

MUSIC, Maximum Likelihood, and Cramer-Rao Bound

PETRE STOICA AND ARYE NEHORAI, MEMBER, IEEE

Abstract—The problem of finding the directions of multiple plane waves with narrow-band arrays of sensors, and the related problem of estimating the parameters of multiple superimposed exponential signals in noise, have attracted considerable interest recently. Several methods, such as the MUSIC and maximum likelihood (ML), have been proposed for solving these problems.

This paper studies the performance of the MUSIC and ML methods, and analyzes their statistical efficiency. It also derives the Cramer-Rao bound (CRB) for the estimation problems mentioned above, and establishes some useful properties of the CRB covariance matrix. The relationship between the MUSIC and ML estimators is investigated as well. Finally, the paper contains a numerical study of the statistical efficiency of the MUSIC estimator for the problem of finding the directions of two plane waves using a uniform linear array.

A more exact description of the results of this paper can be found in the conclusion section.

I. INTRODUCTION AND PRELIMINARIES

SEVERAL important problems in the signal processing field, among them direction finding with narrow-band sensor arrays, estimation of the parameters of multiple superimposed exponential signals in noise, and resolution of overlapping echos (see [1], [2], [9], [12] and the references therein), can be reduced to estimating the parameters in the following model:

$$y(t) = A(\theta)x(t) + e(t) \quad t = 1, 2, \dots, N. \quad (1.1a)$$

In (1.1), $y(t) \in \mathbf{C}^{m \times 1}$ is the noisy data vector, $x(t) \in \mathbf{C}^{n \times 1}$ is the vector of signal amplitudes, $e(t) \in \mathbf{C}^{m \times 1}$ is an additive noise, and the matrix $A(\theta) \in \mathbf{C}^{m \times n}$ has the following special structure:

$$A(\theta) = [a(\omega_1) \cdots a(\omega_n)] \quad (1.1b)$$

where $\{\omega_i\}$ are real parameters, $a(\omega_i) \in \mathbf{C}^{m \times 1}$ is a so-called transfer vector [between the i th signal and $y(t)$], and $\theta = [\omega_1 \cdots \omega_n]^T$. In Section II, we will briefly discuss how the model (1.1) encompasses the data models used in some of the applications mentioned above, and we will introduce the basic assumptions on (1.1).

Manuscript received July 15, 1987; revised July 15, 1988. The work of A. Nehorai was supported by the Air Force Office of Scientific Research under Grant AFOSR-88-0080.

P. Stoica is with the Department of Automatic Control, Polytechnic Institute of Bucharest, Splaiul Independentei 313, R-77 206 Bucharest, Romania.

A. Nehorai is with the Department of Electrical Engineering, Yale University, New Haven, CT 06520.

IEEE Log Number 8926678.

There are three main problems associated with fitting models of the form (1.1) to the data $\{y(1), \dots, y(N)\}$.

a) Estimation of the number of signals n . Methods for estimating n are well documented in the literature (see, e.g., [3], [4], [21], and [22]) and will not be discussed here. In this paper we assume that the number of signals n is given.

b) Estimation of the signal amplitudes $\{x(t)\}$. Once an estimate of θ is available, the estimation of $x(t)$ reduces (under reasonable conditions) to a simple least-squares fit. We will omit any explicit discussion on the problem of estimating $\{x(t)\}$. However, estimates of $\{x(t)\}$ will implicitly appear in the analysis that follows. Note that since it is required to estimate $\{x(t)\}_{t=1}^N$ (and not their "average characteristics," such as their covariance matrix), we will consider these variables to be deterministic (i.e., fixed). This assumption does not exclude the possibility that $x(1), \dots, x(N)$ are samples from a random process. If so, then the distributional results derived in what follows should be interpreted as being conditioned on $\{x(t)\}_{t=1}^N$.

c) Estimation of the parameter vector θ . Methods for accomplishing this task, and their performance, are the main topics to be dealt with in this paper.

A class of methods for estimating θ in (1.1), which has received significant attention, is based on the eigendecomposition of the sample covariance matrix of $y(t)$ [1]–[9], [15]–[22]. A representative member of this class is the MUSIC (Multiple Signal Characterization) algorithm [1], [2]. There has been considerable interest recently in analyzing the statistical performance of the MUSIC. Some interesting and related studies of the resolvability of MUSIC have been reported in [7] and [8] (also see [21]). However, an expression for the covariance matrix, say C , of the MUSIC estimate of θ has not been derived in these papers. A preliminary analysis of the MUSIC performance expressed by C can be found in [5], but it appears to be incomplete. In Section III, after a brief review of the MUSIC, we provide an explicit expression for C that holds for sufficiently large values of N . In this section we also discuss an improved MUSIC estimator introduced in [6], for which we present a performance analysis as well, and a computationally inexpensive modification that should improve its performance.

The expression of C derived in this paper can be used to compare the performance of the MUSIC to the performance achieved by other methods. In particular, compar-

ison to the performance corresponding to the Cramer–Rao bound (CRB) should be of interest. An expression for the Cramer–Rao lower bound on the covariance matrix of any unbiased estimator of the parameters θ in the general model (1.1) does not appear to be available in the literature (only expressions for special cases corresponding to $N = 1$ and a specific “transfer vector” $a(\omega)$ can be found, see, e.g., [24]–[26]). In Section IV, we derive the CRB covariance matrix under reasonable conditions. The behavior of that matrix when m or N , or both, increase is also studied.

The maximum-likelihood (ML) method can also be used, under appropriate assumptions, to estimate the parameter vector θ in (1.1). The ML estimator (MLE) of θ has been the topic of two interesting and related recent papers [10] and [12] (see also [11], [13], and [14] for studies on ML estimation of parameters in special cases of (1.1)). In Section V, we briefly review the ML approach to estimation of θ , study the consistency properties of the MLE, and discuss its statistical (second-order) performance. The main issue of this section is the question of the asymptotic efficiency of the MLE. It is shown that the MLE is *not* statistically efficient if m is small, even if N is large; and that it can achieve the CRB only if m is increased.

In Section VI, we investigate the relationship between the MUSIC and the ML estimators. Only some unsupported claims about this relationship can be found in the literature (see, e.g., [5] and [23]). We show that the MUSIC is a large sample (for $N \gg 0$) realization of the MLE if and only if the signals are uncorrelated.

Finally, in Section VII, we present an analytic comparison between the MUSIC estimation error variance and the CRB. For uncorrelated signals, the MUSIC variance is shown to approach the CRB as m and N increase. For correlated signals, MUSIC is shown to be statistically inefficient. Also presented in this section are the results of a numerical study. Specifically, we perform a detailed comparison of the CRB and MUSIC variance in the case of two complex sine waves, over the set of feasible angular frequencies and for several values of SNR (signal-to-noise ratio), m , and the coefficient of correlation between the two sine waves.

II. NOTATION, BASIC ASSUMPTIONS, AND SPECIAL CASES

Let us first list some notational conventions that will be used in this paper.

- A^T = the transpose of matrix $A \in \mathbf{C}^{k \times p}$
- A^+ = the conjugate of A
- A^* = the conjugate transpose of A
- \bar{A} = the real part of A
- \tilde{A} = the imaginary part of A
- $\text{tr } A$ = the trace of $A \in \mathbf{C}^{k \times k}$
- A_{ij} = the i, j element of A
- $A \circ B$ = the Hadamard product of $A \in \mathbf{C}^{k \times p}$ and $B \in \mathbf{C}^{k \times p}$ ($[A \circ B]_{ij} = A_{ij} B_{ij}$)

$A \geq B$ = the difference matrix $A - B$ is positive semidefinite, with A and B being Hermitian positive semidefinite matrices

$\delta_{k,p}$ = the Dirac delta ($= 1$ if $k = p$, and $= 0$ otherwise)

ω = a generic element of the vector θ ; to avoid a complication of notation, the symbols ω and θ are used to denote both the true and the unknown parameters

$d(\omega) = da(\omega)/d\omega$

E = the expectation operator; for deterministic signals, $E(\cdot) = \lim_{N \rightarrow \infty} (1/N) \sum_{t=1}^N (\cdot)$.

Next, we introduce some basic assumptions on the model (1.1). The MUSIC and the ML methods are based on different sets of assumptions. However, some assumptions are common to both methods. The common assumptions are listed first.

A1: $m > n$, and the vectors $a(\omega)$ corresponding to $(n + 1)$ different values of ω are linearly independent. (This is a weak assumption that guarantees the uniqueness of both the MUSIC and ML estimators.)

A2: $Ee(t) = 0$, $Ee(t)e^*(t) = \sigma I$ and $Ee(t)e^T(t) = 0$. (This is a more restrictive assumption that is essential for the MUSIC algorithm; for the ML method, relaxation of A2 is possible in principle, but would lead to considerable complications.)

The following additional assumption is needed for the MUSIC.

AMU: The matrix

$$P = Ex(t)x^*(t) \quad (2.1)$$

is nonsingular (positive definite), and $N > m$; and the following one is needed for the MLE:

AML: $Ee(t)e^*(s)$

$$= Ee(t)e^T(s) = 0 \quad \text{for } t \neq s,$$

and $e(t)$ is Gaussian distributed.

Assumption AML appears more restrictive than AMU (again AML could in principle be relaxed, but this would introduce significant complications). The distinction made above between the assumptions used by MUSIC and MLE is important for realizing which one of these two estimators, if any, is usable in a certain situation (see below).

Next we describe briefly some applications of the general model (1.1). For other possible applications of (1.1), we refer to [9], [12], and the references therein.

A. Direction Finding with Uniform Linear Sensor Arrays

The problem of determining the directions of n plane waves impinging on a linear uniform narrow-band array of m sensors can be formulated as that of estimating the parameters θ of the model (1.1), where $x(t)$ is the vector of complex wave amplitudes, N is the number of “snapshots,” and

$$a(\omega) = [1 \ e^{i\omega} \ \dots \ e^{i(m-1)\omega}]^T. \quad (2.2)$$

Note that in this case, $A(\theta)$ is a vanderMonde matrix and therefore assumption A1 is satisfied. Assumption A2 and AML mean that the noise is spatially and temporally uncorrelated, and assumption AMU that the plane waves are not "fully coherent" and the number of snapshots is greater than the number of sensors in the array. All these assumptions look reasonable and could be satisfied. Thus, both the MUSIC and the MLE could be usable in this type of application.

B. Estimation of Complex Sine Wave Frequencies from Multiple-Experiment Data

Consider the following signal model:

$$y_k(t) = \sum_{p=1}^n \gamma_p(t) e^{i\omega_p k} + e_k(t) \quad \begin{array}{l} k = 1, \dots, m \\ t = 1, \dots, N \end{array} \quad (2.3)$$

where m denotes the number of samples in an experiment, N is the number of experiments, $\{\gamma_p(t)\}$ and $\{\omega_p\}$ are the amplitudes and frequencies of the complex sine waves, and $\{e_k(t)\}$ is an additive noise. The model (2.3) can be written in the form (1.1) using the following definitions:

$$\begin{aligned} y(t) &= [y_1(t) \cdots y_m(t)]^T \\ x(t) &= [e^{i\omega_1} \gamma_1(t) \cdots e^{i\omega_n} \gamma_n(t)]^T \\ e(t) &= [e_1(t) \cdots e_m(t)]^T \\ a(\omega) &= [1 \ e^{i\omega} \cdots e^{i(m-1)\omega}]^T. \end{aligned} \quad (2.4)$$

The conditions A2 and AML mean that the noise within an experiment is white and that the noises of any two different experiments are uncorrelated, which is plausible. Note that in this case we may well have $m > N$, which implies that the MUSIC may not be usable.

C. Estimation of Complex Sine Wave Frequencies from Single-Experiment Data

The model for this application is given by (2.3) with t dropped:

$$y_k = \sum_{p=1}^n \gamma_p e^{i\omega_p k} + e_k \quad k = 1, \dots, m \quad (2.5)$$

which can be written in the form of (1.1) using the notation (2.4). Assumption A1 is satisfied if $m > n$, A2 means that the noise e_k is white, and AML reduces to the requirement that e_k is Gaussian. Thus, the MLE may be usable. Since $N = 1$, assumption AMU is not satisfied for $n > 1$, and hence the MUSIC is not usable when there are at least two signals and the problem is stated as previously. To be able to use MUSIC we must recast the model in a different form.

Let us denote the number of available samples by M (not m). Let m be some integer greater than n , and define

$$\begin{aligned} y(t) &= [y_t \cdots y_{t+m-1}]^T \\ a(\omega) &= [1 \ e^{i\omega} \cdots e^{i(m-1)\omega}]^T \end{aligned}$$

$$\begin{aligned} x(t) &= [\gamma_1 e^{i\omega_1 t} \cdots \gamma_n e^{i\omega_n t}]^T \\ e(t) &= [e_t \cdots e_{t+m-1}]^T \\ t &= 1, \dots, M - m + 1. \end{aligned}$$

Using the notation above, we can write (2.5) in the form (1.1) with $N = M - m + 1$. In contrast to the multiple-experiment case, here $e(t)$ and $e(s)$ are correlated for $t \neq s$, and thus, the MLE is not applicable. On the other hand, the model written in the above form satisfies the MUSIC assumptions A1, A2, and AMU, provided $2n < 2m < M + 1$, which is readily achieved. Note that m can be chosen rather arbitrarily. This arbitrariness raises the question as to whether an optimal choice exists that minimizes the estimation errors of the MUSIC algorithm. This type of question will be considered in Section VII.

III. THE MUSIC ESTIMATOR

We begin by setting some additional notation (which will be used extensively in this paper) and a brief description of the MUSIC algorithm. Next, we establish the asymptotic distribution of the MUSIC estimator and derive an explicit expression for the covariance matrix of its estimation errors. Finally, we consider the improved MUSIC estimator introduced in [6] for which we also provide a performance analysis.

A. The MUSIC Algorithm

In this subsection, we assume that conditions A1, A2, and AMU hold. Under these assumptions, the covariance matrix of the observation vector $y(t)$ is given by

$$R \triangleq E y(t) y^*(t) = A(\theta) P A^*(\theta) + \sigma I. \quad (3.1)$$

For notational convenience, we will simply write A instead of $A(\theta)$ whenever there is no possibility of confusion. If $\hat{\theta}$ is an estimate of θ , then we will also write \hat{A} instead of $A(\hat{\theta})$.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ denote the eigenvalues of R . Since $\text{rank}(P A^*) = n$, it follows that

$$\begin{aligned} \lambda_i &> \sigma \quad \text{for } i = 1, \dots, n; \quad \text{and} \\ \lambda_i &= \sigma \quad \text{for } i = n + 1, \dots, m. \end{aligned} \quad (3.2)$$

It will be assumed throughout this paper that $\{\lambda_i\}_{i=1}^n$ are distinct.

Denote the unit-norm eigenvectors associated with $\lambda_1, \dots, \lambda_n$ by s_1, \dots, s_n , and those corresponding to $\lambda_{n+1}, \dots, \lambda_m$ by g_1, \dots, g_{m-n} . Also define

$$\begin{aligned} S &= [s_1 \cdots s_n] \quad (m \times n) \\ G &= [g_1 \cdots g_{m-n}] \quad (m \times (m - n)). \end{aligned} \quad (3.3)$$

Next, observe that

$$R G = P A^* G + \sigma G = \sigma G$$

which readily implies

$$A^* G = 0 \quad (3.4a)$$

or, equivalently,

$$a^*(\omega) GG^*a(\omega) = 0 \quad \text{for } \omega = \omega_1, \dots, \omega_n. \quad (3.4b)$$

Since the normalized eigenvectors $\{s_i, g_j\}$ are orthonormal,

$$SS^* + GG^* = I \quad (3.5)$$

it follows that (3.4) can also be written as

$$a^*(\omega)[I - SS^*]a(\omega) = 0 \quad \text{for } \omega = \omega_1, \dots, \omega_n. \quad (3.6)$$

It is not difficult to see that the true parameter values $\{\omega_1, \dots, \omega_n\}$ are the only solutions of (3.4) or (3.6). The proof is by contradiction. Assume that there exists another solution which we denote by ω_{n+1} . The matrix SS^* in (3.6) is the orthogonal projection operator onto the subspace spanned by the columns of S . Thus, it would follow from (3.6) that the linearly independent vectors $\{a(\omega_i)\}_{i=1}^{n+1}$ (by assumption A2) belong to the column space of S . However, this is impossible since that space is of dimension equal to n .

