Estimation Accuracy of Maximum Likelihood Direction Finding Using Large Arrays*

M. Viberg	B. Ottersten	A. Nehorai
Dept of EE	Dept of EE	Dept of EE
Linköping University	Royal Institute of Technology	Yale University
S-581 83 Linköping	S-100 44 Stockholm	New Haven, CT 06520
Sweden	Sweden	USA

To be presented at the 25th Asilomar Conference, California Nov., 1991.

Abstract

This paper analyzes the performance of methods for estimating the parameters of narrowband signals arriving at an array of sensors. The so-called deterministic and stochastic maximum likelihood (ML) methods are considered. A performance analysis is carried out for finite number of snapshots but assuming that the array is composed of a sufficiently large number, m, of sensors. Strong consistency of the parameter estimates is proved and the asymptotic covariance matrix of the estimation error is derived. Unlike the previously studied large (time) sample case, the present analysis shows that the accuracy is the same for the two ML methods. The covariance matrix of the estimation error attains the Cramér-Rao bound (CRB). For many array geometries of practical interest, the array propagation vectors become orthogonal as m is increased. It is shown that also the traditional beamforming method provides consistent (but not necessarily efficient) estimates under this assumption. This is true also in the presence of perfectly correlated emitters.

1 Introduction

Sensor array signal processing has been an active research area for more than three decades. The generic problem is to determine unknown signal parameters from simultaneous measurements of spatially distributed sensors. Several important applications have been reported, for example, interference rejection in radar and radio communication systems and source localization using underwater sonar arrays.

A vast number of methods have been proposed for solving the estimation problem, see e.g. [1, 2, 3]. Asymptotic results for several estimators have recently appeared in the literature, assuming the number of time samples (snapshots), N, to be large, see e.g. [4, 5]. The so-called *stochastic* maximum likelihood (SML) method has been shown to be asymptotically (for large N) efficient, i.e., the estimation error covariance attains the Cramér-Rao bound (CRB). On the other hand, the deterministic maximum likelihood (DML) estimator is not efficient, due to the increasing number of estimated parameters as N increases.

This paper concerns analysis of estimation accuracy for arbitrary N, assuming a large number of sensors, m. Arrays of several hundred antennas are not uncommon in, for example, radar and sonar systems. A condition guaranteeing consistent estimates for $m \to \infty$ is introduced. It is shown that the SML and DML methods are asymptotically identical for large m (unlike the large N case). The asymptotic covariance matrix of the estimation error coincides with the deterministic CRB. Conditions for consistency and statistical efficiency of the traditional beamforming estimates are also discussed.

2 Estimation Problem

This section formulates the DOA estimation problem and briefly describes the methods under consideration.

2.1 Problem Formulation

Consider an array of m sensors, having arbitrary positions and characteristics. Impinging on the array are the waveforms of d ($d < m$) far-field point sources. Under the narrowband assumption, the complex array output, $\mathbf{x}(t)$, is modeled by the following equation

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

*The work of M. Viberg and B. Ottersten was supported in part by the Swedish Research Council for Engineering Sciences. The work of A. Nehorai was supported by the Air Force Office of Scientific Research under Grant no. AFOSR-90-0164 and the Office of Naval Research under Grant no. N00014-91-J-1298.

The columns of the $m \times d$ matrix $\mathbf{A}$ are the so-called *steering vectors*, $\mathbf{a}(\theta_i)$, $i = 1, \ldots, d$. The steering vectors are functions of the directions-of-arrival (DOAs), θ_i and describe the array response to a unit waveform from the direction θ_i. The vector $\mathbf{a}(\theta_i)$ is scaled so that its norm is independent of θ_i. The d-vector $\mathbf{s}(t)$ is composed of the complex emitter waveforms, received at time t, and the m-vector $\mathbf{n}(t)$ accounts for additive measurement noise as well as modeling errors. The array output is sampled at N distinct time instants, and the data matrix is formed

$$\mathbf{X}_N = \mathbf{A}(\boldsymbol{\theta})\mathbf{S}_N + \mathbf{N}_N \ , \qquad (2)$$

where $\mathbf{X}_N = [\mathbf{x}(1), \ldots, \mathbf{x}(N)]$, $\mathbf{S}_N = [\mathbf{s}(1), \ldots, \mathbf{s}(N)]$, and $\mathbf{N}_N = [\mathbf{n}(1), \ldots, \mathbf{n}(N)]$. Based on these measurements, the problem of interest is to determine the DOAs of all emitters. The number of signals, d, is assumed to be known.

