Maximum Likelihood Array Processing for Stochastic Coherent Sources

Petre Stoica, *Fellow, IEEE,* Björn Ottersten, *Member, IEEE,* Mats Viberg, *Member, IEEE,* and Randolph L. Moses, *Senior Member, IEEE*

***Abstract*—Maximum likelihood (ML) estimation in array signal processing for the stochastic *noncoherent* signal case is well documented in the literature. Herein, we focus on the equally relevant case of stochastic *coherent* signals. Explicit large-sample realizations are derived for the ML estimates of the noise power and the (singular) signal covariance matrix. The asymptotic properties of the estimates are examined, and some numerical examples are provided. In addition, we show the surprising fact that the ML estimates of the signal parameters obtained by ignoring the information that the sources are coherent coincide in large samples with the ML estimates obtained by exploiting the coherent source information. Thus, the ML signal parameter estimator derived for the noncoherent case (or its large-sample realizations) asymptotically achieves the lowest possible estimation error variance (corresponding to the *coherent* Cramér-Rao bound).**

I. Introduction

SENSOR array signal processing addresses the problem of estimating the parameters of multiple emitter signals arriving at an array of sensors with known characteristics. A multitude of estimators have been presented for this problem and their asymptotic properties have been investigated [1], [2]. The stochastic maximum likelihood (ML) estimator, derived under the stochastic emitter signal model, has been shown to achieve the Cramér-Rao bound (CRB) in large samples and hence it yields asymptotically efficient parameter estimates [3]–[5]. This important property of the stochastic ML estimator depends on the assumption that the emitter signals are noncoherent (i.e., no two signals are fully correlated). When coherent emitter signals are present, the maximum likelihood estimator must be reformulated and a new Cramér-Rao lower bound on the estimation error variance is also obtained. Note that there are methods for consistently detecting the presence of coherent signals [6],[7].

This paper formulates the stochastic ML problem for coherent emitter signals, under a rank constraint on the emitter covariance matrix. By making use of an appropriate parameterization, we show that explicit large-sample expressions for the ML estimates of the noise power and the signal covariance matrix can be obtained. The asymptotic properties of the ML estimates of the aforementioned parameters are examined. It is shown both analytically and numerically that the ML estimator of these parameters, which makes use of the rank information, outperforms the ML estimator which ignores that information. Regarding the signal parameters (such as the angles-of-arrival), we prove the somewhat intriguing result that the asymptotic estimation accuracy of these parameters is not affected by the knowledge of the rank of the emitter signal covariance matrix. Thus, the original stochastic ML estimator that ignores the rank information, or its large-samples realizations (such as MODE and WSF [4], [6], [8]), achieves the lowest possible estimation error variance for the signal parameters.

In closing this section, we mention the related work by Bresler [9]. The approach in [9] aims at deriving an exact realization of the ML estimate in either coherent or noncoherent scenarios, and it is somewhat intricate. The large-sample ML estimator introduced in the following sections is much simpler both conceptually and computationally.

II. Problem Formulation

The following narrowband processing array model is hypothesized:

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{n}(t) \ . \tag{1}$$

The measurement vector, $\mathbf{y}(t)$, represents the m sensor outputs. The n emitter signals are collected in the vector $\mathbf{x}(t)$ and the additive noise is denoted $\mathbf{n}(t)$. The array response, $\mathbf{A}(\boldsymbol{\theta})$, is a known function of the unknown signal parameters, $\boldsymbol{\theta}$, (these can be, for example, the angles-of-arrival).

The stochastic model assumes that the complex-valued observation vector is zero-mean and circularly Gaussian distributed with

$$\mathrm{E}\{\mathbf{y}(t)\mathbf{y}^*(s)\} = \mathbf{R}\delta_{ts} = (\mathbf{A}\mathbf{P}\mathbf{A}^* + \sigma^2\mathbf{I})\delta_{ts} \tag{2}$$

$$\mathrm{E}\{\mathbf{y}(t)\mathbf{y}^T(s)\} = \mathbf{0} \tag{3}$$

where the superscript $(\cdot)^*$ denotes the conjugate transpose, σ^2 is the power of the (spatially white) sensor noise, $\mathbf{P} = \mathrm{E}\{\mathbf{x}(t)\mathbf{x}^*(t)\}$ is the emitter signal covariance matrix and δ_{ts} denotes the Kronecker delta. To simplify the notation,

Manuscript received August 5, 1994; revised July 17, 1995. The associate editor coordinating the review of this paper and approving it for publication was Dr. R. D. Preuss.

P. Stoica is with Systems and Control Group, Uppsala University, Uppsala, Sweden.

B. Ottersten is with Signal Processing/S3, Royal Institute of Technology, Stockholm, Sweden.

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

R. L. Moses is with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210 USA.

Publisher Item Identifier S 1053-587X(96)00722-0.

the dependence of $\mathbf{A}$ on the signal parameter vector, $\boldsymbol{\theta}$, is suppressed whenever there is no risk for confusion. The objective is to estimate the signal parameters, the emitter signal covariance matrix, and the noise variance based on the sensor measurements. It is assumed that the number of signals, n, is known. When discussing the *coherent* ML estimator and the corresponding CRB, we also assume that the rank of $\mathbf{P}$ is known.

The maximum likelihood (ML) estimator based on the model above consists of solving the following problem [4], [9]–[11]:

$$\min_{\boldsymbol{\theta},\sigma^2,\mathbf{P}} [\ln|\mathbf{R}| + \mathrm{Tr}(\mathbf{R}^{-1}\hat{\mathbf{R}})] \tag{4}$$

where $|\cdot|$ and $\mathrm{Tr}(\cdot)$ denote the determinant and the trace of a matrix, respectively, $\mathbf{R}$ is the covariance matrix of $\mathbf{y}(t)$ introduced above, and $\hat{\mathbf{R}}$ is the sample covariance matrix of the array output vector

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{y}(t)\mathbf{y}^*(t). \tag{5}$$

Hereafter, N denotes the number of available snapshots.

The minimization problem in (4) is greatly simplified by employing a result in a paper by Böhme [10] (see [11] for a simple proof of the result in question). According to [10] and [11], the minimizing arguments of the function in (4) are given by

$$\hat{\sigma}^2(\hat{\boldsymbol{\theta}}) = \frac{1}{m-n}\mathrm{Tr}\left[\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\hat{\mathbf{R}}\right]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{6}$$

$$\hat{\mathbf{P}}(\hat{\boldsymbol{\theta}}) = \mathbf{A}^{\dagger}(\boldsymbol{\theta})\left[\hat{\mathbf{R}} - \hat{\sigma}^2(\boldsymbol{\theta})\mathbf{I}\right]\mathbf{A}^{\dagger *}(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{7}$$

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\left|\mathbf{A}(\boldsymbol{\theta})\hat{\mathbf{P}}(\boldsymbol{\theta})\mathbf{A}^*(\boldsymbol{\theta}) + \hat{\sigma}^2(\boldsymbol{\theta})\mathbf{I}\right| \tag{8}$$

where

$$\mathbf{A}^{\dagger} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^* \tag{9}$$

$$\boldsymbol{\Pi}_{\mathbf{A}}^{\perp} = \mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}. \tag{10}$$

The importance of the above result lies in the fact that it provides explicit expressions for the ML estimates of the noise power and signal covariance matrix, hence reducing the dimension of the original optimization problem (4) by (n^2+1). On the other hand, the result has a drawback: Equations (6)–(8) are derived by considering the minimization of the normalized negative log-likelihood function in (4) with respect to $\mathbf{P}$ over the set of Hermitian matrices, and not over the set of Hermitian positive (semi)definite matrices as it should. As a consequence, the minimizing $\hat{\mathbf{P}}(\hat{\boldsymbol{\theta}})$ so obtained (see (7)) is not guaranteed to be positive (semi)definite, and indeed it is known [9] that $\hat{\mathbf{P}}(\hat{\boldsymbol{\theta}})$ may be indefinite.

The aforementioned drawback is minor when $\mathbf{P}$ is (strictly) positive definite. The reason for this is as follows. Since the ML estimates above are consistent, $\hat{\mathbf{P}}(\hat{\boldsymbol{\theta}})$ tends to $\mathbf{P}$ as $N \to \infty$, which means that $\hat{\mathbf{P}}(\hat{\boldsymbol{\theta}})$ must be positive definite (and hence a "valid" ML estimator) for sufficiently large N. Hence, (6)–(8) provide a *large-sample* realization of the ML processor in the case of a *nonsingular* covariance matrix $\mathbf{P}$ (i.e., the case of *noncoherent* sources).

