Array Signal Processing with Incomplete Data

Erik G Larsson
Petre Stoica

Technical report 1999-012
November 1999
ISSN 1404-3203

Department of Information Technology
Uppsala University

Array Signal Processing with Incomplete Data

Erik Larsson and Petre Stoica

Dept. of Systems and Control
Uppsala University
P.O. Box 27
SE-751 03 Uppsala
Sweden
Email: egl@syscon.uu.se

April 4, 2000

Abstract

This paper considers the problem of estimating the Direction-of-Arrival (DOA) of one or more signals using an array of sensors, where some of the sensors fail to work before the measurement is completed. Methods for estimating the array output covariance matrix are discussed. In particular, the Maximum-Likelihood estimate of this covariance matrix and its asymptotic accuracy are derived and discussed. Different covariance matrix estimates are used for DOA estimation together with the MUSIC algorithm and with a covariance matching technique. In contrast to MUSIC, the covariance matching technique can utilize information on the estimation accuracy of the array covariance matrix, and it is demonstrated that this yields a significant performance gain.

1 Introduction

The problem of estimating the Direction-of-Arrival (DOA) of a number of signals using an array of sensors has been the subject of research for a fairly long time. The DOA estimation problem typically arises in underwater acoustics, RADAR and communication applications. A large number of methods have been proposed for DOA estimation, for instance the Multiple Signal Classification (MUSIC), Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT), Method of Direction Estimation (MODE), Signal Subspace Fitting (SSF), Noise Subspace Fitting (NSF) and variants thereof, see, e.g., [1, 2, 3, 4, 5, 6] and the references therein. These methods determine the DOA by exploiting properties of the second-order statistics of the data, in particular by estimating the array output covariance matrix matrix and its eigendecomposition.

This paper considers the problem of estimating the DOA from incomplete measurements, that is when some of the data are missing. In some applications such as underwater surveillance, large sensor arrays perform measurements during a long time, and in some cases it happens that one or more of the sensors fail before the measurement is complete. This leads to problems for the conventional methods, since they cannot directly be used when the measurement data are incomplete. In particular, most of these methods use the sample covariance matrix as an estimate of the array output covariance matrix, which cannot be computed in the usual way if some of the data are missing.

The approach taken in this paper is to provide the existing methods, such as MUSIC etc., with a statistically sound estimate of the array output covariance matrix, and use this estimate instead of the sample covariance matrix. Furthermore, we investigate the possibility to utilize information on the estimation accuracy of the array covariance matrix by using the covariance matching technique proposed in [7]. It is shown that this is indeed possible, and that doing so gives significant performance gains.

The paper is organized as follows. In Section 2 the conventional sensor array model and a model for a sensor array with failing sensors are introduced. In Section 3 we propose and analyze different ways to estimate the array output covariance matrix. Section 4 treats DOA estimation based on the covariance matrix estimates derived in Section 3. Section 5 presents some numerical examples and Section 6 concludes the paper.

2 Problem Formulation

2.1 Conventional Array Model

Consider the output of an array consisting of m sensors receiving a superposition of d narrow-band plane waves from far-field emitters. The (complex) envelope of the received waveforms at time t can be modeled as

$$\boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{s}(t) + \boldsymbol{n}(t) \in \mathbb{C}^{m\times 1} \tag{1}$$

where the parameter vector $\boldsymbol{\theta}_0 = [\theta_0^1 \dots \theta_0^d]^T$ contains the DOAs of the d signals, the estimation of which is our main problem, the matrix $\boldsymbol{A}(\boldsymbol{\theta}) = \begin{bmatrix}\boldsymbol{a}(\theta_1) & \dots & \boldsymbol{a}(\theta_d)\end{bmatrix}$ contains the *array steering vectors* $\boldsymbol{a}(\theta)$, $\boldsymbol{s}(t) = \begin{bmatrix}s_1(t) & \dots & s_d(t)\end{bmatrix}^T$ is a vector containing the d source signals at time t, and $\boldsymbol{n}(t) = \begin{bmatrix}n_1(t) & \dots & n_m(t)\end{bmatrix}^T$ is noise. The steering vector $\boldsymbol{a}(\theta)$ is the array response for a unit-amplitude signal impinging from the direction θ. For example, for a Uniform Linear Array (ULA), that is an array of m equispaced identical sensors placed on a straight line, we have $\boldsymbol{a}(\theta) = \begin{bmatrix}1 & e^{j\frac{\omega c}{\Delta}\sin\theta} & \dots & e^{j\frac{\omega c}{\Delta}(m-1)\sin\theta}\end{bmatrix}^T$, where ω is the carrier frequency, c the propagation speed and Δ the inter-element spacing expressed in half-wavelength units.

In this paper, the *Gaussian signal model* is adopted which means that the signal $\boldsymbol{s}(t)$ is modeled as a temporally white, zero-mean circularly symmetric Gaussian random variable with a positive definite covariance matrix $\boldsymbol{S}$, that is $E[\boldsymbol{s}(t_1)\boldsymbol{s}^*(t_2)] = \delta(t_1 - t_2)\boldsymbol{S}$ and $E[\boldsymbol{s}(t_1)\boldsymbol{s}^T(t_2)] = \boldsymbol{0}$ for any t_1 and t_2. A similar zero-mean complex Gaussian assumption is made about the noise $\boldsymbol{n}(t)$. In particular, $E[\boldsymbol{n}(t_1)\boldsymbol{n}^*(t_2)] = \delta(t_1 - t_2)\sigma^2\boldsymbol{I}_m$, and $E[\boldsymbol{n}(t_1)\boldsymbol{n}^T(t_2)] = \boldsymbol{0}$ where $\boldsymbol{I}_m$ denotes the $m \times m$ identity matrix and σ^2 is the noise power. As a remark, note that an alternative to the Gaussian signal model is the *deterministic signal* *model* in which $\boldsymbol{s}(t)$ is modeled as a sequence of deterministic vectors instead of random variables. The choice of signal model has been debated in the literature, and for instance [8, 6] provide some justification as to why the Gaussian signal model is more appropriate for the sensor array problem.

Let the array output be sampled at the time instants $t = 1, \ldots, N$, and assume that the so-obtained array output measurements are stacked into the *array output matrix*

$$\boldsymbol{X} \triangleq \begin{bmatrix} \boldsymbol{x}(1) & \ldots & \boldsymbol{x}(N) \end{bmatrix} \in \mathbb{C}^{m \times N} \tag{2}$$

Many algorithms for inference about the parameter vector $\boldsymbol{\theta}$ rely on properties of the *array output covariance matrix* $\boldsymbol{R}$, which is defined through the relation

$$\boldsymbol{R} = E[\boldsymbol{x}(t)\boldsymbol{x}^*(t)] = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^* + \sigma^2\boldsymbol{I} \in \mathbb{C}^{m \times m} \tag{3}$$

As an estimate of $\boldsymbol{R}$, it is common to use the *sample covariance matrix*

$$\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{x}(t)\boldsymbol{x}^*(t) = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^* \in \mathbb{C}^{m \times m} \tag{4}$$

2.2 Model for an Array with Failing Sensors

In this section we present a model of a sensor array in which one or more sensors fail to work after a certain time. Assume that the total number of sensors in the array is m_1, and that data were collected with all m_1 sensors for $t = 1, \ldots, N_1$. After $t = N_1$ at least one sensor fails to work so that m_2 working sensors remain. Using the m_2 sensors, data are collected for $t = N_1 + 1, \ldots, N_1 + N_2$. Similarly, for $t > N_1 + N_2$ at least one more sensor fails to work so that m_3 working sensors remain, and so on. Let $p - 1$ be the number of time instants at which sensor failure occurred, so that the total measurement time is $N_1 + \ldots + N_p$. Note that $m_1 > m_2 > \cdots > m_p$.

We assume that $\{m_1, \ldots, m_p\}$ and $\{N_1, \ldots, N_p\}$ are *known*. This is justified by the fact that the failing sensors and the moments when they stop functioning are typically easy to identify. Without loss of generality, we further assume that the sensors are numbered so that all m_1 sensors work for $t = 1, \ldots, N_1$, the sensors numbered $m_1 - m_2 + 1, \ldots, m_1$ work for $t = N_1 + 1, \ldots, N_1 + N_2$, etc.

Let $\boldsymbol{x}_k(t) \triangleq \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_{m_k} \end{bmatrix} \boldsymbol{x}(t) \in \mathbb{C}^{m_k \times 1}$ be the last m_k elements of $\boldsymbol{x}(t)$. Define the array data matrix $\boldsymbol{X}_k$ for measurement period k analogously with (2) through

$$\begin{aligned}\boldsymbol{X}_k \triangleq \Big[\boldsymbol{x}_k(N_1 + \cdots + N_{k-1} + 1) \quad \ldots \\ \boldsymbol{x}_k(N_1 + \cdots + N_k)\Big] \in \mathbb{C}^{m_k \times N_k}\end{aligned} \tag{5}$$

The snapshot matrix $\boldsymbol{X}_k$ contains all available measurements of $\boldsymbol{x}(t)$ during the time interval in which exactly m_k sensors were functioning.