The main interest herein is to ascertain the accuracy of the ML estimates of $\boldsymbol{\theta}$. Two ML approaches have been studied in the literature. In the so-called *deterministic* model, no restrictions are put on the signal waveforms, i.e., they are modeled as arbitrary sequences, [3, 6]. An alternative is the *stochastic* model, where the signals are assumed to be stationary Gaussian random processes. In both models, the noise term, $\{\mathbf{n}(t)\}$ is assumed to be Gaussian. The noise has zero mean and is spatially and temporally white, i.e.,

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

where δ_t is the Kronecker delta and $(\cdot)^*$ is complex conjugate transpose. Both methods are clearly applicable regardless of the actual distribution of the signal waveforms, although they maximize the likelihood function only under the appropriate assumption. We will here study their performance assuming deterministic signals.

A crucial assumption for most DOA estimation methods is that the functional form of $\mathbf{a}(\cdot)$ is available to the user. For simplicity, it will be assumed that the sensors are omnidirectional and lie in the same plane as the impinging waves. To enable unique identification of the DOAs, it is usually assumed that the array manifold has no ambiguities [1], i.e., that for distinct parameters $\theta_1, \ldots, \theta_m$, the matrix $[\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_m)]$ has full rank. For $m = 2$, the physical interpretation of this assumption is that the sidelobes of the array must be lower than the main lobe. Since the size of the array increases in the present case, the sidelobes may approach the main lobe in height as m increases (but still be lower for any finite m). To prevent this, the following condition is imposed on the array geometry.

Assumption A.1. There exists an $\epsilon > 0$ and an M such that

$$\frac{1}{m}\Big[\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_{2d})\Big]^* \Big[\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_{2d})\Big] \geq \epsilon \, \mathbf{I} \ , \qquad (4)$$

for all $m > M$ and for any set of distinct DOAs $\theta_1, \ldots, \theta_{2d}$. The matrix inequality $\mathbf{X} \geq \mathbf{Y}$ means that $\mathbf{X} - \mathbf{Y}$ is positive semidefinite. □

As will be seen, the ML methods does not require any further assumption on the array geometry to give consistent estimates. However, the conventional beamforming method cannot perfectly separate multiple sources unless the steering vectors become orthogonal when m is increased. This is indeed the case for many array geometries of practical interest, including the linear equi-spaced array. For later reference we thus introduce the following.

Assumption A.2.

$$\lim_{m \to \infty} \frac{1}{m} \mathbf{a}^*(\theta)\mathbf{a}(\eta) = \begin{cases} 1 & , \ \theta = \eta \\ 0 & , \ \theta \neq \eta \end{cases} \qquad (5)$$

□

2.2 Estimation Methods

A systematic approach to many parameter estimation problems is the maximum likelihood method. This technique requires a probabilistic setup for the measurements. If the emitter signals are modeled as unknown deterministic quantities, the observation process is distributed as $\mathbf{x}(t) \in \mathrm{N}(\mathbf{A}\mathbf{s}(t), \sigma^2\mathbf{I})$. The unknown parameters in this model are $\boldsymbol{\theta} = \theta_1, \ldots, \theta_d$ as well as $\mathbf{s}(1), \ldots, \mathbf{s}(N)$, σ^2. Introduce the pseudo-inverse of $\mathbf{A}$ and the orthogonal projector onto the nullspace of $\mathbf{A}^*$ by

$$\mathbf{A}^\dagger(\boldsymbol{\theta}) = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^* \qquad (6)$$

$$\mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta}) = \mathbf{I} - \mathbf{A}\mathbf{A}^\dagger = \mathbf{I} - \mathbf{P}_{\mathbf{A}} \ . \qquad (7)$$