The basic idea of the MUSIC algorithm is the exploitation of the property (3.4), or (3.6), of the true covariance matrix R . In practice, R is unknown, but it can be consistently estimated from the available data. Let

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

Similar to the eigendecomposition of R , let $\{\hat{s}_1, \dots, \hat{s}_n, \hat{g}_1, \dots, \hat{g}_{m-n}\}$ denote the unit-norm eigenvectors of \hat{R} , arranged in the descending order of the associated eigenvalues, and let \hat{S} and \hat{G} denote the matrices S and G made of $\{\hat{s}_i\}$ and, respectively, $\{\hat{g}_i\}$. Define

$$f(\omega) = a^*(\omega) \hat{G} \hat{G}^* a(\omega) \quad (3.7a)$$

$$= a^*(\omega) [I - \hat{S} \hat{S}^*] a(\omega). \quad (3.7b)$$

The MUSIC estimates of $\{\omega_i\}$ are obtained by picking the n values of ω for which $f(\omega)$ is minimized. Minimization of $f(\omega)$ is usually done by evaluating it at the points of a fine grid, using (3.7a) or (3.7b) [(3.7a) is preferred to (3.7b) if $n > m - n$, and vice versa].

There are several variants of the MUSIC algorithm described above, which are currently in use (see the excellent survey paper [6] and [15]–[17]). Several computationally efficient (adaptive or batch) implementations are also available [18]–[20], [30]. For the sake of conciseness, in this paper we will concentrate on the basic MUSIC algorithm and its improved version introduced in [6] (to be described later). Other variants of the MUSIC can be analyzed similarly with respect to their statistical properties [37].

B. MUSIC Asymptotic Distribution

In this and the next subsection, we assume that conditions A1, A2, AMU, and AML hold. As already ex-

plained above, assumption AML is not necessary for the application of the MUSIC algorithm. However, it will be used in the analysis of the MUSIC estimator.

To establish the distribution of the MUSIC estimator, we need the following result on the statistics of the eigenvectors of the sample covariance matrix \hat{R} .

Lemma 3.1: a) The estimation errors $(\hat{s}_i - s_i)$ are asymptotically (for large N) jointly Gaussian distributed with zero means and covariance matrices given by

$$\begin{aligned} E(\hat{s}_i - s_i)(\hat{s}_j - s_j)^* &= \frac{\lambda_i}{N} \left[\sum_{\substack{k=1 \\ k \neq i}}^n \frac{\lambda_k}{(\lambda_k - \lambda_i)^2} s_k s_k^* + \sum_{k=1}^{m-n} \frac{\sigma}{(\sigma - \lambda_i)^2} g_k g_k^* \right] \\ &\cdot \delta_{i,j} \triangleq W_i \delta_{i,j} \end{aligned} \quad (3.8a)$$

$$\begin{aligned} E(\hat{s}_i - s_i)(\hat{s}_j - s_j)^T &= -\frac{\lambda_i \lambda_j}{N(\lambda_i - \lambda_j)^2} s_j s_i^T (1 - \delta_{i,j}) \triangleq V_{i,j}. \end{aligned} \quad (3.8b)$$

b) The orthogonal projections of $\{\hat{g}_i\}$ onto the column space of S are asymptotically (for large N) jointly Gaussian distributed with zero means and covariance matrices given by

$$\begin{aligned} E(SS^* \hat{g}_i)(SS^* \hat{g}_j)^* &= \frac{\sigma}{N} \left[\sum_{k=1}^n \frac{\lambda_k}{(\sigma - \lambda_k)^2} s_k s_k^* \right] \delta_{i,j} \\ &\triangleq \frac{1}{N} U \delta_{i,j} \end{aligned} \quad (3.9)$$

$$E(SS^* \hat{g}_i)(SS^* \hat{g}_j)^T = 0 \quad \text{for all } i, j. \quad (3.10)$$

Proof: The results (3.8) are standard, see [7], [8] and the references therein. Result (3.9) has apparently been introduced in [5] and [6] (based on results for the real case, presented in [31]) but a proof was not provided. Result (3.10) appears to be new. Proofs of (3.9) and (3.10) can be found in Appendix A. ■

We can now state and prove the result on the asymptotic distribution of the MUSIC estimator.

Theorem 3.1: The MUSIC estimation errors $\{\hat{\omega}_i - \omega_i\}$ are asymptotically (for large N) jointly Gaussian distributed with zero means and variances-covariances given by

$$\begin{aligned} E(\hat{\omega}_i - \omega_i)(\hat{\omega}_j - \omega_j) &= \frac{1}{2N} \frac{\text{Re} \{d^*(\omega_j) GG^* d(\omega_j) \cdot a^*(\omega_j) U a(\omega_j)\}}{h(\omega_i) h(\omega_j)} \end{aligned} \quad (3.11a)$$

where U is defined in (3.9), $d(\omega) = da(\omega)/d\omega$, and

$$h(\omega) = d^*(\omega) GG^* d(\omega). \quad (3.11b)$$

Proof: See Appendix B. ■

For the variance of the estimation error $(\hat{\omega}_i - \omega_i)$, we obtain from (3.11) the following expression:

$$\begin{aligned}
E(\hat{\omega}_i - \omega_i)^2 &= \frac{1}{2N} \frac{a^*(\omega_i) U a(\omega_i)}{h(\omega_i)} \\
&= \frac{\sigma}{2N} \left[\sum_{k=1}^n \frac{\lambda_k}{(\sigma - \lambda_k)^2} |a^*(\omega_i) s_k|^2 \right] \Bigg/ \\
&\quad \left[\sum_{k=1}^{m-n} |d^*(\omega_i) g_k|^2 \right]. \quad (3.12)
\end{aligned}$$

It is interesting to note that the MUSIC variance may take large values when some of the eigenvalues $\{\lambda_k\}_{k=1}^n$ are close to σ . This case corresponds to closely spaced signals (when the matrix A is almost rank deficient), to low signal-to-noise ratio, or to highly correlated signals (when the signal covariance matrix P is nearly singular). The variance may also be large when the vector $d(\omega_i)$ is close to the column space of A (or S) (and therefore, quasi-orthogonal to $\{g_k\}$). In such a case, the transfer vector $a(\omega)$ is relatively insensitive to variations of ω around ω_i , which means that $f(\omega)$ has a flat minimum at $\omega = \omega_i$.

In Section VII, we will derive an alternative formula to (3.12). Using that formula, we will reinforce the conclusions above and will show that the MUSIC estimator variance has the tendency to decrease with increasing m (which is intuitively expected).

C. An Improved MUSIC Estimator and Its Asymptotic Distribution

An appealing (valid for large N) maximum likelihood approach has been used recently in [6] to derive an improved MUSIC estimator. More exactly, the improved estimator is obtained by maximizing the likelihood of the vector

$$\epsilon_i \triangleq a^*(\omega) \hat{g}_i \quad i = 1, \dots, m-n. \quad (3.13)$$

Note that the MUSIC estimator minimizes $\sum_{i=1}^{m-n} |\epsilon_i|^2$. It is claimed in [6] that the estimator that maximizes the asymptotic (for $N \gg 0$) likelihood of (3.13) is given by the minimizer of the following function:

$$\alpha(\omega) = f(\omega)/r(\omega). \quad (3.14a)$$

In (3.14a), $f(\omega)$ is the MUSIC function given by (3.7), and

$$r(\omega) = a^*(\omega) \hat{U} a(\omega) \quad (3.14b)$$

where \hat{U} is the matrix U , (3.9), made of $\{\hat{\lambda}_k\}$ and $\{\hat{s}_k\}$. As an aside, observe that $r(\omega)$ is related to the numerator in the variance formula (3.12). It is shown in the following that the estimates obtained by minimizing (3.14) for any function $r(\omega)$ have the same asymptotic (for large N) distribution as the MUSIC estimator.

Theorem 3.2: Assume that the function $r(\omega)$ satisfies the regularity condition $r(\omega_i) \neq 0$ for $i = 1, \dots, n$, but is otherwise arbitrary. Then the estimates minimizing $f(\omega)$ and $\alpha(\omega)$ have the same asymptotic (for large N) distribution.

Proof: See Appendix C. ■

The result above may explain the similarity in the performances of the MUSIC and the estimator minimizing

(3.14), observed in some of the simulations in [6]. Note that since the result holds for a general function $r(\omega)$, improvement of the MUSIC performance should not be attempted by modifying the MUSIC function as in (3.14a) (at least, not for a "sufficiently large" N). In fact, we show in the following that an "exact" ML approach based on the "data" (3.13) does *not* lead to minimization of (3.14a) and (3.14b).

In Appendix D, we show that the asymptotic (for large N) negative log-likelihood function of (3.13) is given by

$$\begin{aligned}
-\ln L &= \text{const} + (m-n) \ln [a^*(\omega) U a(\omega)] \\
&\quad + \frac{Nf(\omega)}{a^*(\omega) U a(\omega)}. \quad (3.15)
\end{aligned}$$

Observe that (3.15) is $O(1)$ for ω close to a true value, and $O(N)$ otherwise. Thus, for ω close to ω_i for some i (which is the case of great interest), the dominant term of $-\ln L$ is affected if we neglect the second term in (3.15) as was implicitly done in [6]. On the other hand, the dominant term of (3.15) is not affected if U in the second and third terms of $-\ln L$ is replaced by a consistent estimate. We replace U by \hat{U} defined previously. Thus, we propose to determine the estimates of $\{\omega_i\}$ by minimizing the following function with respect to ω :

$$\beta(\omega) = \frac{m-n}{N} \ln r(\omega) + \frac{f(\omega)}{r(\omega)} \quad (3.16)$$

where $r(\omega)$ is defined in (3.14b).

Since the derivative of the first term in (3.16) with respect to ω is $O(1/N)$, it is not difficult to see that the asymptotic (for $N \gg 0$) distribution of the estimator that minimizes (3.16) is also identical to that of the MUSIC. However, in the finite sample case, use of (3.16) may lead to improved performance as compared to (3.14), since (3.16) corresponds to a more exact (and computationally inexpensive) approximation of the likelihood function. Furthermore, when grid methods are used to approximately locate the minima of (3.16) (which is what is usually done), the first term in (3.16) may not be negligible, as explained above. Numerical experience with the new eigenanalysis estimator introduced above will be reported elsewhere.

IV. THE CRAMER-RAO BOUND

In this section, we assume that conditions A1, A2, and AML hold. Under these conditions, we derive the CRB on the covariance matrix of any unbiased estimator of θ and σ . In the following sections, we compare the performance of the MUSIC and the ML estimators (the MLE is discussed in Section V) to the ultimate performance corresponding to the CRB. The usefulness of the CRB formula derived in the sequel is not, of course, limited to the performance studies reported in this paper. It may also be used to establish the relative efficiency of other estimators for θ and σ proposed in the literature (see, e.g., [15]–[17], [20]).

The first result of this section is contained in the following theorem.

Theorem 4.1: Under the assumptions stated, the CRB for θ and σ is given by

$$\text{CRB}(\theta) = \frac{\sigma}{2} \left\{ \sum_{t=1}^N \text{Re} [X^*(t) D^* \cdot [I - A(A^*A)^{-1}A^*] DX(t)] \right\}^{-1} \quad (4.1)$$

and

$$\text{var}_{\text{CRB}}(\sigma) = \frac{\sigma^2}{mN} \quad (4.2)$$

where

$$X(t) = \begin{bmatrix} x_1(t) & 0 \\ & \ddots \\ 0 & x_n(t) \end{bmatrix}$$

$$D = [d(\omega_1) \cdots d(\omega_n)]$$

(recall that $d(\omega) = da(\omega)/d\omega$).

Proof: See Appendix E. ■

In the following, we drop the dependence of the CRB on θ for notational convenience. Instead, we stress the dependence of the CRB on m and N . We would expect the CRB to “decrease” when m or N increases. This intuitively expected result holds indeed, as shown in the next theorem.

Theorem 4.2: The CRB covariance matrix (4.1) satisfies the following order relations:

$$\text{CRB}(N) \geq \text{CRB}(N + 1) \quad (4.3a)$$

$$\text{CRB}(m) \geq \text{CRB}(m + 1). \quad (4.3b)$$

Proof: See Appendix F. ■

The results above are valid for a general transfer vector $a(\omega)$. In the following, we present some specialized results for transfer vectors of the form (2.2), which appear in several signal processing applications (see Section II for some examples).

A. CRB for $n = 1$, $N = 1$ and $a(\omega)$ Given by (2.2)

In this case, we have

$$A = [1 \ e^{i\omega} \cdots e^{i(m-1)\omega}]^T$$

$$D = [0 \ ie^{i\omega} \cdots i(m-1)e^{i(m-1)\omega}]^T$$

which gives

$$A^*A = m$$

$$D^*D = 1 + 2^2 + \cdots + (m-1)^2$$

$$= \frac{m(m-1)(2m-1)}{6}$$

$$D^*A = -i[1 + 2 + \cdots + (m-1)]$$

$$= -i \frac{m(m-1)}{2}. \quad (4.4)$$

Inserting (4.4) into the expression (4.1) of the CRB, we obtain

$$\text{CRB} = \frac{6\sigma}{|x|^2} \frac{1}{m(m^2-1)} = \frac{6}{\text{SNR}} \frac{1}{m(m^2-1)} \approx \frac{6}{m^3 \text{SNR}} \quad (4.5)$$

which agrees with the result for this specialized case derived in [24] (see also [25]). In (4.5), $\text{SNR} = |x|^2/\sigma$.

B. Asymptotic CRB for $a(\omega)$ Given by (2.2)

According to Theorem 4.2, for increasing m or N , $\text{CRB}(m, N)$ forms a sequence of monotonically nonincreasing positive definite matrices. In particular, this implies that $\text{CRB}(m, N)$ has a well-defined limit when either m or N tends to infinity. In the following, we evaluate the limit (or asymptotic) matrices $\text{CRB}(m, \infty)$ and $\text{CRB}(\infty, \infty)$. The formula derived for $\text{CRB}(m, \infty)$ holds for general $a(\omega)$. However, to obtain a formula for $\text{CRB}(\infty, \infty)$ we need to specify $a(\omega)$. The formula for $\text{CRB}(\infty, \infty)$ provided in the following holds for $a(\omega)$ given by (2.2).

Theorem 4.3: a) For sufficiently large N , the CRB is given by

$$\text{CRB}(m, \infty) = \frac{\sigma}{2N} \left\{ \text{Re} \left[\left\{ D^* [I - A(A^*A)^{-1}A^*] D \right\} \circ P^T \right] \right\}^{-1} \quad (4.6)$$

where P is defined in (2.1), and \circ denotes the Hadamard matrix product.

b) Let $a(\omega)$ be given by (2.2). Then, for sufficiently large m and N , the CRB is given by

$$\text{CRB}(\infty, \infty) = \frac{6}{m^3 N} \begin{bmatrix} 1/\text{SNR}_1 & & 0 \\ & \ddots & \\ 0 & & 1/\text{SNR}_n \end{bmatrix} \quad (4.7)$$

where $\text{SNR}_i = P_{ii}/\sigma$ is the signal-to-noise ratio for the i th signal.

