



Exact and Large Sample ML Techniques for Parameter Estimation and Detection in Array Processing

B. Ottersten, M. Viberg, P. Stoica, A. Nehorai

February 23, 1993

TRITA-SB-9302

ROYAL INSTITUTE
OF TECHNOLOGY
Department of
Signals, Sensors & Systems
Signal Processing
S-100 44 STOCKHOLM

KUNGL TEKNISKA HÖGSKOLAN
Institutionen för
Signaler, Sensorer & System
Signalbehandling
100 44 STOCKHOLM

Chapter 4

Exact and Large Sample ML Techniques for Parameter Estimation and Detection in Array Processing

B. OTTERSTEN¹ M. VIBERG¹ P. STOICA² A. NEHORAI³

To appear in “Radar Array Processing”,
Simon Haykin (ed.), Springer-Verlag.

¹Björn Ottersten and Mats Viberg are with the Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden

²Petre Stoica is with the Department of Control and Computers, Polytechnic Institute of Bucharest, Splaiul Independentei 313, R-77 206 Bucharest, Romania and his work was partially supported by grants from the Swedish Intitute and the Royal Swedish Academy of Engineering Sciences.

³Arye Nehorai is with the Department of Electrical Engineering, Yale University, P.O. Box 2157 Yale Station, New Haven, CT 06520, USA and his work was supported by the AFOSR under Grant No. AFOSR-90-0164 and by the Office of Naval Research.

4.1 Introduction

Sensor array signal processing deals with the problem of extracting information from a collection of measurements obtained from sensors distributed in space. The number of signals present is assumed to be finite, and each signal is parametrized by a finite number of parameters. Based on measurements of the array output, the objective is to estimate the signals and their parameters. This research area has attracted considerable interest for several years. A vast number of algorithms has appeared in the literature for estimating unknown signal parameters from the measured output of a sensor array.

The interest in sensor array processing stems from the many applications where emitter waveforms are measured at several points in space and/or time. In radar applications, as discussed in Chapters 2 and 3, the objective is to determine certain parameters associated with each return signal; these may include bearing, elevation, doppler shift, etc. By using an array of antenna receivers, the accuracy with which the signal parameters can be determined is greatly improved as compared to single-receiver systems.

As the signal environment becomes more dense, both in space and frequency, sensor arrays will play a more significant role in the field of radio and microwave communications. Communication satellites placed in geostationary orbits are faced with an increasing amount of channel interference because of limited space and a limited frequency band. By utilizing the spatial selectivity that may be obtained with an array of sensors, satellites operating in the same frequency band can be placed close together without unacceptable interference. Mobile communication systems are faced with the same problem. An increasing demand for mobile telephones and a limited bandwidth has increased the number of base stations serving an area. If a base station is equipped with a collection of receiving antennas, the spatial dimension can be used to distinguish emitters. By appropriately processing the output of the antenna array, undesirable noise and interference can be suppressed in favor of the signal-of-interest.

Underwater arrays with acoustical sensors (hydrophones) are frequently used in surveillance. Several hydrophones are either towed behind a vessel or dropped in the ocean to passively measure acoustical energy from other vessels. The objective is to detect and estimate incoming signals in order to locate and identify vessels. In geophysical

exploration the ground is often excited by a detonation and waves reflected from boundaries between layers in the earth are sensed by an array of geophones. By identifying the wavefronts arriving at the array, information about the structure of the earth is inferred.

A common feature in the applications mentioned above is that several superimposed wavefields are sampled both in space and in time. The same basic data model is therefore useful in all of these problems. Each application does, however, have aspects that either complicate or simplify the estimation and detection procedures. In an underwater environment, the propagation medium is often quite inhomogeneous and the wavefront is dispersed and degraded. On the other hand, electromagnetic propagation in the atmosphere is often well modeled as homogeneous propagation. An important complication in, for example, radar systems arises from *multipath* propagation, i.e., scenarios in which scaled and delayed versions of the same wavefront impinge on the array from different directions.

It should also be noted that there are strong connections between array signal processing and the harmonic retrieval problem, statistical factor analysis, and identification of linear systems as well.

4.1.1 Background

Classical *beamforming* techniques were the first attempts to utilize an array of sensors to increase antenna aperture and directivity. By forming a weighted sum of the sensor outputs, signals from certain directions are coherently added while incoherently adding the noise and signals from other directions. The antenna is steered to different directions by altering the sensor weights. *Adaptive antennas* were later developed, in which the sensor weights are updated on-line attempting to maximize desired signal energy while suppressing interference and noise. Beamforming and adaptive antenna techniques have been extensively studied in many books and papers, see for example [1]–[8]. Details of implementation are found in Chapter 5 of this book, whereas Chapter 6 is devoted to beamforming in two dimensions (e.g., azimuth and elevation). Chapter 7 presents an application to an imaging system. In the presence of multiple closely spaced sources, beamforming techniques cannot provide consistent parameter estimates. The accuracy of

the estimates is not constrained by the amount of data, but rather by the array geometry and sensor responses, which limit the resolution capabilities of these methods.

The availability of accurate and inexpensive analog to digital converters allows the design of arrays where each sensor output is digitized individually. This greatly expands the signal processing possibilities for the array data. Identification methods adopted from time series analysis were applied to the sensor array processing problem and demonstrated a performance advantage relative to beamforming techniques, [9]–[11].

The introduction of the MUSIC (MUltiple SIgnal Classification) algorithm [12]–[14] (see also [15]), a generalization of Pisarenko’s harmonic retrieval method [16], provided a new geometric interpretation for the sensor array processing problem. This was an attempt to more fully exploit the correct underlying model of the parametrized sensor array problem. The MUSIC data model is formulated using concepts of complex vector spaces, and powerful matrix algebra tools are applied to the problem. Currently, the vector space formulation of the sensor array problem is used extensively. This has resulted in a large number of algorithms, often referred to as *eigenvector* or *subspace* techniques, see e.g. [17]–[20].

The MUSIC algorithm is based on a known parametrization of the array response, which leads to a parametrized model of the sensor array output. When cast in an appropriate statistical framework, the Maximum Likelihood (ML) parameter estimate can be derived for the sensor array problem. ML estimation is a systematic approach to many parameter estimation problems, and ML techniques for the sensor array problem have been studied by a number of researchers, see for example [21]–[27]. Within the vector space formulation, different probabilistic models have been used for the emitter signals. Performance bounds (Cramér-Rao bounds) on the estimation error covariance can be derived based on the different model assumptions, see [21] and [28]–[32].

Unfortunately, the ML estimation method generally requires a multidimensional non-linear optimization at a considerable computational cost. Reduced computational complexity is usually achieved by use of a suboptimal estimator, in which case the quality of the estimate is an issue. Determining what is an accurate estimator is in general a very difficult task. However, an analysis is possible, for example, when assuming a large amount of data. Although asymptotic in character, these results are often useful for

algorithm comparison as well as for predicting estimation accuracy in realistic scenarios. Therefore, there has recently been a great interest in analytic performance studies of array processing algorithms. The MUSIC algorithm, because of its generality and popularity, has received the most attention, see e.g. [25, 30, 31] and [33]–[37]. Analysis of other subspace based algorithms is found in, for example [27, 32] and [38]–[45].

4.1.2 Chapter Outline

This chapter attempts to establish a connection between the classical ML estimation techniques and the more recent eigenstructure or subspace based methods. The estimators are related in terms of their asymptotic behavior and also compared to performance bounds. Computational schemes for calculating the estimates are also presented.

An optimal subspace based technique termed WSF (weighted subspace fitting) is discussed in some more detail. The detection problem is also addressed and a hypothesis testing scheme, based on the WSF criterion, is presented. The estimation and detection methods discussed herein are applicable to arbitrary array configurations and emitter signals. The case of highly correlated or even coherent sources (specular multipath), which is particularly relevant in radar applications, is given special attention.

The chapter is organized as follows: Section 4.2 presents a data model and the underlying assumptions. In Section 4.3, two probabilistic models are formulated as well as the corresponding ML estimators. Multidimensional subspace based methods, their asymptotic properties, and the relation to the ML estimators is the topic of Section 4.4. In Section 4.5, a Newton type algorithm is applied to the various cost functions and its computational complexity is discussed. Section 4.6 presents schemes for detecting the number of signals. The performance of the estimation and detection procedures is examined through numerical examples and simulations in Section 4.7.

4.2 Sensor Array Processing

This section presents the mathematical model of the array output and introduces the basic assumptions. Consider the scenario of Figure 4.1. An array of m sensors arranged

Figure 4.1: *A passive sensor array receiving emitter signals from point sources.*

in an arbitrary geometry, receives the waveforms generated by d point sources. The output of each sensor is modeled as the response of a linear time-invariant system. Let $h_{ki}(t)$ be the impulse response of the k^{th} sensor to a signal $\bar{s}_i(t)$ impinging on the array. The impulse response depends on the physical antenna structure, the receiver electronics, other antennas in the array through mutual coupling, as well as the signal parameters. The time delay of the i^{th} signal at the k^{th} array element, relative to some fixed reference point, is denoted τ_{ki} . The output of the k^{th} array element can then be written as a superposition

$$\bar{x}_k(t) = \sum_{i=1}^d h_{ki}(t) * \bar{s}_i(t - \tau_{ki}) + \bar{n}_k(t) , \quad (4.1)$$

where $(*)$ denotes convolution and $\bar{n}_k(t)$ is an additive noise term independent of $\bar{s}_i(t)$.

The sensor outputs are collected in the m -vector

$$\bar{\mathbf{x}}(t) = \begin{bmatrix} \bar{x}_1(t) \\ \vdots \\ \bar{x}_m(t) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^d h_{1i} * \bar{s}_i(t - \tau_{1i}) \\ \vdots \\ \sum_{i=1}^d h_{mi} * \bar{s}_i(t - \tau_{mi}) \end{bmatrix} + \begin{bmatrix} \bar{n}_1(t) \\ \vdots \\ \bar{n}_m(t) \end{bmatrix} . \quad (4.2)$$

4.2.1 Narrowband Data Model

Consider an emitter signal $\bar{s}(t)$ and express this signal in terms of a center frequency ω

$$\bar{s}(t) = \alpha(t) \cos(\omega t + \phi(t)) . \quad (4.3)$$

If the amplitude, $\alpha(t)$, and phase, $\phi(t)$, of the signal vary slowly relative to the propagation time across the array τ , i.e., if

$$\alpha(t - \tau) \approx \alpha(t) \quad \phi(t - \tau) \approx \phi(t) , \quad (4.4)$$

the signal is said to be *narrowband*. The narrowband assumption implies that

$$\bar{s}(t - \tau) = \alpha(t - \tau) \cos(\omega(t - \tau) + \phi(t - \tau)) \approx \alpha(t) \cos(\omega t - \omega\tau + \phi(t)) . \quad (4.5)$$

In other words, the narrowband assumption on $\bar{s}(t)$ allows the time delay of the signal to be modeled as a simple phase shift of the carrier frequency. Now, the stationary response of the k^{th} sensor to $\bar{s}(t)$, may be expressed as

$$\begin{aligned} \bar{x}_k(t) &= h_k(t) * \bar{s}(t - \tau_k) = h_k(t) * \alpha(t) \cos(\omega t - \omega\tau_k + \phi(t)) \\ &\simeq |H_k(\omega)| \alpha(t) \cos(\omega t - \omega\tau_k + \phi(t) + \arg H_k(\omega)) , \end{aligned} \quad (4.6)$$

where $H_k(\omega)$ is the Fourier transform of the impulse response $h_k(t)$. The narrowband assumption is implicitly used also here. It is assumed that the support of $\bar{s}(t)$ in the frequency domain, is small enough to model the receiver response, H_k , as constant over this range.

It is notationally convenient to adopt a complex signal representation for $\bar{x}_k(t)$. The time delay of the narrowband signal is then expressed as a multiplication by a complex number. The *complex envelope* of the noise-free signal has the form

$$\begin{aligned} x_k(t) &= x_k^c(t) + jx_k^s(t) \\ &= H_k(\omega) e^{-j\omega\tau_k} \alpha(t) e^{j\phi(t)} \\ &= H_k(\omega) e^{-j\omega\tau_k} s(t) , \end{aligned} \quad (4.7)$$

where signal $s(t)$ is the complex envelope of $\bar{s}(t)$ and

$$x_k^c(t) = |H_k(\omega)|\alpha(t) \cos(\phi(t) + \arg H_k(\omega) - \omega\tau_k) \quad (4.8)$$

$$x_k^s(t) = |H_k(\omega)|\alpha(t) \sin(\phi(t) + \arg H_k(\omega) - \omega\tau_k) \quad (4.9)$$

are the low-pass *in-phase* and *quadrature* components of $\bar{x}_k(t)$, respectively. In practice, these are generated by a *quadrature detector* in which the signal is multiplied by $\sin(\omega t)$ and $\cos(\omega t)$ and low-passed filtered

$$x_k^c(t) = [2\bar{x}_k(t) \cos(\omega t)]_{\text{LP}} \quad (4.10)$$

$$x_k^s(t) = -[2\bar{x}_k(t) \sin(\omega t)]_{\text{LP}} . \quad (4.11)$$

When the narrowband assumption does not hold, temporal filtering of the signals is required for the model (4.7) to be valid.

4.2.2 Parametric Data Model

We will now discuss the *parametrized* model that forms the basis for the later developments. A collection of unknown parameters is associated with each emitter signal. These parameters may include bearing, elevation, range, polarization, carrier frequency, etc.. The p parameters associated with the i^{th} signal are collected in the parameter vector $\boldsymbol{\theta}_i$. The k^{th} sensor's response and the time delay of propagation for the i^{th} signal are denoted by, $H_k(\boldsymbol{\theta}_i)$ and $\tau_k(\boldsymbol{\theta}_i)$ respectively. The following *parametrized* data model is obtained (from (4.2) and (4.7))

$$\begin{aligned} \mathbf{x}(t) &= \begin{bmatrix} x_1(t) \\ \vdots \\ x_m(t) \end{bmatrix} = \sum_{i=1}^d \begin{bmatrix} a_1(\boldsymbol{\theta}_i) \\ \vdots \\ a_m(\boldsymbol{\theta}_i) \end{bmatrix} s_i(t) + \begin{bmatrix} n_1(t) \\ \vdots \\ n_m(t) \end{bmatrix} \\ &= [\mathbf{a}(\boldsymbol{\theta}_1) \dots \mathbf{a}(\boldsymbol{\theta}_d)] [s_1(t) \dots s_d(t)]^T + \mathbf{n}(t) \\ &= \mathbf{A}(\boldsymbol{\theta}_0) \mathbf{s}(t) + \mathbf{n}(t) , \end{aligned} \quad (4.12)$$

where the response of the k^{th} sensor to the i^{th} signal is $a_k(\boldsymbol{\theta}_i) = H_k(\boldsymbol{\theta}_i)e^{-j\omega\tau_k(\boldsymbol{\theta}_i)}$. The vector $\mathbf{x}(t)$ belongs to an m -dimensional complex vector space, $\mathbf{x}(t) \in \mathcal{C}^{m \times 1}$. The pd -vector of the “true signal parameters” is denoted $\boldsymbol{\theta}_0 = [\theta_{11}, \dots, \theta_{1p}, \dots, \theta_{d1}, \dots, \theta_{dp}]^T$.

To illustrate the parametrization of the array response, consider a uniform linear array (ULA) with identical sensors and uniform spacing Δ . Assume that the sources are in the far field of the array and that the medium is non-dispersive, so that the wavefronts can be approximated as planar. Then, the parameter of interest is the direction of arrival (DOA) of the wavefronts, θ , measured relative to the normal of the array. The propagation delay τ between two adjacent sensors is related to the DOA by the following equation

$$\sin \theta = \frac{c\tau}{\Delta}, \quad (4.13)$$

where c is the wave propagation velocity. Hence, the array response vector is given by

$$\mathbf{a}(\theta) = a(\theta) \begin{bmatrix} 1 \\ e^{-j\frac{\omega\Delta \sin \theta}{c}} \\ \vdots \\ e^{-j(m-1)\frac{\omega\Delta \sin \theta}{c}} \end{bmatrix}, \quad (4.14)$$

where $a(\theta)$ is the response of the first element. Vectors with this special structure are commonly referred to as *Vandermonde* vectors.

The *array response* vector, $\mathbf{a}(\boldsymbol{\theta}_i)$, is an element of a complex m -dimensional vector space and will, in general, describe a p -dimensional manifold parametrized by $\boldsymbol{\theta}_i$. In the following, we will at times specialize to the case when each emitter signal has one unknown parameter, e.g., the direction of arrival. In this case, the array response vector traces a one-dimensional manifold in the space as the angle, θ , varies over the parameter range of interest, see Figure 4.2. The array manifold, \mathcal{A} , is defined as the collection of all array response vectors over the parameter range of interest.

$$\mathcal{A} = \{\mathbf{a}(\boldsymbol{\theta}_i) \mid \boldsymbol{\theta}_i \in \Theta\}. \quad (4.15)$$

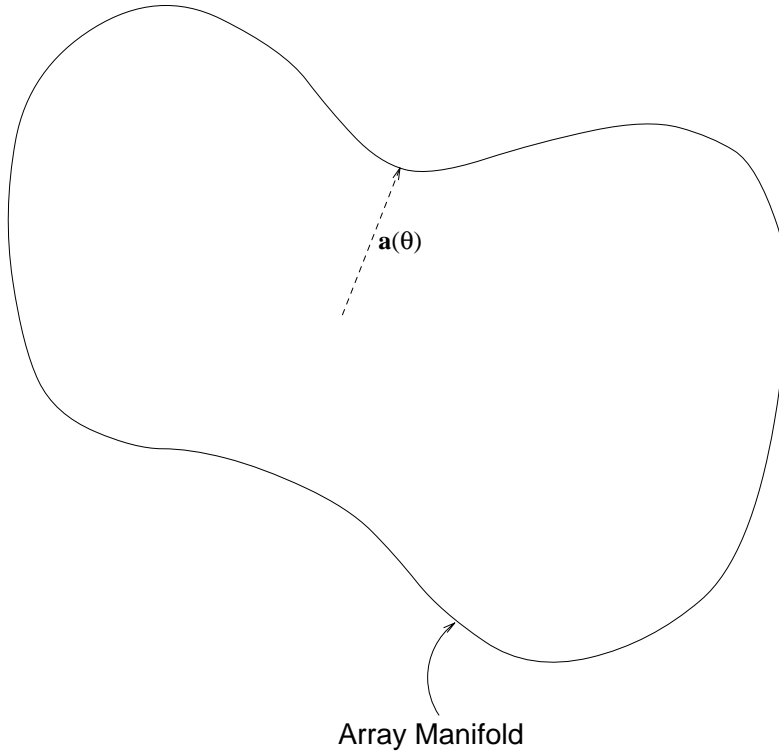


Figure 4.2: *Array manifold.*

Let the process $\mathbf{x}(t)$ be observed at N time instants, $\{t_1, \dots, t_N\}$. Each vector observation is called a *snapshot* of the array output and the *data matrix* is the collection of the array snapshots

$$\mathbf{X}_N = [\mathbf{x}(t_1) \dots \mathbf{x}(t_N)] = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{S}_N + \mathbf{N}_N, \quad (4.16)$$

where the matrices \mathbf{S}_N and \mathbf{N}_N are defined analogously to \mathbf{X}_N .