The partial array output covariance matrix for measurement period k, $\boldsymbol{R}_k$, is the $m_k \times m_k$ lower right corner of the array output covariance matrix $\boldsymbol{R}$ defined in (3):

$$\boldsymbol{R}_k \triangleq E[\boldsymbol{x}_k(t)\boldsymbol{x}_k^*(t)] = [\boldsymbol{0} \quad \boldsymbol{I}_{m_k}]\boldsymbol{R}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{I}_{m_k}\end{bmatrix} \in \mathbb{C}^{m_k \times m_k} \tag{6}$$

Similarly, the sample covariance matrix $\hat{\boldsymbol{R}}_k$ based on data from measurement period k is defined analogously with (4) through the relation

$$\hat{\boldsymbol{R}}_k = \frac{1}{N_k}\boldsymbol{X}_k\boldsymbol{X}_k^* \in \mathbb{C}^{m_k \times m_k} \tag{7}$$

3 Estimation of the Array Output Covariance Matrix

In this section, two different ways of estimating the array output covariance matrix for the sensor array problem with failing sensors are considered. The simplest way to estimate $\boldsymbol{R}$ is to use data from the first measurement period only, that is when all sensors were working, and hence discard data from measurement periods $2, \ldots, p$. In other words, take $\hat{\boldsymbol{R}}_0 = \frac{1}{N_1}\boldsymbol{X}_1\boldsymbol{X}_1^*$. Owing to obvious reasons, this is expected to be a poor estimate.

In the following we first present a relatively simple *ad-hoc* estimate. Next we derive the Maximum-Likelihood estimate of the covariance matrix and analyze its asymptotic properties.

3.1 The ad-hoc Estimate

Element i, j of the *ad-hoc* estimate $\hat{\boldsymbol{R}}_{ah}$ is an (unstructured) estimate of the covariance of sensor outputs i and j based on data from the measurement periods where both sensor i and j were working. We can express this as

$$\hat{\boldsymbol{R}}_{ah}^{i,j} = \frac{1}{|\Omega_{i,j}|}\sum_{t\in\Omega_{i,j}} \boldsymbol{x}_i(t)\boldsymbol{x}_j^*(t) \tag{8}$$

where $\Omega_{i,j}$ is a set containing the time instants at which both sensor i and j were working:

$$t \in \Omega_{i,j} \quad \Leftrightarrow \quad t \leq N_1 + \cdots + N_{\min\{k_i,k_j\}} \tag{9}$$

where k_i and k_j are the largest integers such that $m_{k_i} \geq i$ and $m_{k_j} \geq j$, respectively, and $|\Omega_{i,j}| = N_1 + \cdots + N_{\min\{k_i,k_j\}}$ is the number of elements in $\Omega_{i,j}$.

We note in passing that $\hat{\boldsymbol{R}}_{ah}$ is guaranteed to be Hermitian but not necessarily positive definite. Indeed, it is easy to find examples where $\hat{\boldsymbol{R}}_{ah}$ becomes indefinite, which is an undesired property.

3.2 The Maximum-Likelihood Estimate

It is well-known that for the classical sensor array problem the sample covariance matrix $\hat{\boldsymbol{R}}$ is equal to the Maximum-Likelihood (ML) estimate of the array output covariance matrix. In this section we consider the ML estimation of the array covariance matrix $\boldsymbol{R}$ for the problem with failing sensors.

Following [9], it turns out that a fruitful approach for the derivation of the MLE of $\boldsymbol{R}$ is to introduce a parameterization of $\boldsymbol{R}$ in terms of the Cholesky factorization of its inverse. Let the lower triangular $m_1 \times m_1$ matrix

$$\boldsymbol{H} \triangleq \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ & 1 & 0 & & & \vdots \\ & & 1 & \ddots & & \vdots \\ \boldsymbol{h}_{m_1} & & & \ddots & \ddots & \vdots \\ & \boldsymbol{h}_{m_1-1} & & & 1 & 0 \\ & & & & \boldsymbol{h}_2 & 1 \end{bmatrix} \tag{10}$$

and the $m_1 \times m_1$ diagonal matrix with real-valued strictly positive diagonal elements

$$\boldsymbol{D} \triangleq \begin{bmatrix} d_{m_1} & 0 & \cdots & \cdots & 0 \\ 0 & d_{m_1-1} & 0 & & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & d_1 \end{bmatrix} \tag{11}$$

constitute the Cholesky factors of $\boldsymbol{R}^{-1}$, so that $\boldsymbol{HDH}^* = \boldsymbol{R}^{-1}$. This factorization is a one-to-one mapping between the set of positive definite matrices $\boldsymbol{R}$ and the set of Cholesky factors $\boldsymbol{H}$ and $\boldsymbol{D}$ [10].

The derivation of the ML estimate relies on the following simple but important observation (see also [9]).

Lemma 1. *Assume that $\boldsymbol{R}^{-1} = \boldsymbol{HDH}^*$ is the Cholesky factorization of $\boldsymbol{R}^{-1}$. Then*

$$\boldsymbol{R}_k^{-1} = \boldsymbol{H}_k \boldsymbol{D}_k \boldsymbol{H}_k^* \quad \textit{for} \quad k = 1, \ldots, p \tag{12}$$

where $\boldsymbol{H}_k = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_{m_k} \end{bmatrix} \boldsymbol{H} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_{m_k} \end{bmatrix}$ is the $m_k \times m_k$ lower right corner of $\boldsymbol{H}$, and $\boldsymbol{D}_k$ is defined analogously; by convention $\boldsymbol{H}_1 = \boldsymbol{H}$ and $\boldsymbol{R}_1 = \boldsymbol{R}$.

Proof. See Appendix A. □

It follows from the assumptions made in Section 2.2 that the negative log-likelihood function for the observed data can be written as, to within an additive constant,

$$
\begin{aligned}
l &= \sum_{k=1}^{p}\Big(N_k \log|\boldsymbol{R}_k| + \sum_{t=1+N_1+\cdots+N_{k-1}}^{N_1+\cdots+N_k} \boldsymbol{x}_k^*(t)\boldsymbol{R}_k^{-1}\boldsymbol{x}_k(t)\Big) \\
&= \sum_{k=1}^{p} N_k\Big(\log|\boldsymbol{R}_k| + \operatorname{Tr}\left\{\boldsymbol{R}_k^{-1}\hat{\boldsymbol{R}}_k\right\}\Big)
\end{aligned}
\tag{13}
$$

where $|\cdot|$ denotes the determinant and $\operatorname{Tr}\{\cdot\}$ denotes the trace. Using Lemma 1 and the fact that $\log|\boldsymbol{R}_k| = -2\log|\boldsymbol{H}_k| - \sum_{j=1}^{m_k}\log d_j$ where $|\boldsymbol{H}_k| = 1$, yields

$$
l = \sum_{k=1}^{p} N_k\Big(\operatorname{Tr}\left\{\boldsymbol{D}_k\boldsymbol{H}_k^*\hat{\boldsymbol{R}}_k\boldsymbol{H}_k\right\} - \sum_{j=1}^{m_k}\log d_j\Big) \tag{14}
$$

For the sake of notational convenience, let

$$
\boldsymbol{g}_j \triangleq \begin{bmatrix} 1 \\ \boldsymbol{h}_j \end{bmatrix} \tag{15}
$$

which permits us to write

$$
N_k \operatorname{Tr}\left\{\boldsymbol{D}_k\boldsymbol{H}_k^*\hat{\boldsymbol{R}}_k\boldsymbol{H}_k\right\} = \sum_{j=1}^{m_k} d_j\boldsymbol{g}_j^*\boldsymbol{\Gamma}_{k,j}\boldsymbol{g}_j \tag{16}
$$

where

$$
\boldsymbol{\Gamma}_{k,j} \triangleq N_k \begin{bmatrix}\boldsymbol{0} & \boldsymbol{I}_j\end{bmatrix} \hat{\boldsymbol{R}}_k \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{I}_j\end{bmatrix} \in \mathbb{C}^{j\times j} \tag{17}
$$

for $j = 1, \ldots, m_{k-1}$ and $\boldsymbol{\Gamma}_{k,m_k} = N_k\hat{\boldsymbol{R}}_k$.

Now (14) can be rewritten as follows:

$$
\begin{aligned}
l &= \sum_{k=1}^{p}\Big(- N_k \sum_{j=1}^{m_k} \log d_j + \sum_{j=1}^{m_k} d_j \boldsymbol{g}_j^* \boldsymbol{\Gamma}_{k,j} \boldsymbol{g}_j \Big) \\
&= -N_1 \sum_{j=1}^{m_1} \log d_j + \sum_{j=1}^{m_1} d_j \boldsymbol{g}_j^* \boldsymbol{\Gamma}_{1,j} \boldsymbol{g}_j \\
&\quad - N_2 \sum_{j=1}^{m_2} \log d_j + \sum_{j=1}^{m_2} d_j \boldsymbol{g}_j^* \boldsymbol{\Gamma}_{2,j} \boldsymbol{g}_j - \dots \\
&\quad - N_p \sum_{j=1}^{m_p} \log d_j + \sum_{j=1}^{m_p} d_j \boldsymbol{g}_j^* \boldsymbol{\Gamma}_{p,j} \boldsymbol{g}_j \\
&= -N_1 \sum_{j=m_2+1}^{m_1} \log d_j + \sum_{j=m_2+1}^{m_1} d_j \boldsymbol{g}_j^* \boldsymbol{\Gamma}_{1,j} \boldsymbol{g}_j \\
&\quad - (N_1 + N_2) \sum_{j=m_3+1}^{m_2} \log d_j \\
&\quad + \sum_{j=m_3+1}^{m_2} d_j \boldsymbol{g}_j^* (\boldsymbol{\Gamma}_{1,j} + \boldsymbol{\Gamma}_{2,j}) \boldsymbol{g}_j \quad - \quad \dots \\
&\quad - (N_1 + \dots + N_p) \sum_{j=1}^{m_p} \log d_j \\
&\quad + \sum_{j=1}^{m_p} d_j \boldsymbol{g}_j^* (\boldsymbol{\Gamma}_{1,j} + \dots + \boldsymbol{\Gamma}_{p,j}) \boldsymbol{g}_j
\end{aligned}
\tag{18}
$$