Proof: See Appendix G. ■

The usefulness of the above asymptotic CRB formulas lies in the fact that they are (much) easier to evaluate than the exact finite-case formula (4.1), yet they may provide good approximations to the exact CRB for reasonably large values of m and N .

V. THE MAXIMUM LIKELIHOOD ESTIMATOR

In this section, we assume that conditions A1, A2, and AML hold. The log-likelihood function of the observations $\{y(t)\}_{t=1}^N$ is then given by

$$L = \text{const} - mN \ln \sigma - \frac{1}{\sigma} \sum_{t=1}^N [y(t) - Ax(t)]^* \cdot [y(t) - Ax(t)] \quad (5.1)$$

(see (E.1) in Appendix E). The likelihood (5.1) can be concentrated with respect to σ and $\{x(t)\}$. Some straightforward calculations show that the ML estimators of these parameters are given by (see, for example, (E.2a)–(E.2c) in Appendix E)

$$\hat{x}(t) = [\hat{A}^* \hat{A}]^{-1} [\hat{A}^* y(t)] \quad t = 1, \dots, N \quad (5.2)$$

$$\begin{aligned} \hat{\sigma} &= \frac{1}{mN} \sum_{t=1}^N [y(t) - \hat{A} \hat{x}(t)]^* [y(t) - \hat{A} \hat{x}(t)] \\ &= \frac{1}{m} \text{tr} [I - \hat{A} (\hat{A}^* \hat{A})^{-1} \hat{A}^*] \hat{R} \end{aligned} \quad (5.3)$$

where \hat{A} denotes the ML estimate of A (i.e., $\hat{A} = A(\hat{\theta})$, where $\hat{\theta}$ is the MLE of θ). Inserting (5.2) and (5.3) (with \hat{A} , which is not yet determined, replaced by A) into (5.1), we obtain the concentrated likelihood function

$$\text{const} - mN \ln F(\theta)$$

where

$$F(\theta) = \text{tr} [I - A(A^* A)^{-1} A^*] \hat{R}. \quad (5.4)$$

Thus, the ML estimate of θ is given by the minimizer of $F(\theta)$. Detailed discussions on the implementation of the MLE can be found in [10]–[12], [33]–[36], and [42]. Here, we are interested in the properties of the MLE.

From the general theory of estimation, it is known that an MLE is asymptotically (as the number of data points tends to infinity) efficient, under some “regularity” conditions (see, e.g., [27]). In other words, its asymptotic covariance matrix attains the CRB. However, what should “asymptotic” mean in our case where m or N or both could be large? Furthermore, does our estimation problem satisfy the “regularity” conditions? The essential regularity condition is that the MLE is consistent. In our case, there are mN “data” from which $1 + n(N + 1)$ parameters are estimated. The ratio number of data to number of estimated parameters remains bounded if $m < \infty$, even if $N \rightarrow \infty$; this ratio increases without bound if and only if $m \rightarrow \infty$. This observation suggests that the CRB cannot be achieved by increasing N ; the essential requirement for attaining the CRB should be to increase m . The heuristical discussion above is made more precise in the following (see also [37]).

A. The Case of Small m

We begin the analysis by studying the consistency properties of the MLE when $N \rightarrow \infty$ and $m < \infty$. Since \hat{R} tends to R as $N \rightarrow \infty$, it follows that the MLE of θ tends to the minimizer of the following (asymptotic) criterion function:

$$\begin{aligned} &\text{tr} [I - \hat{A} (\hat{A}^* \hat{A})^{-1} \hat{A}^*] R \\ &= \text{tr} [I - \hat{A} (\hat{A}^* \hat{A})^{-1} \hat{A}^*] [A P A^* + \sigma I] \\ &= \text{tr} [I - \hat{A} (\hat{A}^* \hat{A})^{-1} \hat{A}^*] A P A^* + \sigma(m - n) \\ &\geq \sigma(m - n). \end{aligned} \quad (5.5)$$

Clearly, (5.5) is minimized by $\hat{A} = A$, which shows that the MLE of θ is consistent when $N \rightarrow \infty$ (and $m < \infty$). The MLE of $\{x(t)\}$ and σ , however, are inconsistent. This can be seen as follows. When $N \rightarrow \infty$, $\hat{x}(t)$ and $\hat{\sigma}$ tend to the following limits:

$$\begin{aligned} \hat{x}(t) &\rightarrow (A^* A)^{-1} A^* y(t) \\ &= x(t) + (A^* A)^{-1} A^* e(t) \end{aligned} \quad (5.6a)$$

$$\begin{aligned} \hat{\sigma} &\rightarrow \frac{1}{m} \text{tr} [I - A(A^* A)^{-1} A^*] R \\ &= \frac{m - n}{m} \sigma \end{aligned} \quad (5.6b)$$

which, for $m < \infty$, differ from the true values $x(t)$ and σ . The inconsistency of $\hat{x}(t)$ and $\hat{\sigma}$ in the case of $m < \infty$ (anticipated in the previous discussion) implies that the MLE of θ does not achieve the CRB for large N if m is small. The following example illustrates this fact.

Consider the case of a single complex sine wave. The MLE of the sine wave frequency ω is given by the minimizer of [see (5.4)]

$$F(\theta) = \text{tr} [I - a(\omega) [a^*(\omega) a(\omega)]^{-1} a^*(\omega)] \hat{R} \quad (5.7a)$$

where $a(\omega)$ is given by (2.2). Since $a^*(\omega) a(\omega) \equiv m$, minimization of (5.7a) is equivalent to maximization of the following function (which, we note in passing, can be interpreted as an averaged periodogram):

$$g(\theta) = a^*(\omega) \hat{R} a(\omega). \quad (5.7b)$$

We want to determine the asymptotic (for $N \gg 0$) variance of the estimate $\hat{\omega}$ that maximizes (5.7b), and to show that this variance is strictly greater than the CRB for $m < \infty$. After some calculations presented in Appendix H, we obtain

$$\text{var}_{\text{ML}}(\hat{\omega}) = \frac{6\sigma}{N} \frac{mP + \sigma}{m^2(m^2 - 1)P^2}. \quad (5.8)$$

The asymptotic (for large N) CRB is given by [see (4.5) and (4.6)]

$$\text{var}_{\text{CR}}(\hat{\omega}) = \frac{6\sigma}{N} \frac{1}{Pm(m^2 - 1)}.$$

Thus,

$$\text{var}_{\text{ML}}(\hat{\omega}) / \text{var}_{\text{CR}}(\hat{\omega}) = 1 + \frac{\sigma}{mP} = 1 + \frac{1}{m \text{SNR}} \quad (5.9)$$

which shows that *the MLE is inefficient for $m < \infty$, even though $N \rightarrow \infty$* . The example above will be significantly generalized in Section VII.

B. The Case of Large m

We now assume that both N and m tend to infinity. Then the consistency of $\hat{\theta}$ and $\hat{\sigma}$ follows from (5.5) and (5.6b). To establish the consistency of $\hat{x}(t)$, we need to introduce the following additional assumption:

$$a^*(\omega) a(\omega) \rightarrow \infty \quad \text{as } m \rightarrow \infty. \quad (5.10)$$

Observe that $a(\omega)$ given by (2.2) satisfies (5.10). Under (5.10), the covariance matrix of the bias term in (5.6a),

$$E(A^*A)^{-1}A^*e(t)e^*(t)A(A^*A)^{-1} = \sigma(A^*A)^{-1}$$

tends to zero as $m \rightarrow \infty$, which establishes the consistency of $\hat{x}(t)$.

The condition (5.10) is not only sufficient but also necessary for the consistency of $\hat{x}(t)$. In fact, without this condition, the analysis of consistency would be meaningless. Indeed, the signals that do not satisfy (5.10) must be damped in some way (for example, exponentially damped). For such transient signals the behavior for large m is of no interest.

Once the consistency of the MLE has been established, its asymptotic efficiency essentially follows from the general theory of ML estimation [27]. Thus, *under (5.10) the MLE of θ will achieve the CRB for large m and N* . As a simple illustration of this property, observe from (5.9) that the asymptotic (for large N) ratio $\text{var}_{\text{ML}}(\hat{\omega})/\text{var}_{\text{CR}}(\hat{\omega})$ tends to one as m increases.

Remark: For uniformly spaced linear arrays or uniformly sampled undamped exponential signals, the above requirement on N to be large is not necessary. Indeed, the consistency and asymptotic (for large m) efficiency of the ML estimates of sine wave parameters have been established in [26] for the single-experiment case ($N = 1$). ■

Let us summarize the main results of this section. The MLE of θ converges to the true values when N increases. However, if m is small, then the MLE does not achieve the CRB even if N is increased without bound! For damped signal models, there is no remedy to this situation and the CRB cannot in general be attained. For undamped signal models [which satisfy (5.10)], the MLE achieves the CRB as m becomes large. See also Section VII and [37] for an analysis that reinforces the conclusions above.

Remark: It is worth noting that the inefficiency of the ML estimator of θ in the case of a small m is a direct consequence of the requirement to estimate the amplitude values $\{x(1), \dots, x(N)\}$. If $\{x(t)\}_{t=1}^N$ can be assumed to be a sample from a Gaussian white process *and* if it is required to estimate the covariance matrix P only (a far less demanding requirement), then one can conjecture that in such a case the MLE of θ will be statistically efficient for large N and any $m > n$. ■

The previous results provide some guidelines for choosing the values of m and N in a given application of the MLE (assuming that selection of m and N is at the disposal of the user, which, for some applications, such as array processing, where the number m of sensors is fixed, may not be the case). For damped signals, one should proceed in the following rather obvious manner: m should be chosen according to some guess of the signal damping period and N should be increased as much as possible (under restrictions on computer and measurement time). For undamped signals, one should select the values of m and N by a compromise between statistical

efficiency and computational complexity. When m is increased, the MLE performance approaches the CRB performance. Furthermore, the CRB for undamped signal models is expected to decrease (much) faster with m than with N (for example, note from (4.7) that for complex sine waves the CRB decreases as $1/m^3N$ as m and N increase). Thus, from the viewpoint of statistical efficiency, the tendency should be to increase m rather than N . On the other hand, the computational burden associated with the MLE increases faster with m than with N (for example, evaluation of $F(\theta)$ for a given θ requires $O(m^2N)$ arithmetic operations). Thus, the need for the compromise mentioned above when selecting m and N is clearly seen.

VI. THE RELATIONSHIP BETWEEN THE MUSIC AND ML ESTIMATORS

In this section we assume that conditions A1, A2, AMU, and AML all hold, such that both the MUSIC and ML estimators are usable. We want to investigate possible relationships between MUSIC and MLE.

By invoking the invariance principle of the ML estimators, it has been claimed by some authors (see, e.g., [5]) that MUSIC is a realization of the MLE. More precisely, it was claimed that since under the conditions stated, the sample eigenvectors $\{\hat{s}_i\}$ are ML estimates of the true eigenvectors $\{s_i\}$ (see, e.g., [28] and [31]), the MUSIC estimate $\hat{\theta}$ of θ obtained from $\{\hat{s}_i\}$ should be the ML estimate by the invariance principle. However, this line of argument is not quite correct. Briefly stated, the reason is as follows. When $\{\theta\}$ span the set of feasible values D_θ , $\{s_i\}$ span a set that let us denote by D_s . Every point from D_s can be mapped back to a point in D_θ by using the so-called inverse function. Existence of this inverse function is a key condition for the validity of the invariance principle. Now, due to estimation errors, $\{\hat{s}_i\}$ will in general not belong to D_s (which is a ‘‘thin’’ set in $\mathcal{C}^{m \times 1}$); as a consequence, the inverse function cannot be used to determine the point in D_θ that corresponds to $\{\hat{s}_i\}$ for the simple reason that there is no such point (the mapping from $\{\hat{s}_i\}$ to $\hat{\theta}$ employed by MUSIC is only an approximation of the correct inverse function). Thus, the invariance principle fails to be applicable. More details on the invariance principle of ML estimators, and its failure to apply to the type of problem discussed above, can be found in [29].

In Section III, we have seen that the MUSIC estimator is a large sample (for $N \gg 0$) realization of an ML estimator obtained from the approximate distribution of the statistic (3.13). This property does not imply any immediate relationship between the MUSIC and the MLE of Section V, obtained from the exact distribution of the raw data statistic. However, it suggests that the MUSIC estimator possesses some ‘‘optimality’’ property and therefore that a relationship to the MLE is likely to exist. That this is indeed the case is shown in the following theorem.

Theorem 6.1: Under the assumptions stated (A1, A2, AMU, and AML), the MUSIC estimator is a large sample

($N \gg 0$) realization of the MLE of Section V, if and only if the signal covariance matrix P is diagonal.

Proof: See Appendix I. ■

The result of Theorem 6.1 is pleasantly intuitive. When the n signals are uncorrelated, it should indeed be possible (at least for N) to decouple the n -dimensional search problem implied by the MLE into the n one-dimensional problems solved by MUSIC. When the signals are correlated, this should not be possible. Note that it is this decoupling that makes the MUSIC estimator much more attractive computationally than the MLE.

We may also remark that the theorem above provides a theoretical explanation for the good performance of the MUSIC, observed in many experiments with uncorrelated signals, as well as for a degradation of performance when the signals are highly correlated. More quantitative results on these aspects of the MUSIC performance can be found in the next section (see also the companion paper [37]).

VII. AN ANALYTIC AND NUMERICAL STUDY OF PERFORMANCE

Our aim in this section is to study in more detail the MUSIC estimation error variance and to compare it to the CRB. We will use the asymptotic (for large N) formulas (3.12) and (4.6) for the variance of the MUSIC estimator and the CRB, respectively. Thus, our results will be valid for a sufficiently large number N of experiments or snapshots.

A. An Analytic Study

We begin by developing a more convenient formula for the MUSIC error variance. From (3.12) we obtain

$$\begin{aligned} \text{var}_{\text{MU}}(\hat{\omega}_i) &= \frac{\sigma}{2N} \left[\sum_{k=1}^n \left(\frac{1}{\lambda_k - \sigma} + \frac{\sigma}{(\lambda_k - \sigma)^2} \right) |a^*(\omega_i) s_k|^2 \right] / \\ &\quad \left[\sum_{k=1}^{m-n} |d^*(\omega_i) g_k|^2 \right] \\ &= \frac{\sigma}{2N} [a^*(\omega_i) (S\hat{\Lambda}^{-1}S^* + \sigma S\hat{\Lambda}^{-2}S^*) a(\omega_i)] / \\ &\quad [d^*(\omega_i) GG^* d(\omega_i)] \end{aligned} \quad (7.1)$$

where

$$\hat{\Lambda} \triangleq \Lambda - \sigma I \triangleq \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} - \sigma I.$$

Using (3.1) and the eigendecomposition of R , we can write

$$R = APA^* + \sigma I = SAS^* + \sigma GG^* = S\hat{\Lambda}S^* + \sigma I$$

which implies

$$\begin{aligned} APA^* &= S\hat{\Lambda}S^* \\ APA^*APA^* &= S\hat{\Lambda}^2S^* \end{aligned}$$

and therefore

$$\begin{aligned} (S^*A)P(A^*S) &= \hat{\Lambda} \\ (S^*A)P(A^*A)P(A^*S) &= \hat{\Lambda}^2. \end{aligned} \quad (7.2)$$

Since the columns of A lie in the column space of S , and A has full rank, it follows that the matrix S^*A is nonsingular and

$$SS^* = A(A^*A)^{-1}A. \quad (7.3)$$

The nonsingularity of (S^*A) and (7.2) gives

$$\begin{aligned} (A^*S)\hat{\Lambda}^{-1}(S^*A) &= P^{-1} \\ (A^*S)\hat{\Lambda}^{-2}(S^*A) &= P^{-1}(A^*A)^{-1}P^{-1}. \end{aligned} \quad (7.4)$$

Using (7.3) and (7.4) in (7.1), we get

$$\begin{aligned} \text{var}_{\text{MU}}(\hat{\omega}_i) &= \frac{\sigma}{2N} \left\{ [P^{-1}]_{ii} + \sigma [P^{-1}(A^*A)^{-1}P^{-1}]_{ii} \right\} / \\ &\quad h(\omega_i) \end{aligned} \quad (7.5a)$$

where

$$h(\omega) = d^*(\omega) [I - A(A^*A)^{-1}A^*] d(\omega) \quad (7.5b)$$

and $(\cdot)_{ij}$ denotes the i, j element of the matrix in question. The variance (7.5a) may take relatively large values if the signals are highly correlated (i.e., P is nearly singular) or they are closely spaced (i.e., A^*A is nearly singular), or $d(\omega_i)$ is close to the column space of A for some i [a similar conclusion has been drawn in Section III using (3.12)].