4.2.3 Assumptions and Problem Formulation

The received signal waveforms are assumed to be narrowband, see (4.4). When deriving the ML estimator, the following assumptions regarding the statistical properties of the data are further imposed. The signal, $\mathbf{s}(t_i)$, and noise, $\mathbf{n}(t_i)$, terms are independent, zero-mean, complex, Gaussian random processes with second-order moments

$$\mathbb{E}\{\mathbf{s}(t_i)\mathbf{s}^H(t_j)\} = \mathbf{S} \delta_{ij} \quad \mathbb{E}\{\mathbf{n}(t_i)\mathbf{n}^H(t_j)\} = \sigma^2 \mathbf{I} \delta_{ij} \quad (4.17)$$

$$\mathbf{E}\{\mathbf{s}(t_i)\mathbf{s}^T(t_j)\} = \mathbf{0} \quad \mathbf{E}\{\mathbf{n}(t_i)\mathbf{n}^T(t_j)\} = \mathbf{0} , \quad (4.18)$$

where $(\cdot)^H$ denotes complex conjugate transpose, \mathbf{S} is the unknown signal covariance matrix, δ_{ij} represents the Kronecker delta and \mathbf{I} is the identity matrix. In many applications, the assumption of Gaussian emitter signals is not realistic. A common alternative model assumes that the emitter signals are unknown deterministic wave forms, see Section 3.6. In the next section, the deterministic (or conditional) model is discussed and strong justification is provided for why the Gaussian model, which we refer to as the stochastic model, is appropriate for the signal parameter estimation problem at hand. The detection and estimation schemes derived from the Gaussian model is found to yield superior performance, *regardless of the actual emitter signals*.

As seen from (4.17), the noise is assumed to be *temporally white* and also independent from sensor to sensor, i.e., *spatially white*. The noise power is assumed to be identical in all sensors, and its value, σ^2 , is unknown. If the spatial whiteness condition is not met, the covariance matrix of the noise must be known, e.g. from measurements with no signals present. In such a case, a *whitening* transformation is performed on the data, rendering the transformed noise to be spatially white as described in Section 3.5.

Under the assumptions above the covariance matrix of the measured array output, referred to as the *array covariance*, takes the following familiar form

$$\mathbf{R} = \mathbf{E}\{\mathbf{x}(t_i)\mathbf{x}^H(t_i)\} = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{S}\mathbf{A}^H(\boldsymbol{\theta}_0) + \sigma^2\mathbf{I} . \quad (4.19)$$

A critical assumption for parameter based array processing techniques is that the functional form of the parametrization of \mathcal{A} is known. If the array is carefully designed, deriving an analytic expression for the array response may be tractable. However, in most practical applications only array calibration data are available, and the issue of generating an appropriate array manifold is crucial. It is further required that the manifold vectors are continuously differentiable w.r.t. the parameters, and that for any collection of m distinct $\boldsymbol{\theta}_i \in \Theta$, the matrix $[\mathbf{a}(\boldsymbol{\theta}_1), \dots, \mathbf{a}(\boldsymbol{\theta}_m)]$ has full rank. An array satisfying the latter assumption is said to be *unambiguous*. Due to the Vandermonde structure of $\mathbf{a}(\boldsymbol{\theta})$ in the ULA case, it is simple to show that the ULA is unambiguous if the parameter set is $\Theta = (-\pi/2, \pi/2)$.

With these preliminaries the sensor array problem can now be formulated as follows:

Given the observations \mathbf{X}_N and a model for the array response $\mathbf{a}(\boldsymbol{\theta}_i)$, estimate the number of incoming signals d , the signal parameters $\boldsymbol{\theta}_0$, the signal covariance matrix \mathbf{S} (or alternatively the waveforms \mathbf{S}_N) and the noise variance σ^2 .

The emphasis here is on the estimation of d and $\boldsymbol{\theta}_0$. Estimation of the other unknowns is only briefly discussed. Detailed treatments on the estimation of \mathbf{S} is given in [24] and [46], whereas estimation of \mathbf{S}_N is the subject of [47].

4.2.4 Parameter Identifiability

Under the assumption of independent Gaussian distributed observations, all information in the measured data is contained in the second-order moments. The question of parameter identifiability is thus reduced to investigating under what conditions the array covariance uniquely determines the signal parameters. Let $\boldsymbol{\eta}$ represent the unknown parameters of the covariance \mathbf{R} . It is assumed that no *a priori* information on the signal covariance is available. Noting that \mathbf{S} is a Hermitian matrix, $\boldsymbol{\eta}$ contains the following $d^2 + pd + 1$ real parameters

$$\boldsymbol{\eta} = [\theta_{11}, \dots, \theta_{dp}, s_{11}, \dots, s_{dd}, \bar{s}_{21}, \tilde{s}_{21}, \dots, \bar{s}_{d,d-1}, \tilde{s}_{d,d-1}, \sigma^2]^T, \quad (4.20)$$

where $\bar{s}_{ij} = \text{Re}\{s_{ij}\}$ and $\tilde{s}_{ij} = \text{Im}\{s_{ij}\}$. Let \mathbf{R}_1 and \mathbf{R}_2 be two covariance matrices associated with the parameter vectors $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$, and let $\mathbf{A}_1, \mathbf{A}_2$ and $\mathbf{S}_1, \mathbf{S}_2$ be the corresponding response matrices and emitter covariances. We will distinguish between *system identifiability* (SI)

$$\mathbf{R}_1 = \mathbf{R}_2 \iff (\mathbf{A}_1 = \mathbf{A}_2 \text{ and } \mathbf{S}_1 = \mathbf{S}_2), \quad (4.21)$$

and *unambiguous parametrization* (UP)

$$(\mathbf{A}_1 = \mathbf{A}_2 \text{ and } \mathbf{S}_1 = \mathbf{S}_2) \iff \boldsymbol{\eta}_1 = \boldsymbol{\eta}_2. \quad (4.22)$$

The problem is parameter identifiable (PI), i.e.,

$$\mathbf{R}_1 = \mathbf{R}_2 \iff \boldsymbol{\eta}_1 = \boldsymbol{\eta}_2 \quad (4.23)$$

if and only if it is both SI and UP. Let \mathbf{A}_1 and \mathbf{A}_2 be two $m \times d$ matrices of full rank, whereas \mathbf{S}_1 and \mathbf{S}_2 are arbitrary $d \times d$ matrices. Consider the relation

$$\mathbf{A}_1 \mathbf{S}_1 \mathbf{A}_1^H + \sigma_1^2 \mathbf{I} = \mathbf{A}_2 \mathbf{S}_2 \mathbf{A}_2^H + \sigma_2^2 \mathbf{I} . \quad (4.24)$$

If the number of signals is less than the dimension of $\mathbf{x}(t_i)$, i.e. $d < m$, then (4.24) implies $\sigma_1^2 = \sigma_2^2$ since the smallest eigenvalue of the two covariances must be equal. It also follows that the matrices \mathbf{S}_1 and \mathbf{S}_2 have the same rank, denoted d' . In the case of coherent signals (specular multipath or “smart jamming”), the emitter covariance matrix is singular, i.e., d' is strictly less than d . Let

$$\mathbf{S}_i = \mathbf{L}_i \mathbf{L}_i^H, \quad i = 1, 2 \quad (4.25)$$

denote the Cholesky factorization of \mathbf{S}_i , where \mathbf{L}_i , $i = 1, 2$ are $d \times d'$ matrices of full rank. Clearly, (4.24) is equivalent to

$$\mathbf{A}_1 \mathbf{L}_1 = \mathbf{A}_2 \mathbf{L}_2 \mathbf{L}_3 , \quad (4.26)$$

for some $d' \times d'$ unitary matrix \mathbf{L}_3 . In [48], it is shown that a sufficient condition for (4.26) to imply $\mathbf{A}_1 = \mathbf{A}_2$ (subject to the order of the columns) and, hence, $\mathbf{S}_1 = \mathbf{S}_2$ is that

$$d < (m + d')/2 . \quad (4.27)$$

By assuming an unambiguous array, (4.22) is trivially satisfied and we conclude that the problem is PI. Notice that the requirement that $[\mathbf{a}(\boldsymbol{\theta}_1), \dots, \mathbf{a}(\boldsymbol{\theta}_m)]$ has full rank for distinct signal parameters is problem-dependent and, therefore, has to be established for the specific array under study.

It has been recently shown in [49] that the condition (4.27) is not only sufficient but also essentially necessary for PI. Note that if (4.27) is replaced by the weaker condition

$d < 2d'm/(2d' + p)$, then PI is guaranteed except for a set of parameter values having zero measure, see [48]. Thus, the latter condition guarantees what is commonly called generic parameter identifiability. However, the concept of generic identifiability is of limited practical value, since for a non-zero measure set of scenarios with parameters close to the non-identifiable set, the accuracy of any parameter estimator will be very poor, [49]. Hence, we will assume in the following that (4.27) is satisfied.

4.3 Exact Maximum Likelihood Estimation

The Maximum Likelihood (ML) method is a standard technique in statistical estimation theory. To apply the ML method, the *likelihood function* (see e.g. [50, 51]) of the observed data has to be specified. The ML estimates are calculated as the values of the unknown parameters that maximize the likelihood function. This can be interpreted as selecting the set of parameters that make the observed data most probable.

When applying the ML technique to the sensor array problem, two main methods have been considered, depending on the model assumption on the signal waveforms. When the emitter signals are modeled as Gaussian random processes, a *stochastic* ML (SML) method is obtained, see also Sections 2.4.4 and 3.6. If on the other hand the emitter signals are modeled as unknown deterministic quantities, the resulting estimator is referred to as the *deterministic* (or conditional) ML (DML) estimator, see also Sections 2.3.3 and 3.6.

4.3.1 The Stochastic Maximum Likelihood Method

In many applications it is appropriate to model the signals as stationary stochastic processes, possessing a certain probability distribution. The by far most commonly advocated distribution is the Gaussian one. Not only is this for the mathematical convenience of the resulting approach, but the Gaussian assumption is also often motivated by the Central Limit Theorem.

Under the assumptions of Section 4.2.3, the observation process, $\mathbf{x}(t_i)$, constitutes a stationary, zero-mean Gaussian random process having second-order moments

$$\mathbb{E}\{\mathbf{x}(t_i)\mathbf{x}^H(t_j)\} = \mathbf{R} \delta_{ij} = (\mathbf{A}(\boldsymbol{\theta})\mathbf{S}\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2 \mathbf{I}) \delta_{ij} \quad (4.28)$$

$$\mathbb{E}\{\mathbf{x}(t_i)\mathbf{x}^T(t_j)\} = \mathbf{0} . \quad (4.29)$$

In most applications, no a-priori information on the signal covariance matrix is available. Since \mathbf{S} is a Hermitian matrix, it can be uniquely parametrized by d^2 real parameters, namely the real diagonal elements and the real and imaginary parts of the lower (or

upper) off-diagonal entries, cf. (4.20). Other possible assumptions are that \mathbf{S} is completely known or unknown but diagonal (uncorrelated signals). Herein, $\boldsymbol{\theta}$, \mathbf{S} and σ^2 are all considered to be completely unknown, resulting in a total of $d^2 + pd + 1$ unknown parameters.

The likelihood function of a single observation, $\mathbf{x}(t_i)$, is

$$p_i(\mathbf{x}) = \frac{1}{\pi^m |\mathbf{R}|} e^{-\mathbf{x}^H \mathbf{R}^{-1} \mathbf{x}}, \quad (4.30)$$

where $|\mathbf{R}|$ denotes the determinant of \mathbf{R} . This is the *complex m -variate* Gaussian distribution, see e.g. [52]. Since the snapshots are independent and identically distributed, the likelihood of the complete data set $\mathbf{x}(t_1), \dots, \mathbf{x}(t_N)$ is given by

$$p(\mathbf{x}(t_1), \dots, \mathbf{x}(t_N) | \boldsymbol{\theta}, \mathbf{S}, \sigma^2) = \prod_{i=1}^N \frac{1}{\pi^m |\mathbf{R}|} e^{-\mathbf{x}^H(t_i) \mathbf{R}^{-1} \mathbf{x}(t_i)}. \quad (4.31)$$

Maximizing $p(\boldsymbol{\theta}, \sigma^2, \mathbf{S})$ is equivalent to minimizing the negative *log-likelihood function*,

$$\begin{aligned} -\log(p(\boldsymbol{\theta}, \sigma^2, \mathbf{S})) &= -\sum_{i=1}^N \log \left[\frac{1}{\pi^m |\mathbf{R}|} e^{-\mathbf{x}^H(t_i) \mathbf{R}^{-1} \mathbf{x}(t_i)} \right] \\ &= mN \log \pi + N \log |\mathbf{R}| + \sum_{i=1}^N \mathbf{x}^H(t_i) \mathbf{R}^{-1} \mathbf{x}(t_i). \end{aligned} \quad (4.32)$$

Ignoring the constant term and normalizing by N , the SML estimate is obtained by solving the following optimization problem,

$$[\hat{\boldsymbol{\theta}}, \hat{\mathbf{S}}, \hat{\sigma}^2] = \arg \min_{\boldsymbol{\theta}, \mathbf{S}, \sigma^2} l(\boldsymbol{\theta}, \mathbf{S}, \sigma^2), \quad (4.33)$$

where the criterion function is defined as

$$l(\boldsymbol{\theta}, \mathbf{S}, \sigma^2) = \log |\mathbf{R}| + \frac{1}{N} \sum_{i=1}^N \mathbf{x}^H(t_i) \mathbf{R}^{-1} \mathbf{x}(t_i). \quad (4.34)$$

By well-known properties of the trace operator⁴ $\text{Tr}\{\cdot\}$, the normalized negative log-likelihood function can be expressed as

$$l(\boldsymbol{\theta}, \mathbf{S}, \sigma^2) = \log |\mathbf{R}| + \text{Tr}\{\mathbf{R}^{-1} \hat{\mathbf{R}}\}, \quad (4.35)$$

⁴For a scalar a , $\text{Tr}\{a\} = a$, and for matrices \mathbf{A} and \mathbf{B} of appropriate dimensions, $\text{Tr}\{\mathbf{A}\mathbf{B}\} = \text{Tr}\{\mathbf{B}\mathbf{A}\}$ and $\text{Tr}\{\mathbf{A}\} + \text{Tr}\{\mathbf{B}\} = \text{Tr}\{\mathbf{A} + \mathbf{B}\}$.

where $\hat{\mathbf{R}}$ is the sample covariance

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}(t_i) \mathbf{x}^H(t_i) . \quad (4.36)$$

With some algebraic effort the ML criterion function can be concentrated with respect to \mathbf{S} and σ^2 [24, 46, 53], thus, reducing the dimension of the required numerical optimization to pd . The SML estimates of the signal covariance matrix and the noise power are obtained by inserting the SML estimates of $\boldsymbol{\theta}$ in the following expressions

$$\hat{\mathbf{S}}(\boldsymbol{\theta}) = \mathbf{A}^\dagger(\boldsymbol{\theta})(\hat{\mathbf{R}} - \hat{\sigma}^2(\boldsymbol{\theta}) \mathbf{I})\mathbf{A}^{\dagger H}(\boldsymbol{\theta}) \quad (4.37)$$

$$\hat{\sigma}^2(\boldsymbol{\theta}) = \frac{1}{m-d} \text{Tr}\{\mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta})\hat{\mathbf{R}}\} , \quad (4.38)$$

where \mathbf{A}^\dagger is the pseudo-inverse of \mathbf{A} and $\mathbf{P}_{\mathbf{A}}^\perp$ is the orthogonal projector onto the null space of \mathbf{A}^H , i.e.

$$\mathbf{A}^\dagger = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \quad (4.39)$$

$$\mathbf{P}_{\mathbf{A}} = \mathbf{A} \mathbf{A}^\dagger \quad (4.40)$$

$$\mathbf{P}_{\mathbf{A}}^\perp = \mathbf{I} - \mathbf{P}_{\mathbf{A}} . \quad (4.41)$$

The concentrated form of the SML criterion is now obtained by substituting (4.37)–(4.38) into (4.35). The signal parameter estimates are obtained by solving the following optimization problem

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} V_{SML}(\boldsymbol{\theta}) \quad (4.42)$$

$$V_{SML}(\boldsymbol{\theta}) = \log \left| \mathbf{A}(\boldsymbol{\theta}) \hat{\mathbf{S}}(\boldsymbol{\theta}) \mathbf{A}^H(\boldsymbol{\theta}) + \hat{\sigma}^2(\boldsymbol{\theta}) \mathbf{I} \right| . \quad (4.43)$$

Remark 4.1 It is possible to include the obvious a-priori information that \mathbf{S} is positive semidefinite. Since (4.37) may be indefinite, this yields potentially a different ML estimator, see [46]. In general, if the rank of \mathbf{S} is known to be d' , a different parametrization should be employed, for instance, $\mathbf{S} = \mathbf{L}\mathbf{L}^H$, where \mathbf{L} is a $d \times d'$ “lower triangular”

matrix. If $d' = d$, these possible modifications will have no effect for “large enough N ”, since (4.37) is a consistent estimate of \mathbf{S} . Even if $d' < d$, it can be shown that the asymptotic (for large N) statistical properties of the SML estimate cannot be improved by the square-root parametrization. Since the latter leads to a significantly more complicated optimization problem, the unrestricted parametrization of \mathbf{S} appears to be preferable. We will therefore always refer to the minimizer of (4.43) as being the SML estimate of the signal parameters. \square

Although the dimension of the parameter space is reduced substantially, the form of the resulting criterion function (4.43) is complicated and the minimizing $\boldsymbol{\theta}$ can, in general, not be found analytically. In Section 4.5, a numerical procedure is described for carrying out the required optimization.