In the case of *coherent* sources, however, the covariance matrix $\mathbf{P}$ is singular, whereas $\hat{\mathbf{P}}(\hat{\boldsymbol{\theta}})$ given by (7) is nonsingular (w.p. 1) for any finite N. In such a case, $\hat{\mathbf{P}}(\hat{\boldsymbol{\theta}})$ cannot represent a valid ML estimate of $\mathbf{P}$ (not even for $N \to \infty$). As a consequence, $\hat{\boldsymbol{\theta}}$ given by (8) may not be a realization of the "true" ML signal parameter estimate which exploits the knowledge that the sources are coherent. In fact, according to the parsimony principle of parameter estimation (see, e.g., [12]), incorporation of any *a priori* knowledge into the ML problem should lead to improved accuracy compared with the situation where that knowledge is not incorporated (such as is the case in (6)–(8)). This means that $\hat{\boldsymbol{\theta}}$ given by (8) might be expected to be less accurate than the ML signal parameter estimate based on the information that $\mathbf{P}$ is singular. Interestingly enough, though, numerical evaluations of the Cramér-Rao bounds derived by ignoring and, respectively, exploiting the information that $\mathrm{rank}(\mathbf{P}) < n$ have shown that the asymptotic signal parameter bounds corresponding to the two cases are identical [4]. Thus, there is numerical evidence supporting the conjecture that (8) is a large-sample realization of the ML signal parameter estimator, even in the coherent source case. However, a proof of this conjecture was not available.

The goal of this paper is threefold:

1) to derive explicit large-sample expressions that are comparable with (6)–(8) for the estimators that minimize (4) in the case of coherent sources
2) to prove the conjecture that (8) still provides a large-sample realization of the ML signal parameter estimate even in scenarios where it is known *a priori* that $\mathbf{P}$ is rank deficient
3) to establish the large-sample properties of the ML estimates of the noise power and signal covariance matrix derived under the rank constraint on $\mathbf{P}$ and to show that they outperform the corresponding ML estimates that ignore the information that $\mathbf{P}$ is rank deficient.

III. ML Estimation for Coherent Sources

Let the rank of the emitter signal covariance matrix be

$$p \triangleq \mathrm{rank}(\mathbf{P}) \leq n\ . \tag{11}$$

When there are coherent emitter signals present, $p < n$. To guarantee that the estimation problem under study is "parameter identifiable," we make the assumption that the array manifold is unambiguous and that $m > 2n - p$ [13]. Let

$$\mathbf{P} = \mathbf{L}\,\mathbf{L}^* \tag{12}$$

denote the Cholesky factorization of the signal covariance matrix ($\mathbf{L}$ is $n \times p$). It readily follows from (12) that $\mathbf{P}$ can uniquely be written as

$$\mathbf{P} = \begin{bmatrix}\mathbf{I}\\ \mathbf{C}\end{bmatrix}\mathbf{S}\,[\mathbf{I}\ \ \mathbf{C}^*] \tag{13}$$

where $\mathbf{S}$ is $p \times p$, $\mathbf{C}$ is $(n-p) \times p$, and $\mathbf{I}$ is $p \times p$. Note that this square root parameterization requires $p^2 + 2(n-p)p = n^2 - (n-p)^2$ real parameters, whereas the full parameterization

(4) requires n^2 parameters. In what follows, it is assumed that $\mathbf{S}$ is a nonsingular matrix. As the source signals can be arbitrarily ordered, the previous assumption on $\mathbf{S}$ does not introduce any restriction. Define

$$\mathbf{A}_c = \mathbf{A}\begin{bmatrix}\mathbf{I}\\ \mathbf{C}\end{bmatrix}. \tag{14}$$

With this notation, the matrix $\mathbf{R}$ in (2) can be written as

$$\mathbf{R} = \mathbf{A}_c\mathbf{S}\mathbf{A}_c^* + \sigma^2\mathbf{I}. \tag{15}$$

Let $\boldsymbol{\alpha}$ denote the parameter vector composed of the elements of $\boldsymbol{\theta}$ along with the real and imaginary parts of the elements of $\mathbf{C}$. We parameterize the negative log-likelihood function in (4) by $\boldsymbol{\alpha}$, $\mathbf{S}$, and σ^2. Using this parameterization, the matrix $\mathbf{P}$ is guaranteed to be positive (semi)definite and to satisfy the rank constraint (11).

Since $\mathbf{S}$ in the representation (15) of $\mathbf{R}$ is positive definite, it follows from the discussion on (6)–(8) that large-sample realizations of the ML estimates of σ^2, $\mathbf{S}$, and $\boldsymbol{\alpha}$ are given by

$$\hat{\sigma}^2(\hat{\boldsymbol{\alpha}}) = \frac{1}{m-p}\operatorname{Tr}\left[\boldsymbol{\Pi}^{\perp}_{\mathbf{A}_c}(\boldsymbol{\alpha})\hat{\mathbf{R}}\right]\Big|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}} \tag{16}$$

$$\hat{\mathbf{S}}(\hat{\boldsymbol{\alpha}}) = \mathbf{A}_c^{\dagger}(\boldsymbol{\alpha})\left[\hat{\mathbf{R}} - \hat{\sigma}^2(\boldsymbol{\alpha})\mathbf{I}\right]\mathbf{A}_c^{\dagger *}(\boldsymbol{\alpha})\Big|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}} \tag{17}$$

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}}\left|\mathbf{A}_c(\boldsymbol{\alpha})\hat{\mathbf{S}}(\boldsymbol{\alpha})\mathbf{A}_c^*(\boldsymbol{\alpha}) + \hat{\sigma}^2(\boldsymbol{\alpha})\mathbf{I}\right|. \tag{18}$$

Next, we derive an explicit expression for the ML estimate of the part of $\boldsymbol{\alpha}$ corresponding to $\mathbf{C}$. To this end, we note the following result proved in [3]–[5], and [8].

Theorem 1: A large-sample realization of $\hat{\boldsymbol{\alpha}}$ in (18) (i.e., an estimate of $\boldsymbol{\alpha}$ having the same asymptotic distribution as $\hat{\boldsymbol{\alpha}}$) is given by the maximizer of the function

$$f(\boldsymbol{\alpha}) = \operatorname{Tr}\left[\boldsymbol{\Pi}_{\mathbf{A}_c}(\boldsymbol{\alpha})\mathbf{W}\right] \tag{19}$$

where

$$\boldsymbol{\Pi}_{\mathbf{A}_c} = \mathbf{A}_c\mathbf{A}_c^{\dagger} \tag{20}$$

$$\mathbf{W} = \hat{\mathbf{E}}_s\hat{\hat{\boldsymbol{\Lambda}}}^2\hat{\boldsymbol{\Lambda}}^{-1}\hat{\mathbf{E}}_s^* \tag{21}$$

and where $\hat{\mathbf{E}}_s, \hat{\boldsymbol{\Lambda}}$ and $\hat{\hat{\boldsymbol{\Lambda}}}$ are defined as follows. Let $\{\hat{\lambda}_k\}_{k=1}^p$ and $\{\hat{\mathbf{e}}_k\}_{k=1}^p$ denote the p principal (largest) eigenvalues of $\hat{\mathbf{R}}$ and their associated orthonormal eigenvectors, respectively. Then,

$$\hat{\mathbf{E}}_s = [\hat{\mathbf{e}}_1, \cdots, \hat{\mathbf{e}}_p] \tag{22}$$

$$\hat{\boldsymbol{\Lambda}} = \operatorname{diag}(\hat{\lambda}_1, \cdots, \hat{\lambda}_p) \tag{23}$$

$$\hat{\hat{\boldsymbol{\Lambda}}} = \operatorname{diag}(\hat{\lambda}_1 - \hat{\sigma}^2, \ldots, \hat{\lambda}_p - \hat{\sigma}^2) \tag{24}$$

with

$$\hat{\sigma}^2 = \frac{1}{m-p}\left(\operatorname{Tr}\hat{\mathbf{R}} - \operatorname{Tr}\hat{\boldsymbol{\Lambda}}\right). \tag{25}$$