To proceed, let us define

$$
\boldsymbol{S}_{k,j} \triangleq \sum_{i=1}^{k} \boldsymbol{\Gamma}_{i,j} \in \mathbb{C}^{j \times j} \tag{19}
$$

Then (18) can be written

$$
l = \sum_{k=1}^{p} \sum_{j=m_{k+1}+1}^{m_k} \left\{ -(N_1 + \dots + N_k) \log d_j + d_j \boldsymbol{g}_j^* \boldsymbol{S}_{k,j} \boldsymbol{g}_j \right\} \tag{20}
$$

where by convention $m_{p+1} = 0$. The minimization of l w.r.t. d_j is now straightforward and yields

$$
\hat{d}_j^{-1} = \frac{\hat{\boldsymbol{g}}_j^* \boldsymbol{S}_{k,j} \hat{\boldsymbol{g}}_j}{N_1 + \dots + N_k} \tag{21}
$$

for $j = m_{k+1} + 1, \dots, m_k$, $k = 1, \dots, p$. In (21) $\hat{\boldsymbol{g}}_j$ is the minimizer of $\boldsymbol{g}_j^* \boldsymbol{S}_{k,j} \boldsymbol{g}_j$ subject to $\boldsymbol{g}_j^* \boldsymbol{u} = 1$ with $\boldsymbol{u} \triangleq \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$, viz.

$$
\hat{\boldsymbol{g}}_j = \frac{\boldsymbol{S}_{k,j}^{-1} \boldsymbol{u}}{\boldsymbol{u}^* \boldsymbol{S}_{k,j}^{-1} \boldsymbol{u}} \tag{22}
$$

for $j = m_{k+1}+1, \ldots, m_k$ and $k = 1, \ldots, p$.

To summarize, we have shown the following.

Theorem 1. *The Maximum-Likelihood estimate of the array output covariance matrix* $\boldsymbol{R}$ *is* $\hat{\boldsymbol{R}}_{ml} = \hat{\boldsymbol{H}}^{-*}\hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{H}}^{-1}$, *where* $\hat{\boldsymbol{H}}$ *and* $\hat{\boldsymbol{D}}$ *are computed as follows:*

1. *Compute* $\hat{\boldsymbol{R}}_k$ *for* $k = 1, \ldots, p$ *according to (7).*
2. *Compute* $\boldsymbol{S}_{k,j}$ *for* $j = 1, \ldots, m_k$ *and* $k = 1, \ldots, p$ *according to (17) and (19).*
3. *Compute* $\hat{\boldsymbol{g}}_j$ *for* $j = m_{k+1}+1, \ldots, m_k$ *and* $k = 1, \ldots, p$ *according to (22).*
4. *Compute* $\hat{d}_j$ *according to (21) for* $j = 1, \ldots, m_1$.
5. *Let* $\hat{\boldsymbol{D}} = \operatorname{diag}\left\{\hat{d}_{m_1}, \ldots, \hat{d}_1\right\}$.
6. *Construct* $\hat{\boldsymbol{H}}$ *similarly to (10) by using (15) (note that the first element of* $\hat{\boldsymbol{g}}_j$ *is unity by construction).*

Remark 1. *The ML estimate of the covariance matrix,* $\hat{\boldsymbol{R}}_{ml}$, *is positive semidefinite by construction. Furthermore,* $\hat{d}_j > 0$ *with probability one so that* $\hat{\boldsymbol{R}}_{ml}$ *is positive definite with probability one. As noted above, the* ad-hoc *estimate* $\hat{\boldsymbol{R}}_{ah}$ *can be indefinite, which in particular implies that the ML estimate and the* ad-hoc *estimate are not in general equal to another.*

We next turn our attention to the properties of $\hat{\boldsymbol{R}}_{ml}$. We have the following result.

Theorem 2. *Let* $\hat{\boldsymbol{r}}_{ml} = \operatorname{vec}(\hat{\boldsymbol{R}}_{ml})$ *where* $\operatorname{vec}(\cdot)$ *denotes the operation of forming a vector by stacking the columns of* $(\cdot)$ *on top of each other. Furthermore, introduce for notational simplicity the lower triangular matrices* $\tilde{\boldsymbol{H}} = \boldsymbol{H}\boldsymbol{D}^{1/2}$ *and* $\tilde{\boldsymbol{H}}_k = \boldsymbol{H}_k\boldsymbol{D}_k^{1/2}$. *According to Lemma 4 in Appendix B, there are matrices* $\boldsymbol{J}, \tilde{\boldsymbol{J}} \in \mathbb{C}^{m_1^2 \times m_1^2}$ *such that* $\tilde{\boldsymbol{H}}$ *can be uniquely parameterized by a parameter vector* $\boldsymbol{\eta} \in \mathbb{R}^{m_1^2 \times 1}$ *according to* $\operatorname{vec}(\tilde{\boldsymbol{H}}) = \boldsymbol{J}\boldsymbol{\eta}$, *and* $\operatorname{vec}(\tilde{\boldsymbol{H}}^*) = \tilde{\boldsymbol{J}}\boldsymbol{\eta}$. *Choose* $\boldsymbol{J}$ *and* $\tilde{\boldsymbol{J}}$ *this way, and define*

$$\boldsymbol{J}_k \triangleq \left\{\begin{bmatrix}\boldsymbol{0} & \boldsymbol{I}_{m_k}\end{bmatrix} \otimes \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{I}_{m_k}\end{bmatrix}\right\}\boldsymbol{J} \in \mathbb{C}^{m_k^2 \times m_1^2} \tag{23}$$

where $\otimes$ *denotes the Kronecker product (note that by definition* $\boldsymbol{J}_1 = \boldsymbol{J}$*). Furthermore let* $\boldsymbol{G}_k \triangleq \tilde{\boldsymbol{H}}_k^T \otimes \boldsymbol{I}_{m_k}$. *Then the asymptotic (large-sample) covariance matrix of* $\hat{\boldsymbol{r}}_{ml}$ *is*

$$\boldsymbol{\Sigma}(\boldsymbol{R}) \triangleq E\big[(\boldsymbol{r} - \hat{\boldsymbol{r}}_{ml})(\boldsymbol{r} - \hat{\boldsymbol{r}}_{ml})^*\big] = \frac{d\boldsymbol{r}}{d\boldsymbol{\eta}}\boldsymbol{I}_{\boldsymbol{\eta}}^{-1}\left(\frac{d\boldsymbol{r}}{d\boldsymbol{\eta}}\right)^* \tag{24}$$

where

$$\begin{aligned}\boldsymbol{I}_{\boldsymbol{\eta}} = \sum_{k=1}^{p} 2N_k \operatorname{Re}\Big\{&\boldsymbol{J}_k^T\boldsymbol{G}_k^c(\boldsymbol{r}_k^c\boldsymbol{r}_k^*)^{T_{m_k}}\boldsymbol{G}_k^*\boldsymbol{J}_k \\ &+ \boldsymbol{J}_k^*\boldsymbol{G}_k(\boldsymbol{R}_k^T \otimes \boldsymbol{R}_k)\boldsymbol{G}_k^*\boldsymbol{J}_k\Big\}\end{aligned} \tag{25}$$

and

$$\frac{d\boldsymbol{r}}{d\boldsymbol{\eta}} = (\boldsymbol{R}^T \otimes \boldsymbol{R})\big((\tilde{\boldsymbol{H}}^c \otimes \boldsymbol{I}_{m_1})\boldsymbol{J} + (\boldsymbol{I}_{m_1} \otimes \tilde{\boldsymbol{H}})\tilde{\boldsymbol{J}}\big) \tag{26}$$

Here $(\cdot)^c$ *denotes complex conjugate and* $(\cdot)^{T_N}$ *denotes transposing within each consecutive* $N \times N$ *block of* $(\cdot)$. *In other words if, for instance,* $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ *where* A_{ij} *are* $N \times N$,
then $A^{T_N} = \begin{pmatrix} A_{11}^T & A_{12}^T \\ A_{21}^T & A_{22}^T \end{pmatrix}$.

Proof. See Appendix B. □

Remark 2. *An estimate of* $\boldsymbol{\Sigma}(\boldsymbol{R})$ *is later formed by evaluating (24) at* $\hat{\boldsymbol{R}}_{ml}$ *instead of at the true* $\boldsymbol{R}$. *This estimate will be denoted by* $\boldsymbol{\Sigma}(\hat{\boldsymbol{R}}_{ml})$.

As an illustration of the above result, consider two uncorrelated signals with equal powers and DOAs of 0° and 45°, respectively, arriving at a linear array with 5 equispaced elements. Assume that $N_1 = 100$ samples are collected with all $m_1 = 5$ sensors, and that a varying number N_2 of samples are collected with only the last $m_2 = 2$ sensors. Consider the estimation of the *received power* in sensor 1, 2 and 5, that is the estimation of the elements $\boldsymbol{R}_{1,1}$, $\boldsymbol{R}_{2,2}$ and $\boldsymbol{R}_{5,5}$ of the array output covariance matrix $\boldsymbol{R}$. At a first glance it might seem natural that the estimation accuracy of the elements $\boldsymbol{R}_{1,1}$ and $\boldsymbol{R}_{2,2}$ should be the same and independent of N_2. This is however not the case as Figure 1 demonstrates. The figure shows the asymptotic (large-sample) variance of the estimation error of $\boldsymbol{R}_{1,1}$, $\boldsymbol{R}_{2,2}$, $\boldsymbol{R}_{5,5}$ provided by Theorem 2, together with results obtained by Monte-Carlo simulation. It can be noted that the estimation accuracy of $\boldsymbol{R}_{5,5}$ increases with increasing N_2, which is obvious, but also that the estimation accuracies of $\boldsymbol{R}_{1,1}$ and $\boldsymbol{R}_{2,2}$ are different and increase slightly with increasing N_2. In other words, data from sensor 4 and 5 improves the estimation of the received power in sensor 1 and 2. The small deviation between the asymptotic variance and the empirical variance for sensors 1 and 2 is attributed to the fact that Theorem 2 provides a large-sample result only.

