," *IEEE Trans. Acoust., Speech Signal Processing*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.
- [28] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, no. 11, pp. 2436–2449, Nov. 1991.



**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 was in 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 and 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.