□

A straightforward calculation shows that $f(\boldsymbol{\alpha})$, which was defined in the previous theorem, can be rewritten as

$$\begin{aligned} f(\boldsymbol{\alpha}) &= \operatorname{Tr}\left\{\left([\mathbf{I}\ \ \mathbf{C}^*](\mathbf{A}^*\mathbf{A})\begin{bmatrix}\mathbf{I}\\ \mathbf{C}\end{bmatrix}\right)^{-1}\left([\mathbf{I}\ \ \mathbf{C}^*](\mathbf{A}^*\mathbf{W}\mathbf{A})\begin{bmatrix}\mathbf{I}\\ \mathbf{C}\end{bmatrix}\right)\right\} \\ &= \operatorname{Tr}\left\{[\mathbf{X}^*(\mathbf{A}^*\mathbf{A})\mathbf{X}]^{-1}[\mathbf{X}^*(\mathbf{A}^*\mathbf{W}\mathbf{A})\mathbf{X}]\right\} \end{aligned} \tag{26}$$

where

$$\mathbf{X} = \begin{bmatrix}\mathbf{I}\\ \mathbf{C}\end{bmatrix}. \tag{27}$$

Let $\{\mu_k\}_{k=1}^p$ and $\{v_k\}_{k=1}^p$ denote the p principal eigenvalues and their associated eigenvectors of the matrix $(\mathbf{A}^*\mathbf{A})^{-1}(\mathbf{A}^*\mathbf{W}\mathbf{A})$, and define

$$\hat{\mathbf{V}}(\boldsymbol{\theta}) = [v_1, \ldots, v_p]. \tag{28}$$

Observe that the function $f(\boldsymbol{\alpha})$ is invariant to the post-multiplication of $\mathbf{X}$ by any nonsingular matrix. It then follows from the extended Rayleigh quotient result in [14] or from the Poincaré separation theorem (see, e.g., [15]) that

$$f(\boldsymbol{\alpha}) \le \sum_{k=1}^{p}\mu_k \tag{29}$$

where the upper bound is obtained for

$$\mathbf{X}\mathbf{G} = \hat{\mathbf{V}}(\boldsymbol{\theta}) \tag{30}$$

and where $\mathbf{G}$ is an arbitrary nonsingular matrix. If $\hat{\mathbf{V}}(\boldsymbol{\theta})$ is partitioned as

$$\hat{\mathbf{V}}(\boldsymbol{\theta}) = \begin{bmatrix}\hat{\mathbf{V}}_1(\boldsymbol{\theta})\\ \hat{\mathbf{V}}_2(\boldsymbol{\theta})\end{bmatrix}\begin{matrix}\}\ p\\ \}\ n-p\end{matrix} \tag{31}$$

then (30) implies

$$\begin{bmatrix}\mathbf{I}\\ \mathbf{C}\end{bmatrix}\mathbf{G} = \begin{bmatrix}\hat{\mathbf{V}}_1(\boldsymbol{\theta})\\ \hat{\mathbf{V}}_2(\boldsymbol{\theta})\end{bmatrix} \tag{32}$$

from which we immediately obtain the following explicit expression for the minimizing $\mathbf{C}$ matrix:

$$\hat{\mathbf{C}}(\boldsymbol{\theta}) = \hat{\mathbf{V}}_2(\boldsymbol{\theta})\,\hat{\mathbf{V}}_1^{-1}(\boldsymbol{\theta}). \tag{33}$$

To arrive at a concentrated form for $f(\boldsymbol{\alpha})$, observe that

$$\operatorname{rank}[(\mathbf{A}^*\mathbf{A})^{-1}(\mathbf{A}^*\mathbf{W}\mathbf{A})] = \operatorname{rank}[\mathbf{A}^*(\boldsymbol{\theta})\,\hat{\mathbf{E}}_s] \tag{34}$$

where now $\boldsymbol{\theta}$ denotes the indeterminate signal parameter vector. Let $\mathbf{E}_s$ denote the limit of $\hat{\mathbf{E}}_s$ as $N \to \infty$ ($\mathbf{E}_s$ is defined from the p principal eigenvectors of $\mathbf{R}$, similarly to $\hat{\mathbf{E}}_s$ in (22)). It is well known that the subspace generated by $\mathbf{E}_s$ is included in the range of the true array matrix $\mathbf{A}$. This means that there exists an $n \times p$ matrix $\mathbf{Q}$ such that

$$\mathbf{E}_s = \mathbf{A}\mathbf{Q}. \tag{35}$$

From (35), we obtain

$$(\mathbf{E}_s^*\mathbf{A})\mathbf{Q} = \mathbf{I} \qquad (p \times p) \tag{36}$$

which implies that

$$\operatorname{rank}(\mathbf{E}_s^*\mathbf{A}) = p\,. \tag{37}$$

Under a weak analyticity condition on the matrix $\mathbf{A}(\boldsymbol{\theta})$, viewed as a function of $\boldsymbol{\theta}$, it then follows that

$$\operatorname{rank}\left[\mathbf{A}^*(\boldsymbol{\theta})\hat{\mathbf{E}}_s\right] = p \tag{38}$$

for finite but large N values and for almost any $\boldsymbol{\theta}$ vector in a compact set including the true parameters (see, e.g., [12]). Hence, for sufficiently large N and generically in $\boldsymbol{\theta}$, we have

$$\operatorname{rank}\left[(\mathbf{A}^*\mathbf{A})^{-1}(\mathbf{A}^*\mathbf{W}\mathbf{A})\right] = p \tag{39}$$

which implies that the function of $\boldsymbol{\theta}$, left after the concentration of $f(\boldsymbol{\alpha})$ with respect to $\mathbf{C}$, is given by (cf. (29))

$$\sum_{k=1}^{p} \mu_k = \operatorname{Tr}\left[(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{W}\mathbf{A}\right]. \tag{40}$$

In other words, a large-sample realization of the ML estimator of $\boldsymbol{\theta}$, under the *a priori* information that $\operatorname{rank}(\mathbf{P}) = p$, is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \operatorname{Tr}\left[\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{W}\right]. \tag{41}$$

This estimate, known as the MODE/WSF estimate, was shown in [6] and [8] to be a large-sample realization of the stochastic ML estimator in (8) (see also Theorem 1).

We summarize the previously derived results in the following theorem.

Theorem 2: The ML signal parameter estimator (8), which is derived by ignoring any *a priori* information about the matrix $\mathbf{P}$, is a large-sample realization of the ML signal parameter estimator that does exploit the information that $\mathbf{P}$ is a positive (semi)definite matrix of rank p. A large-sample realization of (8) can be obtained by minimizing the MODE/WSF function in (41).

Once $\boldsymbol{\theta}$ is estimated, from either (8) or (41), large-sample solutions to the ML estimation problem for σ^2 and $\mathbf{P}$, under the condition that $\mathbf{P}$ is positive semidefinite and its rank is p, are given by (16) and, respectively, (13) and (17), where $\hat{\boldsymbol{\alpha}}$ is composed from $\hat{\boldsymbol{\theta}}$ in either (8) or (41) and $\hat{\mathbf{C}}(\hat{\boldsymbol{\theta}})$ in (33). □

As an illustration, we specialize the ML estimator previously derived to the case of $p = 1$, which is important for radar applications and multipath propagation studies. We present a step-by-step summary of the proposed ML technique for this case.

Step 1: Compute the principal eigenvector $\hat{\mathbf{e}}_1$ of $\hat{\mathbf{R}}$, and the large-sample ML estimator of the signal parameters as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left[\hat{\mathbf{e}}_1^*\boldsymbol{\Pi}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\hat{\mathbf{e}}_1\right].$$

Step 2: Evaluate

$$\hat{\mathbf{v}}_1 = (\hat{\mathbf{A}}^*\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^*\hat{\mathbf{e}}_1 \triangleq \begin{bmatrix}\hat{\mathbf{v}}_{1,1}\\ \hat{\mathbf{v}}_{1,2}\end{bmatrix}\begin{matrix}\}\ 1\\ \}\ n-1\end{matrix}$$

Then, compute $\hat{\mathbf{C}} = \hat{\mathbf{v}}_{1,2}/\hat{\mathbf{v}}_{1,1}$ and

$$\hat{\mathbf{A}}_c = \hat{\mathbf{A}}\begin{bmatrix}1\\ \hat{\mathbf{C}}\end{bmatrix}.$$

Here, $\hat{\mathbf{A}} = \mathbf{A}(\hat{\boldsymbol{\theta}})$, with $\hat{\boldsymbol{\theta}}$ provided by Step 1.

Step 3: Compute the large-sample ML estimates of the noise power and signal covariance matrix by using (16) and, respectively, (13) and (17), where $\mathbf{A}_c$ is replaced by $\hat{\mathbf{A}}_c$ and $\mathbf{C}$ by $\hat{\mathbf{C}}$ determined in Step 2.

The asymptotic properties of the ML estimates derived above are examined in the next section.