4.3.2 The Deterministic ML Method

In some applications, for example radar and radio communication, the signal waveforms are often far from being Gaussian random variables. The deterministic model is then a natural one, since it makes no assumptions at all on the signals. Instead $\mathbf{s}(t_i)$, $i = 1, \dots, N$ are regarded as unknown parameters that have to be estimated. In fact, in some applications, such as microwave communications, estimation of $\mathbf{s}(t_i)$ is of more interest than estimation of $\boldsymbol{\theta}$. The ML estimator for this model is termed the DML method.

Similar to the stochastic signal model, the observation process, $\mathbf{x}(t_i)$, is Gaussian distributed given the unknown quantities. The first- and second-order moments are different though:

$$\mathbf{E}\{\mathbf{x}(t_i)\} = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t_i) \quad (4.44)$$

$$\mathbf{E}\{(\mathbf{x}(t_i) - \mathbf{E}\{\mathbf{x}(t_i)\})(\mathbf{x}(t_j) - \mathbf{E}\{\mathbf{x}(t_j)\})^H\} = \sigma^2 \mathbf{I} \delta_{ij} \quad (4.45)$$

$$\mathbf{E}\{(\mathbf{x}(t_i) - \mathbf{E}\{\mathbf{x}(t_i)\})(\mathbf{x}(t_j) - \mathbf{E}\{\mathbf{x}(t_j)\})^T\} = \mathbf{0} . \quad (4.46)$$

The unknown parameters are in this case, $\boldsymbol{\theta}$, $\mathbf{s}(t_i)$, $i = 1, \dots, N$, and σ^2 .

The joint probability distribution of the observations is formed by conditioning on the parameters of the deterministic model: the signal parameters, the noise variance,

and the waveforms. As the snapshots are independent, the conditional density is given by

$$p(\mathbf{x}(t_1) \dots \mathbf{x}(t_N) | \boldsymbol{\theta}, \sigma^2, \mathbf{S}_N) = \prod_{i=1}^N \frac{1}{|\pi \sigma^2 \mathbf{I}|} e^{-\sigma^{-2} (\mathbf{x}(t_i) - \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}(t_i))^H (\mathbf{x}(t_i) - \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}(t_i))} , \quad (4.47)$$

and the negative log likelihood function has the following form

$$\begin{aligned} -\log(p(\boldsymbol{\theta}, \sigma^2, \mathbf{S}_N)) &= Nm \log(\pi \sigma^2) + \sigma^{-2} \text{Tr} \{ (\mathbf{X}_N - \mathbf{A}(\boldsymbol{\theta}) \mathbf{S}_N)^H (\mathbf{X}_N - \mathbf{A}(\boldsymbol{\theta}) \mathbf{S}_N) \} \\ &= Nm \log(\pi \sigma^2) + \sigma^{-2} \|\mathbf{X}_N - \mathbf{A}(\boldsymbol{\theta}) \mathbf{S}_N\|_F^2 , \end{aligned} \quad (4.48)$$

where $\|\cdot\|_F$ is the Frobenius norm⁵ of a matrix, and \mathbf{X}_N and \mathbf{S}_N are defined in (4.16). The deterministic maximum likelihood estimates are the minimizing arguments of (4.48). For fixed $\boldsymbol{\theta}$ and \mathbf{S}_N , the minimum with respect to σ^2 is readily derived as

$$\hat{\sigma}^2 = \frac{1}{m} \text{Tr} \{ \mathbf{P}_\mathbf{A}^\perp(\boldsymbol{\theta}) \hat{\mathbf{R}} \} . \quad (4.49)$$

Substituting (4.49) into (4.48) shows that $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{S}}_N$ are obtained by solving the non-linear least-squares problem

$$[\hat{\boldsymbol{\theta}}, \hat{\mathbf{S}}_N] = \arg \min_{\boldsymbol{\theta}, \mathbf{S}_N} \|\mathbf{X}_N - \mathbf{A}(\boldsymbol{\theta}) \mathbf{S}_N\|_F^2 . \quad (4.50)$$

Since the above criterion function is quadratic in the signal waveform parameters, it is easy to minimize with respect to \mathbf{S}_N , see [22, 23, 54]. This results in the following estimates

$$\hat{\mathbf{S}}_N = \mathbf{A}^\dagger(\boldsymbol{\theta}) \mathbf{X}_N \quad (4.51)$$

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} V_{DML}(\boldsymbol{\theta}) \quad (4.52)$$

$$V_{DML}(\boldsymbol{\theta}) = \text{Tr} \{ \mathbf{P}_\mathbf{A}^\perp(\boldsymbol{\theta}) \hat{\mathbf{R}} \} . \quad (4.53)$$

⁵The Frobenius norm of a matrix is given by the square root of the sum of the squared moduli of the elements.

Comparing (4.43) and (4.53), we see that the DML criterion depends on $\boldsymbol{\theta}$ in a simpler way than does the SML criterion. It is, however, important to note that also (4.52)–(4.53) is a non-linear multidimensional minimization problem, and the criterion function often possesses a large number of local minima.

4.3.3 Bounds on the Estimation Accuracy

In any practical application involving the estimation of signal parameters, it is of utmost importance to assess the performance of various estimation procedures. However, any accuracy measure may be of limited interest unless one has an idea of what the best possible performance is. An important measure of how well a particular method performs is the covariance matrix of the estimation errors. Several lower bounds on the estimation error covariance are available in the literature, see e.g., [55]–[57]. Of these, the Cramér-Rao Lower Bound (CRLB) is by far the most commonly used. The main reason for this is its simplicity, but also the fact that it is often (asymptotically) tight, i.e., there exists an estimator that (asymptotically) achieves the CRLB. Such an estimator is said to be (asymptotically) *efficient*. Unless otherwise explicitly stated, the word “asymptotically” is used throughout this chapter to mean that the amount of data (N) is large.

Theorem 4.1 *Let $\hat{\boldsymbol{\eta}}$ be an unbiased estimate of the real parameter vector $\boldsymbol{\eta}_0$, i.e., $E\{\hat{\boldsymbol{\eta}}\} = \boldsymbol{\eta}_0$, based on the observations \mathbf{X}_N . The Cramér-Rao lower bound on the estimation error covariance is then given by*

$$E\{(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)^T\} \geq \left[-E \left\{ \frac{\partial^2 \log p(\mathbf{X}_N | \boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} \right\} \right]^{-1}. \quad (4.54)$$

□

The matrix within square brackets in (4.54) (i.e., the inverse of the CRLB) is referred to as the *Fisher Information Matrix* (FIM).

The Cramér-Rao Lower Bound

The CRLB based on the Gaussian signal model is discussed, for example in [21, 28], and is easily derived from the normalized negative log likelihood function in (4.35). Let $\boldsymbol{\eta}$

represent the vector of unknown parameters in the stochastic model

$$\boldsymbol{\eta} = [\boldsymbol{\theta}^T, s_{11}, \dots, s_{dd}, \bar{s}_{21}, \tilde{s}_{21}, \dots, \bar{s}_{d,d-1}, \tilde{s}_{d,d-1}, \sigma^2]^T, \quad (4.55)$$

where $\bar{s}_{ij} = \text{Re}\{s_{ij}\}$ and $\tilde{s}_{ij} = \text{Im}\{s_{ij}\}$. Introduce the short hand notation

$$\mathbf{R}_i = \frac{\partial \mathbf{R}}{\partial \boldsymbol{\eta}_i} \quad (4.56)$$

and recall the following differentiation rules [58]

$$\frac{\partial}{\partial \boldsymbol{\eta}_i} \log |\mathbf{R}| = \text{Tr}\{\mathbf{R}^{-1} \mathbf{R}_i\} \quad (4.57)$$

$$\frac{\partial}{\partial \boldsymbol{\eta}_i} \text{Tr}\{\mathbf{R}^{-1} \hat{\mathbf{R}}\} = -\text{Tr}\{\mathbf{R}^{-1} \mathbf{R}_i \mathbf{R}^{-1} \hat{\mathbf{R}}\}. \quad (4.58)$$

$$(4.59)$$

The first derivative of (4.35) with respect to the i^{th} component of the parameter vector is given by

$$\frac{\partial l(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}_i} = \text{Tr}\{\mathbf{R}^{-1} \mathbf{R}_i (\mathbf{I} - \mathbf{R}^{-1} \hat{\mathbf{R}})\}. \quad (4.60)$$

The following then gives the ij^{th} element of the inverse of the CRLB

$$\begin{aligned} \{FIM\}_{ij} &= N \text{E} \left\{ \frac{\partial^2 l(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}_j \partial \boldsymbol{\eta}_i} \right\} \Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} \\ &= N \text{E} \left\{ \text{Tr} \left[\left((\mathbf{R}^{-1})_j \mathbf{R}_i + \mathbf{R}^{-1} \mathbf{R}_{ij} \right) (\mathbf{I} - \mathbf{R}^{-1} \hat{\mathbf{R}}) - \mathbf{R}^{-1} \mathbf{R}_i (\mathbf{R}^{-1})_j \hat{\mathbf{R}} \right] \right\} \\ &= N \text{E} \left\{ \text{Tr} \left[-\mathbf{R}^{-1} \mathbf{R}_i (\mathbf{R}^{-1})_j \hat{\mathbf{R}} \right] \right\} \\ &= N \text{Tr} \left\{ \mathbf{R}^{-1} \mathbf{R}_i \mathbf{R}^{-1} \mathbf{R}_j \right\}. \end{aligned} \quad (4.61)$$

The appearance of N on the right hand side above is due to the normalization of (4.35). In many applications, only the signal parameters are of interest. However, the above formula involves derivatives with respect to all $d^2 + pd + 1$ components of $\boldsymbol{\eta}$, and in general none of the elements of (4.61) vanish. A compact expression for the CRLB on

the covariance matrix of *the signal parameters only* is presented in [29, 32, 59]. The Cramér-Rao inequality for $\boldsymbol{\theta}$ is given by

$$\mathbb{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\} \geq \mathbf{B}_{STO} , \quad (4.62)$$

where

$$\{\mathbf{B}_{STO}^{-1}\}_{ij} = \frac{2N}{\sigma^2} \text{Re} \left[\text{Tr} \left\{ \mathbf{A}_j^H \mathbf{P}_{\mathbf{A}}^\perp \mathbf{A}_i \mathbf{S} \mathbf{A}^H \mathbf{R}^{-1} \mathbf{A} \mathbf{S} \right\} \right] \quad i, j = 1, \dots, pd . \quad (4.63)$$

For the special case when there is one parameter associated with each signal ($p = 1$), the CRLB for the signal parameters can be put in a simple matrix form

$$\mathbf{B}_{STO} = \frac{\sigma^2}{2N} \left[\text{Re} \left\{ (\mathbf{D}^H \mathbf{P}_{\mathbf{A}}^\perp \mathbf{D}) \odot (\mathbf{S} \mathbf{A}^H \mathbf{R}^{-1} \mathbf{A} \mathbf{S})^T \right\} \right]^{-1} , \quad (4.64)$$

where \odot denotes the Hadamard (or Schur) product, i.e., element-wise multiplication, and

$$\mathbf{D} = \left[\left. \frac{\partial \mathbf{a}(\theta)}{\partial \theta} \right|_{\theta=\theta_1} , \dots , \left. \frac{\partial \mathbf{a}(\theta)}{\partial \theta} \right|_{\theta=\theta_d} \right] . \quad (4.65)$$

The Deterministic Cramér-Rao Lower Bound

The CRLB for *deterministic* signals is derived in [30, 31], and is restated here in its asymptotic form, and for the case $p = 1$. The emitter signals are arbitrary second-order ergodic sequences, and with some abuse of notation, the limiting signal sample covariance matrix is denoted

$$\mathbf{S} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbf{s}(t_i) \mathbf{s}^H(t_i) . \quad (4.66)$$

If the signal waveforms happen to be realizations of stationary stochastic processes, the limiting signal sample covariance will indeed coincide with the signal covariance matrix under mild assumptions (e.g., bounded fourth-order moments).

Let $\hat{\boldsymbol{\theta}}$ be an asymptotically unbiased estimate of the true parameter vector $\boldsymbol{\theta}_0$. For large N , the Cramér-Rao inequality for the signal parameters can then be expressed as

$$\mathbb{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\} \geq \mathbf{B}_{DET} , \quad (4.67)$$

where

$$\{\mathbf{B}_{DET}^{-1}\}_{ij} = \frac{2N}{\sigma^2} \text{Re} \left[\text{Tr} \left\{ \mathbf{A}_j^H \mathbf{P}_A^\perp \mathbf{A}_i \mathbf{S} \right\} \right] . \quad (4.68)$$

It should be noted that the above inequality assumes implicitly that asymptotically unbiased estimates of *all* unknown parameters in the deterministic model, i.e., $\boldsymbol{\theta}$, \mathbf{S}_N and σ^2 are available. No assumptions on the signal waveforms are, however, made so the inequality applies also if they happen to be realizations of Gaussian processes. One may therefore guess that \mathbf{B}_{STO} is tighter than \mathbf{B}_{DET} . We shall prove this statement later in this section.

4.3.4 Asymptotic Properties of the ML Estimates

The ML estimator has a number of attractive properties that hold for general, sufficiently regular, likelihood functions. The most interesting one for our purposes states that if the ML estimates are *consistent*⁶, then they are also asymptotically *efficient*. In this respect, the ML method has the best asymptotic properties possible.

Stochastic Maximum Likelihood

The SML likelihood function is regular and the general theory of ML estimation can be applied to yield the following result.

Theorem 4.2 *Under the Gaussian signal assumption, the SML parameter estimates are consistent and the normalized estimation error, $\sqrt{N}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$, has a limiting zero-mean, normal distribution with covariance matrix equal to N times the CRLB on $\hat{\boldsymbol{\eta}}$.*

Proof See for example Chapter 6.4 in [56] for a proof. □

From the above we conclude that for the SML method, the asymptotic distribution of $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is $\mathcal{N}(\mathbf{0}, \mathbf{C}_{SML})$, where

$$\mathbf{C}_{SML} = N \mathbf{B}_{STO} , \quad (4.69)$$

and where \mathbf{B}_{STO} is given by (4.63).

⁶An estimate is consistent if it converges to the true value as the amount of data tends to infinity.

Deterministic Maximum Likelihood

The deterministic model for the sensor array problem has an important drawback. Since the signal waveforms themselves are regarded as unknown parameters, it follows that the dimension of the parameter vector grows without bound with increasing N . For this reason, consistent estimation of all model parameters is impossible. More precisely, the DML estimate of $\boldsymbol{\theta}$ is consistent, whereas the estimate of \mathbf{S}_N is inconsistent. To verify the consistency of $\hat{\boldsymbol{\theta}}$, observe that under mild conditions, the criterion function (4.53) converges w.p.1 and uniformly in $\boldsymbol{\theta}$ to the limit function

$$\bar{V}_{DML}(\boldsymbol{\theta}) = \text{Tr} \left\{ \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \mathbf{R} \right\} = \text{Tr} \left\{ \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \left(\mathbf{A}(\boldsymbol{\theta}_0) \mathbf{S} \mathbf{A}^H(\boldsymbol{\theta}_0) + \sigma^2 \mathbf{I} \right) \right\} \quad (4.70)$$

as N tends to infinity. Hence, $\hat{\boldsymbol{\theta}}$ converges to the minimizing argument of $\bar{V}_{DML}(\boldsymbol{\theta})$. It is readily verified that $\bar{V}_{DML}(\boldsymbol{\theta}) \geq \sigma^2 \text{Tr} \left\{ \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \right\} = \sigma^2(m-d) = \bar{V}(\boldsymbol{\theta}_0)$ (recall that the trace of a projection matrix equals the dimension of the subspace onto which it projects). Let $\mathbf{S} = \mathbf{L}\mathbf{L}^H$ be the Cholesky-factorization of the signal covariance matrix, where \mathbf{L} is $d \times d'$. Clearly $\bar{V}_{DML}(\boldsymbol{\theta}) = \sigma^2(m-d)$ holds if and only if $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta}_0) \mathbf{L} = \mathbf{0}$, in which case

$$\mathbf{A}(\boldsymbol{\theta}) \mathbf{L}_1 = \mathbf{A}(\boldsymbol{\theta}_0) \mathbf{L} , \quad (4.71)$$

for some $d \times d'$ matrix \mathbf{L}_1 of full rank. By the UP assumption and (4.27), the relation (4.71) is possible if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Thus, we conclude that the minimizer of (4.53) converges w.p.1 to the true value $\boldsymbol{\theta}_0$.

The signal waveform estimates are, however, inconsistent since

$$\hat{\mathbf{S}}_N = \mathbf{A}^{\dagger}(\hat{\boldsymbol{\theta}}) \mathbf{X}_N \longrightarrow \mathbf{S}_N + \mathbf{A}^{\dagger}(\boldsymbol{\theta}_0) \mathbf{N}_N \text{ as } N \rightarrow \infty . \quad (4.72)$$

Owing to the inconsistency of $\hat{\mathbf{S}}_N$, the general properties of ML estimators are not valid here. Thus, as observed in [30], the asymptotic covariance matrix of the signal parameter estimate does not coincide with the deterministic CRLB. Note that the deterministic Cramér-Rao inequality (4.67) is indeed applicable, as the DML estimate of \mathbf{S}_N can be shown to be asymptotically unbiased (with some effort) in spite of its inconsistency.

The asymptotic distribution of the DML signal parameter estimates is derived in [39, 44] and is given next for the case of one parameter per emitter signal.

Theorem 4.3 *Let $\hat{\boldsymbol{\theta}}$ be obtained from (4.53). Then, $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to $\mathbf{N}(\mathbf{0}, \mathbf{C}_{DML})$, where*

$$\frac{1}{N} \mathbf{C}_{DML} = \mathbf{B}_{DET} + 2N \mathbf{B}_{DET} \text{Re} \left\{ (\mathbf{D}^H \mathbf{P}_{\mathbf{A}}^{\perp} \mathbf{D}) \odot (\mathbf{A}^H \mathbf{A})^{-T} \right\} \mathbf{B}_{DET}, \quad (4.73)$$

where \mathbf{B}_{DET} is the asymptotic deterministic CRLB as defined in (4.68).