The so-called DML estimate of $\boldsymbol{\theta}$ is obtained from, [3, 6],

$$\hat{\boldsymbol{\theta}}_{DML} = \arg\min_{\boldsymbol{\theta}} \mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta})\hat{\mathbf{R}}\} \ , \qquad (8)$$

where $\mathrm{Tr}\{\cdot\}$ denotes the trace and where $\hat{\mathbf{R}}$ is the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}(t)\mathbf{x}^*(t) \ . \qquad (9)$$

A different ML estimate is obtained by instead modeling the emitter signals as stationary, temporally white Gaussian random processes with covariance $\mathbf{S}$. The observation process is now distributed as $\mathbf{x}(t) \in \mathrm{N}(\mathbf{0}, \mathbf{ASA}^* + \sigma^2\mathbf{I})$.

The unknown parameters in this *stochastic* model are $\boldsymbol{\theta}$, $\mathbf{S}$ and σ^2. For $N \geq d$, the SML estimate has been shown, [2] to be

$$\hat{\boldsymbol{\theta}}_{SML} = \arg\min_{\boldsymbol{\theta}} \log \left| \mathbf{A}\hat{\mathbf{S}}(\boldsymbol{\theta})\mathbf{A}^* + \hat{\sigma}^2(\boldsymbol{\theta})\mathbf{I} \right| \quad (10)$$

$$\hat{\mathbf{S}}(\boldsymbol{\theta}) = \mathbf{A}^\dagger(\hat{\mathbf{R}} - \hat{\sigma}^2\mathbf{I})\mathbf{A}^{\dagger *} \quad (11)$$

$$\hat{\sigma}^2(\boldsymbol{\theta}) = \frac{1}{m-d}\mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}^{\perp}\hat{\mathbf{R}}\}, \quad (12)$$

where $|\cdot|$ denotes the determinant.

The exact ML criteria presented above, both suffer from the requirement of a non-linear multidimensional optimization to calculate the estimates. This is particularly prominent when m is large, since the computational cost is at least proportional to m^2. To overcome this difficulty, several suboptimal techniques have been proposed. Perhaps the most natural of these is the traditional "delay-and-sum" beamforming method. This technique can be derived from (8) by setting $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{a}(\theta)$ and searching for d distinct minima in the so-obtained one-dimensional criterion. The beamforming criterion function is thus

$$\mathrm{Tr}\{\mathbf{P}_{\mathbf{a}}^{\perp}(\theta)\hat{\mathbf{R}}\} \propto -\frac{\mathbf{a}^*(\theta)\hat{\mathbf{R}}\mathbf{a}(\theta)}{\mathbf{a}^*(\theta)\mathbf{a}(\theta)} \propto -\mathbf{a}^*(\theta)\hat{\mathbf{R}}\mathbf{a}(\theta) \quad (13)$$

It is easy to show that (13) is in fact also identical to (10) for $d = 1$.

3 Convergence analysis

The analysis of the described methods for large m is different from the large N case in some important respects. A complication is that the steering vectors are often orthogonal for m tending to infinity. This causes a discontinuity in the limiting criterion function. Consider, for example, the beamforming method under Assumption 2. It is easy to show that

$$\lim_{m\to\infty} \frac{1}{m} - \mathbf{a}^*(\theta)\hat{\mathbf{R}}\mathbf{a}(\theta) = \begin{cases} -\hat{\mathbf{S}}_{ii} & , \ \theta = \theta_i \\ 0 & , \ \theta \neq \theta_k \text{ for any } k \end{cases}$$

where $\hat{\mathbf{S}}_{ii} = \frac{1}{N}\sum_{t=1}^{N}|s_i(t)|^2$. Since the limiting criterion function is discontinuous, the convergence of the sequence of cost functions cannot be uniform in θ. This makes it difficult to establish consistency of the estimates, since these do not necessarily converge to the minimizing arguments of the limiting criteria. To overcome this, we shall use a convergence result that will be useful for proving consistency despite the lack of uniform convergence.