IV. Statistical Analysis

The results of the previous section show that the estimation accuracy of the signal parameters (such as the angles-of-arrival) is not affected by the choice of parameterization for the signal covariance matrix $\mathbf{P}$. The accuracy of the estimates of the emitter signal covariance matrix and the noise power is however affected when a low-rank parameterization of $\mathbf{P}$ is used. In several applications, estimation of both $\boldsymbol{\theta}$ and $\mathbf{P}$ is of interest. For instance, this is the case when $\mathbf{P}$ carries information about the target signatures or the specular multipath parameters. In such a case, it is important to use the most accurate estimate not only for $\boldsymbol{\theta}$ but also for $\mathbf{P}$, as given by the ML estimator of Theorem 2 above. In the following discussion we provide a means of evaluating the asymptotic covariance matrix of this estimator, and also make comparisons with the statistical performance corresponding to the ML estimator which ignores the rank information (11).

Consider the following two parameterizations of the emitter signal covariance matrix $\mathbf{P}$. The full parameterization of the Hermitian matrix $\mathbf{P}$ formed by stacking the real part of the entries on and below the diagonal followed by the imaginary part of the entries below the diagonal. The corresponding n^2 parameters are arranged in the vector

$$\boldsymbol{\eta}_p = [\bar{P}_{11}\ \ \bar{P}_{21}\ldots\bar{P}_{n1}\ \ \bar{P}_{22}\ldots\bar{P}_{nn}\ \ \tilde{P}_{21}\ldots\tilde{P}_{n,n-1}]^T \tag{42}$$

where $\bar{(\cdot)}$ denotes the real part, and $\tilde{(\cdot)}$ denotes the imaginary part. Introduce the following notation:

$$\boldsymbol{\eta}_p = \operatorname{vech}\mathbf{P} = \begin{bmatrix}Q_p \operatorname{vec}\bar{\mathbf{P}}\\ \Gamma_p \operatorname{vec}\tilde{\mathbf{P}}\end{bmatrix} \tag{43}$$

where $\operatorname{vec}\mathbf{P}$ denotes the vector obtained by stacking the columns of $\mathbf{P}$ on top of each other, and where Q_p is a $(n^2+n)/2 \times n^2$ matrix selecting only the entries of $\operatorname{vec}\bar{\mathbf{P}}$ corresponding to the elements on and below the diagonal of $\bar{\mathbf{P}}$. The other $(n^2-n)/2 \times n^2$ selection matrix Γ_p selects the entries of $\operatorname{vec}\tilde{\mathbf{P}}$ corresponding to the elements below the diagonal of $\tilde{\mathbf{P}}$.

The alternative parameterization is based on a square root factorization of the emitter covariance matrix as presented in the previous section. Let $\mathbf{P} = \mathbf{L}\mathbf{L}^*$ in (12) be the Cholesky factorization of the emitter covariance matrix; thus, $\mathbf{L}$ is a lower triangular matrix with real elements on the diagonal. When we have prior knowledge of the rank of $\mathbf{P}$, p, $\mathbf{L}$ is chosen to be $n \times p$, and the corresponding parameter vector $\boldsymbol{\eta}_l$ has $n^2 - (n-p)^2 = 2np - p^2$ entries. In a similar fashion as above, the following notation is introduced:

$$\boldsymbol{\eta}_l = \text{vech}\mathbf{L} = \begin{bmatrix} Q_l \,\text{vec}\, \bar{\mathbf{L}} \\ \Gamma_l \,\text{vec}\, \tilde{\mathbf{L}} \end{bmatrix} \tag{44}$$

where Q_l is a $(2np - p^2 + p)/2 \times (2np - p^2)$ selection matrix selecting the entries of $\text{vec}\,\bar{\mathbf{L}}$ corresponding to the elements on and below the diagonal of $\bar{\mathbf{L}}$, whereas the $(2np - p^2 - p)/2 \times (2np - p^2)$ selection matrix Γ_l selects the entries of $\text{vec}\,\tilde{\mathbf{L}}$ corresponding to the elements below the diagonal of $\tilde{\mathbf{L}}$.

In order to compare the estimation accuracies corresponding to the two parameterizations introduced above, the Jacobi matrix relating them must be derived. Let

$$\mathbf{J} = \frac{\partial\,\text{vech}\,\mathbf{P}}{\partial\,\text{vech}\,\mathbf{L}} = \begin{bmatrix} Q_p \frac{\partial\,\text{vec}\,\bar{\mathbf{P}}}{\partial\,\text{vec}\,\bar{\mathbf{L}}} Q_l^T & Q_p \frac{\partial\,\text{vec}\,\bar{\mathbf{P}}}{\partial\,\text{vec}\,\tilde{\mathbf{L}}} \Gamma_l^T \\ \Gamma_p \frac{\partial\,\text{vec}\,\tilde{\mathbf{P}}}{\partial\,\text{vec}\,\bar{\mathbf{L}}} Q_l^T & \Gamma_p \frac{\partial\,\text{vec}\,\tilde{\mathbf{P}}}{\partial\,\text{vec}\,\tilde{\mathbf{L}}} \Gamma_l^T \end{bmatrix}. \tag{45}$$

When examining the blocks in the matrix above, observe that

$$\bar{\mathbf{P}} = \bar{\mathbf{L}}\bar{\mathbf{L}}^T + \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T \tag{46}$$

$$\tilde{\mathbf{P}} = \tilde{\mathbf{L}}\bar{\mathbf{L}}^T - \bar{\mathbf{L}}\tilde{\mathbf{L}}^T\,. \tag{47}$$

First, consider the $(1, 1)$ block of (45). Using (46), we obtain

$$\begin{aligned} \frac{\partial\,\text{vec}\,\bar{\mathbf{P}}}{\partial\,\text{vec}\,\bar{\mathbf{L}}} &= \frac{\partial\,\text{vec}(\bar{\mathbf{L}}\bar{\mathbf{L}}^T + \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T)}{\partial\,\text{vec}\,\bar{\mathbf{L}}} \\ &= (\bar{\mathbf{L}} \otimes \mathbf{I}) \frac{\partial\,\text{vec}\,\bar{\mathbf{L}}}{\partial\,\text{vec}\,\bar{\mathbf{L}}} + (\mathbf{I} \otimes \bar{\mathbf{L}}) \frac{\partial\,\text{vec}(\bar{\mathbf{L}}^T)}{\partial\,\text{vec}\,\bar{\mathbf{L}}} \end{aligned} \tag{48}$$

where $\otimes$ denotes the Kronecker matrix product. Introduce the permutation matrix $\boldsymbol{\Sigma}^{mn}$ such that for any $m \times n$ matrix $\mathbf{A}$

$$\text{vec}(\mathbf{A}^T) = \boldsymbol{\Sigma}^{mn}\,\text{vec}\,\mathbf{A}. \tag{49}$$

Making use of this notation along with (48) results in

$$\frac{\partial\,\text{vec}\,\bar{\mathbf{P}}}{\partial\,\text{vec}\,\bar{\mathbf{L}}} = (\bar{\mathbf{L}} \otimes \mathbf{I}) + (\mathbf{I} \otimes \bar{\mathbf{L}})\boldsymbol{\Sigma}^{np}. \tag{50}$$

Similarly, for the other blocks in (45), we obtain

$$\frac{\partial\,\text{vec}\,\bar{\mathbf{P}}}{\partial\,\text{vec}\,\tilde{\mathbf{L}}} = (\tilde{\mathbf{L}} \otimes \mathbf{I}) + (\mathbf{I} \otimes \tilde{\mathbf{L}})\boldsymbol{\Sigma}^{np} \tag{51}$$

$$\frac{\partial\,\text{vec}\,\tilde{\mathbf{P}}}{\partial\,\text{vec}\,\bar{\mathbf{L}}} = -(\tilde{\mathbf{L}} \otimes \mathbf{I}) + (\mathbf{I} \otimes \tilde{\mathbf{L}})\boldsymbol{\Sigma}^{np} \tag{52}$$

$$\frac{\partial\,\text{vec}\,\tilde{\mathbf{P}}}{\partial\,\text{vec}\,\tilde{\mathbf{L}}} = (\bar{\mathbf{L}} \otimes \mathbf{I}) - (\mathbf{I} \otimes \bar{\mathbf{L}})\boldsymbol{\Sigma}^{np}\,. \tag{53}$$

A compact matrix expression for $\mathbf{J}$ may thus be written as

$$\mathbf{J} = \frac{\partial\,\text{vech}\,\mathbf{P}}{\partial\,\text{vech}\,\mathbf{L}} = \text{Re}\{\mathbf{Q}_p^c(\mathbf{L} \otimes \mathbf{I})\mathbf{Q}_l^T + \mathbf{Q}_p(\mathbf{I} \otimes \mathbf{L})\boldsymbol{\Sigma}^{np}\mathbf{Q}_l^T\} \tag{54}$$

where $(\cdot)^c$ denotes the complex conjugate and

$$\mathbf{Q}_p = \begin{bmatrix} Q_p \\ -j\Gamma_p \end{bmatrix} \qquad \mathbf{Q}_l = \begin{bmatrix} Q_l \\ -j\Gamma_l \end{bmatrix}. \tag{55}$$

Since the two estimators described in the previous section are asymptotically statistically efficient under appropriate model assumptions, their asymptotic covariance matrices are given by the corresponding CRB's. This fact is used here when comparing the performances of the estimators. As proved in Theorem 2, the ML signal parameter estimators obtained in the two parameterizations under discussion are asymptotically identical. In other words their associated CRB matrices coincide. In view of this fact, the following discussion focuses on the CRB matrices corresponding to the $\mathbf{P}$ and σ^2 parameters.