4 DOA Estimation

A large number of algorithms for DOA estimation have appeared in the literature: Maximum-Likelihood methods [8, 6], Multiple Signal Classification (MUSIC) [1, 6], Subspace Fitting Methods [4, 5], Method of Direction Estimation (MODE) [11] and Covariance Matching Techniques (COMET) [7]. In this paper, the usage of certain modified versions of MUSIC and COMET will be considered. We also note here that modifying the Stochastic ML method for the DOA estimation problem with failing sensors appears to be nontrivial, since the estimation of the DOAs and the signal parameters, that is $\boldsymbol{S}$, appears difficult to separate. This implies that one would end up with a nonlinear optimization problem with $d^2 + d + 1$ parameters, which is far from trivial to handle.

4.1 MUSIC

Consider the array output covariance matrix (3). Since $\boldsymbol{S}$ has full rank, it follows that the $m-d$ smallest eigenvalues of $\boldsymbol{R}$ are equal to σ^2, and the d largest eigenvalues are strictly greater than σ^2. Thus, we can form the eigendecomposition

$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^* + \sigma^2\boldsymbol{I} = \boldsymbol{E}_s\boldsymbol{\Lambda}\boldsymbol{E}_s^* + \sigma^2\boldsymbol{E}_n\boldsymbol{E}_n^* \tag{27}$$

where $\boldsymbol{\Lambda}$ is a $d \times d$ diagonal matrix containing the d largest eigenvalues of $\boldsymbol{R}$, $\boldsymbol{E}_s \in \mathbb{C}^{m\times d}$ contains the eigenvectors corresponding to the d largest eigenvalues, $\boldsymbol{E}_n \in \mathbb{C}^{m\times(m-d)}$ contains the eigenvectors corresponding to the eigenvalue σ^2, $\boldsymbol{E}_s^*\boldsymbol{E}_s = \boldsymbol{I}_d$, $\boldsymbol{E}_n^*\boldsymbol{E}_n = \boldsymbol{I}_{m-d}$ and $\boldsymbol{E}_s^*\boldsymbol{E}_n = \boldsymbol{0}$.

In order to estimate the DOAs, the MUSIC algorithm computes the eigendecomposition of the sample covariance matrix $\hat{\boldsymbol{R}}$, viz.

$$\hat{\boldsymbol{R}} = \hat{\boldsymbol{E}}_s\hat{\boldsymbol{\Lambda}}_s\hat{\boldsymbol{E}}_s^* + \hat{\boldsymbol{E}}_n\hat{\boldsymbol{\Lambda}}_n\hat{\boldsymbol{E}}_n^* \tag{28}$$

The matrices $\hat{\boldsymbol{E}}_s$ and $\hat{\boldsymbol{E}}_n$ are estimates of $\boldsymbol{E}_s$ and $\boldsymbol{E}_n$ (note that $\hat{\boldsymbol{\Lambda}}_n$ is no longer a scaled identity matrix). The DOAs are estimated as the locations of the d largest (usually very sharp) peaks of the *MUSIC spectrum*

$$P(\theta) = \frac{\boldsymbol{a}^*(\theta)\boldsymbol{a}(\theta)}{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{E}}_n\hat{\boldsymbol{E}}_n^*\boldsymbol{a}(\theta)} \tag{29}$$

4.2 COMET

Assume that a random vector has a covariance matrix $\boldsymbol{R}(\boldsymbol{\theta})$ that is parameterized by a parameter vector $\boldsymbol{\theta}$, and consider the problem of estimating $\boldsymbol{\theta}$ given an estimate $\hat{\boldsymbol{R}}$ of $\boldsymbol{R}$ together with a measure of the estimation error $\boldsymbol{\Sigma}_{\boldsymbol{r}} = E[(\hat{\boldsymbol{r}}-\boldsymbol{r})(\hat{\boldsymbol{r}}-\boldsymbol{r})^*]$ where $\boldsymbol{r} = \operatorname{vec}(\boldsymbol{R})$ and $\hat{\boldsymbol{r}} = \operatorname{vec}(\hat{\boldsymbol{R}})$, respectively. The idea of COMET is to determine $\boldsymbol{\theta}$ by minimizing a Weighted Least-Squares criterion of the form $(\hat{\boldsymbol{r}}-\boldsymbol{r}(\boldsymbol{\theta}))^*\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}}^{-1}(\hat{\boldsymbol{r}}-\boldsymbol{r}(\boldsymbol{\theta}))$, where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}}$ is an estimate of $\boldsymbol{\Sigma}_{\boldsymbol{r}}$.

In order to carry out the minimization of this criterion it turns out to be useful to introduce a real parameterization of $\boldsymbol{R}$ or equivalently of $\boldsymbol{r}$. A key tool for doing so is provided by Lemma 4 in Appendix B, which shows that there are invertible matrices $\boldsymbol{J}_d \in \mathbb{C}^{d^2\times d^2}$ and $\boldsymbol{J}_{m_1} \in \mathbb{C}^{m_1^2\times m_1^2}$ such that $\operatorname{vec}(\boldsymbol{S}) = \boldsymbol{J}_d\boldsymbol{\xi}$ and $\boldsymbol{r} = \operatorname{vec}(\boldsymbol{R}) = \boldsymbol{J}_{m_1}\boldsymbol{\gamma}$, where $\boldsymbol{\xi} \in \mathbb{R}^{d^2\times 1}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{m_1^2\times 1}$ are the respective parameter vectors. We can write using Lemma 4 and some matrix algebra results from [12],

$$\begin{aligned}\boldsymbol{\gamma} &= \boldsymbol{J}_{m_1}^{-1}\boldsymbol{r} = \boldsymbol{J}_{m_1}^{-1}\operatorname{vec}\Big(\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}\boldsymbol{A}(\boldsymbol{\theta})^* + \sigma^2\boldsymbol{I}\Big)\\ &= \boldsymbol{J}_{m_1}^{-1}\Big(\big(\boldsymbol{A}^c(\boldsymbol{\theta})\otimes\boldsymbol{A}(\boldsymbol{\theta})\big)\operatorname{vec}(\boldsymbol{S}) + \sigma^2\operatorname{vec}(\boldsymbol{I}_{m_1})\Big)\\ &= \boldsymbol{J}_{m_1}^{-1}\left[\big(\boldsymbol{A}^c(\boldsymbol{\theta})\otimes\boldsymbol{A}(\boldsymbol{\theta})\big)\boldsymbol{J}_d \quad \operatorname{vec}(\boldsymbol{I}_{m_1})\right]\begin{bmatrix}\boldsymbol{\xi}\\ \sigma^2\end{bmatrix} \triangleq \boldsymbol{\phi}(\boldsymbol{\theta})\boldsymbol{\alpha}\end{aligned} \tag{30}$$

COMET estimates $\boldsymbol{\theta}$ as the minimizer of

$$\begin{aligned}g(\boldsymbol{\theta}) &= (\hat{\boldsymbol{r}}-\boldsymbol{r})^*\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}}^{-1}(\hat{\boldsymbol{r}}-\boldsymbol{r}) = (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}^{-1}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\\ &= (\hat{\boldsymbol{\gamma}}-\boldsymbol{\phi}(\boldsymbol{\theta})\boldsymbol{\alpha})^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}^{-1}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\phi}(\boldsymbol{\theta})\boldsymbol{\alpha})\end{aligned} \tag{31}$$

where $\hat{\boldsymbol{\Sigma}}_\gamma = \boldsymbol{J}_{m_1}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}}\boldsymbol{J}_{m_1}^{-*}$ and $\hat{\boldsymbol{\gamma}} = \boldsymbol{J}_{m_1}^{-1}\hat{\boldsymbol{r}}$. The result needed to proceed is the following.

Lemma 2. *Assume that the asymptotic covariance matrix in Theorem 2 evaluated at $\hat{\boldsymbol{R}}_{ml}$ is used as weighting in (31), i.e., $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{R}}_{ml})$. Then $\hat{\boldsymbol{\Sigma}}_\gamma$ is real.*

Proof. The result follows from the proof of Theorem 2 by noting that $\boldsymbol{I}_{\boldsymbol{\eta}}$ is real and $\frac{d\boldsymbol{r}}{d\boldsymbol{\eta}} = \boldsymbol{J}_{m_1}\frac{d\boldsymbol{\gamma}}{d\boldsymbol{\eta}}$, where the latter matrix is the derivative of a real quantity w.r.t. a real variable. □

Applying the lemma we find that all quantities in (31) are *real*. Thus (31) can be minimized explicitly w.r.t. $\boldsymbol{\alpha}$, yielding

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = \left(\boldsymbol{\phi}^*(\boldsymbol{\theta})\hat{\boldsymbol{\Sigma}}_\gamma^{-1}\boldsymbol{\phi}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{\phi}^*(\boldsymbol{\theta})\hat{\boldsymbol{\Sigma}}_\gamma^{-1}\hat{\boldsymbol{\gamma}} \in \mathbb{R}^{(d^2+1)\times 1} \tag{32}$$

Plugging this expression into (31) yields, after some simplifications, the *real* optimization problem

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\arg\min}\, \hat{\boldsymbol{\gamma}}^T\hat{\boldsymbol{\Sigma}}_\gamma^{-1/2}\Pi^{\perp}_{\hat{\boldsymbol{\Sigma}}_\gamma^{-1/2}\boldsymbol{\phi}(\boldsymbol{\theta})}\hat{\boldsymbol{\Sigma}}_\gamma^{-1/2}\hat{\boldsymbol{\gamma}} \tag{33}$$

where $\Pi^{\perp}_{\boldsymbol{X}} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$.