Evaluation of formula (7.5) for the MUSIC variance can be done directly from the original parameters σ , P , and $\theta = \{\omega_i\}$ of the problem. The eigendecomposition of R is not necessary, which is in contrast to the evaluation of (3.12). Furthermore, (7.5) can be conveniently used to analytically compare the MUSIC error variance and the CRB. For this, note from (4.6) that the CRB is given by

$$\begin{aligned} \text{var}_{\text{CR}}(\hat{\omega}_i) &= \frac{\sigma}{2N} \left\{ [D^*[I - A(A^*A)^{-1}A^*]D] \odot P^T \right\}^{-1}_{ii}. \end{aligned} \quad (7.6)$$

First, we consider the case of *uncorrelated signals*. In such a case, the matrix P is diagonal, and (7.5) and (7.6) reduce to

$$\text{var}_{\text{MU}}(\hat{\omega}_i) = \frac{1}{2N \cdot \text{SNR}_i} \left[1 + \frac{[(A^*A)^{-1}]_{ii}}{\text{SNR}_i} \right] / h(\omega_i) \quad (7.7a)$$

and

$$\text{var}_{\text{CR}}(\hat{\omega}_i) = \left(\frac{1}{2N \cdot \text{SNR}_i} \right) / h(\omega_i) \quad (7.7b)$$

where $\text{SNR}_i = P_{ii}/\sigma$. Since in the case of uncorrelated signals, the MUSIC is a large sample (for $N \gg 0$) realization of the MLE (see Theorem 6.1), it follows that (7.7a) also gives the variance of the latter estimator.

From (7.7) we obtain

$$\text{var}_{\text{MU}}(\hat{\omega}_i)/\text{var}_{\text{CR}}(\hat{\omega}_i) = 1 + [(A^*A)^{-1}]_{ii}/\text{SNR}_i. \quad (7.8)$$

It is interesting to note that in this case $\text{var}_{\text{MU}}(\hat{\omega}_i)$ decreases *monotonically* with increasing m . This is so since both $[(A^*A)^{-1}]_{ii}$ and $\text{var}_{\text{CR}}(\hat{\omega}_i)$ (see Theorem 4.2) monotonically decrease when m increases. For reasonably large values of m and SNR, the ratio (7.8) expressing the efficiency of the MUSIC estimator will be close to one. Furthermore, if the signals are undamped so that (5.10) holds, the ratio will tend to one as m increases. Thus, we rediscover in another way the fact shown in Section V, that *the MLE, which for diagonal P is equivalent to MUSIC, achieves the CRB as m becomes large for signal models satisfying (5.10)*. If (5.10) is not satisfied and $\text{SNR} \neq \infty$, then the ratio (7.8) will remain strictly greater than one, thus reinforcing our claim in Section V that *for damped signal models the CRB cannot be attained*.

Next, consider the case of *correlated signals*. In this case, the MUSIC error variance cannot attain the CRB. Furthermore, if the matrix P is nearly singular, then the differences $\{\text{var}_{\text{MU}}(\hat{\omega}_i) - \text{var}_{\text{CR}}(\hat{\omega}_i)\}$ may take substantial values, even if m and SNR increase without bound. We illustrate these facts by considering the practically important case of $a(\omega)$ given by (2.2). Using the results in Appendix G we can then show that, as m increases, the variances (7.5) and (7.6) tend to the following limits:

$$\begin{aligned} \text{var}_{\text{MU}}(\hat{\omega}_i) &= \frac{6\sigma}{Nm^3} (P^{-1})_{ii} \\ \text{var}_{\text{CR}}(\hat{\omega}_i) &= \frac{6\sigma}{Nm^3} (P)_{ii}. \end{aligned}$$

Thus, the ratio $\text{var}_{\text{MU}}(\hat{\omega}_i)/\text{var}_{\text{CR}}(\hat{\omega}_i) = (P)_{ii}(P^{-1})_{ii}$ increases without bound as P approaches a singular matrix.

To conclude, for uncorrelated signals, the MUSIC estimator has an excellent performance for reasonably large values of N , m , and SNR. Furthermore, for undamped uncorrelated signals, the MUSIC error variance attains the CRB for large N and m (or SNR). For correlated signals, however, the MUSIC cannot achieve the CRB. If the signals are highly correlated, then the MUSIC estimator may be very inefficient even for large values of N , m , and SNR.

B. A Numerical Study

Consider the case of two signals ($n = 2$) of equal powers, and let

$$P = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \quad \rho \in [0, 1).$$

Furthermore, let $a(\omega)$ be given by (2.2). We evaluated the efficiency ratio

$$\text{eff} = \text{var}_{\text{CR}}(\hat{\omega}_i)/\text{var}_{\text{MU}}(\hat{\omega}_i) \quad (7.9)$$

for $\rho = 0, 0.5, 0.7$ and 0.9 ; $\sigma = 0.01$ (SNR = 20 dB), and 1 (SNR = 0 dB); $m = 5, 10, 25$, and 100; and varying (ω_1, ω_2) . Note that in this case, when $n = 2$, the efficiency ratio (7.9) can be written as a function of the magnitude of the parameter separation $\Delta\omega = |\omega_1 - \omega_2|$. The results obtained are shown in Figs. 1–4 as a function of $\Delta\omega/\pi$.

Figs. 1 and 2 show the results for SNR values of 0 dB and 20 dB, respectively. These figures verify our theoretical result that, for uncorrelated and not too closely spaced signals, the MUSIC is statistically efficient for sufficiently large values of m . As shown in the figures, the values of m for which *eff* is close to one decrease when the signal-to-noise ratio or $\Delta\omega$ increases. The figures also demonstrate the degradation of the MUSIC efficiency when the correlation factor ρ increases. Note that for correlated signals, the MUSIC is in general inefficient even for high values of m and SNR.

Fig. 3 shows the efficiency ratio (7.9) for $\rho = 0.5$, SNR = 0 dB, and different values of m . Fig. 4 shows the corresponding results for SNR = 20 dB. Note from these figures that in the case of $\rho \neq 0$, *eff* decreases with increasing m , for some values of $\Delta\omega$.

Next we show how our results can be used to evaluate the resolvability of the MUSIC algorithm quantitatively. The MUSIC algorithm is unlikely to resolve the signals when $8[\text{st. dev.}_{\text{MU}}(\omega_i)] > \Delta\omega$ (see [8]). Fig. 5 shows $8\sqrt{N}[\text{st. dev.}_{\text{MU}}(\omega_i)]$ as a function of $\Delta\omega$ for SNR = 0 dB and different values of m and ρ . The straight lines shown in the figure correspond to $\sqrt{N}\Delta\omega$ for $N = 100, 200$, and 500. Fig. 6 shows similar results for SNR = 20 dB. MUSIC is unlikely to resolve the two signals for values of $\Delta\omega$ smaller than those at the intersections between the normalized standard deviation curves and the straight lines. It can be seen from the figures that the resolvability of the MUSIC increases with SNR, N or m , as expected. Note also that the resolvability increases much faster with increasing m than with N (as predicted by the developed theory).

VIII. CONCLUSIONS

There are several new results obtained in this paper that should be mentioned.

- The MUSIC estimator was shown to be Gaussian distributed for sufficiently large N , and two equivalent explicit formulas, (3.12) and (7.5), for its error variance have been provided. Formula (3.12) can be readily used in practical applications to evaluate the MUSIC accuracy, since consistent estimates of the terms appearing in (3.12) are obtained in the course of estimation of θ . Formula (7.5) is more convenient for theoretical studies of performance, as explained in the previous section. Note that formulas (3.12) or (7.5) can also be used to study the resolvability of the MUSIC algorithm.

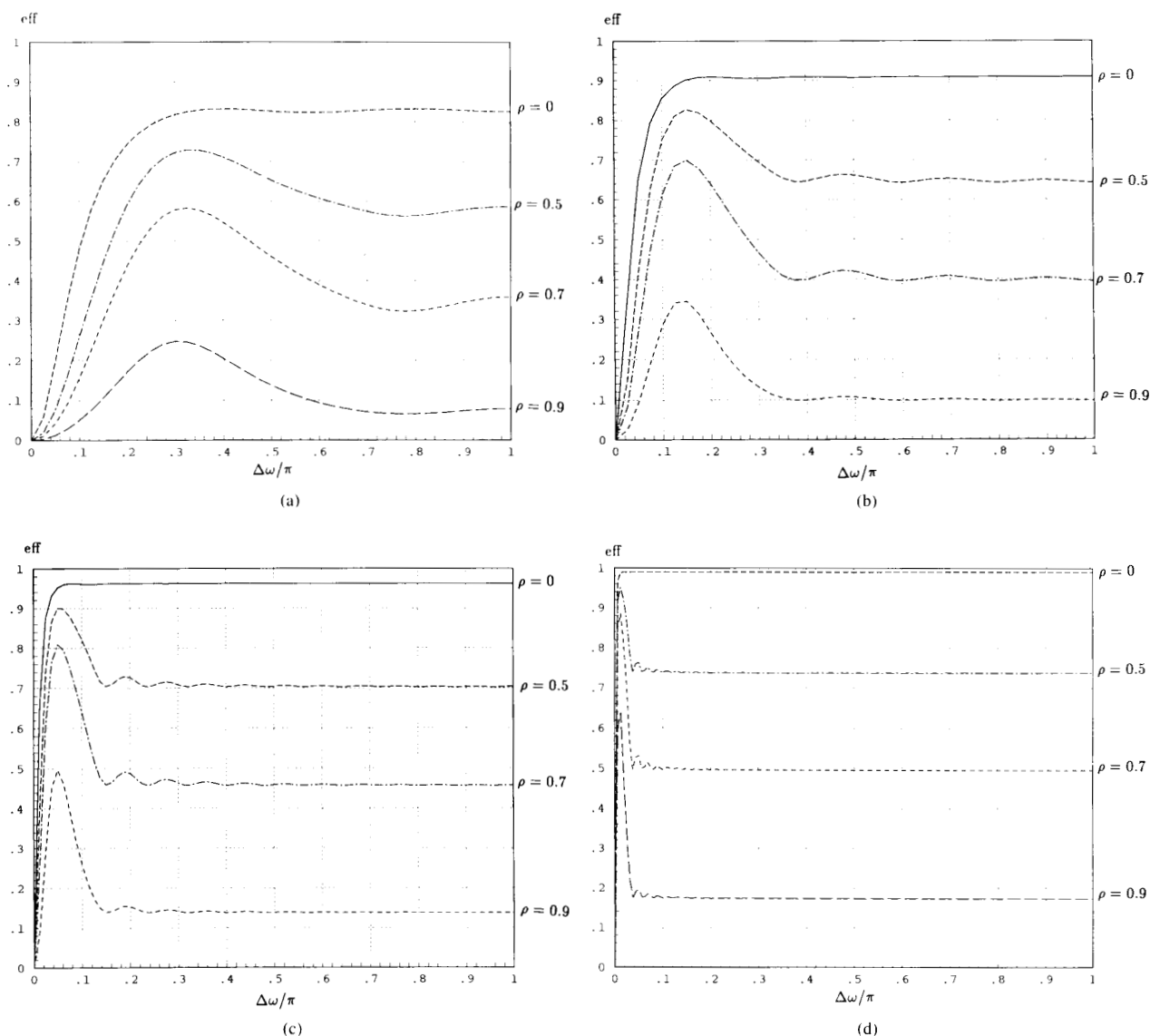


Fig. 1. The relative efficiency (7.9) of the MUSIC estimator for two equal power signals as a function of signal separation $\Delta\omega = |\omega_1 - \omega_2|$. The results shown are for different values of number of sensors m and correlation factor ρ . SNR = 0 dB. (a) $m = 5$, (b) $m = 10$, (c) $m = 25$, (d) $m = 100$.

- The improved MUSIC estimator introduced in [6] has been shown to perform closely to the (basic) MUSIC estimator for large N , thus providing theoretical justification for the empirically observed results of [6]. In fact, the equivalence (for large N) of MUSIC to a whole class of "improved" MUSIC estimators of the form considered in [6] has been proved. A new MUSIC estimator has been obtained by slightly modifying the estimator of [6]; the new estimator is expected to perform better than other MUSIC-type estimators, for reasonably small values of m and N .

- An explicit formula has been derived for the CRB on the covariance matrix of any unbiased estimator of θ . It

was shown that the CRB monotonically decreases with increasing m or N . Simple formulas have been presented for the CRB in the case of large N , and in the case of large m and N and $a(\omega)$ given by (2.2). The CRB formulas derived in this paper should be useful in practical applications and theoretical studies to compare the performance of a given estimator to the ultimate performance corresponding to the CRB.