The ML asymptotic covariance of the estimated signal covariance matrix parameters is given by the corresponding CRB and is denoted by CRB_p when the full parameterization is used (i.e., $\mathbf{P}$ is parameterized by vech$\mathbf{P}$) and by CRB_l for the low-rank parameterization (when $\mathbf{P}$ is parameterized by vech $\mathbf{L}$). Note that we have the following relationship:

$$\overline{\text{CRB}}_p = \mathbf{J}\text{CRB}_l\mathbf{J}^T \tag{56}$$

where $\overline{\text{CRB}}_p$ denotes the asymptotic CRB on the signal covariance matrix elements (i.e., vech $\mathbf{P}$), corresponding to the low-rank square-root parameterization of $\mathbf{P}$. This allows us to transform CRB_l to a lower bound on the estimation accuracy of the emitter covariance matrix elements. In what follows we compare CRB_p with $\overline{\text{CRB}}_p$.

We begin by deriving a closed-form expression for CRB_l. To evaluate CRB_l consider the parameter vector $\boldsymbol{\eta}^T = [\boldsymbol{\theta}^T \quad \boldsymbol{\eta}_l^T \quad \sigma^2]$. The elements of the inverse CRB, that is, the Fisher information matrix, are given by the Bangs' formula [17] (for $N = 1$)

$$\begin{aligned} F_{ij} &= \text{Tr}\left\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \boldsymbol{\eta}_i}\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \boldsymbol{\eta}_j}\right\} \\ &= \text{vec}\left(\mathbf{R}^{-1/2}\frac{\partial \mathbf{R}}{\partial \boldsymbol{\eta}_i}\mathbf{R}^{-*/2}\right)^* \text{vec}\left(\mathbf{R}^{-1/2}\frac{\partial \mathbf{R}}{\partial \boldsymbol{\eta}_j}\mathbf{R}^{-*/2}\right) \\ &= \left(\frac{\partial\,\text{vec}\,\mathbf{R}}{\partial \boldsymbol{\eta}_i}\right)^* (\mathbf{R}^{-c/2} \otimes \mathbf{R}^{-1/2})^* \\ &\quad (\mathbf{R}^{-c/2} \otimes \mathbf{R}^{-1/2})\frac{\partial\,\text{vec}\,\mathbf{R}}{\partial \boldsymbol{\eta}_j} \\ &= \left(\frac{\partial\,\text{vec}\,\mathbf{R}}{\partial \boldsymbol{\eta}_i}\right)^* (\mathbf{R}^{-T} \otimes \mathbf{R}^{-1})\frac{\partial\,\text{vec}\,\mathbf{R}}{\partial \boldsymbol{\eta}_j}\,. \end{aligned} \tag{57}$$

Thus, a matrix expression for the information matrix is given by

$$F = \left(\frac{\partial\,\text{vec}\,\mathbf{R}}{\partial \boldsymbol{\eta}}\right)^* (\mathbf{R}^{-T} \otimes \mathbf{R}^{-1})\frac{\partial\,\text{vec}\,\mathbf{R}}{\partial \boldsymbol{\eta}}\,. \tag{58}$$

The result below provides matrix expressions for the derivatives of vec $\mathbf{R}$ with respect to the three components of the parameter vector $\boldsymbol{\eta}^T = [\boldsymbol{\theta}^T \quad \boldsymbol{\eta}_l^T \quad \sigma^2]$.

Theorem 3: The derivatives of the array output covariance matrix are given by

$$\frac{\partial\,\text{vec}\,\mathbf{R}}{\partial \boldsymbol{\theta}} = (\mathbf{A}^c\mathbf{P}^c \otimes \mathbf{I})\mathbf{D} + (\mathbf{I} \otimes \mathbf{A}\mathbf{P})\boldsymbol{\Sigma}^{mn}\mathbf{D}^c \tag{59}$$

$$\frac{\partial\,\text{vec}\,\mathbf{R}}{\partial \boldsymbol{\eta}_l} = (\mathbf{A}^c \otimes \mathbf{A})(\boldsymbol{\Psi}\mathbf{Q}_p^* + \boldsymbol{\Sigma}^{nn}\mathbf{Q}_p^T)\mathbf{J} \tag{60}$$

$$\frac{\partial\,\text{vec}\,\mathbf{R}}{\partial \sigma^2} = \text{vec}\,\mathbf{I} \tag{61}$$

where $\mathbf{D} = \frac{\partial\,\text{vec}\,\mathbf{A}}{\partial \boldsymbol{\theta}}$ and $\boldsymbol{\Psi} = \mathbf{I} - \text{diag}(\text{vec}\,\mathbf{I})$.

Proof: See Appendix A. □

The matrix expressions for the blocks of F are obtained by using (58) and (59)–(61). Note from (A.2) that an expression for CRB_p is readily obtained in the same fashion as above but letting $\mathbf{J} = \mathbf{I}$ in (60). Note that a similar expression for CRB_p can be found in [16].

An expression for the inverse of CRB_l is obtained by applying block matrix inversion formulas:

$$\text{CRB}_l^{-1} = F_{ll} - F_{l\theta}F_{\theta\theta}^{-1}F_{\theta l} - \frac{(F_{l\sigma^2} - F_{l\theta}F_{\theta\theta}^{-1}F_{\theta\sigma^2})(F_{l\sigma^2} - F_{l\theta}F_{\theta\theta}^{-1}F_{\theta\sigma^2})^*}{F_{\sigma^2\sigma^2} - F_{\sigma^2\theta}F_{\theta\theta}^{-1}F_{\theta\sigma^2}} \quad (62)$$

where F_{ll}, $F_{l\theta}$, etc., denote the blocks of F.

Observe from (60) and (62) that the bound on the square root parameters may be related to the bound on the full parameterization through $\text{CRB}_l^{-1} = \mathbf{J}^T\text{CRB}_p^{-1}\mathbf{J}$. This, together with (56), gives us the relation

$$\overline{\text{CRB}}_p = \mathbf{J}(\mathbf{J}^T\text{CRB}_p^{-1}\mathbf{J})^{-1}\mathbf{J}^T . \quad (63)$$

Note that (63) can also be derived from standard parameter transformation results in the CRB theory (see, e.g., [15]). We previously obtained (63) as a byproduct of the calculations in (56)–(63). The main purpose of the latter calculations has been to provided an explicit formula for CRB_p (as well as CRB_l).

When a full-rank square-root parameterization is used, the bounds are, of course, identical: $\overline{\text{CRB}}_p = \text{CRB}_p$. However, when a low-rank square-root parameterization is applied, we generally have $\overline{\text{CRB}}_p < \text{CRB}_p$ as illustrated in the next section by means of numerical examples. This is in contrast to the bound on the signal parameters, which is not affected by the parameterization of $\mathbf{P}$ (as proved in the previous section).

The accuracy with which the noise variance is estimated is also affected by the choice of ML estimator. The following result relates the two estimators discussed.