Note that the optimization problem (33) is nonlinear and d-dimensional. Given an initial estimate, obtained by e.g. MUSIC, a Newton-type method can be used to minimize the criterion function.

4.3 The Cramér-Rao Bound

The Cramér-Rao inequality gives a lower bound on the achievable variance of any unbiased estimator [13]. Specifically, if Ψ is a set of observations of a random variable depending on a real parameter vector $\boldsymbol{\theta}$ and $l(\boldsymbol{\theta}) \triangleq \log P(\Psi|\boldsymbol{\theta})$ denotes the log-likelihood function of the parameter vector $\boldsymbol{\theta}$ given the observations Ψ, it holds that for any unbiased estimator $\hat{\boldsymbol{\theta}}$

$$E\big[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\big] \geq \left(E\Big[\frac{dl(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\frac{dl(\boldsymbol{\theta})}{d\boldsymbol{\theta}^T}\Big]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^{-1} \triangleq \boldsymbol{F}^{-1} \tag{34}$$

where $\boldsymbol{\theta}_0$ is the true parameter vector and the matrix inequality $\boldsymbol{X} \geq \boldsymbol{Y}$ means that the difference $\boldsymbol{X} - \boldsymbol{Y}$ is positive semidefinite. The matrix $\boldsymbol{F}$ is the *Fisher Information Matrix.*

Let us introduce the *total* parameter vector $\begin{bmatrix}\boldsymbol{\theta}^T & \boldsymbol{\xi}^T & \sigma^2\end{bmatrix}^T$ where $\boldsymbol{\xi}$ is a real vector parameterizing $\boldsymbol{S}$ as in Section 4.2. Then the FIM for the DOA estimation problem has the structure

$$\boldsymbol{F} = \begin{bmatrix}\boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\xi}} & \boldsymbol{F}_{\boldsymbol{\theta}\sigma^2} \\ \boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\xi}}^T & \boldsymbol{F}_{\boldsymbol{\xi}\boldsymbol{\xi}} & \boldsymbol{F}_{\boldsymbol{\xi}\sigma^2} \\ \boldsymbol{F}_{\boldsymbol{\theta}\sigma^2}^T & \boldsymbol{F}_{\boldsymbol{\xi}\sigma^2}^T & \boldsymbol{F}_{\sigma^2\sigma^2}\end{bmatrix} \tag{35}$$

and the CRB for $\boldsymbol{\theta}$ can be expressed as

$$E\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\right] \geq \boldsymbol{F}^{-1}\Big|_{\boldsymbol{\theta\theta}} \tag{36}$$

Extracting the "$\boldsymbol{\theta}$-corner" from $\boldsymbol{F}^{-1}$ is far from trivial. In [8] a closed-form expression for the CRB is obtained for the conventional sensor array problem discussed in Section 2.1 by using results on asymptotic analysis of concentrated likelihood functions; the same result can be found in [14]. Unfortunately, for the problem with missing data, concentrating the likelihood function w.r.t. $\boldsymbol{\theta}$ appears difficult, so the techniques in [14, 8] are not directly applicable. Instead, we evaluate the total FIM $\boldsymbol{F}$, and extract the relevant elements of $\boldsymbol{F}^{-1}$.

Since the measurements are independent, the total FIM equals the sum of the FIMs of the different measurement periods, viz. $\boldsymbol{F} = \sum_{k=1}^{p} \boldsymbol{F}^k$. The evaluation of $\boldsymbol{F}^k$ is straightforward. Let

$$\boldsymbol{D} \triangleq \begin{bmatrix} \frac{d\boldsymbol{a}(\theta)}{d\theta}\Big|_{\theta=\theta_0^1} & \cdots & \frac{d\boldsymbol{a}(\theta)}{d\theta}\Big|_{\theta=\theta_0^d} \end{bmatrix} \tag{37}$$

and define $\boldsymbol{A}_k = [\boldsymbol{0} \quad \boldsymbol{I}_{m_k}]\boldsymbol{A}$, and $\boldsymbol{D}_k = [\boldsymbol{0} \quad \boldsymbol{I}_{m_k}]\boldsymbol{D}$. Then it is not difficult to show that, see, e.g., [15, Section IV]:

$$\begin{aligned}
\boldsymbol{F}^k_{\boldsymbol{\theta\theta}} &= 2\,\mathrm{Re}\left\{(\boldsymbol{S}\boldsymbol{A}_k^*\boldsymbol{R}_k^{-1}\boldsymbol{A}_k\boldsymbol{S}) \odot (\boldsymbol{D}_k^*\boldsymbol{R}_k^{-1}\boldsymbol{D}_k)^T + \right.\\
&\quad \left. + (\boldsymbol{S}\boldsymbol{A}_k^*\boldsymbol{R}_k^{-1}\boldsymbol{D}_k) \odot (\boldsymbol{S}\boldsymbol{A}_k^*\boldsymbol{R}_k^{-1}\boldsymbol{D}_k)^T\right\}\\
\boldsymbol{F}^k_{\boldsymbol{\xi\xi}} &= \boldsymbol{J}_d^*(\boldsymbol{A}_k^T\boldsymbol{R}_k^{-T}\boldsymbol{A}_k^c) \otimes (\boldsymbol{A}_k^*\boldsymbol{R}_k^{-1}\boldsymbol{A}_k)\boldsymbol{J}_d\\
\boldsymbol{F}^k_{\sigma^2\sigma^2} &= \mathrm{Tr}\{\boldsymbol{R}_k^{-2}\}\\
\boldsymbol{F}^k_{\boldsymbol{\theta\xi}} &= \boldsymbol{Q}(\boldsymbol{D}^T\boldsymbol{R}_k^{-T}\boldsymbol{A}^c \otimes \boldsymbol{S}\boldsymbol{A}_k^*\boldsymbol{R}_k^{-1}\boldsymbol{A}_k + \\
&\quad + \boldsymbol{S}^T\boldsymbol{A}_k^T\boldsymbol{R}_k^{-T}\boldsymbol{A}^c \otimes \boldsymbol{D}_k^*\boldsymbol{R}_k^{-1}\boldsymbol{A}_k)\boldsymbol{J}_d\\
\boldsymbol{F}^k_{\boldsymbol{\theta}\sigma^2} &= 2\,\mathrm{Re}\left\{\,\mathrm{diag}\{\boldsymbol{S}\boldsymbol{A}_k^*\boldsymbol{R}_k^{-2}\boldsymbol{D}\}\right\}\\
\boldsymbol{F}^k_{\boldsymbol{\xi}\sigma^2} &= \boldsymbol{J}_d^*(\boldsymbol{A}_k^T\boldsymbol{R}_k^{-T} \otimes \boldsymbol{A}_k^*\boldsymbol{R}_k^{-1})\,\mathrm{vec}(\boldsymbol{I}_{m_k})
\end{aligned} \tag{38}$$

where row i of $\boldsymbol{Q}$ equals $\left(\mathrm{vec}(\boldsymbol{e}_i\boldsymbol{e}_i^T)\right)^T$, $\boldsymbol{e}_i$ is the i^{th} row of the $d \times d$ identity matrix, $\boldsymbol{J}_d$ is defined as in Section 4.2, and $\odot$ denotes elementwise multiplication.

5 Numerical Examples

This section presents some performance results obtained by Monte-Carlo simulation. The following DOA estimation methods were considered.

- *MUSIC – ad-hoc covariance matrix*, and *MUSIC – ML covariance matrix*. MUSIC was applied with the covariance matrix estimates $\hat{\boldsymbol{R}}_{ah}$ and $\hat{\boldsymbol{R}}_{ml}$, respectively.

- *COMET*. The COMET approach was used with the covariance matrix estimate $\hat{\boldsymbol{R}}_{ml}$ and weighting $\boldsymbol{\Sigma}(\hat{\boldsymbol{R}}_{ml})$, i.e. the weighting considered in Lemma 2. A few Newton-steps were taken using a scoring technique

[7], with MUSIC estimates as initial values. In the case that MUSIC failed to resolve the sources, a combination of a grid search and a scoring technique was used. Prior to applying COMET, the weighting matrix was regularized to ensure a conditioning number $< 10^6$.

Figure 2 shows the result of the first experiment. Two sources with DOAs $0°$ and $10°$ were considered, and white noise of different powers was added to the source signals. A uniform linear array with nominally $m_1 = 12$ elements was used for reception and the sources were independent and of unity power, i.e. $\boldsymbol{S} = \boldsymbol{I}$. The number of snapshots were $\begin{bmatrix} N_1 & N_2 & N_3 & N_4 \end{bmatrix} = \begin{bmatrix} 100 & 100 & 200 & 200 \end{bmatrix}$ and $\begin{bmatrix} m_1 & m_2 & m_3 & m_4 \end{bmatrix} = \begin{bmatrix} 12 & 11 & 10 & 9 \end{bmatrix}$. The figure shows the Root-Mean-Square (RMS) error of the DOA estimate for the signal with DOA $0°$ versus the Signal-to-Noise Ratio (SNR) in dB.

In the second experiment, the same model was used but the DOAs were $5°$ and $6°$. This is a much more difficult scenario. Figure 3 shows the results. The performance for *MUSIC – ad-hoc covariance matrix* is not shown since this method failed to resolve the two sources.