- It has been shown that the MLE of the parameter vector θ does not achieve the CRB for $N \rightarrow \infty$, if $m < \infty$. For undamped signal models, however, the MLE approaches the CRB performance if m is increased. Based on this type of result, some guidelines for choosing the

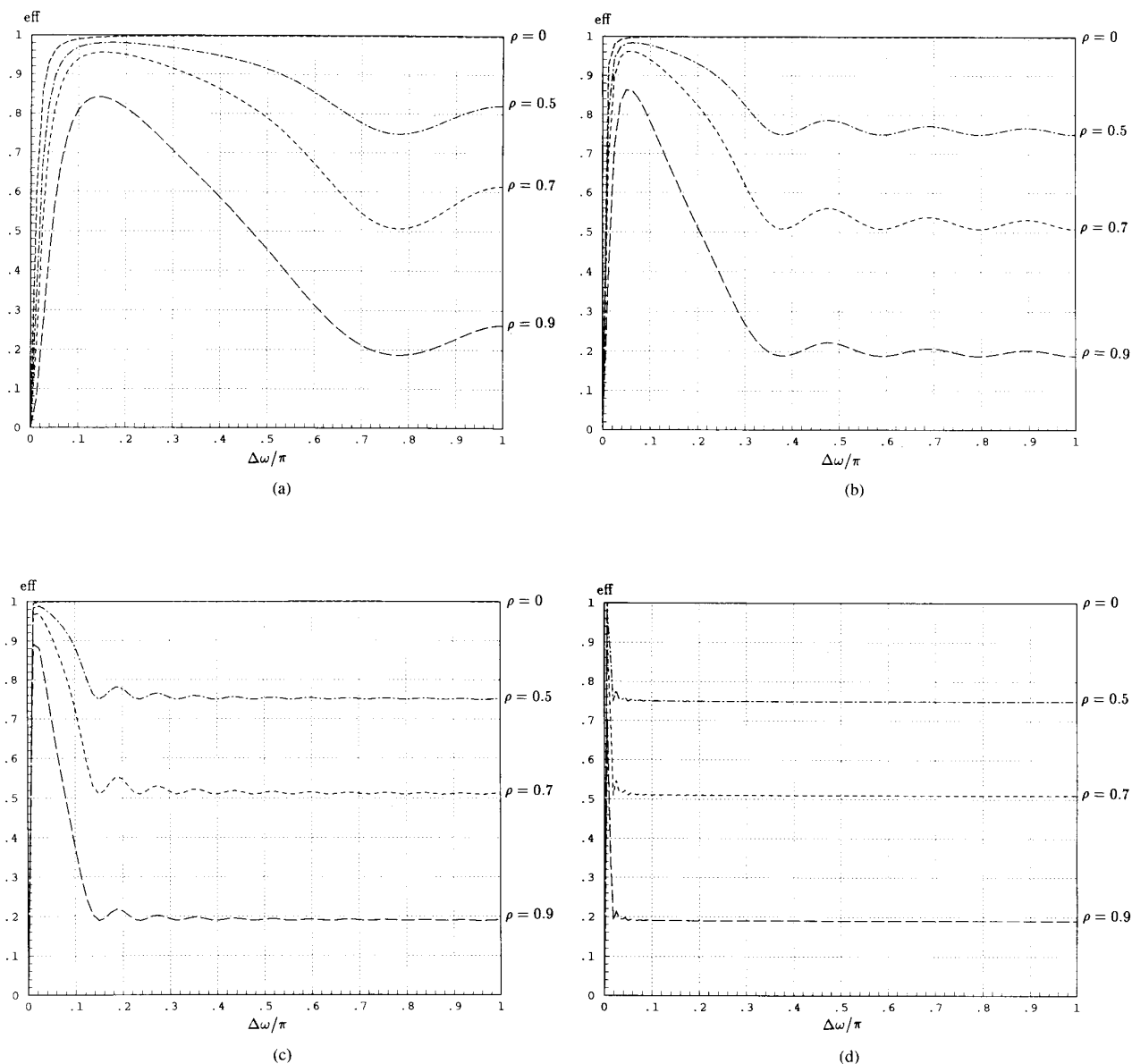


Fig. 2. As in Fig. 1, but for SNR = 20 dB.

values of m and N in a given application (when possible) have been provided.

- The MUSIC estimator has been proven to be a large sample (for $N \gg 0$) realization of the MLE for any $m > n$, if and only if the signals are uncorrelated. A consequence of this result is that the MUSIC should achieve the CRB for uncorrelated undamped signals and large m and N (see above). This property has been shown explicitly to hold for general uncorrelated undamped signals. Furthermore, it was shown that in the case of uncorrelated signals, the MUSIC error variance monotonically decreases with increasing m . In particular, this provides a theoretical justification to the widespread opinion that the computationally efficient Pisarenko algorithm, which cor-

responds to MUSIC with $m = n + 1$ (the smallest possible value), is quite statistically inefficient, and better accuracy may be achieved by increasing m (at the expense of additional computations). For correlated signals, however, the MUSIC performance degrades. It was shown that this degradation of performance can be considerable if the signals are highly correlated (as a remedy, in such cases and for uniform linear arrays, the MUSIC based on a sub-aperture smoothed covariance matrix can be used as proposed, e.g., in [39]). Furthermore, in the case of correlated signals, the MUSIC error variance may occasionally increase when m increases. However, as shown in the numerical examples of Section VII, its general tendency is to decrease with increasing m .

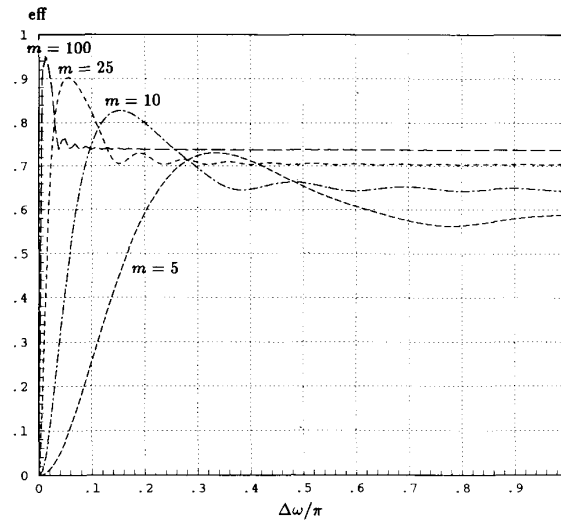


Fig. 3. The relative efficiency (7.9) of the MUSIC estimator for two equal power signals as a function of signal separation $\Delta\omega = |\omega_1 - \omega_2|$. The results shown are for different values of number of sensors m . $\rho = 0.5$ and SNR = 0 dB.

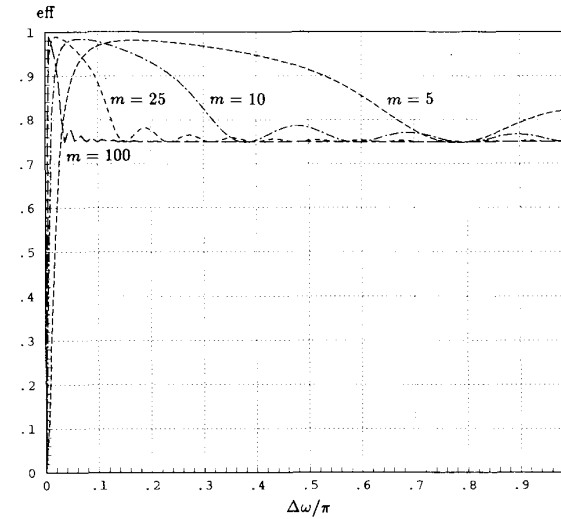


Fig. 4. As in Fig. 3, but for SNR = 20 dB.

APPENDIX A
PROOF OF (3.9) AND (3.10)

Introduce the notation

$$\hat{\Lambda} = \begin{bmatrix} \hat{\lambda}_1 & & 0 \\ & \ddots & \\ 0 & & \hat{\lambda}_n \end{bmatrix} \quad \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} \hat{\lambda}_{n+1} & & 0 \\ & \ddots & \\ 0 & & \hat{\lambda}_m \end{bmatrix}$$

$$\Delta = G^* \hat{R} G - \sigma I$$

$$\Gamma = S^* \hat{R} G.$$

For large N , $\hat{S} = S + O(1/\sqrt{N})$, $\hat{\Lambda} = \Lambda + O(1/\sqrt{N})$, $\hat{\Sigma} = \Sigma + O(1/\sqrt{N})$, and $G^* \hat{S} = G^* S + O(1/\sqrt{N})$. Using these facts and the eigendecomposition of \hat{R} , we get

$$\begin{aligned} & (G^* \hat{G}) (\hat{G}^* G) \\ &= G^* (I - \hat{S} \hat{S}^*) G = I - (G^* \hat{S}) (\hat{S}^* G) \approx I \end{aligned} \tag{A.1}$$

$$\begin{aligned} \Delta &= G^* (\hat{S} \hat{\Lambda} \hat{S}^* + \hat{G} \hat{\Sigma} \hat{G}^*) G - \sigma I \\ &\approx (G^* \hat{G}) \hat{\Sigma} (\hat{G}^* G) - \sigma I \\ &= (G^* \hat{G}) (\hat{\Sigma} - \sigma I) (\hat{G}^* G) - [I - (G^* \hat{G}) (\hat{G}^* G)] \\ &\quad \cdot \sigma \approx (G^* \hat{G}) (\hat{\Sigma} - \sigma I) (\hat{G}^* G) \end{aligned} \tag{A.2}$$

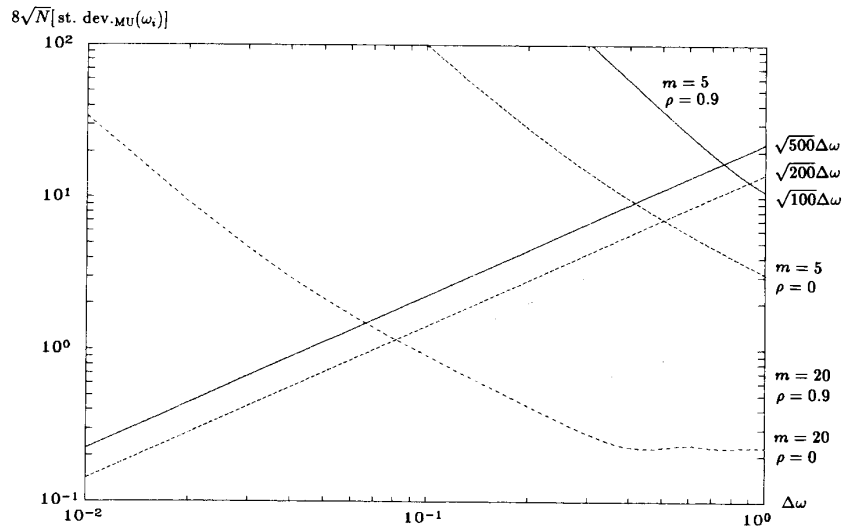


Fig. 5. Curves of $8\sqrt{N}[\text{st. dev.}_{\text{MU}}(\omega_i)]$ as a function of $\Delta\omega$ for two equal power signals, SNR = 0 dB. The straight lines correspond to $\sqrt{N}\Delta\omega$. The MUSIC algorithm is not likely to resolve the signals for values of $\Delta\omega$ smaller than those at the intersection of the standard deviation curves and the straight lines.

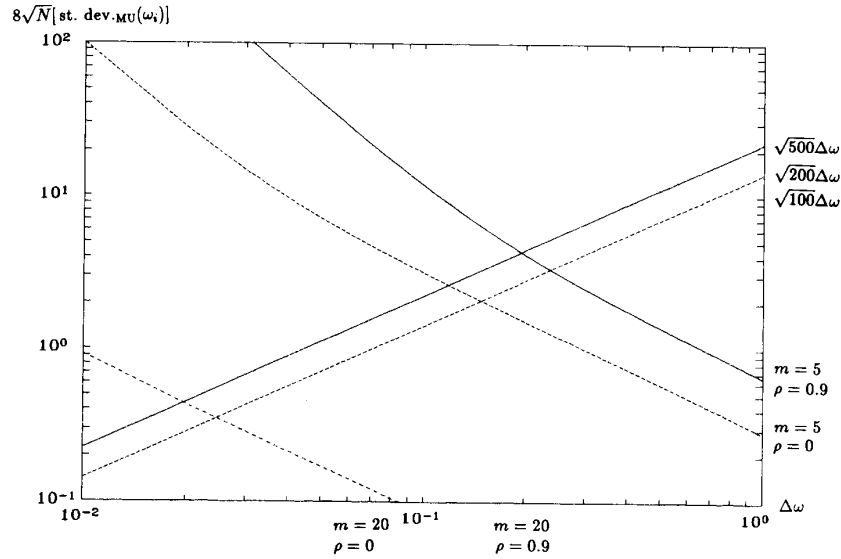


Fig. 6. As in Fig. 5, but for SNR = 20 dB.

$$\begin{aligned} \Gamma &= S^*(\hat{S}\hat{S}^* + \hat{G}\hat{\Sigma}\hat{G}^*)G = (S^*\hat{S})\hat{\Lambda}(\hat{S}^*G) \\ &\quad + (S^*\hat{G})\hat{\Sigma}(\hat{G}^*G) \\ &\approx \Lambda(\hat{S}^*G) + \sigma(S^*\hat{G})(\hat{G}^*G) \end{aligned} \quad (\text{A.3})$$

and

$$\begin{aligned} (S^*\hat{G})(\hat{G}^*G) \\ = S^*(I - \hat{S}\hat{S}^*)G = -(S^*\hat{S})(\hat{S}^*G) \approx -\hat{S}^*G \end{aligned} \quad (\text{A.4})$$

where the terms neglected in the approximations are $O(1/N)$. To proceed we need the following result.

R: The asymptotic distributions of the elements of \hat{S}^*G and of \hat{G}^*G are independent.

Proof: Define

$$u(t) = S^*y(t)$$

$$v(t) = G^*y(t) = G^*e(t)$$

and observe that

$$\Delta_{ij} = \frac{1}{N} \sum_{t=1}^N v_i(t) v_j^*(t) - \sigma\delta_{i,j}$$

$$\Gamma_{kp} = \frac{1}{N} \sum_{t=1}^N u_k(t) v_p^*(t).$$

Since

$$Eu(t) v^*(s) = S^* E y(t) e^*(s) G = \sigma S^* G \delta_{t,s} = 0$$

all t, s

$$Ev(t) v^*(s) = G^* E e(t) e^*(s) G = \sigma I \delta_{t,s}$$

$$Ev(t) v^T(s) = G^* E e(t) e^T(s) G^+ = 0 \quad \text{all } t, s$$

$$Eu(t) v^T(s) = S^* E y(t) e^T(s) G^+ = 0 \quad \text{all } t, s$$

it follows that

$$E\Delta_{ij} = 0 \quad E\Gamma_{kp} = 0.$$

Furthermore, using the following formula [38] for the expectation of the product of four complex Gaussian random variables $\{x_i\}_{i=1}^4$ of which at least one is of zero mean,

$$Ex_1 x_2 x_3 x_4 = (Ex_1 x_2)(Ex_3 x_4) + (Ex_1 x_3)(Ex_2 x_4) \\ + (Ex_1 x_4)(Ex_2 x_3)$$

we get

$$E\Delta_{ij} \Gamma_{kp}^* = \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N E[v_i(t) v_j^*(t) - \sigma \delta_{i,j}] \\ \cdot [u_k^*(s) v_p(s)] = 0 \quad (\text{A.5})$$

$$E\Delta_{ij} \Gamma_{kp} = \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N E[v_i(t) v_j^*(t) - \sigma \delta_{i,j}] \\ \cdot [u_k(s) v_p^*(s)] = 0. \quad (\text{A.6})$$

Since the elements of Δ and Γ are Gaussian distributed by the multivariate central limit theorem, it follows from (A.5) and (A.6) that Δ_{ij} and Γ_{kp} are independent (complex) random variables.

Now, observe from (A.1) and (A.2) that asymptotically the columns of $G^* \hat{G}$ form an orthonormal basis for the eigenspace of Δ . Thus, (A.2) defines $\hat{G}^* G$ uniquely (to within a change of sign) as a continuous function of Δ . It follows that Δ determines the asymptotic distribution of $(\hat{G}^* G)$. Furthermore, from (A.3) and (A.4), we have that $\Gamma \approx (\Lambda - \sigma I)(\hat{S}^* G)$ and, therefore, that Γ determines the asymptotic distribution of $(\hat{S}^* G)$. Since the distributions of Δ and Γ have been shown to be independent, the result follows. ■

It follows from the above result that $(\hat{G}^* G)$ in (A.4) can be considered to be fixed and, therefore, that $S^* \hat{G}$ has the same asymptotic distributions as $-\hat{S}^* G Q$ where Q is some (fixed) unitary matrix (note from (A.1) that $\hat{G}^* G$ is asymptotically unitary). However, the columns of $G Q$ form another set of eigenvectors associated with the repeated eigenvalue σ . Thus, we conclude that $S^* \hat{G}$ and $-\hat{S}^* G$ have the same asymptotic distribution.