Theorem 4: Let $\hat{\sigma}^2(\hat{\boldsymbol{\theta}})$ be the noise variance estimate obtained when no constraints are imposed on the emitter covariance matrix, (6), and let $\hat{\sigma}^2(\hat{\boldsymbol{\alpha}})$ be the noise variance estimate obtained when using the square-root parameterization of the emitter covariance matrix, (16). The asymptotic variances of these two estimates are given by

$$\hat{\sigma}^2(\hat{\boldsymbol{\theta}}) : \frac{\sigma^2}{N(m-n)} \quad (64)$$

$$\hat{\sigma}^2(\hat{\boldsymbol{\alpha}}) : \frac{\sigma^2}{N(m-p)}. \quad (65)$$

Proof: The unconstrained ML estimator is considered in [5], where the result (64) is proved. For the constrained case, consider the model in (15) and the fact that the corresponding ML estimation problem can be decoupled as described in (16)–(18). Noting that the rank of $\mathbf{A}_c$ is p and applying the result in [5] leads directly to (65). □

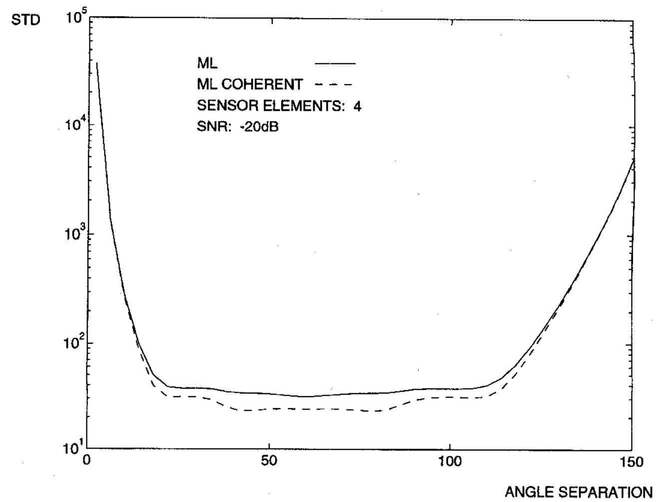


Fig. 1. Bounds on the standard deviation of the signal power estimate versus angle separation.

The ratio of the variances in (64) and, respectively, (65) may be as high as two if $p < n$ and m has its lowest value allowed by the parameter identifiability requirements.

V. Numerical Examples

In order to illustrate the potential gain in estimation accuracy of the proposed coherent ML estimator we present some cases where the asymptotic covariance is numerically evaluated. In the examples below, the CRB for the emitter signal covariance matrix, $\mathbf{P}$, is computed for the two parameterizations under consideration. We will focus on the power estimate of one of the signals. In all cases, the power of this signal of interest is set to one. The bounds are displayed for one snapshot and must be normalized when more snapshots are available.

In the following, a uniform linear array with half wave length spacing is used. The array response vectors have norm $\sqrt{m}$ and the response of the first element is set to 1. All signals arriving at the array are coherent ($p = 1$), and they are related by the vector $\mathbf{L}$. In most cases, two incident signals are considered—one signal of interest and one multipath—in which case, $\mathbf{L} = [1 \ \ \rho e^{j\phi}]^T$. Unless otherwise stated, $\rho = 1$ and $\phi = 0$.

In Fig. 1, the directions of the signals are symmetric about the array broadside. The standard deviation of the estimated power of one of the signals is displayed versus the angle separation in degrees. Four sensors are used (3 dB beamwidth at array broadside is ≈25°) and the signal to noise ratio (SNR) is −20 dB at one sensor. We observe that the relative difference between the estimators is greatest when the array response vectors are well separated. When the angle separation approaches 180°, note that the array response vectors approach each other for this array configuration.

In the following four cases (Figs. 2–5), we choose a scenario in which the incident signal paths are well separated (±20°); for $m \geq 4$, this is more than a beamwidth separation. Unless otherwise stated, the number of sensor elements is four and the SNR is −20 dB. In turn, the dependence of the bounds

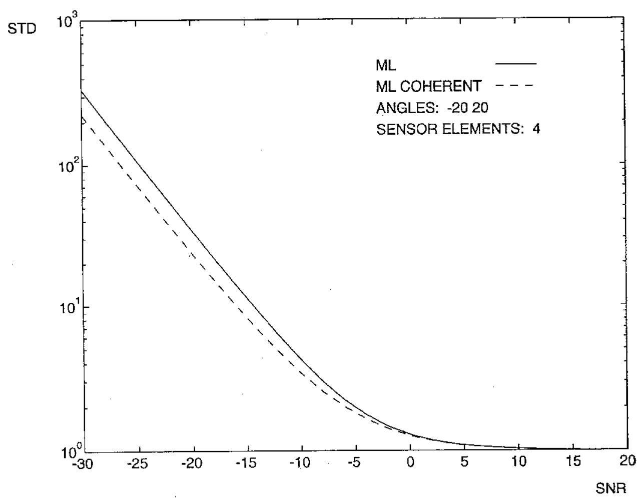


Fig. 2. Bounds on the standard deviation of the signal power estimate versus SNR.

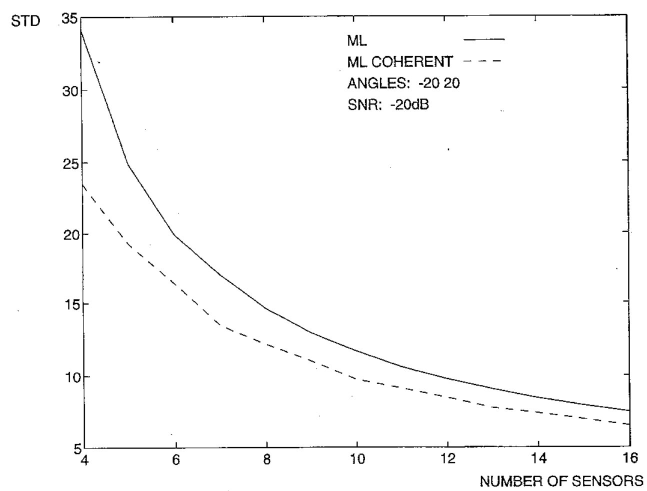


Fig. 3. Bounds on the standard deviation of the signal power estimate versus number of sensors.

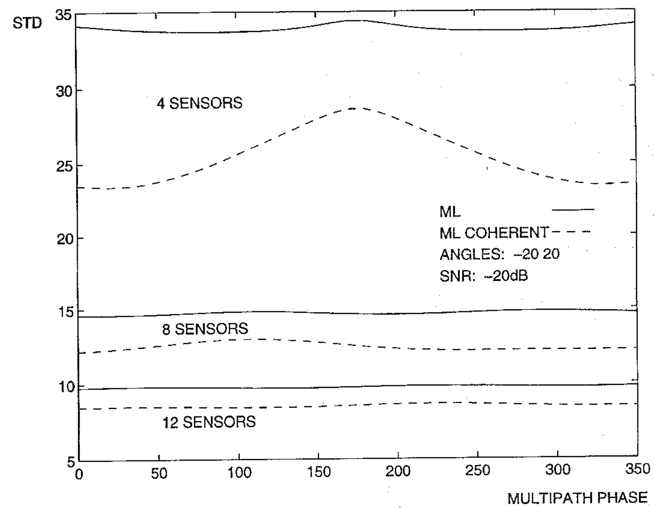


Fig. 4. Bounds on the standard deviation of the signal power estimate versus phase of multipath.

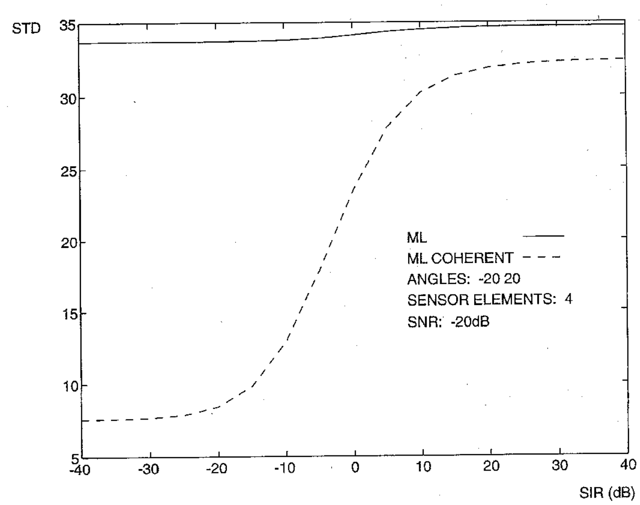


Fig. 5. Bounds on the standard deviation of the signal power estimate versus signal to interference ratio (SIR).

on SNR, number of sensors, relative phase of the signals, and signal to interference ratio is investigated.

Fig. 2 illustrates the standard deviation as a function of SNR. Note that the bounds deviate from each other only in cases with low SNR.

The standard deviation of the power estimate as a function of the number of sensor elements is displayed in Fig. 3. Observe that the bounds approach each other as the number of sensor elements increases (note of course that an increase of m yields a decrease in the array beamwidth).

In Fig. 4, we show the dependence of the bounds on the relative phase of the two signal paths, ϕ. The two bounds are displayed for three different array sizes, $m = 4$, $m = 8$, and $m = 12$. The difference between the bounds is generally modest. It is only in cases with small array aperture that the relative phase significantly affects the accuracy.