Lemma 1 *Let $\hat{\boldsymbol{\theta}}_m$ be obtained by*

$$\hat{\boldsymbol{\theta}}_m = \arg\min_{\boldsymbol{\theta}} V_m(\boldsymbol{\theta}) , \quad (14)$$

where the criterion function consists of a "signal" and a "noise" term

$$V_m(\boldsymbol{\theta}) = f_m(\boldsymbol{\theta}) + \varepsilon_m(\boldsymbol{\theta}) . \quad (15)$$

Assume that these terms satisfy the following conditions

i) $\lim_{m\to\infty} \varepsilon_m(\boldsymbol{\theta}) = 0$, *uniformly in* $\boldsymbol{\theta}$.

ii) there exists a point $\bar{\boldsymbol{\theta}}$ such that for any $\delta > 0$ we can find $\epsilon > 0$ and M, satisfying

$$f_m(\boldsymbol{\theta}) \geq f_m(\bar{\boldsymbol{\theta}}) + \epsilon \quad (16)$$

as soon as $|\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}| > \delta$ and $m > M$, see Figure 1.

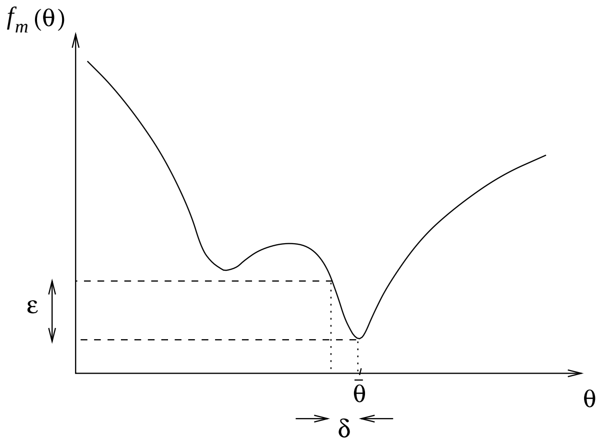

Figure 1: Illustration of condition ii).

Then

$$\lim_{m\to\infty} \hat{\boldsymbol{\theta}}_m = \bar{\boldsymbol{\theta}} . \quad (17)$$

Proof The result is fairly obvious from Figure 1. Without condition *ii)*, the effect of noise could cause a local minimum of $f_m(\theta)$ to become the global minimum of $V_m(\theta)$. To prevent this, the depth of eventual local minima is bounded away from $f_m(\bar{\theta})$ by condition *ii)*. □

Based on the above, let us establish strong consistency of the DML estimate.

Theorem 1 *Under Assumption 1, the DML estimate converges w.p.1 to $\boldsymbol{\theta}_0$ as $m \to \infty$.*

Proof Introduce

$\hat{\mathbf{S}} = N^{-1}\sum_{t=1}^{N}\mathbf{s}(t)\mathbf{s}^*(t) \qquad \hat{\mathbf{Z}} = N^{-1}\sum_{t=1}^{N}\mathbf{s}(t)\mathbf{n}^*(t)$

$\hat{\mathbf{\Sigma}} = N^{-1}\sum_{t=1}^{N}\mathbf{n}(t)\mathbf{n}^*(t)$

and subtract the term $\mathrm{Tr}\{\hat{\mathbf{\Sigma}}\}/m$ from the DML criterion function (8), to yield

$$V_m(\boldsymbol{\theta}) = \frac{1}{m}\mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}^{\perp}\hat{\mathbf{R}}\} - \frac{1}{m}\mathrm{Tr}\{\hat{\mathbf{\Sigma}}\} \ . \qquad (18)$$

By (9) and (3) the sample covariance matrix can be written

$$\hat{\mathbf{R}} = \mathbf{A}(\boldsymbol{\theta}_0)\hat{\mathbf{S}}\mathbf{A}^*(\boldsymbol{\theta}_0) + \mathbf{A}(\boldsymbol{\theta}_0)\hat{\mathbf{Z}} + \hat{\mathbf{Z}}^*\mathbf{A}^*(\boldsymbol{\theta}_0) + \hat{\mathbf{\Sigma}} \ . \ (19)$$