The important implication of the above analysis is that $SS^* \hat{g}_i$ and $-S\hat{S}^* g_i$ have the same limiting distribution. However, the limiting distribution of the latter is the same as that of $-S(\Lambda - \sigma I)^{-1} \Gamma_i$ where Γ_i is the i th column of

Γ [cf. (A.3) and (A.4)]. As explained previously, the asymptotic distribution of $\{\Gamma_i\}$ is Gaussian with zero mean. The covariance matrix of this distribution is derived as follows:

$$\lim_{N \rightarrow \infty} E(\sqrt{N} \Gamma_i \cdot \sqrt{N} \Gamma_j^*) \\ = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \sum_{s=1}^N Eu(t) v_i^*(t) v_j(s) u^*(s) \\ = \sigma \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \sum_{s=1}^N Eu(t) u^*(s) \delta_{t,s} \delta_{i,j} \\ = \sigma \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N S^* [Ax(t) x^*(t) A^* + \sigma I] S \delta_{i,j} \\ = \sigma S^* R S \delta_{i,j} = \sigma \Lambda \delta_{i,j} \quad (\text{A.7})$$

and

$$\lim_{N \rightarrow \infty} E(\sqrt{N} \Gamma_i \cdot \sqrt{N} \Gamma_j^T) \\ = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \sum_{s=1}^N Eu(t) v_i^*(t) v_j^*(s) u^T(s) = 0. \quad (\text{A.8})$$

The results (3.9), (3.10) now follow from (A.7), (A.8), and the observation that $SS^* \hat{g}_i$ and $-S(\Lambda - \sigma I)^{-1} \Gamma_i$ have the same asymptotic distribution.

APPENDIX B PROOF OF (3.11)

As $\{\hat{\omega}_i\}$ is a minimum point of $f(\omega)$, we must have

$$f'(\hat{\omega}_i) = 0$$

where

$$f'(\hat{\omega}_i) \triangleq \left. \frac{df(\omega)}{d\omega} \right|_{\omega=\hat{\omega}_i} \\ = d^*(\hat{\omega}_i) \hat{G} \hat{G}^* a(\hat{\omega}_i) + a^*(\hat{\omega}_i) \hat{G} \hat{G}^* d(\hat{\omega}_i) \\ = 2 \operatorname{Re} [a^*(\hat{\omega}_i) \hat{G} \hat{G}^* d(\hat{\omega}_i)]. \quad (\text{B.1})$$

Following the idea of [5], we use the expression (3.7a) of $f(\omega)$, which appears to be more convenient than (3.7b) for the analysis of the distribution of the estimation errors $\{\hat{\omega}_i - \omega_i\}$. Since $\hat{\omega}_i$ is a consistent estimate of ω_i , we can write for sufficiently large N

$$0 = f'(\hat{\omega}_i) \approx f'(\omega_i) + f''(\omega_i) (\hat{\omega}_i - \omega_i) \\ = 2 \operatorname{Re} [a^*(\omega_i) \hat{G} \hat{G}^* d(\omega_i)] \\ + 2 \operatorname{Re} [d^*(\omega_i) \hat{G} \hat{G}^* d(\omega_i) \\ + a^*(\omega_i) \hat{G} \hat{G}^* d'(\omega_i)] (\hat{\omega}_i - \omega_i) \\ \approx 2 \operatorname{Re} [a^*(\omega_i) \hat{G} \hat{G}^* d(\omega_i)] \\ + 2 [d^*(\omega_i) G G^* d(\omega_i)] (\hat{\omega}_i - \omega_i) \quad (\text{B.2a})$$

where the terms neglected in the approximations are $O(1/N)$. The \hat{G}^* in the first term of (B.2a) can be replaced by G^* without affecting the asymptotic distribu-

tion. Indeed,

$$\begin{aligned} A^* \hat{G} \hat{G}^* &\simeq A^* S S^* \hat{G} \hat{G}^* G G^* = A^* S S^* (I - \hat{S} \hat{S}^*) G G^* \\ &= -A^* S (S^* \hat{S}) (\hat{S}^* G) G^* \simeq -A^* S (\hat{S}^* G) G^* \end{aligned} \quad (\text{B.2b})$$

(where again the terms neglected are $O(1/N)$), and $-\hat{S}^* G$ has been shown in Appendix A to be asymptotically equivalent to $S^* \hat{G}$. Also note that

$$\begin{aligned} a^*(\omega_i) \hat{G} \hat{G}^* d(\omega_i) &= a^*(\omega_i) S S^* \hat{G} \hat{G}^* d(\omega_i) \\ &= a^*(\omega_i) \left[S S^* \hat{g}_1 \cdots S S^* \hat{g}_{m-n} \right] \begin{bmatrix} g_1^* d(\omega_i) \\ \vdots \\ g_{m-n}^* d(\omega_i) \end{bmatrix} \\ &= \sum_{k=1}^{m-n} [g_k^* d(\omega_i)] [a^*(\omega_i) S S^* \hat{g}_k]. \end{aligned} \quad (\text{B.3})$$

From (B.2) and (B.3) we get (neglecting the higher-order terms)

$$\begin{aligned} \hat{\omega}_i - \omega_i &= -\text{Re} \left\{ \sum_{k=1}^{m-n} [g_k^* d(\omega_i)] [a^*(\omega_i) S S^* \hat{g}_k] \right\} / h(\omega_i). \end{aligned} \quad (\text{B.4})$$

The asymptotic zero-mean Gaussian distribution of $\{\hat{\omega}_i - \omega_i\}$ follows from (B.4) and Lemma 3.1, part b. It remains

$$\alpha'(\omega) = \frac{f'(\omega) r(\omega) - f(\omega) r'(\omega)}{r^2(\omega)} \quad (\text{C.2a})$$

$$\alpha''(\omega) = \frac{[f''(\omega) r(\omega) - f(\omega) r''(\omega)] r^2(\omega) - [f'(\omega) r(\omega) - f(\omega) r'(\omega)] 2r(\omega) r'(\omega)}{r^4(\omega)}. \quad (\text{C.2b})$$

to verify formula (3.11) for the variances-covariances of the estimation errors $\{\hat{\omega}_i - \omega_i\}$. Toward this end, note that for two scalar-valued complex variables, u and v , we have

$$\text{Re}(u) \cdot \text{Re}(v) = \frac{1}{2} [\text{Re}(uv) + \text{Re}(uv^*)]. \quad (\text{B.5})$$

Let

$$\begin{aligned} u &= \sum_{k=1}^{m-n} [g_k^* d(\omega_i)] [a^*(\omega_i) S S^* \hat{g}_k] \\ v &= \sum_{p=1}^{m-n} [g_p^* d(\omega_j)] [a^*(\omega_j) S S^* \hat{g}_p]. \end{aligned} \quad (\text{B.6})$$

Using Lemma 3.1 (part b), we obtain

$$\begin{aligned} Euv &= \sum_{k=1}^{m-n} \sum_{p=1}^{m-n} [g_k^* d(\omega_i)] [g_p^* d(\omega_j)] a^*(\omega_i) \\ &\quad \cdot E(SS^* \hat{g}_k) (SS^* \hat{g}_p)^T a^+(\omega_j) = 0 \end{aligned}$$

and

$$\begin{aligned} Euv^* &= \sum_{k=1}^{m-n} \sum_{p=1}^{m-n} [g_k^* d(\omega_i)] [d^*(\omega_j) g_p] a^*(\omega_i) \\ &\quad \cdot \underbrace{E(SS^* \hat{g}_k) (SS^* \hat{g}_p)^*}_{(1/N) U \delta_{k,p}} a(\omega_j) \\ &= \frac{1}{N} [d^*(\omega_j) G G^* d(\omega_i)] [a^*(\omega_i) U a(\omega_j)]. \end{aligned} \quad (\text{B.7})$$

Formula (3.11) now follows from (B.4)–(B.7), and the proof is finished.

Remark: It is worth noting that the asymptotic distribution of the MUSIC estimation errors could be derived from (B.2a), (B.2b), and part (a) of Lemma 3.1, thus circumventing the need of using part (b) of the lemma. However, we preferred to base the derivation on part (b) of Lemma 3.1 to motivate inclusion of the results (3.9), (3.10) which appear to be of independent interest. Note, for example, that results (3.9), (3.10) are used in Appendix D to establish the asymptotic distribution of the random variables (3.13).

APPENDIX C PROOF OF THEOREM 3.2

Let $\hat{\omega}$ denote a generic minimum point of (3.14). A simple Taylor series expansion similar to (B.2) gives

$$0 = \alpha'(\hat{\omega}) \simeq \alpha'(\omega) + \alpha''(\omega) (\hat{\omega} - \omega) \quad (\text{C.1})$$

where

Since $f(\omega) = O(1/N)$ and $f'(\omega) = O(1/\sqrt{N})$, it follows from (C.1) and (C.2) that

$$\begin{aligned} 0 &\simeq \alpha'(\omega) + \alpha''(\omega) (\hat{\omega} - \omega) \\ &= \frac{f'(\omega)}{r(\omega)} + \frac{f''(\omega)}{r(\omega)} (\hat{\omega} - \omega) \end{aligned} \quad (\text{C.3})$$

where the neglected terms go to zero faster than $(\hat{\omega} - \omega)$, when N tends to infinity. From (C.3), we get

$$f'(\omega) + f''(\omega) (\hat{\omega} - \omega) \simeq 0$$

which is exactly (B.2) corresponding to MUSIC, and thus the proof is finished.

APPENDIX D PROOF OF (3.15)

It follows from Lemma 3.1 that the random variables

$$\epsilon_i \triangleq a^*(\omega) \hat{g}_i = a^*(\omega) S S^* \hat{g}_i \quad i = 1, \dots, m-n$$

are Gaussian distributed with zero means and the following variances-covariances:

$$\begin{aligned} E\epsilon_i\epsilon_j^* &= \frac{1}{N} a^*(\omega) Ua(\omega) \delta_{i,j} \\ E\epsilon_i\epsilon_j &= 0 \quad \text{for all } i, j. \end{aligned} \quad (\text{D.1})$$

From (D.1) we get

$$\begin{aligned} E\bar{\epsilon}_k\bar{\epsilon}_p &= \frac{1}{4} E(\epsilon_k + \epsilon_k^*)(\epsilon_p + \epsilon_p^*) \\ &= \frac{1}{2N} a^*(\omega) Ua(\omega) \delta_{k,p} \end{aligned}$$

$$\begin{aligned} E\tilde{\epsilon}_k\tilde{\epsilon}_p &= -\frac{1}{4} E(\epsilon_k - \epsilon_k^*)(\epsilon_p - \epsilon_p^*) \\ &= \frac{1}{2N} a^*(\omega) Ua(\omega) \delta_{k,p} \end{aligned}$$

$$E\bar{\epsilon}_k\tilde{\epsilon}_p = \frac{1}{4i} E(\epsilon_k + \epsilon_k^*)(\epsilon_p - \epsilon_p^*) = 0 \quad \text{for all } k \text{ and } p.$$

Thus, the likelihood function is given by

$$\begin{aligned} L(\epsilon_1, \dots, \epsilon_{m-n}) \\ &\triangleq L(\bar{\epsilon}_1, \dots, \bar{\epsilon}_{m-n}, \tilde{\epsilon}_1, \dots, \tilde{\epsilon}_{m-n}) \\ &= \frac{1}{(2\pi)^{m-n} [(1/2N) a^*(\omega) Ua(\omega)]^{m-n}} \\ &\quad \cdot \exp \left\{ -\frac{N}{a^*(\omega) Ua(\omega)} \sum_{k=1}^{m-n} |\epsilon_k|^2 \right\} \end{aligned}$$

and, therefore, the log-likelihood is

$$\begin{aligned} \ln L &= \text{const} - (m-n) \ln [a^*(\omega) Ua(\omega)] \\ &\quad - \left[N \sum_{k=1}^{m-n} |\epsilon_k|^2 \right] / [a^*(\omega) Ua(\omega)] \end{aligned}$$

which proves (3.15).

APPENDIX E DERIVATION OF THE CRB

The likelihood function of the data is given by

$$\begin{aligned} L(y(1), \dots, y(N)) \\ &= \frac{1}{(2\pi)^{mN} (\sigma/2)^{mN}} \exp \left\{ -\frac{1}{\sigma} \sum_{t=1}^N [y(t) - Ax(t)]^* \right. \\ &\quad \left. \cdot [y(t) - Ax(t)] \right\}. \end{aligned}$$

Thus, the log-likelihood function is

$$\begin{aligned} \ln L &= \text{const} - mN \ln \sigma - \frac{1}{\sigma} \sum_{t=1}^N [y^*(t) - x^*(t)A^*] \\ &\quad \cdot [y(t) - Ax(t)]. \end{aligned} \quad (\text{E.1})$$

First, we calculate the derivatives of (E.1) with respect to σ , $\{\bar{x}(t) \triangleq \text{Re } x(t)\}$, $\{\tilde{x}(t) \triangleq \text{Im } x(t)\}$ and θ . We

have

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{mN}{\sigma} + \frac{1}{\sigma^2} \sum_{t=1}^N e^*(t) e(t) \quad (\text{E.2a})$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \bar{x}(k)} &= \frac{1}{\sigma} [A^*e(k) + A^T e^+(k)] = \frac{2}{\sigma} \text{Re} [A^*e(k)] \\ &k = 1, \dots, N \end{aligned} \quad (\text{E.2b})$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \tilde{x}(k)} &= \frac{1}{\sigma} [-iA^*e(k) + iA^T e^+(k)] = \frac{2}{\sigma} \text{Im} [A^*e(k)] \\ &k = 1, \dots, N \end{aligned} \quad (\text{E.2c})$$

and

$$\begin{aligned} \frac{\partial \ln L}{\partial \omega_i} &= \frac{2}{\sigma} \sum_{t=1}^N \text{Re} \left[x^*(t) \frac{dA^*}{d\omega_i} e(t) \right] \\ &= \frac{2}{\sigma} \sum_{t=1}^N \text{Re} [x_i^*(t) d^*(\omega_i) e(t)] \\ &\quad i = 1, \dots, n \end{aligned}$$

which can be written more compactly as

$$\frac{\partial \ln L}{\partial \theta} = \frac{2}{\sigma} \sum_{t=1}^N \text{Re} [X^*(t) D^* e(t)]. \quad (\text{E.2d})$$

To proceed, we need a number of results. These are stated and proven in the following.

$$\begin{aligned} R1: Ee^*(t) e(t) e^*(s) e(s) \\ &= \begin{cases} m^2 \sigma^2 & \text{for } t \neq s \\ m(m+1) \sigma^2 & \text{for } t = s. \end{cases} \end{aligned} \quad (\text{E.3})$$

Proof: For $t \neq s$,

$$\begin{aligned} Ee^*(t) e(t) e^*(s) e(s) \\ &= [E\bar{e}^T(t) \bar{e}(t) + E\tilde{e}^T(t) \tilde{e}(t)]^2 = m^2 \sigma^2. \end{aligned}$$

For $t = s$,

$$\begin{aligned} Ee^*(t) e(t) e^*(s) e(s) \\ &= E[\bar{e}^T(t) \bar{e}(t) + \tilde{e}^T(t) \tilde{e}(t)]^2 \\ &= E[\bar{e}^T(t) \bar{e}(t)]^2 + 2E[\tilde{e}^T(t) \tilde{e}(t)] E[\bar{e}^T(t) \bar{e}(t)] \\ &\quad + E[\tilde{e}^T(t) \tilde{e}(t)]^2 \\ &= 2E[\bar{e}^T(t) \bar{e}(t)]^2 + \frac{1}{2} m^2 \sigma^2. \end{aligned}$$