In the next case, the power, ρ^2, of the "interfering" multipath is allowed to vary. The bounds in Fig. 5 are computed as a function of the relative power between the signal of interest and the interferer. Note that the relative difference between the bounds significantly decreases when the interference power decreases.

In the following case, we study the effect of an increased number of multipaths. The array is composed of eight sensors ($\approx 15°$ 3 dB beamwidth) and the SNR relative to the first path is -20 dB. In one case, two incident signals are arriving from $\pm 20°$; these results are indicated with $n = 2$ in Fig. 6. In the other case, the number of multipaths is increased to four. Their incident angles are $-20°$, $20°$, $-60°$, and $60°$. These results are indicated with $n = 4$ in Fig. 6.

The signal of interest (with power one and relative phase zero) arrives from $-20°$. The relative phases of all the multipaths are uncorrelated and random. In the second case, the powers are also random (but with a fixed total SIR). The bounds change slightly depending on the realizations of the

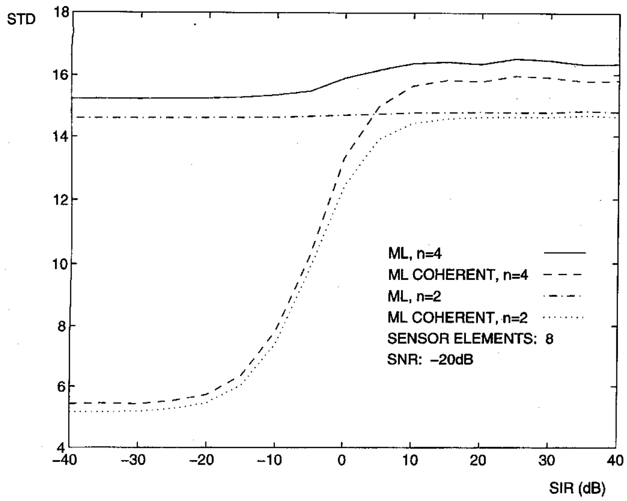


Fig. 6. Bounds on the standard deviation of the signal power estimate versus the relative power between the signal of interest and the power of all the other paths.

mulitpath phases and powers, thus the results are averaged over 50 independent realizations.

The bounds decrease for the case of two paths as expected. Note however that the relative difference between the bounds is quite small when comparing the two cases, $n = 2$ and $n = 4$. This is despite the fact that the relative difference in the number of unknown parameters is large for the two cases, i.e., $n = 2$ giving $2 + 1 + 3 = 6$ and $2 + 1 + 4 = 7$ unknowns respectively for the two estimators and $n = 4$ giving $4 + 1 + 7 = 12$ and $4 + 1 + 16 = 21$ unknowns, respectively, for the two estimators.

Remarks: Based on the numerical examples presented and several other cases not presented herein, we attempt to draw some conclusions.

1) In the examples, the bound on the estimation accuracy of one of the elements of $\mathbf{P}$ is studied. It is our experience that the behavior of the bounds for the other elements of $\mathbf{P}$ is much the same.
2) One may expect that the relative difference between the bounds is greatest when n is much larger than p. However, the difference $(n - p)$ seems to have little impact on the relative difference of the bounds. Rather, the relative difference depends greatly on the number of sensors (i.e., beamwidth) and the signal to interference and noise ratio (SINR). The relative difference increases with decreasing aperture and decreasing SINR. Thus, the power of the multipaths is more critical than the number of multipaths (the identifiability condition must of course be satisfied).
3) The bounds are very similar when the angle separation (in beamwidths) of the signal paths is small. The bounds also approach each other as the SNR (and the SIR) increases.
4) One should note that the examples are chosen to illustrate cases where the difference is relatively large in the bounds. In most cases, the bounds do not differ significantly and we expect the two estimators to provide comparable estimation accuracy. In spite of this fact, we still recommend use of the constrained ML estimator of $\mathbf{P}$ derived in this paper, for reasons detailed in the next section.

VI. Conclusions

The maximum likelihood estimator is formulated for the case of stochastic coherent signals impinging on an antenna array. This assumes that prior knowledge of the rank of the emitter signal covariance matrix is obtained, for example, by some detection procedure. It is shown that the ordinary ML signal parameter estimator, which does not use *a priori* knowledge about the rank of the emitter covariance matrix, is a large-sample realization of the ML signal parameter estimator exploiting the fact that $\mathbf{P}$ is a positive (semi)definite matrix of rank p. The implication of this result, and the main contribution of this paper, is that the knowledge of the rank of the signal covariance matrix does not have to be incorporated in the ML estimation procedure of the signal parameters for large samples. Rather, the signal parameters may be estimated by solving the much simpler unconstrained problem.

Once the signal parameters are estimated, large-sample closed-form solutions to the ML estimation problem for σ^2 and $\mathbf{P}$, under the condition that $\mathbf{P}$ is positive semidefinite and its rank is p, are given. The asymptotic properties of these constrained ML estimates are examined. Although the variance of these parameters is lower than that of the corresponding unconstrained ML estimates, the difference is in general small. However, we recommend the use of the former since computing the two aforementioned estimates involve very similar computational burdens, and in some applications, a small difference between the constrained and unconstrained $\mathbf{P}$ estimates, making the latter indefinite, may lead to a significant deterioration of the signal waveform estimates.

Appendix
Derivatives of the Array Output Covariance Matrix

The three different components of the parameter vector will be considered one at a time. First, consider

$$\frac{\partial \operatorname{vec} \mathbf{R}}{\partial \eta_l} = \frac{\partial \operatorname{vec}(\mathbf{A}\mathbf{P}\mathbf{A}^*)}{\partial \eta_l} = (\mathbf{A}^c \otimes \mathbf{A}) \frac{\partial \operatorname{vec} \mathbf{P}}{\partial \eta_l} \tag{A.1}$$

$$= (\mathbf{A}^c \otimes \mathbf{A}) \frac{\partial \operatorname{vec} \mathbf{P}}{\partial \operatorname{vech} \mathbf{P}} \mathbf{J}. \tag{A.2}$$

Note that

$$\begin{aligned} \frac{\partial \operatorname{vec} \mathbf{P}}{\partial \operatorname{vech} \mathbf{P}} &= \left[\left(\frac{\partial \operatorname{vec} \bar{\mathbf{P}}}{\partial \operatorname{vec} \bar{\mathbf{P}}} + j \frac{\partial \operatorname{vec} \tilde{\mathbf{P}}}{\partial \operatorname{vec} \bar{\mathbf{P}}} \right) Q_p^T \quad \left(\frac{\partial \operatorname{vec} \bar{\mathbf{P}}}{\partial \operatorname{vec} \tilde{\mathbf{P}}} + j \frac{\partial \operatorname{vec} \tilde{\mathbf{P}}}{\partial \operatorname{vec} \tilde{\mathbf{P}}} \right) \Gamma_p^T \right] \\ &= \left[\frac{\partial \operatorname{vec} \bar{\mathbf{P}}}{\partial \operatorname{vec} \bar{\mathbf{P}}} Q_p^T \quad j \frac{\partial \operatorname{vec} \tilde{\mathbf{P}}}{\partial \operatorname{vec} \tilde{\mathbf{P}}} \Gamma_p^T \right]. \end{aligned} \tag{A.3}$$

The derivatives of $\bar{\mathbf{P}}$ and $\tilde{\mathbf{P}}$ are given by

$$\frac{\partial \operatorname{vec} \bar{\mathbf{P}}}{\partial \operatorname{vec} \bar{\mathbf{P}}} Q_p^T = (\mathbf{I} + \Sigma^{nn} - \operatorname{diag}(\operatorname{vec} \mathbf{I}))Q_p^T = (\boldsymbol{\Psi} + \Sigma^{nn})Q_p^T \tag{A.4}$$

$$\frac{\partial \operatorname{vec} \tilde{\mathbf{P}}}{\partial \operatorname{vec} \tilde{\mathbf{P}}} \Gamma_p^T = (\mathbf{I} - \Sigma^{nn} - \operatorname{diag}(\operatorname{vec} \mathbf{I}))\Gamma_p^T = (\boldsymbol{\Psi} - \Sigma^{nn})\Gamma_p^T \tag{A.5}$$

which yields

$$\frac{\partial \operatorname{vec} \mathbf{P}}{\partial \operatorname{vech} \mathbf{P}} = \boldsymbol{\Psi}\mathbf{Q}_p^* + \Sigma^{nn}\mathbf{Q}_p^T. \tag{A.6}$$