Finally, Figure 4 shows the result of a third experiment. The same model as in Figure 3 was used but the source correlation was varied from 0 (uncorrelated sources) to close to 1 (coherent sources). The RMS error of MUSIC increases with increasing source correlation, which is in accordance with the results in [6]. Note however that for source correlation > 0.9, the RMS error of MUSIC does not increase any longer. The reason for this is that the MUSIC spectrum degenerates and exhibits only one peak which is almost always located between $5°$ and $6°$, and the location of this single peak is taken as the DOA estimate for both sources.

From the experiments the following can be concluded:

- Considering usage of the MUSIC algorithm, there can be a significant performance gain when using the ML estimate $\hat{\boldsymbol{R}}_{ml}$ of the covariance matrix instead of the *ad-hoc* estimate $\hat{\boldsymbol{R}}_{ah}$. Specifically, for the case with DOA separation $10°$ there is no significant difference whereas for the more difficult scenario with DOA separation $1°$, MUSIC using $\hat{\boldsymbol{R}}_{ah}$ fails completely to resolve the sources.

- The performance of the COMET algorithm is significantly better than that of MUSIC. The reason for this is that COMET uses information on the estimation accuracy of the array covariance matrix.

- The better performance of the COMET based approach is even more visible when the sources are correlated. This is not surprising since it is well-known that the performance of MUSIC degrades with increasing source correlation [6].

6 Conclusions

This paper has discussed the problem of estimating the parameters in a sensor array model from incomplete data corresponding to the case when one or more of the sensors fail to work before the measurement is complete. Especially the problem of DOA estimation using an array of sensors was considered. First, different ways of estimating the array output covariance matrix were considered. The ML estimate of the covariance matrix and its asymptotic accuracy were derived and discussed. Next, DOA estimation based on the estimated covariance matrices was considered. Two algorithms were discussed, the well-known MUSIC algorithm and a covariance matching technique (COMET). It was demonstrated that COMET, in contrast to MUSIC, can use information on the accuracy of the covariance matrix estimate, and that this can improve the performance when estimating the DOAs from incomplete data.

In our study the best performing method for DOA estimation from incomplete data was COMET using the maximum-likelihood estimate of the covariance matrix $\hat{\boldsymbol{R}}_{ml}$ together with an estimate of the covariance of $\hat{\boldsymbol{R}}_{ml}$ as weighting matrix. For a detailed discussion of the performance results, see Section 5. The proposed COMET approach may therefore be the preferred choice when treating parameter estimation from incomplete data in the sensor array signal processing framework.

A. Proof of Lemma 1

By definition

$$\boldsymbol{R}_k = \begin{bmatrix}\boldsymbol{0} & \boldsymbol{I}_{m_k}\end{bmatrix} \boldsymbol{R} \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{I}_{m_k}\end{bmatrix} = \begin{bmatrix}\boldsymbol{0} & \boldsymbol{I}_{m_k}\end{bmatrix} \boldsymbol{H}^{-*}\boldsymbol{D}^{-1}\boldsymbol{H}^{-1} \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{I}_{m_k}\end{bmatrix} \tag{39}$$

But $\boldsymbol{H}$ has the structure

$$\boldsymbol{H} = \begin{bmatrix}\times & \boldsymbol{0} \\ \times & \boldsymbol{H}_k\end{bmatrix} \tag{40}$$

where $\times$ denotes a block of no interest for our analysis. Then

$$\boldsymbol{H}\begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{H}_k^{-1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{I}_{m_k}\end{bmatrix} \Leftrightarrow \boldsymbol{H}^{-1}\begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{I}_{m_k}\end{bmatrix} = \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{H}_k^{-1}\end{bmatrix} \tag{41}$$

Using this observation twice in (39) yields

$$\begin{aligned}\boldsymbol{R}_k = \begin{bmatrix}\boldsymbol{0} & \boldsymbol{H}_k^{-*}\end{bmatrix} & \begin{bmatrix} \ddots & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & d_{m_k}^{-1} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & d_1^{-1} \end{bmatrix} \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{H}_k^{-1}\end{bmatrix} \\ =& \boldsymbol{H}_k^{-*}\boldsymbol{D}_k^{-1}\boldsymbol{H}_k^{-1}\end{aligned} \tag{42}$$

which completes the proof.

B. Proof of Theorem 2

We first present two useful lemmas. The first lemma quantifies the estimation error when using the sample covariance matrix as an estimate of a covariance matrix.

Lemma 3. *Let* $\boldsymbol{y}(1), \ldots, \boldsymbol{y}(N)$ *be independent, zero-mean circularly symmetric complex Gaussian random vectors with covariance matrix* $\boldsymbol{R} \triangleq E[\boldsymbol{y}(t)\boldsymbol{y}^*(t)]$. *Let* $\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{y}(t)\boldsymbol{y}^*(t)$, $\boldsymbol{r} = \mathrm{vec}(\boldsymbol{R})$ *and* $\hat{\boldsymbol{r}} = \mathrm{vec}(\hat{\boldsymbol{R}})$. *Then*

1. $E[\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}^*] = \boldsymbol{r}\boldsymbol{r}^* + \frac{1}{N}(\boldsymbol{R}^T \otimes \boldsymbol{R})$
2. $E[\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}^T] = \boldsymbol{r}\boldsymbol{r}^T + \frac{1}{N}(\boldsymbol{r}\boldsymbol{r}^T)^{T_N}$

Proof. For a proof of 1), see [7]. To show 2), let

$$\hat{\boldsymbol{r}}_i \triangleq \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{y}(t)y_i^*(t) \tag{43}$$

be the i^{th} column of $\hat{\boldsymbol{R}}$, and similarly $\boldsymbol{r}_i$ the i^{th} column of $\boldsymbol{R}$. This yields

$$E[\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_j^T] = \frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}E[\boldsymbol{y}(t)\boldsymbol{y}^T(s)y_i^*(t)y_j^*(s)] \tag{44}$$

Using the results in [13] for the fourth-order moments of complex Gaussian variables, and the symmetry properties of $\boldsymbol{y}(t)$, the element k, m of $E[\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_j^T]$ can be written, where $t \neq s$

$$\begin{aligned}
E[\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_j^T]\Big|_{k,m} &= \frac{1}{N^2}E[y_k(t)y_m(t)y_i^*(t)y_j^*(t)] \\
&+ \frac{N^2-N}{N^2}E[y_k(t)y_m(s)y_i^*(t)y_j^*(s)] \\
&= \frac{1}{N}\Big(E[y_k(t)y_m(t)]E[y_i^*(t)y_j^*(t)] \\
&+ E[y_k(t)y_i^*(t)]E[y_m(t)y_j^*(t)] \\
&+ E[y_k(t)y_j^*(t)]E[y_m(t)y_i^*(t)]\Big) \\
&+ \frac{N^2-N}{N^2}\Big(E[y_k(t)y_m(s)]E[y_i^*(t)y_j^*(s)] \\
&+ E[y_k(t)y_i^*(t)]E[y_m(s)y_j^*(s)] \\
&+ E[y_k(t)y_j^*(s)]E[y_m(s)y_i^*(t)]\Big) \\
&= \frac{1}{N}\Big(\boldsymbol{R}_{k,i}\boldsymbol{R}_{m,j} + \boldsymbol{R}_{k,j}\boldsymbol{R}_{m,i}\Big) \\
&+ \frac{N^2-N}{N^2}\boldsymbol{R}_{k,i}\boldsymbol{R}_{m,j} \\
&= \boldsymbol{R}_{k,i}\boldsymbol{R}_{m,j} + \frac{1}{N}\boldsymbol{R}_{k,j}\boldsymbol{R}_{m,i}
\end{aligned} \tag{45}$$

Hence, $E[\hat{\boldsymbol{r}}_i\hat{\boldsymbol{r}}_j^T] = \boldsymbol{r}_i\boldsymbol{r}_j^T + \frac{1}{N}\boldsymbol{r}_j\boldsymbol{r}_i^T$, which concludes the proof. □

The second lemma shows that Hermitian matrices and Cholesky factors can be parameterized by a real-valued parameter vector in a neat way.

Lemma 4. *Let* $\boldsymbol{\Omega}_C$ *be the set of all* $n \times n$ *complex matrices with real diagonal elements and zeros above the main diagonal, and let* $\boldsymbol{\Omega}_H$ *be the set of all Hermitian* $n \times n$ *matrices. There there exist*

1. *a* one-to-one *mapping* $f_H : \mathbb{R}^{n^2 \times 1} \to \boldsymbol{\Omega}_H$ *such that* $\text{vec}(f_H(\boldsymbol{\eta})) = \boldsymbol{J}_H \boldsymbol{\eta}$, *where* $\boldsymbol{\eta} \in \mathbb{R}^{n^2 \times 1}$ *is a parameter vector, and* $\boldsymbol{J}_H \in \mathbb{C}^{n^2 \times n^2}$ *is a constant matrix. Furthermore,* $\boldsymbol{J}_H$ *is invertible, and the inverse mapping* f_H^{-1} *has the form* $f_H^{-1}(\boldsymbol{X}) = \boldsymbol{J}_H^{-1} \text{vec}(\boldsymbol{X})$.