From (4.73), it is clearly seen that the covariance of the DML estimate is strictly greater than the deterministic CRLB. However, these two matrices approach the same limit as the number of sensors, m , increases, see [32]. The requirement of the DML method to estimate the signal waveforms has thus a deteriorating effect on the DOA estimates, unless m is large. In many applications the signal waveform estimates may be of importance themselves. Though the DML technique provides such estimates, it should be remarked that they are not guaranteed to be the most accurate ones, unless m is large enough.

4.3.5 Order Relations

As discussed above, the two models for the sensor array problem, corresponding to *deterministic* and *stochastic* modeling of the emitter signals respectively, lead to different ML criteria and CRLB's. The following result, due to [32, 60] relates the covariance matrices of the stochastic and deterministic ML estimates and the corresponding CRLB's.

Theorem 4.4 *Let \mathbf{C}_{SML} and \mathbf{C}_{DML} denote the asymptotic covariances of $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, for the stochastic and deterministic ML estimates, respectively. Furthermore, let \mathbf{B}_{STO} and \mathbf{B}_{DET} be the stochastic CRLB and the deterministic CRLB. The following (in)equalities then hold*

$$\frac{1}{N} \mathbf{C}_{DML} \geq \frac{1}{N} \mathbf{C}_{SML} = \mathbf{B}_{STO} \geq \mathbf{B}_{DET}. \quad (4.74)$$

Proof Theorem 4.2 shows the middle equality above. The left inequality follows by applying the DML method under the Gaussian signal assumption. The Cramér-Rao

inequality then implies that $N^{-1}\mathbf{C}_{DML} \geq \mathbf{B}_{STO}$. To prove the right inequality in (4.74), apply the matrix inversion lemma (e.g., [61] Lemma A.1) to obtain

$$\begin{aligned}\mathbf{S}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A}\mathbf{S} &= \mathbf{S} - \mathbf{S} \left(\mathbf{I} - \mathbf{A}^H(\mathbf{A}\mathbf{S}\mathbf{A}^H + \sigma^2\mathbf{I})^{-1}\mathbf{A}\mathbf{S} \right) \\ &= \mathbf{S} - \mathbf{S} \left(\mathbf{I} + \mathbf{A}^H\sigma^{-2}\mathbf{A}\mathbf{S} \right)^{-1}.\end{aligned}\quad (4.75)$$

Since the matrix $\mathbf{S}(\mathbf{I} + \mathbf{A}^H\sigma^{-2}\mathbf{A}\mathbf{S})^{-1}$ is Hermitian (by the equality above) and positive semi-definite, it follows that $\mathbf{S}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A}\mathbf{S} \leq \mathbf{S}$. Hence, application of [39], Lemma A.2, yields

$$\text{Re} \left\{ (\mathbf{D}^H\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}) \odot (\mathbf{S}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A}\mathbf{S})^T \right\} \leq \text{Re} \left\{ (\mathbf{D}^H\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}) \odot \mathbf{S}^T \right\} \quad (4.76)$$

By inverting both sides of (4.76), the desired inequality follows. If the matrices $\mathbf{D}^H\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{D}$ and \mathbf{S} are both positive definite, the inequality (4.76) is strict showing that the stochastic bound is in this case strictly tighter than the deterministic bound. \square

Remark 4.2 It is of course natural that the SML estimator is more accurate than the DML method under the Gaussian signal assumption. However, this relation remains true for arbitrary second-order ergodic emitter signals, which is more surprising. This is a consequence of the *asymptotic robustness* property of both ML estimators: the asymptotic distribution of the signal parameter estimates is completely specified by $\lim_{N \rightarrow \infty} (1/N) \sum_{i=1}^N \mathbf{s}(t_i)\mathbf{s}^H(t_i)$. As shown in [32, 60], the actual signal waveform sequence (or its distribution) is immaterial. The fact that the SML method always outperforms the DML method, provides strong justification for the stochastic model being appropriate for the sensor array problem. Indeed, the asymptotic robustness and efficiency of the SML method implies that $N\mathbf{B}_{STO} = \mathbf{C}_{SML}$ is a lower bound on the covariance matrix of the normalized estimation error for *any* asymptotically robust method. \square

4.4 Large Sample ML Approximations

The previous section dealt with *optimal* (in the ML sense) approaches to the sensor array problem. Since these techniques are often deemed exceedingly complex, suboptimal methods are of interest. In the present section, several *subspace techniques* are presented based on geometrical properties of the data model.

The focus here is on subspace based techniques where the vector of unknown signal parameters is estimated by performing a multidimensional search on a pd -dimensional criterion. This is in contrast to techniques such as the MUSIC algorithm [12, 13], where the location of d peaks in a p -dimensional so-called *spatial spectrum* determines the signal parameter estimates. Multidimensional versions of the MUSIC approach are discussed in [28, 31, 62, 63]. A multidimensional subspace based technique termed MODE (method of direction estimation) is presented and extensively analyzed in [32, 41, 42]. In [27, 44, 64], a related subspace fitting formulation of the sensor array problem is analyzed and the WSF (weighted subspace fitting) method is proposed.

This section ties together many of the concepts and methods presented in the papers above and discusses the relation of these to the ML techniques of the previous section. A statistical analysis shows that appropriate selections of certain *weighting matrices* give the subspace methods similar (optimal) estimation accuracy as the ML techniques, at a reduced computational cost.

4.4.1 The Subspace Based Approach

All subspace based methods rely on geometrical properties of the spectral decomposition of the array covariance matrix, \mathbf{R} . The early approaches such as MUSIC suffer from a large finite sample bias and are unable to cope with *coherent signals*⁷. This problem is inherent due to the one-dimensional search (p -dimensional search when more than one parameter is associated with each signal) of the parameter space. Means of reducing the susceptibility of these techniques to coherent signals have been proposed for special array structures, see e.g. [65]. In the general case, methods based on a pd -dimensional search need to be employed.

⁷Two signals are said to be coherent if they are identical up to amplitude scaling and phase shift.

If the signal waveforms are non-coherent, the signal covariance matrix, \mathbf{S} , has full rank. However, in radar applications where specular multipath is common, \mathbf{S} may be ill-conditioned or even rank deficient. Let the signal covariance matrix have rank d' . The covariance of the array output is

$$\mathbf{R} = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{S}\mathbf{A}^H(\boldsymbol{\theta}_0) + \sigma^2\mathbf{I} . \quad (4.77)$$

It is clear that any vector in the null space of the matrix $\mathbf{A}\mathbf{S}\mathbf{A}^H$ is an eigenvector of \mathbf{R} with corresponding eigenvalue σ^2 . Since \mathbf{A} has full rank, $\mathbf{A}\mathbf{S}\mathbf{A}^H$ is positive semi-definite and has rank d' . Hence, σ^2 is the smallest eigenvalue of \mathbf{R} with multiplicity $m - d'$. Let $\lambda_1, \dots, \lambda_m$ denote the eigenvalues of \mathbf{R} in non-increasing order, and let $\mathbf{e}_1, \dots, \mathbf{e}_m$ be the corresponding orthonormal eigenvectors. The spectral decomposition of \mathbf{R} then takes the form

$$\mathbf{R} = \sum_{i=1}^m \lambda_i \mathbf{e}_i \mathbf{e}_i^H = \mathbf{E}_s \boldsymbol{\Lambda}_s \mathbf{E}_s^H + \sigma^2 \mathbf{E}_n \mathbf{E}_n^H , \quad (4.78)$$

where

$$\boldsymbol{\Lambda}_s = \text{diag}[\lambda_1, \dots, \lambda_{d'}] , \quad \mathbf{E}_s = [\mathbf{e}_1, \dots, \mathbf{e}_{d'}] , \quad \mathbf{E}_n = [\mathbf{e}_{d'+1}, \dots, \mathbf{e}_m] . \quad (4.79)$$

The diagonal matrix $\boldsymbol{\Lambda}_s$ contains the so-called *signal eigenvalues*, and these are assumed to be distinct. From the above discussion, it is clear that \mathbf{E}_n is orthogonal to $\mathbf{A}\mathbf{S}\mathbf{A}^H$, which implies that the d' -dimensional range space of \mathbf{E}_s is contained in the d -dimensional range space of $\mathbf{A}(\boldsymbol{\theta}_0)$

$$\Re\{\mathbf{E}_s\} \subseteq \Re\{\mathbf{A}(\boldsymbol{\theta}_0)\} . \quad (4.80)$$

If the signal covariance \mathbf{S} has full rank, these subspaces coincide since they have the same dimension, $d' = d$. The range space of \mathbf{E}_s is referred to as the *signal subspace* and its orthogonal complement, the range space of \mathbf{E}_n , is called the *noise subspace*. The signal and noise subspaces can be consistently estimated from the eigendecomposition of the sample covariance

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}(t_i) \mathbf{x}^H(t_i) = \hat{\mathbf{E}}_s \hat{\boldsymbol{\Lambda}}_s \hat{\mathbf{E}}_s^H + \hat{\mathbf{E}}_n \hat{\boldsymbol{\Lambda}}_n \hat{\mathbf{E}}_n^H . \quad (4.81)$$

Signal Subspace Formulation

The relation in (4.80) implies that there exists a $d \times d'$ matrix \mathbf{T} of full rank, such that

$$\mathbf{E}_s = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{T} . \quad (4.82)$$

In general, there is no value of $\boldsymbol{\theta}$ such that $\hat{\mathbf{E}}_s = \mathbf{A}(\boldsymbol{\theta})\mathbf{T}$ when the signal subspace is estimated from the noise-corrupted data. With this observation, a natural estimation criterion is to find the *best weighted least squares fit* of the two subspaces, viz.

$$[\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}] = \arg \min_{\boldsymbol{\theta}, \mathbf{T}} \|\hat{\mathbf{E}}_s - \mathbf{A}(\boldsymbol{\theta})\mathbf{T}\|_{\mathbf{W}}^2, \quad (4.83)$$

where $\|\mathbf{A}\|_{\mathbf{W}}^2 = \text{Tr}\{\mathbf{A}\mathbf{W}\mathbf{A}^H\}$ and \mathbf{W} is a $d' \times d'$ positive definite weighting matrix (to be specified). The optimization (4.83) is a least-squares problem, which is linear in \mathbf{T} and non-linear in $\boldsymbol{\theta}$. Substituting the pseudo-inverse solution, $\hat{\mathbf{T}}_{LS} = \mathbf{A}^\dagger \hat{\mathbf{E}}_s$ (see, e.g. [61], Complement C7.3) in (4.83), we obtain

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} V_{SSF}(\boldsymbol{\theta}) , \quad (4.84)$$

where

$$\begin{aligned} V_{SSF}(\boldsymbol{\theta}) &= \|\hat{\mathbf{E}}_s - \mathbf{A}(\boldsymbol{\theta})\mathbf{A}^\dagger(\boldsymbol{\theta})\hat{\mathbf{E}}_s\|_{\mathbf{W}}^2 \\ &= \|\{\mathbf{I} - \mathbf{P}_{\mathbf{A}}(\boldsymbol{\theta})\}\hat{\mathbf{E}}_s\|_{\mathbf{W}}^2 \\ &= \text{Tr}\{\mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta})\hat{\mathbf{E}}_s\mathbf{W}\hat{\mathbf{E}}_s^H\mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta})\} \\ &= \text{Tr}\{\mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta})\hat{\mathbf{E}}_s\mathbf{W}\hat{\mathbf{E}}_s^H\} . \end{aligned} \quad (4.85)$$

The above equations define the class of subspace fitting (SSF) methods. Different members in this class correspond to specific choices of weighting matrix in (4.85). It should be noted that the choice $\mathbf{W} = \mathbf{I}$ yields Cadzow's method, [63]. A statistical analysis suggests other weightings, leading to superior performance as demonstrated later.

Noise Subspace Formulation

If the emitter signal covariance, \mathbf{S} , has full rank, i.e., $d' = d$, it is clear that the columns of $\mathbf{A}(\boldsymbol{\theta}_0)$ are orthogonal to the noise subspace, i.e.

$$\mathbf{E}_n^H \mathbf{A}(\boldsymbol{\theta}_0) = \mathbf{0} . \quad (4.86)$$

When an estimate of the noise subspace is available, a weighted least-squares measure of orthogonality can be formulated. A natural estimate of $\boldsymbol{\theta}$ is then obtained by minimizing that measure

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} V_{NSF}(\boldsymbol{\theta}) , \quad (4.87)$$

where

$$\begin{aligned} V_{NSF}(\boldsymbol{\theta}) &= \|\hat{\mathbf{E}}_n^H \mathbf{A}(\boldsymbol{\theta})\|_{\mathbf{U}}^2 \\ &= \text{Tr}\{\hat{\mathbf{E}}_n^H \mathbf{A}(\boldsymbol{\theta}) \mathbf{U} \mathbf{A}^H(\boldsymbol{\theta}) \hat{\mathbf{E}}_n\} \\ &= \text{Tr}\{\mathbf{U} \mathbf{A}^H(\boldsymbol{\theta}) \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^H \mathbf{A}(\boldsymbol{\theta})\} , \end{aligned} \quad (4.88)$$

where $\mathbf{U} \geq \mathbf{0}$ is a $d \times d$ weighting matrix. Different choices of weighting, \mathbf{U} , result in signal parameter estimates with different asymptotic properties. Indeed, if \mathbf{U} is a diagonal matrix, the problem decouples and the MUSIC estimator is obtained. Again, a statistical analysis suggests other weightings giving better estimation accuracy.

4.4.2 Relation Between the Subspace Formulations

The subspace techniques above can be shown to be closely related. Before discussing their relationship, it is necessary to define what is meant herein by asymptotic equivalence of estimators.

Definition 4.1 *Two estimates, $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$, of the same true parameter vector are asymptotically equivalent if*

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2) \rightarrow \mathbf{0} \quad (4.89)$$

in probability as $N \rightarrow \infty$.

Notice, in particular, that if $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are asymptotically equivalent, the asymptotic distributions of the normalized estimation errors $\sqrt{N}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)$ and $\sqrt{N}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_0)$ are identical. A first-order Taylor expansion shows that under mild conditions, two consistent parameter estimates obtained from the minimization of two different functions $J(\boldsymbol{\theta})$ and $V(\boldsymbol{\theta})$ are asymptotically equivalent if

$$J'(\boldsymbol{\theta}_0) = V'(\boldsymbol{\theta}_0) + o_p(1/\sqrt{N}) \quad (4.90)$$

$$J''(\boldsymbol{\theta}_0) = V''(\boldsymbol{\theta}_0) + o_p(1), \quad (4.91)$$

where the symbol $o_p(\cdot)$ represent the “in probability” version of the corresponding deterministic notation⁸.

The following theorem establishes the relationship between the signal and noise subspace formulations, (4.85) and (4.88).

Theorem 4.5 *The estimate obtained by minimizing (4.88) with weighting matrix given by*

$$\mathbf{U} = \mathbf{A}^\dagger(\boldsymbol{\theta}_0)\hat{\mathbf{E}}_s\mathbf{W}\hat{\mathbf{E}}_s^H\mathbf{A}^{\dagger H}(\boldsymbol{\theta}_0), \quad (4.92)$$

is asymptotically equivalent to the minimizing argument of (4.85).

Proof To establish (4.90), consider first the derivative of (4.85). Using the formula (A.3) for the derivative of the projection matrix and defining $\mathbf{A}_i = \partial\mathbf{A}/\partial\theta_i$, we have

$$\begin{aligned} \frac{\partial V_{SSF}}{\partial\theta_i} &= -\text{Tr}\left\{\frac{\partial\mathbf{P}_\mathbf{A}}{\partial\theta_i}\hat{\mathbf{E}}_s\mathbf{W}\hat{\mathbf{E}}_s^H\right\} \\ &= -2\text{Re}\left[\text{Tr}\left\{\mathbf{P}_\mathbf{A}^\perp\mathbf{A}_i\mathbf{A}^\dagger\hat{\mathbf{E}}_s\mathbf{W}\hat{\mathbf{E}}_s^H\right\}\right]. \end{aligned} \quad (4.93)$$

Next, examine the derivative of the cost function (4.88),

$$\frac{\partial V_{NSF}}{\partial\theta_i} = 2\text{Re}\left[\text{Tr}\left\{\mathbf{U}\mathbf{A}^H\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^H\mathbf{A}_i\right\}\right]. \quad (4.94)$$

⁸A sequence, x_N , of random variables is $o_p(a_N)$ if the probability limit of $a_N^{-1}x_N$ is zero.

Inserting the expression for the weighting matrix (4.92), the argument of the trace above is

$$\mathbf{A}^\dagger \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^H \mathbf{A}^{\dagger H} \mathbf{A}^H \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^H \mathbf{A}_i = -\mathbf{A}^\dagger \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^H \mathbf{P}_\mathbf{A}^\perp \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^H \mathbf{A}_i \quad (4.95)$$

$$= -\mathbf{A}^\dagger \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^H \mathbf{P}_\mathbf{A}^\perp \mathbf{E}_n \mathbf{E}_n^H \mathbf{A}_i + o_p(1/\sqrt{N}) \quad (4.96)$$

$$= -\mathbf{A}^\dagger \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^H \mathbf{P}_\mathbf{A}^\perp \mathbf{A}_i + o_p(1/\sqrt{N}) . \quad (4.97)$$

In (4.95), the relations $\mathbf{A}^{\dagger H} \mathbf{A}^H = \mathbf{P}_\mathbf{A} = \mathbf{I} - \mathbf{P}_\mathbf{A}^\perp$ and $\hat{\mathbf{E}}_s^H \hat{\mathbf{E}}_n = \mathbf{0}$ are used. Noting that $\hat{\mathbf{E}}_s^H \mathbf{P}_\mathbf{A}^\perp = O_p(1/\sqrt{N})$ and $\hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^H = \mathbf{E}_n \mathbf{E}_n^H + o_p(1)$ leads to (4.96). Finally, the fact that $\mathbf{P}_\mathbf{A}^\perp \mathbf{E}_n \mathbf{E}_n^H = \mathbf{P}_\mathbf{A}^\perp$ is used to obtain (4.97). Substituting (4.97) into (4.94) and comparing with (4.93) shows that (4.90) is satisfied.