Next, consider

$$\begin{aligned}\frac{\partial \operatorname{vec} \mathbf{R}}{\partial \boldsymbol{\theta}_k} &= \operatorname{vec}\left(\frac{\partial \mathbf{A}}{\partial \boldsymbol{\theta}_k}\mathbf{P}\mathbf{A}^*\right) + \operatorname{vec}\left(\mathbf{A}\mathbf{P}\frac{\partial \mathbf{A}^*}{\partial \boldsymbol{\theta}_k}\right) \\ &= (\mathbf{A}^c\mathbf{P}^c \otimes \mathbf{I})\frac{\partial \operatorname{vec} \mathbf{A}}{\partial \boldsymbol{\theta}_k} + (\mathbf{I} \otimes \mathbf{A}\mathbf{P})\frac{\partial \operatorname{vec} \mathbf{A}^*}{\partial \boldsymbol{\theta}_k} \\ &= (\mathbf{A}^c\mathbf{P}^c \otimes \mathbf{I})\frac{\partial \operatorname{vec} \mathbf{A}}{\partial \boldsymbol{\theta}_k} + (\mathbf{I} \otimes \mathbf{A}\mathbf{P})\Sigma^{mn}\frac{\partial \operatorname{vec} \mathbf{A}^c}{\partial \boldsymbol{\theta}_k}.\end{aligned} \tag{A.7}$$

The matrix expression corresponding to (A.7) may be written as

$$\frac{\partial \operatorname{vec} \mathbf{R}}{\partial \boldsymbol{\theta}} = (\mathbf{A}^c\mathbf{P}^c \otimes \mathbf{I})\mathbf{D} + (\mathbf{I} \otimes \mathbf{A}\mathbf{P})\Sigma^{mn}\mathbf{D}^c \tag{A.8}$$

where the $mn \times n$ matrix $\mathbf{D}$ contains the derivatives of the array steering vectors $\mathbf{D} = \frac{\partial \operatorname{vec} \mathbf{A}}{\partial \boldsymbol{\theta}}$.

Finally, note that

$$\frac{\partial \operatorname{vec} \mathbf{R}}{\partial \sigma^2} = \frac{\partial \operatorname{vec} \sigma^2 \mathbf{I}}{\partial \sigma^2} = \operatorname{vec} \mathbf{I}. \tag{A.9}$$

ACKNOWLEDGMENT

The authors would like to thank the reviewers for providing us with several useful comments and suggestions. We are also grateful to Dr. R. Preuss. He is one of the rare Associate Editors who offer personal views on and specific advice for the revision of submitted manuscripts besides the usual ternary editorial decision.

REFERENCES

[1] P. Stoica and A. Nehorai, "MUSIC, Maximum likelihood and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.
[2] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 1110–1121, May 1991.
[3] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1783–1795, Oct. 1990.
[4] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample ML techniques for parameter estimation and detection in array processing," in *Radar Array Processing* Haykin, Litva, and Shepherd, Eds. Berlin: Springer-Verlag, 1993, pp. 99–151.
[5] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Processing*, vol. 40, pp. 590–600, Mar. 1992.
[6] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 2436–2449, Nov. 1991.
[7] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp.1190–1196, Aug. 1989.
[8] P. Stoica and K. Sharman, "Maximum Likelihood Methods for Direction-of-Arrival Estimation," *IEEE Trans. Signal Processing*, vol. 38, pp. 1132–1143, July 1990.
[9] Y. Bresler, "Maximum likelihood estimation of linearly structured covariance with application to antenna array processing," in *Proc. 4th ASSP Workshop Spectrum Estimation Modeling*, Minneapolis, MN, Aug. 1988, pp. 172–175.
[10] J. F. Böhme, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Processing*, vol. 11, pp. 329–337, 1986.
[11] P. Stoica and A. Nehorai, "On the concentrated stochastic likelihood function in array signal processing," Tech. Rep. 9214, Yale Univ., New Haven, 1992.
[12] T. Söderström and P. Stoica, *System Identification.* London: Prentice-Hall International, 1989.
[13] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 996–1000, July 1989.
[14] P. Stoica, "Extension of rayleigh quotient with application," *Bul. Polytechnic Inst., Automat. Contr. Ser.*, vol. XLVIII, pp. 49–53, 1986.
[15] C. R. Rao, *Linear Statistical Inference and its Applications.* New York: Wiley, 1973.
[16] A. J. Weiss and B. Friedlander, "On the Cramér-Rao bound for direction finding of correlated signals," *IEEE Trans. Signal Processing*, vol. 41, pp. 495–499, Jan. 1993.
[17] W. J. Bangs, *Array Processing with Generalized Beamformers*, Ph.D. thesis, Yale Univ., New Haven, 1971.

Petre Stoica (F'94) received the M.S. and D.Sc. degrees in automatic control from Bucharest Polytechnic Institute, Bucharest, Romania, in 1972 and 1979, respectively. In 1993, he was awarded an honorary doctorate by Uppsala University, Uppsala, Sweden.

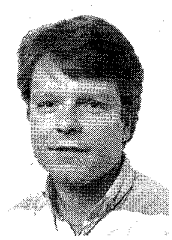
Since 1972, he has been with the Department of Automatic Control and Computers at the Polytechnic Institute of Bucharest, Romania, where he holds the position of Professor of System Identification and Signal Processing. He spent the years 1992 and 1993 and the first half of 1994 with the Systems and Control Group at Uppsala University, Sweden, as a Guest Professor. In the second half of 1994, he held a Chalmers 150th Anniversary Visiting Professorship with the Applied Electronics Department at Chalmers University of Technology, Gothenburg, Sweden. Currently, he is affiliated with the Systems and Control Group of Uppsala University.

Dr. Stoica is a Corresponding Member of the Romanian Academy.

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

From 1984 to 1985, he was with Linköping University as a Research and Teaching Assistant in the Control Theory Group. He was a Visiting Researcher at the Department of Electrical Engineering, Linköping University, during 1988. In 1990, he was a Postdoctoral Fellow associated with the Information Systems Laboratory, Stanford University, and from 1990 to 1991, he was a Research Associate in the Electrical Engineering Department at Linköping University. In 1991, he was appointed Professor of Signal Processing at the Royal Institute of Technology (KTH), Stockholm, and is currently head of the Department of Signals, Sensors, and Systems at KTH. His research interests include stochastic signal processing, sensor array processing, system identification, time series analysis, and wireless communications.

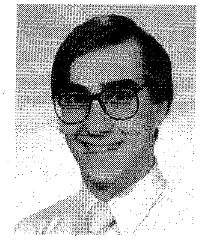
Dr. Ottersten received the Signal Processing Society Paper Award in 1993.



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

He joined the Division of Automatic Control at the Department of Electrical engineering, Linköping University, in 1984, and from November 1989 until August 1993, he was a Research Associate. From October 1988 to March 1989, he was on leave at the Informations Systems Laboratory, Stanford University, Stanford, CA, USA, as a Visiting Scholar. From August 1992 to August 1993, he held a Fulbright–Hayes grant scholarship as a Visiting Researcher at the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT, USA, and at the Informations Systems Laboratory, Stanford University. Since September 1993, he has been a Professor of Signal Processing at the Department of Applied Electronics, Chalmers University of Technology, Sweden. His research interests are in statistical signal processing and its application to sensor array signal processing, system identification, communication, and radar systems.

Dr. Viberg received the IEEE Signal Processing Society's 1993 Paper Award (Statistical Signal and Array Processing Area), for the paper "Sensor Array Processing Based on Subspace Fitting," which he co-authored with B. Ottersten. He is currently a member of the IEEE Signal Processing Society's technical committee on Statistical Signal and Array Processing.



Randolph L. Moses (S'78–M'85–SM'90) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Virginia Polytechnic Institute and State University, Blacksburg, USA, in 1979, 1980, and 1984, respectively.

During the summer of 1983, he was a SCEEE Summer Faculty Research Fellow at Rome Air Development Center, Rome, NY, USA. From 1984 to 1985, he was with the Eindhoven University of Technology, Eindhoven, The Netherlands, as a NATO Postdoctoral Fellow. Since 1985, he has been with the Department of Electrical Engineering, The Ohio State University, Columbus, USA, and is currently and Associate Professor there. From 1994 to 1995, he was on sabbatical leave as a Visiting Professor at the System and Control Group at Uppsala University, Uppsala, Sweden. His research interests are in digital signal processing and include parametric time series analysis, radar signal processing, sensor array processing, and system identification.

Dr. Moses served on the Technical Committee on Statistical Signal and Array Processing of the IEEE Signal Processing Society from 1991 to 1994. He is a member of Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi, and Sigma Xi.