Hence, we can write, $V_m(\boldsymbol{\theta}) = f_m(\boldsymbol{\theta}) + \varepsilon_m(\boldsymbol{\theta})$, where

$$\begin{aligned} f_m(\boldsymbol{\theta}) &= \frac{1}{m}\mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{A}(\boldsymbol{\theta}_0)\hat{\mathbf{S}}\mathbf{A}^*(\boldsymbol{\theta}_0)\} \\ \varepsilon_m(\boldsymbol{\theta}) &= \frac{2}{m}\mathrm{Re}\left(\mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{A}(\boldsymbol{\theta}_0)\hat{\mathbf{Z}}\}\right) + \frac{1}{m}\mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}\hat{\mathbf{\Sigma}}\} \end{aligned}$$

Let us first verify the condition *i)* of Lemma 1. Consider the first term of $\varepsilon_m(\boldsymbol{\theta})$ above. Using the short-hand notation $\mathbf{A}_0 = \mathbf{A}(\boldsymbol{\theta}_0)$ and $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta})$, we have

$$\begin{aligned} &\sup_{\boldsymbol{\theta}}\left|\tfrac{1}{m}\mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{A}_0\hat{\mathbf{Z}}\}\right| \le \left|\mathrm{Tr}\left\{\frac{\hat{\mathbf{Z}}\mathbf{A}_0}{m}\right\}\right| \\ &\quad + \sup_{\boldsymbol{\theta}}\left|\mathrm{Tr}\left\{\frac{\hat{\mathbf{Z}}\mathbf{A}}{m} \times m(\mathbf{A}^*\mathbf{A})^{-1} \times \frac{\mathbf{A}^*\mathbf{A}_0}{m}\right\}\right| \\ &\quad \le C \times \sup_{\boldsymbol{\theta}}\left\|\frac{\hat{\mathbf{Z}}\mathbf{A}}{m}\right\|_F \end{aligned}$$

for some constant C (independent of m). Note that

$$\{\hat{\mathbf{Z}}\mathbf{A}(\boldsymbol{\theta})/m\}_{ij} = \frac{1}{N}\sum_{t=1}^{N}s_i(t) \times \frac{1}{m}\sum_{k=1}^{m}n_k(t)A_{kj}(\boldsymbol{\theta}) \ .$$

The last sum is readily seen to converge to zero w.p.1 as m tends to infinity, for all values of $\boldsymbol{\theta}$, see e.g. [7]. Thus, the first term of $\varepsilon_m(\boldsymbol{\theta})$ converges almost surely to zero, uniformly in $\boldsymbol{\theta}$. By similar arguments it is straightforward to verify that this is true also for the second term.

To verify condition *ii)*, rewrite the signal term as

$$f_m(\boldsymbol{\theta}) = \frac{1}{m}\mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{A}_0\hat{\mathbf{S}}\mathbf{A}_0^*\} = \frac{1}{m}\|\mathbf{A}_0\mathbf{L} + \mathbf{A}\mathbf{T}\|_F^2 \ , \qquad (20)$$

where $\mathbf{L}$ and $\mathbf{T}$ are defined by the relations

$$\hat{\mathbf{S}} = \mathbf{L}\mathbf{L}^* \qquad \mathbf{T} = -\mathbf{A}^{\dagger}\mathbf{A}_0\mathbf{L} \ . \qquad (21)$$

Let k denote the number of components of $\boldsymbol{\theta}$ that are not present in $\boldsymbol{\theta}_0$. If $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \ge \delta > 0$, we have $1 \le k \le d$. Assume, without loss of generality, that the unique DOAs in $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ are $\theta_1, \ldots, \theta_k$ and $\theta_{01}, \ldots, \theta_{0k}$ respectively. Let $\mathbf{l}_i^T$ and $\mathbf{t}_i^T$ denote the rows of $\mathbf{L}$ and $\mathbf{T}$, and introduce the matrices

$$\begin{aligned} \mathbf{Q}^T &= \left[\mathbf{l}_1, .., \mathbf{l}_k, \mathbf{t}_1, .., \mathbf{t}_k, \mathbf{l}_{k+1} + \mathbf{t}_{k+1}, .., \mathbf{l}_d + \mathbf{t}_d\right] \\ \mathbf{B} &= [\mathbf{a}(\theta_{01}), \ldots, \mathbf{a}(\theta_{0k}), \mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_d)] \end{aligned}$$