Consider the second derivative of (4.88)

$$\frac{\partial^2 V_{NSF}}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} = 2\text{Re} \left[\text{Tr} \{ \mathbf{U} \mathbf{A}_i^H \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^H \mathbf{A}_i \} \right] + 2\text{Re} \left[\text{Tr} \{ \mathbf{U} \mathbf{A}^H \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^H \mathbf{A}_{ij} \} \right] . \quad (4.98)$$

Inserting (4.92) and evaluating the probability limit as $N \rightarrow \infty$, we obtain after some manipulations

$$\frac{\partial^2 V_{NSF}}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \rightarrow 2\text{Re} \left[\text{Tr} \{ \mathbf{A}_j \mathbf{P}_\mathbf{A}^\perp \mathbf{A}_i^H \mathbf{A}^\dagger \mathbf{E}_s \mathbf{W} \mathbf{E}_s^* \mathbf{A}^{\dagger H} \} \right] . \quad (4.99)$$

Using (A.5), it is straightforward to verify that the limiting second derivative of (4.85) coincides with (4.99), see also [44]. \square

Note that the weighting matrix in (4.92) depends on the true signal parameters $\boldsymbol{\theta}_0$. However, the following lemma shows that the weighting matrix (for both subspace formulations) can be replaced by any consistent estimate thereof without affecting the asymptotic properties. Thus, a two-step procedure can be applied, where $\boldsymbol{\theta}_0$ is replaced by a consistent estimate as described in [41].

Lemma 4.1 *If the weighting matrix \mathbf{W} in (4.85) is replaced by a consistent estimate, an asymptotically equivalent signal parameter estimate is obtained. If $d' = d$, the same is true for \mathbf{U} in (4.88).*

Proof Let $\hat{\mathbf{W}} = \mathbf{W} + o_p(1)$ and write $V_{SSF}(\mathbf{W})$ to stress its dependence on the weighting matrix. Examine the derivative in (4.93)

$$\frac{\partial V_{SSF}}{\partial \theta_i} = V_{SSF,i}(\mathbf{W}) = -2\text{Re} \left[\text{Tr} \{ \mathbf{A}^\dagger \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^H \mathbf{P}_\mathbf{A}^\perp \mathbf{A}_i \} \right] . \quad (4.100)$$

Since $\hat{\mathbf{E}}_s^H \mathbf{P}_\mathbf{A}^\perp = O_p(1/\sqrt{N})$, it readily follows that $V_{SSF,i}(\mathbf{W}) = V_{SSF,i}(\hat{\mathbf{W}}) + o_p(1/\sqrt{N})$ which shows (4.90). Next, assume that $d' = d$ and let $\hat{\mathbf{U}} = \mathbf{U} + o_p(1)$. From (4.94) we then have $V_{NSF,i}(\mathbf{U}) = V_{NSF,i}(\hat{\mathbf{U}}) + o_p(1/\sqrt{N})$, since $\mathbf{A}^H \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^H = O_p(1/\sqrt{N})$. Condition (4.91) is trivially satisfied for both criteria. \square

This result will be useful when considering different choices of weighting matrices which are data dependent.

4.4.3 Relation to ML Estimation

In the previous section, the subspace based methods are derived from a purely geometrical point of view. The statistical properties of these techniques reveal unexpected relations to the previously described ML methods. The following result gives the asymptotic distribution of the signal parameter estimates obtained using the signal subspace technique and a general (positive definite) weighting matrix.

Theorem 4.6 *Let $\hat{\boldsymbol{\theta}}$ be obtained from (4.84)–(4.85). Then $\hat{\boldsymbol{\theta}}$ converges to the true value, $\boldsymbol{\theta}_0$, w.p.1 as N tends to infinity. Furthermore, the normalized estimation error, $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, has a limiting zero-mean Gaussian distribution*

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \in \text{AsN}(\mathbf{0}, \mathbf{C}) . \quad (4.101)$$

The covariance of the asymptotic distribution has the form

$$\mathbf{C} = \mathbf{H}^{-1} \mathbf{Q} \mathbf{H}^{-1} . \quad (4.102)$$

Here, \mathbf{H} denotes the limiting Hessian matrix and \mathbf{Q} is the asymptotic covariance matrix of the normalized gradient

$$\mathbf{H} = \lim_{N \rightarrow \infty} V''_{SSF}(\boldsymbol{\theta}_0) \quad (4.103)$$

$$\mathbf{Q} = \lim_{N \rightarrow \infty} N \mathbb{E}\{V'_{SSF}(\boldsymbol{\theta}_0)V_{SSF}'(\boldsymbol{\theta}_0)\} . \quad (4.104)$$

The ij^{th} elements of these matrices are given by

$$\mathbf{H}_{ij} = 2\text{Re} \left[\text{Tr} \left\{ \mathbf{A}_j^H \mathbf{P}_A^\perp \mathbf{A}_i \mathbf{A}^\dagger \mathbf{E}_s \mathbf{W} \mathbf{E}_s^H \mathbf{A}^{\dagger H} \right\} \right] \quad (4.105)$$

$$\mathbf{Q}_{ij} = 2\sigma^2 \text{Re} \left[\text{Tr} \left\{ \mathbf{A}_j^H \mathbf{P}_A^\perp \mathbf{A}_i \mathbf{A}^\dagger \mathbf{E}_s \mathbf{W} \boldsymbol{\Lambda}_s \tilde{\boldsymbol{\Lambda}}^{-2} \mathbf{W} \mathbf{E}_s^H \mathbf{A}^{\dagger H} \right\} \right] , \quad (4.106)$$

where

$$\tilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}_s - \sigma^2 \mathbf{I} . \quad (4.107)$$

□

Theorem 4.5 relates the asymptotic properties of the signal and noise subspace formulations. Hence, the above result gives the asymptotic distribution for the latter estimator as well, provided \mathbf{U} is chosen conformally with (4.92). If $d = d'$, this imposes no restriction and an arbitrary $\mathbf{U} > \mathbf{0}$ can be chosen. However, in case $d' < d$ the relation (4.92) is not invertible and Theorem 4.6 gives the distribution only for \mathbf{U} 's having this particular form.

Some algebraic manipulations of the expression for the asymptotic estimation error covariance give the following result.

Corollary 4.1

- a) The weightings $\mathbf{W}_{opt} = \tilde{\boldsymbol{\Lambda}}^2 \boldsymbol{\Lambda}_s^{-1}$ and $\mathbf{U}_{opt} = \mathbf{A}^\dagger(\boldsymbol{\theta}_0) \hat{\mathbf{E}}_s \mathbf{W}_{opt} \hat{\mathbf{E}}_s^H \mathbf{A}^{\dagger H}(\boldsymbol{\theta}_0)$ result in estimators which are large sample realizations of the SML method, i.e., the asymptotic distributions of the normalized estimation errors coincide.
- b) If $d' = d$, the weightings $\mathbf{W} = \tilde{\boldsymbol{\Lambda}}$ and $\mathbf{U} = \mathbf{A}^\dagger(\boldsymbol{\theta}_0) \hat{\mathbf{E}}_s \tilde{\boldsymbol{\Lambda}} \hat{\mathbf{E}}_s^H \mathbf{A}^{\dagger H}(\boldsymbol{\theta}_0)$ result in large sample realizations of the DML method.

Proof To prove **a)**, note that $\mathbf{H}(\mathbf{W}_{opt}) = \sigma^{-2}\mathbf{Q}(\mathbf{W}_{opt})$. Comparing (4.105) and (4.63), it suffices to verify that

$$\mathbf{A}^\dagger \mathbf{E}_s \mathbf{W}_{opt} \mathbf{E}_s^H \mathbf{A}^{\dagger H} = \mathbf{S} \mathbf{A}^H \mathbf{R}^{-1} \mathbf{A} \mathbf{S} . \quad (4.108)$$

Equations (4.77)–(4.78) imply

$$\mathbf{E}_s \mathbf{\Lambda}_s \mathbf{E}_s^H = \mathbf{A} \mathbf{S} \mathbf{A}^H + \sigma^2 \mathbf{E}_s \mathbf{E}_s^H , \quad (4.109)$$

from which it follows that

$$\mathbf{A}^\dagger \mathbf{E}_s \tilde{\mathbf{\Lambda}} \mathbf{E}_s^H \mathbf{A}^{\dagger H} = \mathbf{A}^\dagger \mathbf{A} \mathbf{S} \mathbf{A}^H \mathbf{A}^{\dagger H} = \mathbf{S} . \quad (4.110)$$

Using (4.110) and the eigendecomposition of \mathbf{R}^{-1} , we have

$$\begin{aligned} \mathbf{S} \mathbf{A}^H \mathbf{R}^{-1} \mathbf{A} \mathbf{S} &= \mathbf{A}^\dagger \mathbf{E}_s \tilde{\mathbf{\Lambda}} \mathbf{E}_s^H \mathbf{A}^{\dagger H} \mathbf{A}^H \left(\mathbf{E}_s \mathbf{\Lambda}_s^{-1} \mathbf{E}_s^H + \sigma^{-2} \mathbf{E}_n \mathbf{E}_n^H \right) \mathbf{A} \mathbf{A}^\dagger \mathbf{E}_s \tilde{\mathbf{\Lambda}} \mathbf{E}_s^H \mathbf{A}^{\dagger H} \\ &= \mathbf{A}^\dagger \mathbf{E}_s \tilde{\mathbf{\Lambda}} \mathbf{\Lambda}_s^{-1} \tilde{\mathbf{\Lambda}} \mathbf{E}_s^H \mathbf{A}^{\dagger H} , \end{aligned} \quad (4.111)$$

since $\mathbf{A} \mathbf{A}^\dagger \mathbf{E}_s = \mathbf{P}_A \mathbf{E}_s = \mathbf{E}_s$.

To establish **b)**, observe from (4.105), (4.110), and (4.68) that for $\mathbf{W} = \tilde{\mathbf{\Lambda}}$, \mathbf{H} is given by

$$\mathbf{H}^{-1} = \frac{N}{\sigma^2} \mathbf{B}_{DET} . \quad (4.112)$$

If $d' = d$, (4.110) gives

$$\mathbf{A}^\dagger \mathbf{E}_s \mathbf{\Lambda}_s \mathbf{E}_s^H \mathbf{A}^{\dagger H} = \mathbf{S} + \sigma^2 (\mathbf{A}^H \mathbf{A})^{-1} , \quad (4.113)$$

and (4.106) reads

$$\begin{aligned} \mathbf{Q}_{ij} &= 2\sigma^2 \text{Re} \left[\text{Tr} \left\{ \mathbf{A}_j^H \mathbf{P}_A^\perp \mathbf{A}_i (\mathbf{S} + \sigma^2 (\mathbf{A}^H \mathbf{A})^{-1}) \right\} \right] \\ &= \frac{\sigma^4}{N} (\mathbf{B}_{DET}^{-1})_{ij} + 2\sigma^4 \text{Re} \left[\text{Tr} \left\{ \mathbf{A}_j^H \mathbf{P}_A^\perp \mathbf{A}_i (\mathbf{A}^H \mathbf{A})^{-1} \right\} \right] . \end{aligned} \quad (4.114)$$

Now, combining (4.102), (4.112) and (4.114) leads to the following covariance when the weighting is chosen as $\mathbf{W} = \tilde{\mathbf{\Lambda}}$ and $p = 1$

$$\mathbf{C} = N \mathbf{B}_{DET} + 2 N^2 \mathbf{B}_{DET} \text{Re} \left\{ (\mathbf{D}^H \mathbf{P}_{\mathbf{A}}^{\perp} \mathbf{D}) \odot (\mathbf{A}^H \mathbf{A})^{-T} \right\} \mathbf{B}_{DET} = \mathbf{C}_{DML} \quad (4.115)$$

The last equality follows by comparing with (4.73). \square

The result above states that with appropriate choice of weighting matrices, the subspace based techniques are as accurate as the ML methods. In fact, the corresponding subspace and ML estimates of the signal parameters are *asymptotically equivalent*, which is slightly stronger.

It is interesting to observe that the DML weighting for the noise subspace technique gives $\mathbf{U} \rightarrow \mathbf{A}^{\dagger} \mathbf{E}_s \tilde{\mathbf{\Lambda}} \mathbf{E}_s^H \mathbf{A}^{\dagger H} = \mathbf{S}$. Thus, we rediscover the fact that MUSIC is a large sample realization of the DML method if \mathbf{S} is diagonal (uncorrelated sources). Recall that (4.88) reduces to MUSIC whenever \mathbf{U} is diagonal. The asymptotic equivalence of MUSIC and the DML method for uncorrelated signals was first proved in [30]. It should also be remarked that part **b**) above is also shown in [31].

It is straightforward to show that the weighting matrix that gives the lowest asymptotic estimation error variance is $\mathbf{W}_{opt} = \tilde{\mathbf{\Lambda}}^2 \mathbf{\Lambda}_s^{-1}$. This is not surprising, since the corresponding estimation error covariance coincides with the stochastic CRLB. The optimally weighted signal subspace technique (4.85) is referred to as the *weighted subspace fitting* (WSF) method. Systematic derivations of the method can be found in [29], where the SML method is approximated by neglecting terms that do not affect the asymptotic properties, and in [42, 64], where the asymptotic likelihood function of $\mathbf{P}_{\mathbf{A}}^{\perp} \hat{\mathbf{E}}_s$ is considered.

Remark 4.3 The optimally weighted noise subspace technique is obtained by using $\mathbf{U} = \mathbf{A}^{\dagger}(\boldsymbol{\theta}_0) \hat{\mathbf{E}}_s \mathbf{W}_{opt} \hat{\mathbf{E}}_s^H \mathbf{A}^{\dagger H}(\boldsymbol{\theta}_0)$. However, note that the replacement of $\boldsymbol{\theta}_0$ by a consistent estimate in this expression *does* affect the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ if $d' < d$, cf. the requirement $d' = d$ in Lemma 4.1. In fact, one can show that for $d' < d$, the signal parameter estimates obtained using the optimally weighted noise subspace technique

are *not* asymptotically equivalent with the SML estimates. Modifications to the cost function can be made to obtain equivalence with the SML method, but this results in an increase in the computational cost when minimizing the criterion function. In the following we will therefore only consider the WSF cost function. \square

4.5 Calculating the Estimates

All methods considered herein require a multidimensional non-linear optimization for computing the signal parameter estimates. Analytical solutions are, in general, not available and one has to resort to numerical search techniques. Several optimization methods have appeared in the array processing literature, including expected maximization (EM) algorithms [66, 67], the alternating projection (AP) method [68], iterative quadratic ML (IQML) [69], “global” search techniques [70]–[72], as well as different Newton-type techniques [43, 63, 64], [73]–[75]. In this section, Newton-type algorithms for the SML, DML and WSF techniques are described. For the special case of a uniform linear array, a *non-iterative* scheme for the subspace-based method is presented.

4.5.1 Newton-Type Search Algorithms

Assume that $V(\boldsymbol{\theta})$ is a continuously differentiable function, whose minimizer, $\hat{\boldsymbol{\theta}}$, is to be found. One of the most efficient optimization methods for such problems is the damped Newton method [76, 77]. The estimate is iteratively calculated as

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \mu_k \mathbf{H}^{-1} V', \quad (4.116)$$

where $\boldsymbol{\theta}^k$ is the estimate at iteration k , μ_k is the step length, \mathbf{H} represents the Hessian matrix of the criterion function, and V' is the gradient. The Hessian and gradient are evaluated at $\boldsymbol{\theta}^k$. It is well-known that the Newton method gives locally a quadratic convergence to $\hat{\boldsymbol{\theta}}$.

The step length, μ_k , should be appropriately chosen in order to guarantee convergence to a local minimum. An often used scheme for Newton methods is to choose some $\mu < 1$ and to take $\mu_k = (\mu)^i$ for the smallest integer $i \geq 0$ that causes “sufficient decrease” in the criterion function. The reader is referred to [76] and [77] for more sophisticated algorithms for selecting the step length. A useful modification of the damped Newton method, particularly suited for ill-conditioned problems, is to use $(\mathbf{H} + \lambda_k \mathbf{I})^{-1}$ in lieu of $\mu_k \mathbf{H}^{-1}$, where λ_k is increased from 0 until a decrease in the criterion function is observed. This, in combination with the Gauss-Newton modification of the Hessian [76, 77] for non-linear least squares problems, is referred to as the Levenberg-Marguardt technique.

The iterations (4.116) continue until some prescribed stopping criterion is satisfied. Examples of such criteria are:

- $|\mathbf{H}^{-1}V'|$ is less than a specified tolerance and $\mathbf{H} > 0$,
- no improvement can be found along the search direction (μ_k smaller than a tolerance),
- the number of iterations reaches a maximum limit.

The quality of the convergence point depends on the shape of the criterion function in question. If $V(\boldsymbol{\theta})$ possesses several minima, the iterations must be initialized “sufficiently close” to the global minimum in order to prevent convergence to a local extremum. Possible initialization schemes will be discussed in Section 4.5.4.

4.5.2 Gradients and Approximate Hessians

The idea behind Newton’s method is to approximate the criterion function locally around the stationary point by a quadratic function. The role of the Hessian can be seen as a modification of the gradient direction, to take into account the curvature of the approximation. The Newton method has some drawbacks when applied to the problems of interest herein. Firstly, while the negative gradient is, by definition, a descent direction for the criterion function, the Newton-direction ($-\mathbf{H}^{-1}V'$) can be guaranteed to be a descent direction only if \mathbf{H} is positive definite. This may not be the case further away from the minimum, where $V(\boldsymbol{\theta})$ cannot, in general, be well-approximated by a quadratic function. Hence, there may be no value of μ_k that causes a decrease in the criterion. Secondly, the evaluation of the exact Hessian matrices for the SML, DML, and WSF criteria is computationally cumbersome. A standard technique for overcoming these difficulties is to use a less complex approximation of the Hessian matrix, which is also guaranteed to be positive semidefinite. The techniques to be described use the asymptotic (for large N) form of the Hessian matrices, obtained from the previously described asymptotic analysis. In the statistical literature, this is often referred to as the *scoring* method. It is closely related to the Gauss-Newton technique for non-linear least squares problems. It is interesting to observe that for the SML and WSF methods, the inverse of the limiting Hessian coincides (to within a scale factor) with the asymptotic

covariance matrix of the estimation errors. Consequently, an estimate of the accuracy is readily available when the optimum is reached.

In the following, expressions for the gradients and the approximate Hessians are presented for the methods in question. To keep the notation simple, we shall specialize to the case of one parameter per source, i.e., $p = 1$. The extension to a general p is straightforward.