Since

$$\begin{aligned} E[\bar{e}^T(t) \bar{e}(t)]^2 \\ &= E \sum_{i=1}^m \sum_{j=1}^m \bar{e}_i^2(t) \bar{e}_j^2(t) = \sum_{i=1}^m \sum_{\substack{j=1 \\ j \neq i}}^m E\bar{e}_i^2(t) E\bar{e}_j^2(t) \\ &\quad + \sum_{i=1}^m E\bar{e}_i^4(t) \\ &= (m-1)m \frac{\sigma^2}{4} + 3m \frac{\sigma^2}{4} = m(m+2) \frac{\sigma^2}{4} \end{aligned}$$

the proof is finished. ■

$$R2: Ee^*(t) e(t) e^T(s) = 0 \quad \text{for all } t \text{ and } s. \quad (\text{E.4})$$

Proof: For $t \neq s$, the result is immediate since $e(t)$ and $e(s)$ are independent. For $t = s$, it follows from the fact that the third-order moments of Gaussian random variables are equal to zero. ■

$$\begin{aligned} \text{R3: } \operatorname{Re}(x) \operatorname{Re}(y^T) &= \frac{1}{2} [\operatorname{Re}(xy^T) + \operatorname{Re}(xy^*)] \\ \operatorname{Im}(x) \operatorname{Im}(y^T) &= -\frac{1}{2} [\operatorname{Re}(xy^T) - \operatorname{Re}(xy^*)] \\ \operatorname{Re}(x) \operatorname{Im}(y^T) &= \frac{1}{2} [\operatorname{Im}(xy^T) - \operatorname{Im}(xy^*)]. \end{aligned} \quad (\text{E.5})$$

Proof: The result follows from some straightforward calculations. ■

R4: Let H be a nonsingular complex matrix, and denote its inverse by $G \triangleq H^{-1}$. Then

$$\begin{bmatrix} \bar{H} & -\hat{H} \\ \hat{H} & \bar{H} \end{bmatrix}^{-1} = \begin{bmatrix} \bar{G} & -\hat{G} \\ \hat{G} & \bar{G} \end{bmatrix}. \quad (\text{E.6})$$

Proof: The equality (E.6) can equivalently be written as

$$\begin{aligned} \bar{H}\bar{G} - \hat{H}\hat{G} &= I \\ \bar{H}\hat{G} + \hat{H}\bar{G} &= 0 \end{aligned}$$

which certainly must hold since

$$\begin{aligned} I &= HG = (\bar{H} + i\hat{H})(\bar{G} + i\hat{G}) \\ &= (\bar{H}\bar{G} - \hat{H}\hat{G}) + i(\bar{H}\hat{G} + \hat{H}\bar{G}). \end{aligned} \quad \blacksquare$$

Turn now to the evaluation of the CRB covariance matrix, which is given by

$$\Omega = (E\psi\psi^T)^{-1} \quad (\text{E.7a})$$

where

$$\psi^T \triangleq \partial \ln L / \partial [\sigma \bar{x}^T(1) \bar{x}^T(1) \cdots \bar{x}^T(N) \bar{x}^T(N) \theta^T]. \quad (\text{E.7b})$$

Using R1, we get

$$\begin{aligned} E \left[\frac{\partial \ln L}{\partial \sigma} \right]^2 &= \frac{m^2 N^2}{\sigma^2} - 2 \frac{mN}{\sigma^3} \sum_{t=1}^N E e^*(t) e(t) \\ &\quad + \frac{1}{\sigma^4} \sum_{t=1}^N \sum_{s=1}^N E e^*(t) e(t) e^*(s) e(s) \\ &= \frac{m^2 N^2}{\sigma^2} - 2 \frac{m^2 N^2}{\sigma^2} \\ &\quad + \frac{Nm}{\sigma^2} [(N-1)m + (m+1)] \\ &= \frac{mN}{\sigma^2}. \end{aligned} \quad (\text{E.8a})$$

Using R2, we note that $\partial \ln L / \partial \sigma$ is not correlated with any of the other derivatives in (E.2).

Next, we use R3 and the fact that $E e(t) e^T(s) = 0$ for

all t and s (see assumption A2), to obtain

$$\begin{aligned} E \left[\frac{\partial \ln L}{\partial \bar{x}(k)} \middle| \middle| \frac{\partial \ln L}{\partial \bar{x}(p)} \right]^T &= \frac{4}{\sigma^2} \frac{1}{2} \operatorname{Re} [EA^* e(k) e^*(p) A] \\ &= \frac{2}{\sigma} \operatorname{Re} [A^* A] \delta_{k,p} \end{aligned} \quad (\text{E.8b})$$

$$\begin{aligned} E \left[\frac{\partial \ln L}{\partial \bar{x}(k)} \middle| \middle| \frac{\partial \ln L}{\partial \bar{x}(p)} \right]^T &= -\frac{4}{\sigma^2} \frac{1}{2} \operatorname{Im} [EA^* e(k) e^*(p) A] \\ &= -\frac{2}{\sigma} \operatorname{Im} [A^* A] \delta_{k,p} \end{aligned} \quad (\text{E.8c})$$

$$\begin{aligned} E \left[\frac{\partial \ln L}{\partial \bar{x}(k)} \middle| \middle| \frac{\partial \ln L}{\partial \theta} \right]^T &= \frac{4}{\sigma^2} \sum_{t=1}^N \frac{1}{2} \\ &\quad \cdot \operatorname{Re} [EA^* e(k) e^*(t) DX(t)] \\ &= \frac{2}{\sigma} \operatorname{Re} [A^* DX(k)] \end{aligned} \quad (\text{E.8d})$$

$$\begin{aligned} E \left[\frac{\partial \ln L}{\partial \bar{x}(k)} \middle| \middle| \frac{\partial \ln L}{\partial \bar{x}(p)} \right]^T &= \frac{4}{\sigma^2} \frac{1}{2} \operatorname{Re} [EA^* e(k) e^*(p) A] \\ &= \frac{2}{\sigma} \operatorname{Re} [A^* A] \delta_{k,p} \end{aligned} \quad (\text{E.8e})$$

$$\begin{aligned} E \left[\frac{\partial \ln L}{\partial \bar{x}(k)} \middle| \middle| \frac{\partial \ln L}{\partial \theta} \right]^T &= \frac{4}{\sigma^2} \sum_{t=1}^N \left(-\frac{1}{2} \right) \\ &\quad \cdot \operatorname{Im} [EX^*(t) D^* e(t) e^*(k) A]^T \\ &= -\frac{2}{\sigma} \operatorname{Im} [X^*(k) D^* A]^T \\ &= \frac{2}{\sigma} \operatorname{Im} [A^* DX(k)] \end{aligned} \quad (\text{E.8f})$$

$$\begin{aligned} E \left[\frac{\partial \ln L}{\partial \theta} \middle| \middle| \frac{\partial \ln L}{\partial \theta} \right]^T &= \frac{4}{\sigma^2} \frac{1}{2} \sum_{t=1}^N \sum_{s=1}^N \operatorname{Re} [EX^*(t) D^* \\ &\quad \cdot e(t) e^*(s) DX(s)] \\ &= \frac{2}{\sigma} \sum_{t=1}^N \operatorname{Re} [X^*(t) D^* DX(t)] = \Gamma. \end{aligned} \quad (\text{E.8g})$$

Introduce the following notations:

$$\operatorname{var}_{\text{CRB}}(\sigma) = \sigma^2 / mN$$

$$H = \frac{2}{\sigma} A^* A$$

$$G = H^{-1}$$

$$\Delta_k = \frac{2}{\sigma} A^* DX(k).$$

Observe that since the matrix H is Hermitian, its imaginary part must be skew-symmetric $\hat{H}^T = -\hat{H}$. Inserting

(E.8) into (E.7) and using the notation above, we get

$$\Omega = \begin{bmatrix} \text{var}_{\text{CR}}^{-1}(\sigma) & & & & 0 \\ & \begin{bmatrix} \bar{H} & -\bar{H} \\ \hat{H} & \bar{H} \end{bmatrix} & & & \begin{bmatrix} \bar{\Delta}_1 \\ \hat{\Delta}_1 \\ \vdots \\ \bar{\Delta}_N \\ \hat{\Delta}_N \end{bmatrix} \\ & & 0 & & \\ & & & \ddots & \\ 0 & & & & \begin{bmatrix} \bar{H} & -\bar{H} \\ \hat{H} & \bar{H} \end{bmatrix} \\ & & & & \begin{bmatrix} \bar{\Delta}_1 \\ \hat{\Delta}_1 \\ \vdots \\ \bar{\Delta}_N \\ \hat{\Delta}_N \end{bmatrix} \\ & & & & \Gamma \end{bmatrix}^{-1} \quad (\text{E.9})$$

The expression (4.2) for $\text{var}_{\text{CR}}(\sigma)$ is thus proven. To prove the expression (4.1) for $\text{CRB}(\theta)$, we note that (E.9), a standard result on the inverse of a partitioned matrix, and R4 give

$$\text{CRB}^{-1}(\theta) = \Gamma - \begin{bmatrix} \bar{\Delta}_1^T & \hat{\Delta}_1^T \\ \vdots & \vdots \\ \bar{\Delta}_N^T & \hat{\Delta}_N^T \end{bmatrix} \begin{bmatrix} \bar{G} & -\bar{G} \\ \hat{G} & \bar{G} \\ & 0 \\ & & \ddots \\ & & & \bar{G} & -\bar{G} \\ & & & \hat{G} & \bar{G} \end{bmatrix} \begin{bmatrix} \bar{\Delta}_1 \\ \hat{\Delta}_1 \\ \vdots \\ \bar{\Delta}_N \\ \hat{\Delta}_N \end{bmatrix} \quad (\text{E.10})$$

Next, observe that

$$\begin{bmatrix} \bar{G} & -\bar{G} \\ \hat{G} & \bar{G} \end{bmatrix} \begin{bmatrix} \bar{\Delta} \\ \hat{\Delta} \end{bmatrix} = \begin{bmatrix} \bar{G}\bar{\Delta} - \bar{G}\hat{\Delta} \\ \hat{G}\bar{\Delta} + \bar{G}\hat{\Delta} \end{bmatrix} = \begin{bmatrix} \bar{G}\bar{\Delta} \\ \widetilde{G}\bar{\Delta} \end{bmatrix} \quad (\text{E.11a})$$

and that

$$\begin{bmatrix} \bar{\Delta}_1^T & \hat{\Delta}_1^T \\ \vdots & \vdots \\ \bar{\Delta}_N^T & \hat{\Delta}_N^T \end{bmatrix} \begin{bmatrix} \bar{G}\bar{\Delta} \\ \widetilde{G}\bar{\Delta} \end{bmatrix} = \text{Re} [\Delta^* G \Delta]. \quad (\text{E.11b})$$

From (E.10) and (E.11), it follows that

$$\begin{aligned} \text{CRB}^{-1}(\theta) &= \Gamma - \sum_{t=1}^N \text{Re} [\Delta_t^* G \Delta_t] = \frac{2}{\sigma} \sum_{t=1}^N \text{Re} \left\{ X^*(t) D^* DX(t) \right. \\ &\quad \left. - X^*(t) D^* A(A^*A)^{-1} A^* DX(t) \right\} \\ &= \frac{2}{\sigma} \sum_{t=1}^N \text{Re} \left\{ X^*(t) D^* [I - A(A^*A)^{-1} A^*] DX(t) \right\} \end{aligned}$$

which completes the proof.

Remark: A slightly more compressed derivation of the CRB formula (4.1) can be obtained if one uses the extension of Bangs' formula [40] for the CRB matrix with

information in the mean and covariance, see [41], instead of the standard general formula (E.7).

APPENDIX F PROOF OF (4.3)

We have

$$\text{CRB}^{-1}(N+1) = \text{CRB}^{-1}(N) + \frac{2}{\sigma} \text{Re} \left\{ X^*(N+1) D^* \cdot [I - A(A^*A)^{-1} A^*] DX(N+1) \right\}. \quad (\text{F.1})$$

The matrix in braces is Hermitian positive definite and thus its real part is symmetric positive semidefinite. This observation and (F.1) prove (4.3a).

To prove (4.3b), let us introduce for convenience the following notation:

$$H = (A_m^* A_m)^{-1}$$

$$G = A_m^* D_m$$

$$u^* = \text{the last row of } A_{m+1}$$

$$v^* = \text{the last row of } D_{m+1}.$$

Making use of the nested structures of A_{m+1} and D_{m+1} ,

$$A_{m+1} = \begin{bmatrix} A_m \\ u^* \end{bmatrix} \quad D_{m+1} = \begin{bmatrix} D_m \\ v^* \end{bmatrix}$$

and of the matrix inversion lemma (see, e.g., [32]), we can write

$$\begin{aligned} D_{m+1}^* [I - A_{m+1} (A_{m+1}^* A_{m+1})^{-1} A_{m+1}^*] D_{m+1} &= D_m^* D_m + v v^* - (G^* + v u^*) (H^{-1} + u u^*)^{-1} \\ &\quad \cdot (G + u v^*) \\ &= D_m^* D_m + v v^* - (G^* + v u^*) \left(H - \frac{H u u^* H}{1 + u^* H u} \right) \\ &\quad \cdot (G + u v^*) \\ &= D_m^* [I - A_m (A_m^* A_m)^{-1} A_m^*] D_m + Q \end{aligned} \quad (\text{F.2})$$

where

$$\begin{aligned} Q &\triangleq v v^* + \frac{G^* H u u^* H G}{1 + u^* H u} - G^* \left[H - \frac{H u u^* H}{1 + u^* H u} \right] u v^* \\ &\quad - v u^* \left[H - \frac{H u u^* H}{1 + u^* H u} \right] G \\ &\quad - v u^* \left[H - \frac{H u u^* H}{1 + u^* H u} \right] u v^* \\ &= (v v^* + v v^* u^* H u + G^* H u u^* H G - G^* H u v^* \\ &\quad - v u^* H G - v u^* H u v^*) / (1 + u^* H u) \\ &= (v - G^* H u) (v - G^* H u)^* / (1 + u^* H u). \end{aligned}$$

Since the matrix Q is evidently Hermitian positive semi-definite, the inequality (4.3b) follows from (F.2) and the expression (4.1) of the CRB.

APPENDIX G
PROOF OF THEOREM 4.3

Let

$$H \triangleq D^* [I - A(A^*A)^{-1}A^*] D.$$

Then, the i, j element of the matrix whose inverse appears in (4.1) can be written as

$$\begin{aligned} \operatorname{Re} \left\{ d^*(\omega_i) [I - A(A^*A)^{-1}A^*] d(\omega_j) \sum_{t=1}^N x_i^*(t) x_j(t) \right\} \\ = N \operatorname{Re} \left[H_{ij} \cdot \frac{1}{N} \sum_{t=1}^N x_i^*(t) x_j(t) \right]. \end{aligned}$$

Since by definition

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N x_i^*(t) x_j(t) = P_{ji}$$

it readily follows that for sufficiently large N , the CRB is given by

$$\frac{\sigma}{2N} [\operatorname{Re} (H \circ P^T)]^{-1}$$

which proves part (a).

Next consider part (b). To prove (4.7) we make use of the following result [26]:

$$\frac{1}{m^{k+1}} \sum_{t=1}^m t^k e^{it(\omega_1 - \omega_2)} \xrightarrow{m \rightarrow \infty} \begin{cases} 1/(k+1) & \text{for } \omega_1 = \omega_2 \\ 0 & \text{for } \omega_1 \neq \omega_2. \end{cases} \quad (\text{G.1})$$

Using (G.1), we can write

$$\begin{aligned} \frac{1}{m^3} [D^*D]_{kp} &= \frac{1}{m^3} \sum_{t=1}^{m-1} t^2 e^{it(\omega_p - \omega_k)} \xrightarrow{m \rightarrow \infty} \frac{1}{3} \delta_{k,p} \\ \frac{1}{m^2} [A^*D]_{kp} &= \frac{i}{m^2} \sum_{t=1}^{m-1} t e^{it(\omega_p - \omega_k)} \xrightarrow{m \rightarrow \infty} \frac{i}{2} \delta_{k,p} \\ \frac{1}{m} [A^*A]_{kp} &= \frac{1}{m} \sum_{t=1}^m e^{i(t-1)(\omega_p - \omega_k)} \xrightarrow{m \rightarrow \infty} \delta_{k,p} \end{aligned}$$

which readily give

$$\begin{aligned} \frac{1}{m^3} D^* [I - A(A^*A)^{-1}A^*] D \\ = \frac{1}{m^3} D^* D - \left(\frac{1}{m^2} D^* A \right) \left(\frac{1}{m} A^* A \right)^{-1} \\ \cdot \left(\frac{1}{m^2} A^* D \right) \xrightarrow{m \rightarrow \infty} \frac{1}{12} I. \quad (\text{G.2}) \end{aligned}$$

Inserting (G.2) into (4.6), we obtain (4.7), and the proof is finished.