We can then rewrite (20) as

$$\begin{aligned} f_m(\boldsymbol{\theta}) &= \frac{1}{m}\|\mathbf{B}\mathbf{Q}\|_F^2 = \mathrm{Tr}\{\mathbf{Q}^*\frac{1}{m}\mathbf{B}^*\mathbf{B}\mathbf{Q}\} \\ &\ge \epsilon\mathrm{Tr}\{\mathbf{Q}^*\mathbf{Q}\} \ge \epsilon\hat{\mathbf{S}}_{11} > 0 = f_m(\boldsymbol{\theta}_0) \qquad (22) \end{aligned}$$

In the last expression, $\hat{\mathbf{S}}_{11} = \mathbf{l}_1\mathbf{l}_1^*$ denotes the upper left element of $\hat{\mathbf{S}}$ which is nonzero, and in the first inequality we have used Assumption **A1**. Now, the inequality (22) holds for all m (sufficiently large), which shows that condition ii) of Lemma 1 is satisfied with $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$. The assertion of Theorem 1 follows. □

Using the above technique, the strong consistency of the SML and beamforming methods can also be verified.

Theorem 2 *Under Assumption 1 the SML estimates converge to the true value w.p.1. If Assumption 2 holds, this is also true for the beamforming estimates.*

Proof The idea of the proof for the SML method is to rewrite the normalized criterion function as

$$\log\frac{1}{m-d}\mathrm{Tr}\{\mathbf{P}_{\mathbf{A}}^{\perp}\hat{\mathbf{R}}\} + \frac{1}{m-d}\log|\mathbf{A}^{\dagger}\hat{\mathbf{R}}\mathbf{A}| \qquad (23)$$

The first term obeys condition i) of Lemma 1, by the proof of Theorem 1. It is straightforward to verify that the second term above satisfies condition ii). The proof for the beamforming estimates is similar to the proof for DML. □

It should be noted that although beamforming is a spectral-based method, its consistency for m tending to infinity does not require any assumption of non-coherent sources. This is again an important difference from the large N case. It is well-known that spectral-based techniques loose resolution in the presence of coherent sources. However, the present analysis shows that the resolution is regained by increasing m. In fact, this should not be too surprising; recall the case $N = 1$ and a uniform linear array. This is equivalent to the sinusoids-in-noise problem, for which the DFT (corresponding to beamforming) is known to yield even efficient estimates for large m. When $N = 1$, the "sample signal covariance" matrix is of rank one, which corresponds to perfectly coherent signals.

4 Asymptotic Efficiency

It is well-known that if an ML estimator provides consistent estimates (of all its unknown parameters), it is also asymptotically *efficient*, i.e., the asymptotic covariance of the parameter estimates coincides with the Cramér-Rao bound (CRB). In the deterministic signal model, the unknown parameters are $\boldsymbol{\theta}, \mathbf{S}_N$ and σ^2. We have already shown that $\hat{\boldsymbol{\theta}}_{DML}$ is consistent. For the signal waveforms, we have

$$\hat{\mathbf{S}}_N = \mathbf{A}^{\dagger}(\hat{\boldsymbol{\theta}})\mathbf{X}_N = \mathbf{S}_N + \mathbf{A}^{\dagger}(\hat{\boldsymbol{\theta}})\mathbf{N}_N \to \mathbf{S}_N \ , \qquad (24)$$

since $m(\mathbf{A}^*\mathbf{A})^{-1}$ is bounded and $\mathbf{A}^*\mathbf{N}_N/m$ tends to zero w.p.1. Similarly, it is straightforward to show that also the noise variance estimate is consistent. Hence, the general theory of ML estimation gives the following.