Stochastic Maximum Likelihood

The SML cost function is obtained from (4.43). Evaluation of this expression requires $O(m^3)$ complex floating point operations (flops). Making use of the determinant rule $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$ (see e.g. [61], Appendix A) to rewrite (4.43), the computational cost can be reduced to $O(m^2d)$ flops as follows.

$$\begin{aligned} V_{SML}(\boldsymbol{\theta}) &= \log \left| \mathbf{A} \hat{\mathbf{S}}(\boldsymbol{\theta}) \mathbf{A}^H + \hat{\sigma}^2(\boldsymbol{\theta}) \mathbf{I} \right| \\ &= \log \hat{\sigma}^{2m}(\boldsymbol{\theta}) \left| \hat{\sigma}^{-2}(\boldsymbol{\theta}) \hat{\mathbf{S}}(\boldsymbol{\theta}) \mathbf{A}^H \mathbf{A} + \mathbf{I} \right| \\ &= \log \hat{\sigma}^{2(m-d)}(\boldsymbol{\theta}) \left| \mathbf{A}^\dagger \hat{\mathbf{R}} \mathbf{A} \right| , \end{aligned} \quad (4.117)$$

where, in the last equality, we have used (4.37)–(4.38).

Next, introduce the matrix

$$\mathbf{G} = \mathbf{A} \left[(\mathbf{A}^H \hat{\mathbf{R}} \mathbf{A})^{-1} - \hat{\sigma}^{-2}(\boldsymbol{\theta}) (\mathbf{A}^H \mathbf{A})^{-1} \right] . \quad (4.118)$$

The SML gradient can then be expressed as [32],

$$V'_{SML}(\boldsymbol{\theta}) = 2\text{Re} \left(\text{Diag} \left[\mathbf{G}^H \hat{\mathbf{R}} \mathbf{P}_{\mathbf{A}}^\perp \mathbf{D} \right] \right) , \quad (4.119)$$

where the matrix \mathbf{D} is defined in (4.65) and the notation $\text{Diag}[\mathbf{Y}]$, where \mathbf{Y} is a square matrix, means a column vector formed from the diagonal elements of \mathbf{Y} . The approximate Hessian is obtained from the CRLB. From (4.64) we have

$$\mathbf{H}_{SML}(\boldsymbol{\theta}) = \frac{2}{\sigma^2} \text{Re} \left\{ (\mathbf{D}^H \mathbf{P}_{\mathbf{A}}^\perp \mathbf{D}) \odot (\mathbf{S} \mathbf{A}^H \mathbf{R}^{-1} \mathbf{A} \mathbf{S})^T \right\} . \quad (4.120)$$

Replacing the unknown quantities \mathbf{S} and σ^2 by their SML estimates (4.37)–(4.38) and \mathbf{R} by $\mathbf{A}\hat{\mathbf{S}}(\boldsymbol{\theta})\mathbf{A}^H + \hat{\sigma}^2(\boldsymbol{\theta})\mathbf{I}$ in (4.120), leads after some manipulations to

$$\hat{\mathbf{S}}(\boldsymbol{\theta})\mathbf{A}^H \left(\mathbf{A}\hat{\mathbf{S}}(\boldsymbol{\theta})\mathbf{A}^H + \hat{\sigma}^2(\boldsymbol{\theta})\mathbf{I} \right)^{-1} \mathbf{A}\hat{\mathbf{S}}(\boldsymbol{\theta}) = \hat{\sigma}^4(\boldsymbol{\theta})\mathbf{G}^H\hat{\mathbf{R}}\mathbf{G} . \quad (4.121)$$

Inserting (4.121) into (4.120) yields the asymptotic Hessian matrix

$$\mathbf{H}_{SML} = 2\hat{\sigma}^2(\boldsymbol{\theta})\text{Re} \left\{ (\mathbf{D}^H\mathbf{P}_{\mathbf{A}}^\perp\mathbf{D}) \odot (\mathbf{G}^H\hat{\mathbf{R}}\mathbf{G})^T \right\} . \quad (4.122)$$

Deterministic Maximum Likelihood

The DML cost function is given by (4.53)

$$V_{DML}(\boldsymbol{\theta}) = \text{Tr} \{ \mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta})\hat{\mathbf{R}} \} . \quad (4.123)$$

Differentiation with respect to $\boldsymbol{\theta}$ gives the gradient expression [30, 44],

$$V'_{DML}(\boldsymbol{\theta}) = -2\text{Re} \left\{ \text{Diag}[\mathbf{A}^\dagger \hat{\mathbf{R}}\mathbf{P}_{\mathbf{A}}^\perp\mathbf{D}] \right\} . \quad (4.124)$$

The asymptotic DML Hessian is derived in [30, 44]

$$\mathbf{H}_{DML} = 2\text{Re} \left\{ (\mathbf{D}^H\mathbf{P}_{\mathbf{A}}^\perp\mathbf{D}) \odot \mathbf{S}^T \right\} . \quad (4.125)$$

A difficulty when using the above is that the DML method does not explicitly provide an estimate of the signal covariance matrix. One could construct such an estimate from the estimated signal waveforms (4.51) as

$$\hat{\mathbf{S}} = \frac{1}{N}\hat{\mathbf{S}}_N\hat{\mathbf{S}}_N^H = \mathbf{A}^\dagger\hat{\mathbf{R}}\mathbf{A}^{\dagger H} . \quad (4.126)$$

Using this in (4.125) results in the modified Gauss-Newton method, as suggested in [64]. However, (4.126) does not provide a consistent estimate of \mathbf{S} . Thus, we propose here to use the corresponding SML estimate (4.37) in (4.125) to give

$$\mathbf{H}_{DML} = 2\text{Re} \left\{ (\mathbf{D}^H\mathbf{P}_{\mathbf{A}}^\perp\mathbf{D}) \odot \left[\mathbf{A}^\dagger(\hat{\mathbf{R}} - \hat{\sigma}^2(\boldsymbol{\theta})\mathbf{I})\mathbf{A}^{\dagger H} \right]^T \right\} . \quad (4.127)$$

Subspace Fitting Criterion

The WSF cost function is given by (4.85).

$$V_{WSF}(\boldsymbol{\theta}) = \text{Tr}\{\mathbf{P}_{\hat{\mathbf{A}}}^{\perp} \hat{\mathbf{E}}_s \hat{\mathbf{W}}_{opt} \hat{\mathbf{E}}_s^H\}, \quad (4.128)$$

where the data dependent weighting matrix

$$\hat{\mathbf{W}}_{opt} = (\hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2 \mathbf{I})^2 \hat{\mathbf{\Lambda}}_s^{-1} \quad (4.129)$$

is used. Here, $\hat{\sigma}^2$ refers to a consistent estimate of the noise variance, for example the average of the $m - d'$ smallest eigenvalues of $\hat{\mathbf{R}}$. Observe the similarity of (4.128) and the DML criterion (4.123). The first derivative is immediate from (4.124)

$$V'_{WSF}(\boldsymbol{\theta}) = -2\text{Re}\left\{\text{Diag}\left[\mathbf{A}^{\dagger} \hat{\mathbf{E}}_s \hat{\mathbf{W}}_{opt} \hat{\mathbf{E}}_s^H \mathbf{P}_{\hat{\mathbf{A}}}^{\perp} \mathbf{D}\right]\right\}. \quad (4.130)$$

The asymptotic expression for the WSF Hessian is obtained from (4.105) as

$$\mathbf{H}_{WSF} = 2\text{Re}\left\{(\mathbf{D}^H \mathbf{P}_{\hat{\mathbf{A}}}^{\perp} \mathbf{D}) \odot (\mathbf{A}^{\dagger} \hat{\mathbf{E}}_s \hat{\mathbf{W}}_{opt} \hat{\mathbf{E}}_s^H \mathbf{A}^{\dagger H})^T\right\}. \quad (4.131)$$

It is interesting to observe that the scoring method for the WSF technique coincides with the modified Gauss-Newton method used in [64].

4.5.3 Uniform Linear Arrays

For the special case where the array elements are identical and equidistantly spaced along a line, and $p = 1$, the subspace-based method, (4.85), can be implemented in a *non-iterative* fashion, [41, 42].

This simplification is based on a reparametrization of the loss function in terms of the coefficients of the following polynomial

$$b(z) = b_0 z^d + b_1 z^{d-1} + \dots + b_d = b_0 \prod_{k=1}^d (z - e^{-j\omega \Delta \sin \theta_k / c}). \quad (4.132)$$

Thus, the signal parameters can easily be extracted from the roots of the polynomial once the latter is available. The array propagation matrix has a Vandermonde structure in the ULA case, see (4.14). Defining $\phi_k = -\omega\Delta \sin \theta_k/c$, we have

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & \dots & 1 \\ e^{j\phi_1} & \dots & e^{j\phi_d} \\ \vdots & \ddots & \vdots \\ e^{j(m-1)\phi_1} & \dots & e^{j(m-1)\phi_d} \end{bmatrix} . \quad (4.133)$$

Introduce the $m \times (m - d)$ matrix \mathbf{B}

$$\mathbf{B}^H = \begin{bmatrix} b_d & b_{d-1} & \dots & b_0 & \mathbf{0} \\ & \ddots & \ddots & & \ddots \\ \mathbf{0} & & b_d & b_{d-1} & \dots & b_0 \end{bmatrix} , \quad (4.134)$$

and observe that

$$\mathbf{B}^H \mathbf{A} = \mathbf{0} . \quad (4.135)$$

Since \mathbf{B} has rank $(m - d)$, it follows that \mathbf{B} spans the null space of \mathbf{A}^H , i.e.,

$$\mathbf{P}_A^\perp = \mathbf{P}_B = \mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H . \quad (4.136)$$

Hence, (4.85) can be rewritten in terms of the polynomial coefficients as

$$V(\boldsymbol{\theta}) = \text{Tr} \left\{ \mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^H \right\} . \quad (4.137)$$

Since $\mathbf{B}^H \hat{\mathbf{E}}_s$ is of order $O_p(1/\sqrt{N})$ locally around the “true” \mathbf{B} , it can easily be shown that the term $(\mathbf{B}^H \mathbf{B})^{-1}$ in (4.137) can be replaced by any consistent estimate without affecting the asymptotic properties of the minimizing \mathbf{B} . This observation leads to the minimization of the function

$$\text{Tr} \left\{ \mathbf{B}(\hat{\mathbf{B}}^H \hat{\mathbf{B}})^{-1} \mathbf{B}^H \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^H \right\} = \text{Tr} \left\{ (\hat{\mathbf{B}}^H \hat{\mathbf{B}})^{-1} \mathbf{B}^H \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^H \mathbf{B} \right\} , \quad (4.138)$$

which is quadratic in \mathbf{B} . A consistent initial estimate of \mathbf{B} is obtained, for example by using $\hat{\mathbf{B}}^H \hat{\mathbf{B}} = \mathbf{I}$ above. Some care must be exercised concerning the constraint on the polynomial coefficients: all roots of $b(z)$ should lie on the unit circle. This constraint is extensively discussed in [41, 42], and an algorithm for minimizing (4.138) involving only the solution of a linear system of equations is proposed.

4.5.4 Practical Aspects

In this section, the computational requirements for obtaining the various estimates are briefly discussed. A detailed implementation of the scoring method for the WSF algorithm is given.

Initial Data Treatment

The computational complexity of an algorithm depends clearly on the application and the availability of parallel processors. For sequential batch processing, and for $N \gg m$, it is clear that the computational cost is dominated by forming the sample covariance (4.36), which takes Nm^2 flops. However, most applications involve the tracking of a time-varying scenario. The sample covariance is then recursively updated⁹ at a high rate (preferably using parallel processing), and estimates of $\hat{\mathbf{R}}$ are delivered to the estimation routine at a relatively low rate. This is because the scenario varies usually rather slowly as compared to the sampling rate. To be able to track time-variations, the sample covariance estimate $\hat{\mathbf{R}}$ could be reset to zero after each delivery, or alternatively a forgetting factor (e.g., [61], Chapter 9) could be used when forming $\hat{\mathbf{R}}$. The speed at which signal parameter estimates must be calculated is clearly related to how quickly the scenario is changing.

The subspace-based methods require the eigenvalues and the d' principal eigenvectors of $\hat{\mathbf{R}}$. The signal subspace dimension, d' , is typically determined by testing the multiplicity of the smallest eigenvalue of $\hat{\mathbf{R}}$ [28, 79, 80]. See also Chapters 2 and 3. A standard routine for calculating the required eigendecomposition costs approximately $6m^3$ flops [81]. However, much attention has recently been paid to utilize the structure

⁹It is numerically more reliable to update the Cholesky factor of the sample covariance to avoid squaring the data, see e.g., [78].

of \mathbf{R} to reduce this cost. In [82], a “batch-type” procedure for computing an approximation of the eigendecomposition is proposed. The referenced technique requires $O(m^2)$ operations and includes a scheme for estimating d' . Fast algorithms for updating the signal subspace after a rank-one modification of $\hat{\mathbf{R}}$ (i.e., after collecting one additional snapshot) are discussed in, e.g., [83]–[86].

Estimator Implementation

Given the sample covariance or its eigendecomposition, an iteration of the scoring method involves the evaluation of the current values of the criterion function V , gradient V' , approximate Hessian \mathbf{H} , and search direction, $\mathbf{s} = -\mathbf{H}^{-1}V'$.

All techniques require the calculation of \mathbf{A}^\dagger and $\mathbf{P}_\mathbf{A}^\perp$. This is best done by QR-decomposition of \mathbf{A} using, for example, Householder transformations [81]

$$\begin{aligned} \mathbf{A} &= \mathbf{QR} = [\mathbf{Q}_1, \mathbf{Q}_2] \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} \Rightarrow \\ \mathbf{A}^\dagger &= \mathbf{R}_1^{-1} \mathbf{Q}_1^H, \quad \mathbf{P}_\mathbf{A}^\perp = \mathbf{Q}_2 \mathbf{Q}_2^H. \end{aligned}$$

Performing a QR-decomposition using Householder transformations requires $d^2(m-d/3)$ operations. The matrix \mathbf{Q} is stored as a product of Householder transformations, which allows forming $\mathbf{Q}^H \mathbf{x}$, where \mathbf{x} is an m -vector, with $O(md)$ flops, see [81].

We shall describe a detailed implementation of the WSF-scoring method. For the stochastic and deterministic ML techniques, only an approximate operation count is presented. The eigendecomposition of $\hat{\mathbf{R}}$ is assumed available, as well as \mathbf{A} and \mathbf{D} , as evaluated at the current iterate.

WSF-scoring implementation

- Weighted signal subspace matrix

$$\mathbf{M} = \hat{\mathbf{E}}_s (\hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2 \mathbf{I}) \hat{\mathbf{\Lambda}}_s^{-1/2},$$

where $\hat{\sigma}^2 = (\sum_{i=d'+1}^m \hat{\lambda}_i) / (m - d')$.

- QR-decomposition using Householder transformations

$$\mathbf{A} = [\mathbf{Q}_1, \mathbf{Q}_2] \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}$$

- Intermediate variables

$$\mathbf{\Phi} = \mathbf{Q}_2^H \mathbf{D}, \quad \mathbf{\Psi} = \mathbf{M}^H \mathbf{Q}_2, \quad \mathbf{\Gamma} = \mathbf{R}_1^{-1} \mathbf{Q}_1^H \mathbf{M}$$

- Criterion function, gradient, and Hessian

$$V = \text{Tr}\{\mathbf{\Psi}\mathbf{\Psi}^H\}$$

$$V' = -2\text{Re}\{\text{Diag}(\mathbf{\Gamma}\mathbf{\Psi}\mathbf{\Phi})\}$$

$$\mathbf{H} = 2\text{Re}\{(\mathbf{\Phi}^H \mathbf{\Phi}) \odot (\mathbf{\Gamma}\mathbf{\Gamma}^H)^T\}$$

- Search direction: Solve $\mathbf{H}\mathbf{s} = -V'$

The above scheme takes about $3md^2 + 2mdd' + O(d^3)$ flops. Similar implementations of the SML- and DML-scoring techniques require each approximately $(3d + 1)m^2 + 4md^2 + O(d^3)$ flops per iteration. When $m \gg d$, as is common in radar applications, the WSF iterations can thus be implemented using significantly less operations than the SML and DML methods.

Initialization

For the Newton type techniques discussed above, initialization is of course a crucial issue. A natural way to obtain initial estimates is to use a (consistent) suboptimal estimator. When the array has a special invariance structure, the Total Least Squares (TLS) version of the ESPRIT algorithm, [87] is an attractive choice. This algorithm is applicable to any array that consists of two identical subarrays. For more general array geometries, the MUSIC algorithm [12, 13] could be used. An important drawback of most suboptimal techniques is the severe degradation of the estimation accuracy in the presence of highly correlated signals. For regular array structures, spatial smoothing techniques can be applied to cope with coherent signals [65]. However, for a general array structure other methods for initialization have to be considered.

A promising possibility is the initialization scheme from the alternating projection (AP) algorithm [68], see also Section 3.6.2. This technique is an application of the relaxed optimization principle (“search for one parameter while keeping the others fixed”) to the optimization problem in question. The algorithm was originally proposed for the DML method, but it is obviously applicable also to the WSF method, simply by replacing the sample covariance matrix by $\hat{\mathbf{E}}_s \hat{\mathbf{W}}_{opt} \hat{\mathbf{E}}_s^H$. The AP algorithm is applicable to a general array structure and it converges to a local minimum. The convergence can, however, be very slow when the Hessian matrix is ill-conditioned. One possibility is to perform only one step (the initial step) of the AP algorithm and to feed the result to the scoring iterations. This initialization procedure is expected to work also for highly correlated or fully coherent signal waveforms, at least if the INR (Information-to-Noise Ratio) is “high enough”. When the number of sources is large ($d \geq 3$), it may be advisable to perform more than the initial step of the AP algorithm before starting the local Newton search.

4.6 Detection of Coherent/Non-Coherent Signals

Any parametric method requires knowledge of the number of signals d . Determining the number of signals is often referred to as *detection*. The eigenstructure-based techniques require, in addition, information on the dimension of the signal subspace, d' . If the signal covariance has full rank, i.e., there are no coherent signals present, d and d' are identical. Detection schemes, based on a test of the multiplicity of the smallest eigenvalue of $\hat{\mathbf{R}}$ have been proposed in several papers, see e.g., [28, 79, 80] and are discussed in Chapters 2 and 3. In the general case, these techniques can be used only for determining d' . In this section we describe two techniques for determining the number of signals, d , for arbitrary emitter covariances.