APPENDIX H
PROOF OF (5.8)

A ‘‘standard’’ deviation of $\operatorname{var}_{\text{ML}}(\hat{\omega})$, which begins by developing $dg(\omega)/d\omega|_{\omega=\hat{\omega}}$ in a Taylor series around the true frequency, appears to be rather lengthy. In the following, we provide a simpler derivation, which makes use of Theorem 6.1 proven in Appendix I. That result states that

$$\operatorname{var}_{\text{ML}}(\hat{\omega}) = \operatorname{var}_{\text{MU}}(\hat{\omega}) = \frac{\sigma}{2N} \frac{\lambda_1}{(\sigma - \lambda_1)^2} |a^*(\omega) s_1|^2 / [d^*(\omega) (I - s_1 s_1^*) d(\omega)] \quad (\text{H.1})$$

where the expression for $\operatorname{var}_{\text{MU}}(\hat{\omega})$ (i.e., the MUSIC variance) follows from (3.12). Since for the case under discussion, R is given by

$$R = Pa(\omega) a^*(\omega) + \sigma I$$

it readily follows that

$$\begin{aligned} \lambda_1 &= mP + \sigma \\ s_1 &= a(\omega)/\sqrt{m}. \quad (\text{H.2}) \end{aligned}$$

Inserting (H.2) into (H.1), we get (see also (4.4))

$$\begin{aligned} \operatorname{var}_{\text{ML}}(\hat{\omega}) &= \frac{\sigma}{2N} \frac{mP + \sigma}{m^2 P^2} m \left/ \left[\frac{m(m-1)(2m-1)}{6} \right. \right. \\ &\quad \left. \left. - \frac{m(m-1)^2}{4} \right] \right. = \frac{6\sigma}{N} \frac{mP + \sigma}{m^2(m^2 - 1)P^2} \end{aligned}$$

and the proof is finished.

APPENDIX I
PROOF OF THEOREM 6.1

The idea of the proof can be explained as follows. As shown in Section V, the ML estimate $\hat{\theta}$ of θ , which is defined by [see (5.4)]

$$g(\hat{\theta}) \triangleq \operatorname{tr} [\hat{A}(\hat{A}^* \hat{A})^{-1} \hat{A}^* \hat{R}] = \max \quad (\text{I.1})$$

converges to the true values when N tends to infinity. Furthermore, the estimation errors $(\hat{\theta} - \theta)$ are $O(1/\sqrt{N})$ for large N . According to this fact, we can neglect the terms in (I.1) whose derivative with respect to $\hat{\theta}$ is $O(1/N)$, without affecting the asymptotic (for $N \gg 0$) distribution of the MLE that maximizes (I.1).

To implement the idea above, we need to introduce some additional notation. Let

$$\Lambda = \begin{bmatrix} \hat{\lambda}_1 & & 0 \\ & \ddots & \\ 0 & & \hat{\lambda}_n \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \hat{\lambda}_{n+1} & & 0 \\ & \ddots & \\ 0 & & \hat{\lambda}_m \end{bmatrix}.$$

Then, the eigendecomposition of \hat{R} can be written as follows:

$$\hat{R} = \hat{S}\Lambda\hat{S}^* + \hat{G}\Sigma\hat{G}^*. \quad (I.2)$$

Inserting (I.2) into (I.1), we obtain

$$g(\hat{\theta}) = \text{tr} [(\hat{S}^*\hat{A})(\hat{A}^*\hat{A})^{-1}(\hat{A}^*\hat{S})\Lambda] \\ + \text{tr} [(\hat{A}^*\hat{A})^{-1}(\hat{A}^*\hat{G})\Sigma(\hat{G}^*\hat{A})]. \quad (I.3)$$

Since, for large N ,

$$\hat{A}^*\hat{G} = O(1/\sqrt{N})$$

and

$$\Sigma = \sigma I + O(1/\sqrt{N}), \quad (I.4)$$

it follows that we can replace Σ in (I.3) by σI

$$g(\hat{\theta}) \approx \text{tr} [(\hat{S}^*\hat{A})(\hat{A}^*\hat{A})^{-1}(\hat{A}^*\hat{S})\Lambda] \\ + \text{tr} [(\hat{A}^*\hat{A})^{-1}\hat{A}^*\hat{G}\hat{G}^*\hat{A}]\sigma \\ = \text{tr} \Lambda + \text{tr} [(\hat{S}^*\hat{A})(\hat{A}^*\hat{A})^{-1}(\hat{A}^*\hat{S}) - I]\Lambda \\ - \text{tr} [(\hat{A}^*\hat{A})^{-1}\hat{A}^*\hat{S}\hat{S}^*\hat{A} - I]\sigma \\ = \text{tr} \Lambda + \text{tr} [(\hat{S}^*\hat{A})(\hat{A}^*\hat{A})^{-1}(\hat{A}^*\hat{S}) - I] \\ \cdot [\Lambda - \sigma I]. \quad (I.5)$$

Define

$$\Delta = \hat{G}\hat{G}^*\hat{A} = \hat{A} - \hat{S}\hat{S}^*\hat{A}$$

and observe from (I.4) that $\Delta = O(1/\sqrt{N})$. Since $\hat{S}^*\Delta = 0$, we can write

$$\hat{A}^*\hat{A} = [\Delta^* + \hat{A}^*\hat{S}\hat{S}^*][\Delta + \hat{S}\hat{S}^*\hat{A}] \\ = \Delta^*\Delta + (\hat{A}^*\hat{S})(\hat{S}^*\hat{A}). \quad (I.6)$$

Next, note that the matrix S^*A is nonsingular. This follows from the fact that $A = SQ$ for some matrix Q , and Q must be nonsingular since A has full rank by assumption. Thus, the matrix $\hat{S}^*\hat{A}$ which, for large N , is close to S^*A , must also be nonsingular. Using this observation and (I.6), we get

$$(\hat{S}^*\hat{A})(\hat{A}^*\hat{A})^{-1}(\hat{A}^*\hat{S}) - I \\ = [I + (\hat{A}^*\hat{S})^{-1}\Delta^*\Delta(\hat{S}^*\hat{A})^{-1}]^{-1} - I \\ \approx -(\hat{A}^*\hat{S})^{-1}\Delta^*\Delta(\hat{S}^*\hat{A})^{-1} \quad (I.7)$$

where we also used the fact that for some matrix Γ of subunitary norm $(I + \Gamma)^{-1} = I - \Gamma + \Gamma^2 - \Gamma^3 + \dots$. Inserting (I.7) into (I.5), we obtain

$$g(\hat{\theta}) \approx \text{const} - \text{tr} \Delta^*\Delta(\hat{S}^*\hat{A})^{-1}[\Lambda - \sigma I](\hat{A}^*\hat{S})^{-1}. \quad (I.8)$$

To complete the proof, we only need to show that the matrix that multiplies $\Delta^*\Delta$ in (I.8) can be replaced by P . Since

$$R = APA^* + \sigma I = \hat{A}P\hat{A}^* + \sigma I + O(1/\sqrt{N}) \\ = \hat{S}\Lambda\hat{S}^* + \hat{G}\Sigma\hat{G}^* + O(1/\sqrt{N})$$

it follows that

$$(\hat{S}^*\hat{A})P(\hat{A}^*\hat{S}) + \sigma I = \Lambda + O(1/\sqrt{N})$$

or yet

$$P = (\hat{S}^*\hat{A})^{-1}(\Lambda - \sigma I)(\hat{A}^*\hat{S})^{-1} + O(1/\sqrt{N}). \quad (I.9)$$

From (I.8) and (I.9), we obtain the following large sample (for $N \gg 0$) approximation to $g(\hat{\theta})$:

$$g(\hat{\theta}) \approx \text{const} - \text{tr} [\hat{A}^*\hat{G}\hat{G}^*\hat{A}][P]. \quad (I.10)$$

Maximization of (I.10) with respect to $\hat{\theta}$ is equivalent to minimization of the MUSIC function (3.7) if, and essentially only if, the matrix P is diagonal. Thus, the proof is finished.

ACKNOWLEDGMENT

The authors are grateful to Prof. M. Kaveh for very helpful comments on a first version of this paper.

REFERENCES

- [1] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectral Estimation Workshop*, Rome, NY, 1979, pp. 243-258.
- [2] G. Bienvenu and L. Kopp, "Adaptivity to background noise spatial coherence for high resolution passive methods," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Denver, CO, Apr. 1980, pp. 307-310.
- [3] —, "Optimality of high resolution array processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1235-1248, Oct. 1983.
- [4] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387-392, Apr. 1985.
- [5] K. Sharman, T. S. Durrani, M. Wax, and T. Kailath, "Asymptotic performance of eigenstructure spectral analysis methods," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, San Diego, CA, Mar. 1984, pp. 45.5.1-45.5.4.
- [6] K. Sharman and T. S. Durrani, "A comparative study of modern eigenstructure methods for bearing estimation—A new high performance approach," in *Proc. 25th IEEE Conf. Dec. Contr.*, Athens, Greece, Dec. 1986, pp. 1737-1742.
- [7] D. J. Jeffries and D. R. Farrer, "Asymptotic results for eigenvector methods," *IEE Proc.*, vol. 132, pt. F, pp. 589-594, 1985.
- [8] 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; corrections in *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, p. 633, June 1986.
- [9] A. M. Bruckstein, T.-J. Shan, and T. Kailath, "The resolution of overlapping echos," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1357-1367, Dec. 1985.
- [10] R. Kumaresan and A. K. Shaw, "High resolution bearing estimation without eigendecomposition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tampa, FL, Mar. 1985, pp. 576-579.
- [11] R. Kumaresan, L. L. Scharf, and A. K. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 637-640, June 1986.
- [12] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1081-1089, Oct. 1986.
- [13] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975-989, Sept. 1982.
- [14] S. Parthasarathy and D. W. Tufts, "Maximum-likelihood estimation of parameters of exponentially damped sinusoids," *Proc. IEEE*, vol. 73, pp. 1528-1530, Oct. 1985.
- [15] S. J. Orfanidis, "A reduced MUSIC algorithm," in *Proc. 3rd ASSP*

- Workshop on Spectrum Estimation and Modeling*, Boston, MA, Nov. 1986, pp. 165-167.
- [16] R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, pp. 134-139, Jan. 1983.
- [17] G. Xu and Y.-H. Pao, "Single vector approaches to eigenstructure analysis for harmonic retrieval," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, Apr. 1986, pp. 173-176.
- [18] K. C. Sharman, "Adaptive algorithms for estimating the complete covariance eigenstructure," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, Apr. 1986, pp. 1401-1404.
- [19] R. Schreiber, "Implementation of adaptive array algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1038-1045, Oct. 1986.
- [20] D. W. Tufts and C. D. Melissinos, "Simple, effective computation of principal eigenvectors and their eigenvalues and application to high-resolution estimation of frequencies," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1046-1053, Oct. 1986.
- [21] H. Wang and M. Kaveh, "On the performance of signal-subspace processing—Part I: Narrow-band systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1201-1209, Oct. 1986.
- [22] F. Tuteur and Y. Rockah, "The covariance difference method in signal detection," in *Proc. 3rd ASSP Workshop on Spectrum Estimation and Modeling*, Boston, MA, Nov. 1986, pp. 120-122.
- [23] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—A subspace rotation approach to estimation of parameters of sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1340-1342, Oct. 1986.
- [24] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 591-598, Sept. 1974.
- [25] —, "Multiple tone parameter estimation from discrete-time observations," *Bell Syst. Tech. J.*, vol. 55, pp. 1389-1410, Nov. 1976.
- [26] P. Stoica, R. Moses, B. Friedlander, and T. Söderström, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 378-392, Mar. 1989.
- [27] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*. New York: Hafner, 1961.
- [28] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1984.
- [29] P. Stoica and T. Söderström, "On reparameterization of loss function: used in estimation and the invariance principle," Tech. Rep. UPTec 87113R, Uppsala Univ., Sweden, 1987.
- [30] A. A. Beex, "Fast recursive/iterative Toeplitz eigenspace decomposition," in *Signal Processing III: Theories and Applications*, I. T. Young *et al.*, Eds. New York: North-Holland, 1986, pp. 1001-1004.
- [31] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Stat.*, vol. 34, no. 1, pp. 122-148, 1963.
- [32] T. Söderström and P. Stoica, *System Identification*. London: Prentice-Hall, 1988.
- [33] K. Sharman, "Maximum likelihood parameter estimation by simulated annealing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, New York, NY, Apr. 1988, pp. 2741-2744.
- [34] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 477-489, Apr. 1988.
- [35] I. Ziskind and M. Wax, "Maximum likelihood estimation via the alternating projection maximization algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1553-1560, Oct. 1988.
- [36] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," in preparation.
- [37] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramer-Rao bound: Further results and comparisons," submitted for publication. Also to appear in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Glasgow, Scotland, May 1989.
- [38] P. Janssen and P. Stoica, "On the expectation of the product of four matrix-valued Gaussian random variables," EUT Rep. 87-E-178. Eindhoven Univ. Technol., Eindhoven, The Netherlands, 1987; also *IEEE Trans. Automat. Contr.*, vol. 33, pp. 867-870, Sept. 1988.
- [39] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-to-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 806-811, Aug. 1985.
- [40] W. J. Bangs, "Array processing with generalized beamformers," Ph.D. dissertation, Yale Univ., New Haven, CT, 1971.
- [41] H. Hung and M. Kaveh, "On the statistical sufficiency of the coherently averaged covariance matrix for the estimation of the parameters of wideband sources," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Dallas, TX, Mar. 1987, pp. 33-36.
- [42] D. Starer and A. Nehorai, "Maximum likelihood estimation of exponential signals in noise using a Newton algorithm," in *Proc. 4th ASSP Workshop on Spectrum Estimation and Modeling*, Minneapolis, MN, Aug. 1988, pp. 240-245.

Petre Stoica, for a photograph and biography, see p. 391 of the March 1989 issue of this TRANSACTIONS.



Arye Nehorai (S'80-M'83) was born in Haifa, Israel. He received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion-Israel Institute of Technology, in 1976 and 1979, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1983.

From 1983 to 1984 he was a Research Associate at Stanford University. From 1984 to 1985 he was a Research Engineer at Systems Control Technology Inc., in Palo Alto, CA. Since 1985 he has been an Assistant Professor in the Department of Electrical Engineering at Yale University, New Haven, CT. His areas of interest are signal processing and system identification.

Dr. Nehorai is an Associate Editor of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, a member of the Technical Committee on Spectrum Estimation and Modeling in the IEEE ASSP Society, and the Chairman of the Connecticut IEEE ASSP Chapter. He received the Rothschild and Willer Fellowships and the Most Outstanding Graduate Student Award in the Department of Electrical Engineering, Technion. He is a member of Sigma Xi.