Theorem 3 *Let $\hat{\boldsymbol{\theta}}$ be obtained from (8). Then, asymptotically in m*

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \in \mathrm{AsN}(\mathbf{0}, CRB) \ , \qquad (25)$$

where CRB is the deterministic CRB [4, 8]

$$CRB = \frac{\sigma^2}{2N}\left[\mathrm{Re}\left\{(\mathbf{D}^*\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}) \odot \hat{\mathbf{S}}^T\right\}\right]^{-1} \ , \qquad (26)$$

and where $\odot$ means elementwise multiplication.

The unknown parameters in the stochastic model include the signal covariance, which obviously cannot be consistently estimated unless $N \to \infty$. Hence, one might be tempted to guess that the DML method should outperform the SML method for large m and small N. However, this is not the case.

Theorem 4 *The DML and SML estimates of $\boldsymbol{\theta}$ are asymptotically (for large m) identical.*

Proof Again, the SML criterion function is rewritten as (23). Using a first order Taylor expansion, it can be verified that the second term does not influence the asymptotic (for large m) behaviour of the estimates. □

The asymptotic efficiency of the beamforming method is intuitively clear if the CRB is asymptotically diagonal. In view of (26), this is true if the matrices

$$\mathbf{A}^*\mathbf{A} \ , \quad \mathbf{A}^*\mathbf{D} \ , \quad \mathbf{D}^*\mathbf{D}$$

are all diagonal for large m. If this is not the case, the CRB is still diagonal if $N \gg 1$ and the emitter signals are uncorrelated. It is an open problem under what conditions the diagonality of the matrix $\mathbf{A}^*\mathbf{A}$ implies diagonality of the other matrices above.

5 Conclusions

The consistency of the DML and SML methods as the number of sensors tends to infinity is proved. An "identifiability" condition is introduced, which essentially says that the sidelobes of the array must not approach the level of the main lobe for increasing m. The asymptotic efficiency of the DML method follows from its consistency, and it is shown that the DML and SML estimates are asymptotically equivalent.

It is shown that the traditional beamforming method gives consistent estimates if steering vectors corresponding to different DOAs are asymptotically orthogonal. Since this technique coincides with the DML method for one signal, the beamforming estimates are asymptotically efficient if the CRB is asymptotically diagonal.

Interesting questions that deserve further study include a more precise classification of the set of array configurations for which the matrices

$$\mathbf{A}^*\mathbf{A} \ , \quad \mathbf{A}^*\mathbf{D} \ , \quad \mathbf{D}^*\mathbf{D}$$

(with appropriate normalization) are all asymptotically diagonal.

References

[1] R. O. Schmidt. *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation.* PhD thesis, Stanford Univ., Stanford, CA, Nov. 1981.

[2] J.F. Böhme. "Estimation of Spectral Parameters of Correlated Signals in Wavefields". *Signal Processing*, 10:329–337, 1986.

[3] M. Wax. *Detection and Estimation of Superimposed Signals.* PhD thesis, Stanford Univ., Stanford, CA, March 1985.

[4] P. Stoica and A. Nehorai. "MUSIC, Maximum Likelihood and Cramér-Rao Bound". *IEEE Trans. ASSP*, ASSP-37:720–741, May 1989.

[5] M. Viberg and B. Ottersten. "Sensor Array Processing Based on Subspace Fitting". *IEEE Trans. SP*, SP-39(5):1110–1121, May 1991.

[6] J.F. Böhme. "Estimation of Source Parameters By Maximum Likelihood and Nonlinear Regression". In *Proc. ICASSP 84*, pages 7.3.1–7.3.4, 1984.

[7] W.F. Stout. *Almost Sure Convergence.* Academic Press, New York, 1974.

[8] H. Clergeot, S. Tressens, and A. Ouamri. "Performance of High Resolution Frequencies Estimation Methods Compared to the Cramér-Rao Bounds". *IEEE Trans. ASSP*, 37(11):1703–1720, November 1989.