Optimal approaches to sensor array processing involve solving the detection and estimation problems simultaneously. A standard technique in the statistical estimation literature is the generalized likelihood ratio test (GLRT), see e.g. [55]. Another important approach to the detection problem is the minimum description length (MDL) principle, [88]. This technique has been used for estimating the signal subspace dimension [79], and also for detecting the number of sources [89].

Application of the GLRT technique to the Gaussian signal model is described in Section 4.6.1, whereas Section 4.6.2 describes a subspace-based approach to the detection problem. Simulations, comparing the proposed techniques to the MDL based approach, are included in Section 4.7.

4.6.1 GLRT Based Detection

A natural scheme for detecting the number of signals is to formulate a likelihood ratio test based on the maximum likelihood estimator described in Section 4.3. Such a test is often referred to as a generalized likelihood ratio test (GLRT). The sensor array data model has strong similarities with the factor analysis model used in the statistical literature, [50, 51]. We will follow a standard “goodness of fit” test, where the hypothesis that d signals are adequate to describe the observed data is tested against the alternative that \mathbf{R} has an arbitrary structure. More specifically, the hypotheses are that the $\mathbf{x}(t_i)$:s are i.i.d., $N(\mathbf{0}, \mathbf{R})$, with

$H_0 : \mathbf{R} = \mathbf{A}\mathbf{S}\mathbf{A}^H + \sigma^2\mathbf{I}$, where \mathbf{A} is $m \times d$,

$H_1 : \mathbf{R}$ is arbitrary.

Under H_1 , the ML estimate of \mathbf{R} is simply given by $\hat{\mathbf{R}}$, the sample covariance matrix, [51]. Under H_0 , a *structured* estimate of \mathbf{R} is obtained from

$$\hat{\mathbf{R}}(\hat{\boldsymbol{\theta}}) = \mathbf{A}(\hat{\boldsymbol{\theta}})\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})\mathbf{A}^H(\hat{\boldsymbol{\theta}}) + \hat{\sigma}^2(\hat{\boldsymbol{\theta}})\mathbf{I}, \quad (4.139)$$

where $\hat{\mathbf{S}}(\hat{\boldsymbol{\theta}})$, $\hat{\sigma}^2(\hat{\boldsymbol{\theta}})$ are obtained from (4.37)–(4.38) and $\hat{\boldsymbol{\theta}}$ from (4.42)–(4.43). The log-likelihood ratio for this test can be expressed as

$$LR = \log |\hat{\mathbf{R}}(\hat{\boldsymbol{\theta}})| - \log |\hat{\mathbf{R}}|. \quad (4.140)$$

Standard GLRT theory gives the following result.

Theorem 4.7 *Under H_0 , the asymptotic distribution of the normalized log-likelihood ratio*

$$2N \times LR \quad (4.141)$$

is χ^2 with $m^2 - d^2 - pd - 1$ degrees of freedom.

Proof Observing that \mathbf{R} is Hermitian, it is uniquely parametrized by m^2 parameters under H_1 . The normalized log-likelihood ratio is asymptotically χ^2 distributed with degrees of freedom equal to the number of constraints imposed by the parametrization of \mathbf{R} under H_0 rather than H_1 , see [55]. The dimension of the full parameter vector (4.55) is $d^2 + pd + 1$, and hence there are $m^2 - d^2 - pd - 1$ such constraints. \square

The asymptotic distribution of the normalized log-likelihood ratio can be utilized to devise a sequential test procedure as follows.

GLRT Detection Scheme

1. Set $\hat{d} = 0$
2. Null hypothesis, $H_0 : d = \hat{d}$.
3. Choose a threshold, γ , for the hypothesis test, based on the tail area of the Chi-square distribution with $m^2 - d^2 - pd - 1$ degrees of freedom.
4. Compute $V_{SML}(\hat{\boldsymbol{\theta}})$ under H_0 .
5. If $2N \left(V_{SML}(\hat{\boldsymbol{\theta}}) - \log |\hat{\mathbf{R}}| \right) > \gamma$, reject H_0 . Let $\hat{d} = \hat{d} + 1$ and return to 2
6. Accept H_0 and stop.

4.6.2 Subspace Based Detection

Estimation of the Subspace Dimension

Essentially all existing approaches to the determination of d' are based on the observation that the smallest eigenvalue of the array output covariance has multiplicity $m - d'$. A statistical test is then devised to test the multiplicity of the smallest eigenvalue. In [28] a likelihood ratio test is formulated for this purpose. In [79], information theoretic approaches based on the AIC (An Information theoretic Criterion) and MDL criteria are presented. Strong consistency of the MDL method and of a generalized version of the scheme was proved in [80]. Estimation of d' is discussed in some detail in Chapters 2 and 3.

Detection

As noted in several papers, e.g. [28, 62, 63], the presence of coherent sources ($d > d'$) can be detected by observing the distance between the signal subspace and the array manifold. If this distance is *too large*, we have an indication that there are more signals present than was assumed. Below, a systematic procedure for estimating the number of signals based on this observation is developed.

The cost function (4.85) approaches zero at the global minimum as the number of

data becomes large, provided that the correct dimensions, d and d' are chosen. If d is chosen too small the cost increases. A statistical test is necessary to decide when the cost function is close enough to zero for concluding that the choice of d is correct. The test can be formulated as choosing between

H_0 : There are d signals

H_1 : There are more than d signals

Clearly, H_0 should be accepted when the value of the WSF criterion function is small. The following result provides the necessary basis for appropriately setting the threshold.

Theorem 4.8 *Under H_0 , the normalized WSF cost function*

$$\frac{2N}{\hat{\sigma}^2} V_{WSF}(\boldsymbol{\theta}) = \frac{2N}{\hat{\sigma}^2} \text{Tr}\{\mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta}) \hat{\mathbf{E}}_s \hat{\mathbf{W}}_{opt} \hat{\mathbf{E}}_s^H\} , \quad (4.142)$$

evaluated at $\boldsymbol{\theta}_0$ is asymptotically Chi-squared distributed with $2d'(m - d)$ degrees of freedom. When the criterion is evaluated at $\hat{\boldsymbol{\theta}}$, the asymptotic distribution is still Chi-square, but with $2d'(m - d) - pd$ degrees of freedom. Here $\hat{\sigma}^2$ refers to any consistent estimate of the noise variance.

Proof See Appendix B. □

An algorithm for detection, based on Theorem 4.8, is outlined below. The signal subspace dimension or an estimate thereof is assumed available.

WSF Detection Scheme

1. Set $\hat{d} = d'$.
2. Null hypothesis, $H_0 : d = \hat{d}$.
3. Choose a threshold, γ , for the hypothesis test, based on the tail area of the Chi-square distribution with $2d'(m - d) - pd$ degrees of freedom.
4. Compute $V_{WSF}(\hat{\boldsymbol{\theta}})$ under H_0 .
5. If $\frac{2N}{\hat{\sigma}^2} V_{WSF}(\hat{\boldsymbol{\theta}}) > \gamma$, reject H_0 . Let $\hat{d} = \hat{d} + 1$, and return to 2
6. Accept H_0 and stop.

By appropriately choosing the threshold, γ , the user can determine the (asymptotic) error-rate of the test. According to Theorem 4.8, the probability of rejecting H_0 when it is true, is equal to the tail area of the corresponding χ^2 -distribution. Consequently, the above technique does not yield a consistent estimate of d . To guarantee correct detection for sufficiently large sample sizes, the threshold must be allowed to depend on N . In [64] it is shown that if $\gamma = \gamma(N)$ is chosen to satisfy

$$\lim_{N \rightarrow \infty} \frac{\gamma(N)}{N} = 0 \quad (4.143)$$

$$\lim_{N \rightarrow \infty} \frac{\gamma(N)}{\log \log N} = \infty, \quad (4.144)$$

the estimate of the number of signals is strongly consistent. In practice, consistent detection may be of little interest, since only a fixed number of snapshots are available.

Note that the WSF criterion does not require a consistent estimate of the signal subspace dimension d' to ensure a consistent detection and estimation procedure. This is seen by letting $d' = m - 1$ (step 1 above is then replaced by $\hat{d} = 1$) and estimating $\hat{\sigma}^2$ by the smallest eigenvalue of the sample covariance matrix. The smallest eigenvalue is a consistent estimate of the noise variance, and asymptotically $\hat{\mathbf{W}}_{opt} = (\hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2 \mathbf{I})^2 \hat{\mathbf{\Lambda}}_s^{-1}$ provides a rank d' matrix with the correct weights. However, for obtaining efficient estimates it is crucial that the signal subspace dimension is not underestimated, see also [49].

For small sample sizes the situation is rather different. When the signal waveforms are highly correlated but not coherent, the signal covariance is nearly rank deficient. The small sample behavior of the detection/estimation scheme can then often be improved by underestimating d' . For sufficiently high signal correlation, this will in fact be done by any of the well known tests of the multiplicity of the smallest eigenvalue, see e.g. [79, 80].

4.7 Numerical Examples and Simulations

The analysis presented in this chapter has assumed that a “sufficient” amount of data is available to the user. It is of utmost importance to investigate in what region of information-to-noise ratio the asymptotic results can predict the actual performance. This aspect is addressed in this section, using computer simulations and numerical evaluations of the asymptotic variance expressions for the various estimators.

We shall concentrate on a scenario where $d = 2$ planar wavefronts impinge on a linear array of $m = 6$ equidistantly spaced sensors. The separation between adjacent elements is equal to half a wavelength. The array response vectors thus take the form (4.14), with $\omega\Delta/c = \pi$. The signals consist of a strong direct-path at the angle $\theta_1 = 0^\circ$ relative to array broadside, and a weaker multipath, whose DOA is varied from $\theta_2 = 2^\circ$ to $\theta_2 = 10^\circ$. It should be noted that the 3dB beamwidth of the array in question is about 8° . The Signal-to-Noise Ratio (SNR) is 6 dB for the direct path and 3 dB for the multipath, and the signal waveforms have a 99% correlation with correlation phase 0° at the first sensor. The normalized signal covariance matrix is thus

$$\mathbf{S}/\sigma^2 = \begin{bmatrix} 4.0 & 2.8 \\ 2.8 & 2.0 \end{bmatrix}. \quad (4.145)$$

In the simulations, $N = 100$ independent snapshots are generated according to the model (4.12) and the displayed results are based on 500 independent trials (unless otherwise explicitly stated).

Example 4.1 Estimation Accuracy

In the first example, the estimation accuracy is considered for the scenario described above. Figures 4.3 depict the theoretical and empirical RMS (Root-Mean-Square) errors for $\hat{\theta}_1$ and $\hat{\theta}_2$ versus the location of the second source. The theoretical RMS error asserts the quality of the local minimum that is closest to the true DOA values. Hence, the scoring techniques are initialized at the true values in this example. A more realistic initialization procedure is considered in Example 4.3. Notice that the empirical RMS errors are virtually identical for the SML and WSF methods and that these techniques

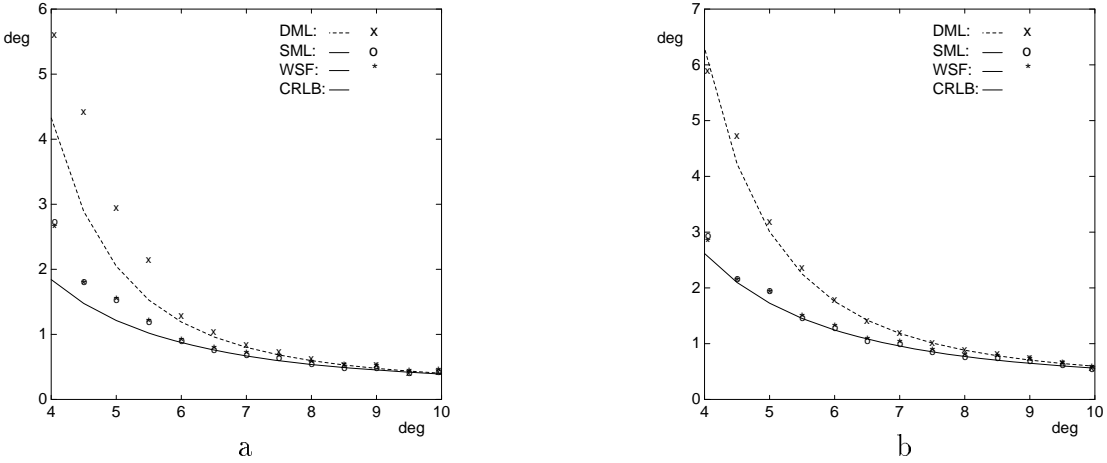


Figure 4.3: RMS error of $\hat{\theta}_1$ (a) and $\hat{\theta}_2$ (b) versus θ_2 for the SML, ML, and WSF methods. The lines represent theoretical values and the symbols 'x', 'o', and '*' are simulation results.

clearly outperform the DML method in this scenario. The difference is particularly prominent in the threshold region. This is due to the high signal correlation – the methods have nearly identical performance for uncorrelated signals, [44]. Notice also the good agreement between the theoretical and empirical curves. It has been observed from several simulations that the asymptotic formulas predict the actual variance accurately whenever the theoretical standard deviation of nearby sources is less than half the angle separation. \square

Example 4.2 Global Properties

Next, the global shape of the different cost functions is examined. The scoring algorithms described in Section 4.5 are initialized on a coarse grid of θ_1 - and θ_2 -values to examine which starting points lead to the global minimum. It has empirically been observed that for a ULA, this “domain of attraction” depends essentially only upon the number of sensors in the array. Hence, we fix $\theta_1 = 0^\circ$, $\theta_2 = 4^\circ$ and vary m in this case. To obtain reproducible results, the asymptotic ($N \rightarrow \infty$) criteria are used, noting that the results

will be similar for all values of N above a certain threshold. Figure 4.4 shows the results for $m = 4$ and $m = 6$, whereas the cases of $m = 8$ and $m = 10$ are presented in Figure 4.5. The region of attraction is identified as the “completely filled” area surrounding the global minimum. From the figures we see that the SML and WSF methods have similar regions of attraction in the scenarios in question and the region for the DML method is smaller in all cases. It is also interesting to observe that for the WSF algorithm, there are a large number of starting points further away from the minimum that also lead to global convergence. However, the number of such points is reduced as m is increased. It is quite clear from this example that proper initialization of the search methods becomes more crucial with increasing array aperture. In [90], this effect is studied in some detail for the related problem of estimating the frequencies of superimposed sinusoids. The asymptotic (for large m) radius of the domain of attraction for the DML algorithm is found to be inversely proportional to m . This is in good agreement with the empirical results presented in this example. This observation suggests that for large m , the initial estimates should have a variance of order $o(1/m^2)$ (for a ULA) to guarantee global convergence of the search algorithm. Simulation results presented in [64] (see also below) indicate that this is indeed the case for the AP initialization scheme. \square

Example 4.3 Initialization and Number of Iterations

In this example, the scoring techniques for minimizing the different cost functions are studied in some more detail. The alternating projection initialization scheme suggested in [68] is used for initializing the scoring methods described in Section 4.5. The original version is used for the SML and DML methods, whereas the same scheme with $\hat{\mathbf{R}}$ replaced by $\hat{\mathbf{E}}_s \hat{\mathbf{W}}_{opt} \hat{\mathbf{E}}_s^H$ is used for the WSF technique. The overall performance is assessed by displaying the probability of reaching the global minimum (Figure 4.6a) and the average number of iterations necessary to fulfill the stopping condition (Figure 4.6b) versus θ_2 . The iterations terminated successfully when $|\mathbf{H}^{-1}V'| < 0.01^\circ$ in the vast majority of the trials (more than 96% for $\theta_2 \geq 3^\circ$). In a small number of trials in the threshold region, the maximum number of iterations (set to 30) was reached or the search terminated at a point where no improvement could be found along the scoring direction ($\mu_k < 0.0001$).

From Figure 4.6a, it is clear that the AP initialization scheme performs excellent for the WSF method – the global minimum was reached in *all* trials. However, a large number of iterations is necessary in the difficult cases where the angle separation is less than 3° . The AP initialization performs satisfactory also for the DML method and somewhat worse, but still acceptable for the SML technique. Recall that the RMS error for $\hat{\theta}_2$ is more than half the angle separation when $\theta_2 < 4.5^\circ$. Notice also that for $\theta_2 > 4^\circ$, the WSF-scoring technique converges in a smaller number of iterations than the DML and SML scoring methods. Still, the DML-scoring technique performs significantly better than the modified Gauss-Newton technique suggested in [64]. \square

Example 4.4 Detection Performance

In the final example, we compare the performance of the various detection procedures outlined in Section 4.6. The scenario is the same as in the previous examples. The estimates are calculated as in Example 4.3. Figure 4.7 displays the probability of correct detection versus the location of the second source for the different detection schemes. The detection results from the MDLC method of [89] are included for comparison. This scheme is based on the DML cost function and relies on the DML estimates. The WSF detection scheme requires information of the signal subspace dimension. The MDL test of [79] is used for this purpose, which results in the value $d' = 1$ in *all* trials. As seen from Figure 4.7, the WSF detection scheme has a slightly lower threshold than the GLRT based technique, whereas the MDLC criterion performs clearly worse. Note that as θ_2 increases, the detection probabilities for the WSF and GLR tests approach the value preset by the threshold. Another important observation is that the threshold value of the angle separation for the WSF and GLRT detection schemes (about 4° for WSF and 4.5° for GLRT) is lower than the value at which reliable estimates is obtained (e.g., when the standard deviation of both angle estimates is lower than half the angle separation), cf. Figure 4.3b. On the other hand, the value at which the global minimum can be reached using the AP initialization scheme is still lower for the WSF algorithm (less than 2°), whereas it is similar to the detection threshold for the SML method and the GLR test, cf. Figure 4.6a. One may suspect that the initialization procedure indeed is the limiting

factor for the GLR test. This is, however, not the case, since for $\theta_2 \leq 3.5^\circ$ essentially all errors are due to *underestimation* of d , so the detection performance cannot be improved upon by using a more efficient initialization procedure. In summary, this observation suggests that the estimation accuracy is indeed the limiting factor for both the WSF and the SML methods if the proposed schemes for calculating the estimates and determining the number of sources are used. However, the estimation accuracy is the best possible in terms of estimation error variance. Of course, no general conclusions can be drawn from the rather limited scenario presented in this study, but similar observations have been made also in other cases, see e.g., [64]. \square