2. *a mapping* $f_C : \mathbb{R}^{n^2 \times 1} \to \boldsymbol{\Omega}_C$ *such that* $\text{vec}(f_C(\boldsymbol{\eta})) = \boldsymbol{J}_C \boldsymbol{\eta}$ *and* $\text{vec}(f_C^*(\boldsymbol{\eta})) = \tilde{\boldsymbol{J}}_C \boldsymbol{\eta}$ *, where* $\boldsymbol{\eta} \in \mathbb{R}^{n^2 \times 1}$ *is the parameter vector, and* $\boldsymbol{J}_C, \tilde{\boldsymbol{J}}_C \in \mathbb{C}^{n^2 \times n^2}$ *are constant matrices. The mapping* f_C *is one-to-one, thus invertible, but in contrast to the case above, the* matrices $\boldsymbol{J}_C$ *and* $\tilde{\boldsymbol{J}}_C$ *are singular.*

Proof. The proof of 1) consists of verifying that all Hermitian matrices can be *uniquely* parameterized by $n + 2\frac{n(n-1)}{2} = n^2$ real parameters. That $\boldsymbol{J}_H$ is invertible and that the inverse mapping has the claimed form is shown in [7]. The proof of 2) relies on the same observation. That e.g. $\boldsymbol{J}_C$ is singular is obvious since it must have at least one row of zeros. □

We now proceed to derive the asymptotic covariance matrix for $\hat{\boldsymbol{r}}_{ml} = \text{vec}\left(\hat{\boldsymbol{R}}_{ml}\right)$. According to general properties of ML estimates, the ML estimate is consistent and the asymptotic covariance matrix of the estimate equals the Cramér-Rao bound, which thus is sufficient to compute. We note first that $\boldsymbol{R}^{-1} = \tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^*$ and by Lemma 1 $\boldsymbol{R}_k^{-1} = \tilde{\boldsymbol{H}}_k \tilde{\boldsymbol{H}}_k^*$.

By a statistical linearization argument [16], the asymptotic covariance matrix can be written in the form (24), where $\boldsymbol{I}_{\boldsymbol{\eta}} \triangleq E[\frac{dl}{d\boldsymbol{\eta}} \frac{dl}{d\boldsymbol{\eta}^T}]$ is the FIM for $\boldsymbol{\eta}$, which we next compute.

Since $\tilde{\boldsymbol{H}}_k = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_{m_k} \end{bmatrix} \tilde{\boldsymbol{H}} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_{m_k} \end{bmatrix}$, it follows, using some results of [12], that $\text{vec}(\tilde{\boldsymbol{H}}_k) = \boldsymbol{J}_k \boldsymbol{\eta} \in \mathbb{C}^{m_k^2 \times 1}$ where $\boldsymbol{J}_k$ was defined in (23).

The log-likelihood function for the observed data is, according to (13), $l = \sum_{k=1}^{p} l_k$, where

$$l_k = N_k \Big(\log |\boldsymbol{R}_k| + \text{Tr} \left\{ \boldsymbol{R}_k^{-1} \hat{\boldsymbol{R}}_k \right\} \Big) \tag{46}$$

The derivative w.r.t. η_i becomes, using some matrix calculus results of [13]

$$\begin{aligned} \frac{dl_k}{d\eta_i} = N_k \, \text{Tr} \Big\{ &- 2\tilde{\boldsymbol{H}}_k^{-1} \frac{d}{d\eta_i} \tilde{\boldsymbol{H}}_k \\ &+ (\frac{d}{\eta_i} \tilde{\boldsymbol{H}}_k) \tilde{\boldsymbol{H}}_k^* \hat{\boldsymbol{R}}_k + (\frac{d}{\eta_i} \tilde{\boldsymbol{H}}_k^*) \hat{\boldsymbol{R}}_k \tilde{\boldsymbol{H}}_k \Big\} \end{aligned} \tag{47}$$

Clearly, $\frac{d}{d\eta_i} \operatorname{vec} \tilde{\boldsymbol{H}}_k = \boldsymbol{J}_k \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is the i^{th} column of the identity matrix. This yields, after some algebraic manipulations,

$$\begin{aligned} \frac{dl_k}{d\boldsymbol{\eta}} &= \begin{bmatrix} \frac{dl_k}{\eta_1} & \vdots & \frac{dl_k}{\eta_{m_1^2}} \end{bmatrix}^T \\ &= 2N_k \left(-\boldsymbol{J}_k^T \operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T}) + \operatorname{Re}\left\{ \boldsymbol{J}_k^T \operatorname{vec}(\hat{\boldsymbol{R}}_k \tilde{\boldsymbol{H}}_k)^c \right\} \right) \end{aligned} \tag{48}$$

Since $E[\hat{\boldsymbol{R}}_k \tilde{\boldsymbol{H}}_k] = \boldsymbol{R}_k \tilde{\boldsymbol{H}}_k = (\tilde{\boldsymbol{H}}_k \tilde{\boldsymbol{H}}_k^*)^{-1} \tilde{\boldsymbol{H}}_k = \tilde{\boldsymbol{H}}_k^{-*}$, it follows that $E[\frac{dl_k}{d\boldsymbol{\eta}}] = 0$, and consequently that $E[\frac{dl}{d\boldsymbol{\eta}}] = 0$. Together with the fact that the observations during different measurement periods are independent, this shows that the Fisher information can be written

$$\boldsymbol{I}_{\boldsymbol{\eta}} = E\left[\frac{dl}{d\boldsymbol{\eta}} \frac{dl}{d\boldsymbol{\eta}^T} \right] = \sum_{k=1}^{p} E\left[\frac{dl_k}{d\boldsymbol{\eta}} \frac{dl_k}{d\boldsymbol{\eta}^T} \right] \tag{49}$$

To evaluate this expression, note first that $\boldsymbol{J}_k^T \operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})$ is the derivative of a real quantity w.r.t. a real variable, thus indeed *real-valued.* Furthermore, it holds that $\operatorname{vec}(\hat{\boldsymbol{R}}_k \tilde{\boldsymbol{H}}_k) = (\tilde{\boldsymbol{H}}_k^T \otimes \boldsymbol{I}_{m_k}) \operatorname{vec}(\hat{\boldsymbol{R}}_k) = \boldsymbol{G}_k \hat{\boldsymbol{r}}_k$, and $\operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-*}) = \boldsymbol{G}_k \boldsymbol{r}_k$.

We get

$$
\begin{aligned}
&E\left[\frac{dl_k}{d\boldsymbol{\eta}}\frac{dl_k}{d\boldsymbol{\eta}^T}\right] = \\
&= 4N_k^2 E\Big[\Big(\boldsymbol{J}_k^T \operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T}) - \operatorname{Re}\left\{\boldsymbol{J}_k^T \operatorname{vec}(\hat{\boldsymbol{R}}_k\tilde{\boldsymbol{H}}_k)^c\right\}\Big) \times \\
&\quad \Big(\big(\operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})\big)^T\boldsymbol{J}_k - \operatorname{Re}\left\{\big(\operatorname{vec}(\hat{\boldsymbol{R}}_k\tilde{\boldsymbol{H}}_k)\big)^*\boldsymbol{J}_k\right\}\Big)\Big] \\
&= 4N_k^2\Big\{\boldsymbol{J}_k^T \operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})(\operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T}))^T\boldsymbol{J}_k \\
&\quad + E\Big[\operatorname{Re}\left\{\boldsymbol{J}_k^T \operatorname{vec}(\hat{\boldsymbol{R}}_k^c\tilde{\boldsymbol{H}}_k^c)\right\}\operatorname{Re}\left\{\big(\operatorname{vec}(\hat{\boldsymbol{R}}_k\tilde{\boldsymbol{H}}_k)\big)^*\boldsymbol{J}_k\right\}\Big] \\
&\quad - \boldsymbol{J}_k^T \operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})E\Big[\operatorname{Re}\left\{\big(\operatorname{vec}(\hat{\boldsymbol{R}}_k\tilde{\boldsymbol{H}}_k)\big)^*\boldsymbol{J}_k\right\}\Big] \\
&\quad - E\Big[\operatorname{Re}\left\{\boldsymbol{J}_k^T \operatorname{vec}\big(\hat{\boldsymbol{R}}_k^c\tilde{\boldsymbol{H}}_k^c\big)\right\}\Big]\big(\operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})\big)^T\boldsymbol{J}_k\Big\} \\
&= 2N_k^2\Big\{2\boldsymbol{J}_k^T \operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})\big(\operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})\big)^T\boldsymbol{J}_k \\
&\quad + \operatorname{Re}\Big\{E\big[\boldsymbol{J}_k^T \operatorname{vec}(\hat{\boldsymbol{R}}_k^c\tilde{\boldsymbol{H}}_k^c)\big(\operatorname{vec}(\hat{\boldsymbol{R}}_k\tilde{\boldsymbol{H}}_k)\big)^*\boldsymbol{J}_k\big]\Big\} \\
&\quad + \operatorname{Re}\Big\{E\big[\boldsymbol{J}_k^* \operatorname{vec}(\hat{\boldsymbol{R}}_k\tilde{\boldsymbol{H}}_k)\big(\operatorname{vec}(\hat{\boldsymbol{R}}_k\tilde{\boldsymbol{H}}_k)\big)^*\boldsymbol{J}_k\big]\Big\} \\
&\quad - 2\boldsymbol{J}_k^T \operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})\operatorname{Re}\left\{\big(\operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-*})\big)^*\boldsymbol{J}_k\right\} \\
&\quad - 2\operatorname{Re}\left\{\boldsymbol{J}_k^T \operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})\right\}\big(\operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})\big)^T\boldsymbol{J}_k\Big\} \\
&= 2N_k^2\Big\{-2\boldsymbol{J}_k^T \operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})\big(\operatorname{vec}(\tilde{\boldsymbol{H}}_k^{-T})\big)^T\boldsymbol{J}_k \\
&\quad + \operatorname{Re}\Big\{E\big[\boldsymbol{J}_k^T\boldsymbol{G}_k^c\hat{\boldsymbol{r}}_k^c\hat{\boldsymbol{r}}_k^*\boldsymbol{G}_k^*\boldsymbol{J}_k + \boldsymbol{J}_k^*\boldsymbol{G}_k\hat{\boldsymbol{r}}_k\hat{\boldsymbol{r}}_k^*\boldsymbol{G}_k^*\boldsymbol{J}_k\big]\Big\}\Big\}
\end{aligned}
\tag{50}
$$

where we used the fact that $\operatorname{Re}\{\boldsymbol{X}\}\operatorname{Re}\{\boldsymbol{Y}\} = \frac{1}{2}\operatorname{Re}\{\boldsymbol{X}\boldsymbol{Y} + \boldsymbol{X}\boldsymbol{Y}^c\}$. Using Lemma 3, this gives (25) after some simplifications.

We now turn to compute the derivative $\frac{d\boldsymbol{r}}{d\boldsymbol{\eta}}$. Using that $\boldsymbol{R} = (\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^*)^{-1}$, and results for the derivative of a matrix inverse, it is found that

$$
\begin{aligned}
\frac{d\boldsymbol{r}}{d\eta_i} &= \operatorname{vec}\Big(\frac{d\boldsymbol{R}}{d\eta_i}\Big) = -\operatorname{vec}\Big(\boldsymbol{R}(\frac{d}{d\eta_i}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^*)\boldsymbol{R}\Big) \\
&= -(\boldsymbol{R}^T\otimes\boldsymbol{R})\frac{d}{d\eta_i}\operatorname{vec}(\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^*) \\
&= -(\boldsymbol{R}^T\otimes\boldsymbol{R})\big((\tilde{\boldsymbol{H}}^c\otimes\boldsymbol{I}_{m_1})\frac{d}{d\eta_i}\operatorname{vec}(\tilde{\boldsymbol{H}}) \\
&\quad + (\boldsymbol{I}_{m_1}\otimes\tilde{\boldsymbol{H}})\frac{d}{d\eta_i}\operatorname{vec}(\tilde{\boldsymbol{H}}^*)\big)
\end{aligned}
\tag{51}
$$

But $\frac{d}{d\eta_i}\operatorname{vec}(\tilde{\boldsymbol{H}}) = \boldsymbol{J}\boldsymbol{e}_i$, and $\frac{d}{d\eta_i}\operatorname{vec}(\tilde{\boldsymbol{H}}^*) = \tilde{\boldsymbol{J}}\boldsymbol{e}_i$, which gives (26) and concludes the proof.

References

[1] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. thesis, Stanford University, Stanford, CA, USA, 1981.

[2] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1132–1141, July 1990.

[3] R. Roy and T. Kailath, "ESPRIT — estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Signal Processing*, vol. ASSP-37, no. 7, pp. 984–995, 1989.

[4] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Transactions on Signal Processing*, vol. 39, pp. 2436–2449, Dec. 1991.

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

[6] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood, and Cramér-Rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 720–741, May 1989.

[7] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, pp. 185–210, 1998.

[8] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Transactions on Signal Processing*, vol. 40, pp. 590–600, March 1992.

[9] C. Liu, "Efficient ML estimation of the multivariate normal distribution from incomplete data," *Journal of Multivariate Analysis*, vol. 69, pp. 206–217, 1999.

[10] G. H. Golub and C. F. van Loan, *Matrix Computations.* Maryland, USA: The Johns Hopkins University Press, 1989.

[11] P. Stoica and K. C. Sharman, "A novel eigenanalysis method for direction estimation," *Proceedings of the Institute of Electrical Engineering*, pp. 19–26, Feb. 1990.

[12] A. Graham, *Kronecker Products and Matrix Calculus with Applications.* New York: John Wiley and Sons, 1981.

[13] P. Stoica and R. Moses, *Introduction to Spectral Analysis.* Upper Saddle River, NJ: Prentice Hall, 1997.

[14] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1783–1795, Oct. 1990.

[15] B. Friedlander and A. Weiss, "Direction finding using noise covariance modeling," *IEEE Transactions on Signal Processing*, vol. 43, pp. 1557–1567, 1995.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs: Prentice Hall, 1993.

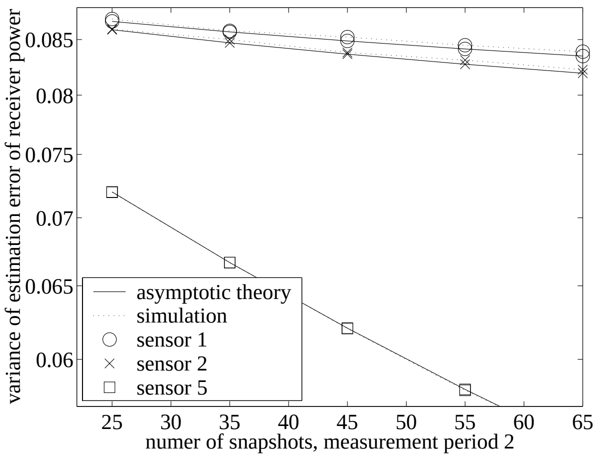

Figure 1: Estimation error variance of the received power in the sensors 1, 2 and 5 for a 5-element ULA with $m_1 = 5$, $m_2 = 2$, $N_1 = 100$ and varying N_2, using the ML estimate of the covariance matrix provided by Theorem 1. Two uncorrelated source signals of equal power were present, with DOAs equal to $0°$ and $45°$, respectively. The Signal-to-Noise Ratio (SNR) was 0 dB.

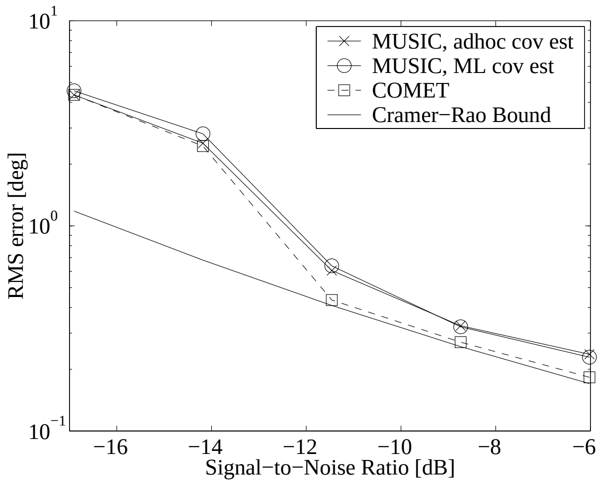

Figure 2: Simulation of DOA estimation with a 12-element array and uncorrelated sources. The parameters were $\boldsymbol{\theta}_0 = [0° \; 10°]^T$, $\boldsymbol{S} = \boldsymbol{I}$, $[N_1 \; N_2 \; N_3 \; N_4] = [100 \; 100 \; 200 \; 200]$ and $[m_1 \; m_2 \; m_3 \; m_4] = [12 \; 11 \; 10 \; 9]$. The figure shows the Root-Mean-Square (RMS) of the DOA estimation error for the signal with DOA $0°$.

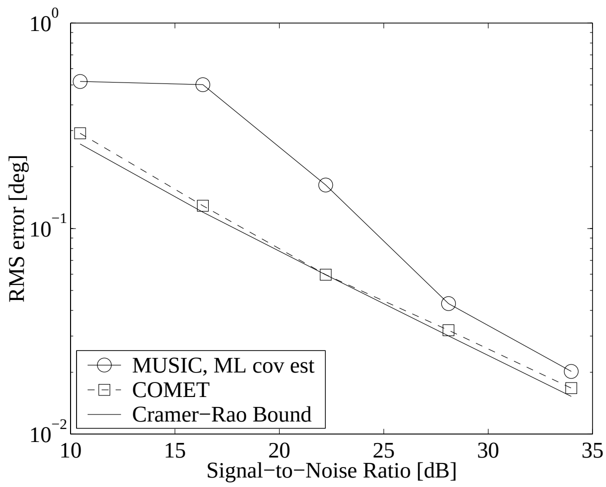

Figure 3: Simulation of DOA estimation with a 12-element array and uncorrelated sources. The parameters were $\boldsymbol{\theta}_0 = [5^\circ\ 6^\circ]^T$, $\boldsymbol{S} = \boldsymbol{I}$, $[N_1\ N_2\ N_3\ N_4] = [100\ 100\ 200\ 200]$ and $[m_1\ m_2\ m_3\ m_4] = [12\ 11\ 10\ 9]$. The figure shows the Root-Mean-Square (RMS) of the DOA estimation error for the signal with DOA 5°.

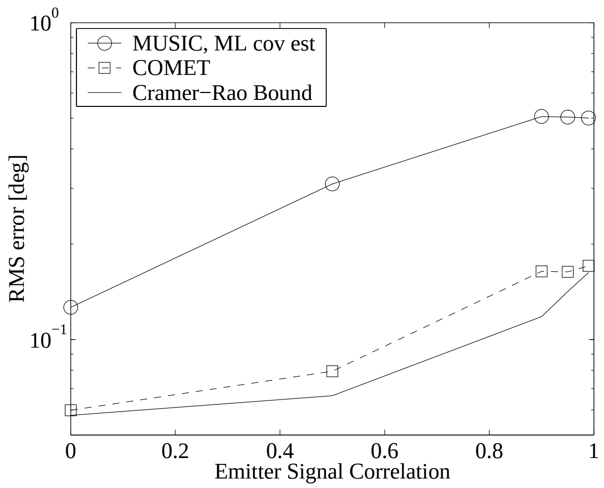

Figure 4: Simulation of DOA estimation with a 12-element array and varying source correlation. The parameters were $\boldsymbol{\theta}_0 = [5^\circ\ 6^\circ]^T$, SNR= 22.5 dB, $[N_1\ N_2\ N_3\ N_4] = [100\ 100\ 200\ 200]$ and $[m_1\ m_2\ m_3\ m_4] = [12\ 11\ 10\ 9]$. The figure shows the Root-Mean-Square (RMS) of the DOA estimation error for the signal with DOA 5°.