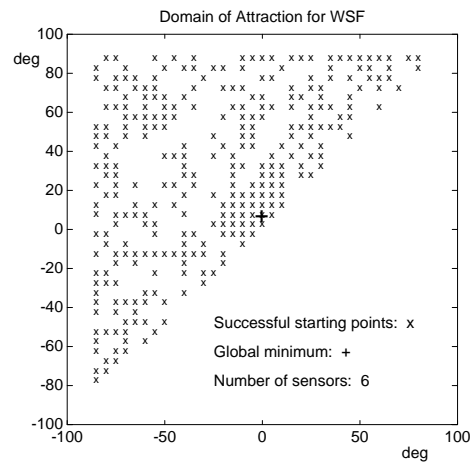
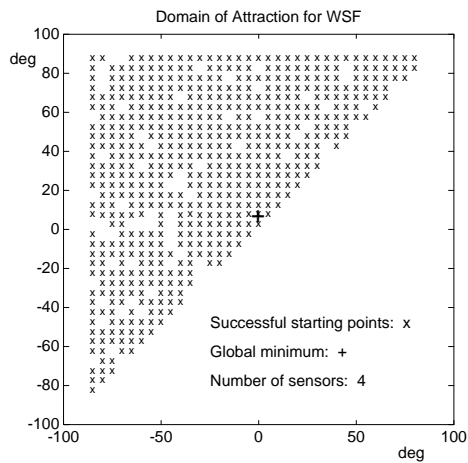
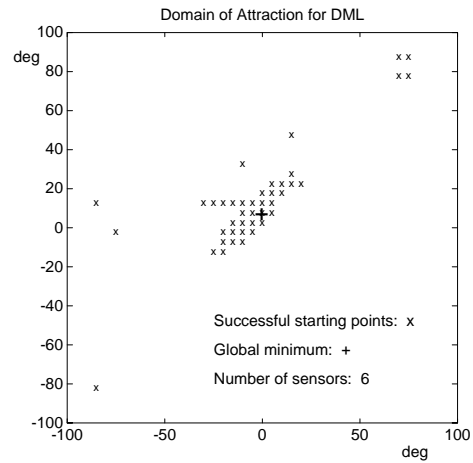
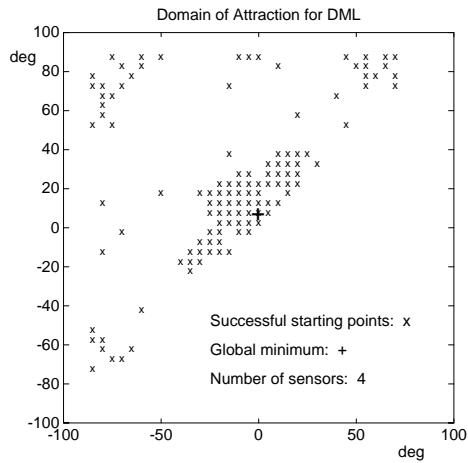
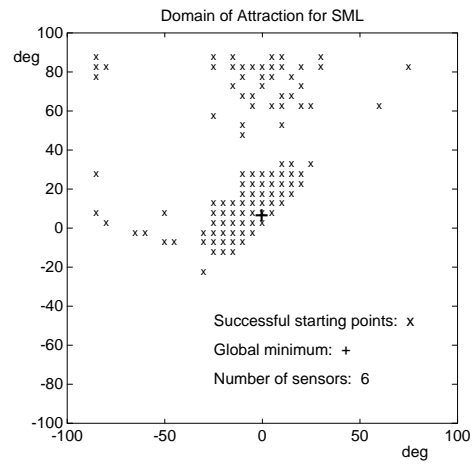
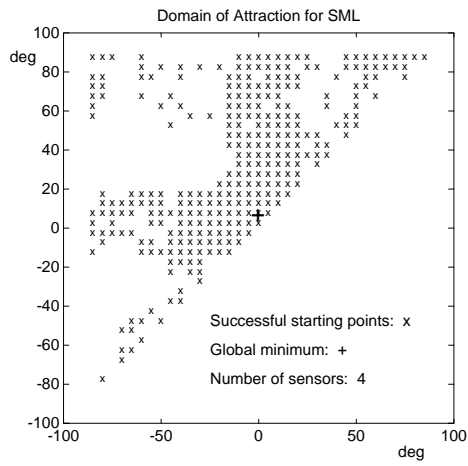


Figure 4.4: Starting values that lead to the global minimum for the scoring technique applied to the SML, DML, and WSF methods. The horizontal axis represents θ_1 and the vertical axis is θ_2 .

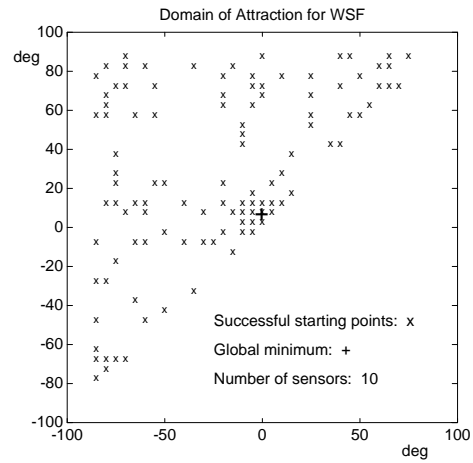
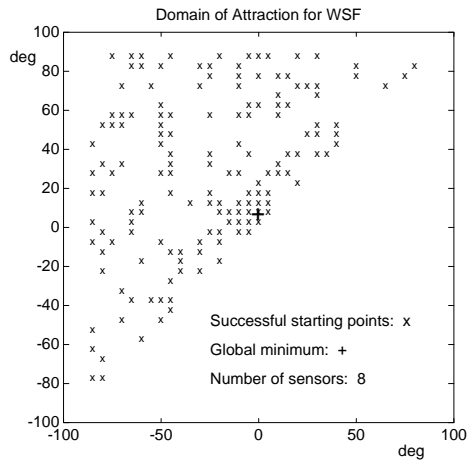
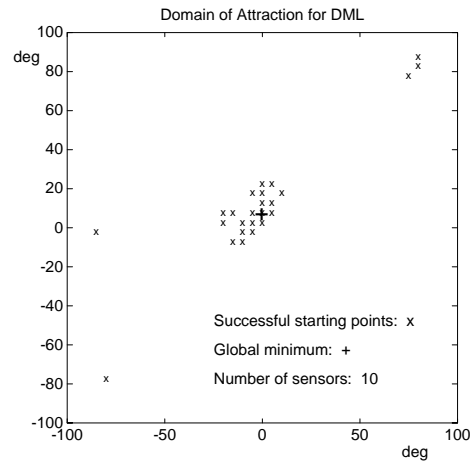
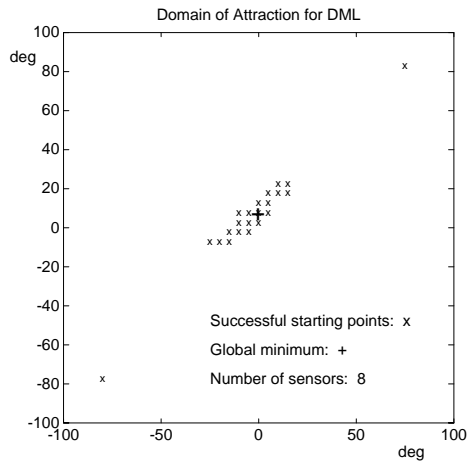
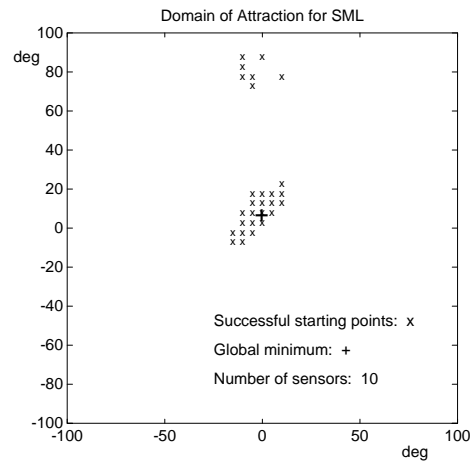
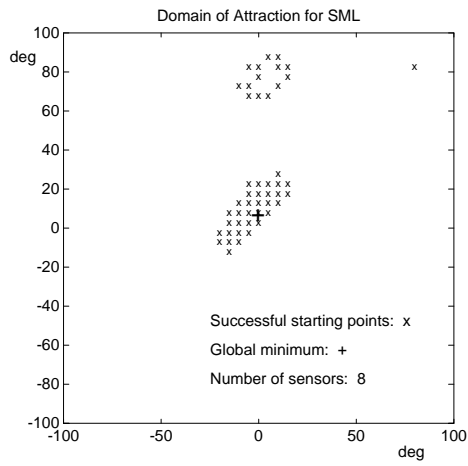


Figure 4.5: Starting values that lead to the global minimum for the scoring technique applied to the SML, DML, and WSF methods. The horizontal axis represents θ_1 and the vertical axis is θ_2 .

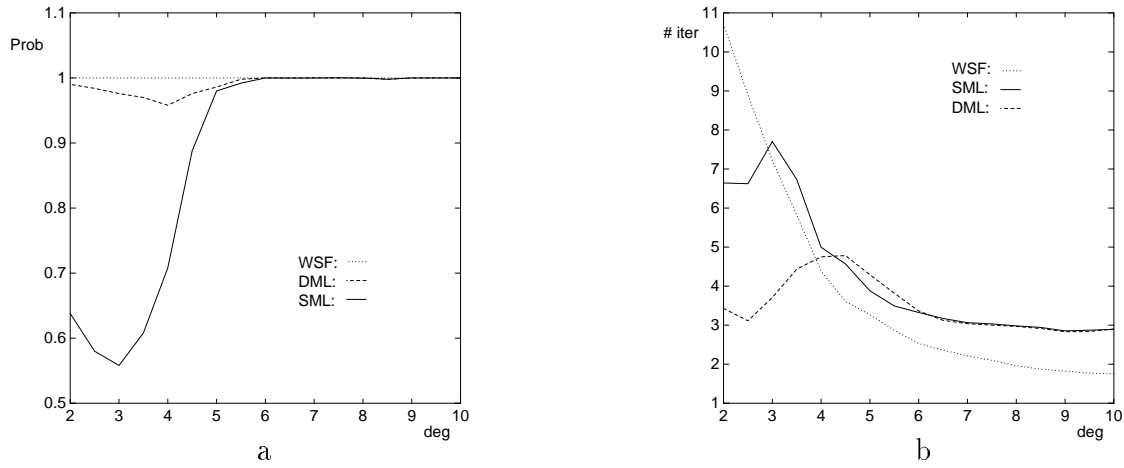


Figure 4.6: Probability of global convergence (a) and average number of iterations (b) versus θ_2 for the SML, DML, and WSF methods.

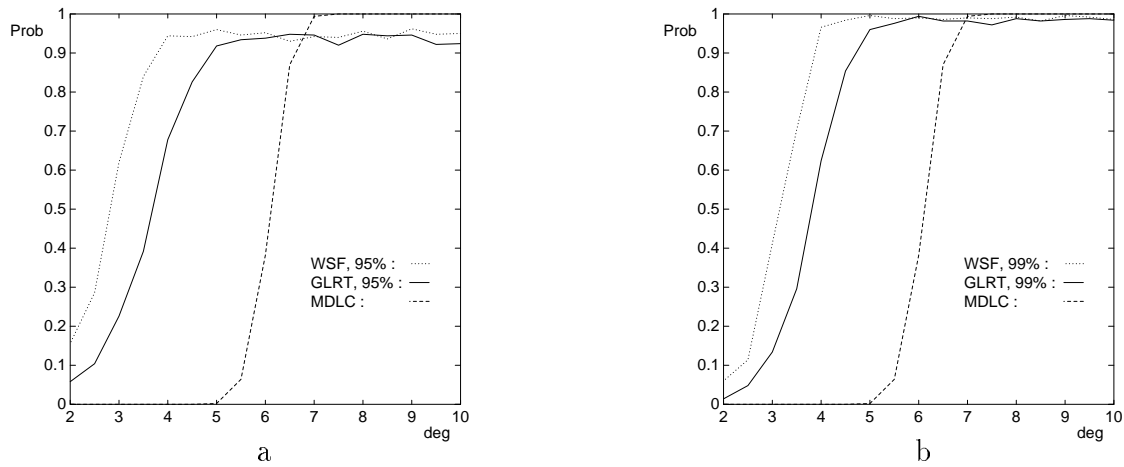


Figure 4.7: Probability of correct detection using 95% (a) and 99% (b) confidence levels versus θ_2 for the GLRT and WSF detection schemes. The results for the MDLC technique is the same in both plots, since it requires no threshold setting.

4.8 Conclusions

This chapter presents exact and large sample realizations of ML estimation techniques for signal parameter estimation from array data. Two different signal models (deterministic vs. stochastic) are considered and the ML estimator is formulated for each model. The asymptotic properties of these methods are examined and related to the corresponding Cramér-Rao lower bounds. The stochastic ML (SML) method is found to provide more accurate estimates than the deterministic ML (DML) technique, independent of the actual emitter signal waveforms. Under the Gaussian signal assumption, the SML method is asymptotically efficient, i.e., the estimation error attains the CRLB.

Two classes of multidimensional subspace or eigenstructure techniques are presented, the noise and signal subspace formulations. The relation between these is established by comparing their asymptotic distributions. The subspace methods that are asymptotically identical to the ML techniques are identified. In the case of coherent signals (specular multipath), only the signal subspace version can yield optimal (SML) performance. This optimal signal subspace method is termed WSF (Weighted Subspace Fitting).

All methods considered herein require a multidimensional search to find the parameter estimates. Newton-type descent algorithms for calculating the ML and signal subspace estimates are presented, and their implementation and initialization is discussed. The Newton-type algorithm for the signal subspace formulation requires $O(md^2)$ flops per iteration, whereas both ML implementations cost $O(m^2d)$. Consequently, the subspace technique offers a computational advantage when the number of emitters (d) is less than the number of sensors (m).

A generalized likelihood ratio test (GLRT) for determining the number of emitters is derived, as well as a detection scheme based on the WSF criterion. The methods are similar in structure, and require a threshold which sets the (asymptotic) probability of detection.

Numerical examples and simulations are presented, comparing the various detection and estimation algorithms. Examples studying empirical vs. theoretical estimation accuracy, global convergence properties, average number of iterations, and detection

performance are included. The performance of the WSF and SML-GLRT techniques is found to be limited only by the estimation accuracy. In other words, global convergence and correct detection is obtained with high probability whenever accurate parameter estimation is possible. In these cases, the estimation error variance is found to be in good agreement with the theoretical variance, based on the asymptotic covariance matrix.

Appendix A Differentiation of the projection matrix

In this appendix, the expressions for the first and second derivatives of the projection matrix $\mathbf{P}_{\mathbf{A}}(\boldsymbol{\theta}) = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H = \mathbf{A} \mathbf{A}^\dagger$ with respect to the elements in $\boldsymbol{\theta}$ are derived. These results can also be found in [54]. For ease of notation we write \mathbf{P} instead of $\mathbf{P}_{\mathbf{A}}(\boldsymbol{\theta})$.

Consider the first derivative

$$\mathbf{P}_i = \frac{\partial \mathbf{P}}{\partial \theta_i} = \mathbf{A}_i \mathbf{A}^\dagger + \mathbf{A} \mathbf{A}_i^\dagger. \quad (\text{A.1})$$

After some algebraic manipulations the following is obtained for the pseudo-inverse

$$\mathbf{A}_i^\dagger = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}_i^H \mathbf{P}^\perp - \mathbf{A}^\dagger \mathbf{A}_i \mathbf{A}^\dagger. \quad (\text{A.2})$$

Combining (A.1) and (A.2) gives

$$\mathbf{P}_i = \mathbf{P}^\perp \mathbf{A}_i \mathbf{A}^\dagger + (\dots)^H, \quad (\text{A.3})$$

where the notation $(\dots)^H$ means that the same expression appears again with complex conjugate and transpose. From (A.3) it is easily verified that $\text{Tr}(\mathbf{P}_i) = 0$ as expected since the trace of a projection matrix depends only on the dimension of the subspace onto which it projects.

The second derivative is given by

$$\mathbf{P}_{ij} = \mathbf{P}_j^\perp \mathbf{A}_i \mathbf{A}^\dagger + \mathbf{P}^\perp \mathbf{A}_{ij} \mathbf{A}^\dagger + \mathbf{P}^\perp \mathbf{A}_i \mathbf{A}_j^\dagger + (\dots)^H. \quad (\text{A.4})$$

Using (A.2) and that $\mathbf{P}_j^\perp = -\mathbf{P}_j$ gives

$$\begin{aligned} \mathbf{P}_{ij} = & -\mathbf{P}^\perp \mathbf{A}_j \mathbf{A}^\dagger \mathbf{A}_i \mathbf{A}^\dagger - \mathbf{A}^{+H} \mathbf{A}_j^H \mathbf{P}^\perp \mathbf{A}_i \mathbf{A}^\dagger + \mathbf{P}^\perp \mathbf{A}_{ij} \mathbf{A}^\dagger + \\ & + \mathbf{P}^\perp \mathbf{A}_i (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}_j^H \mathbf{P}^\perp - \mathbf{P}^\perp \mathbf{A}_i \mathbf{A}^\dagger \mathbf{A}_j \mathbf{A}^\dagger + (\dots)^H. \end{aligned} \quad (\text{A.5})$$

Appendix B Asymptotic Distribution of the WSF Criterion

A proof of Theorem 4.8 is provided in this appendix. The distribution of $V(\boldsymbol{\theta}_0)$ is based on the asymptotic distribution of the signal eigenvectors of the sample covariance. From the asymptotic theory of principal components [91, 92], we have the following result.

Lemma 4.2 *The d' largest eigenvectors of $\hat{\mathbf{R}}$ are asymptotically (for large N) normally distributed with means and covariances given by*

$$\mathbb{E}\{\hat{\mathbf{e}}_k\} = \mathbf{e}_k + O(N^{-1}) \quad (\text{B.1})$$

$$\mathbb{E}\{(\hat{\mathbf{e}}_k - \mathbf{e}_k)(\hat{\mathbf{e}}_l - \mathbf{e}_l)^H\} = \delta_{kl} \frac{\lambda_k}{N} \sum_{i=1, i \neq k}^m \frac{\lambda_i}{(\lambda_i - \lambda_k)^2} \mathbf{e}_i \mathbf{e}_i^H + o(N^{-1}) \quad (\text{B.2})$$

$$\mathbb{E}\{(\hat{\mathbf{e}}_k - \mathbf{e}_k)(\hat{\mathbf{e}}_l - \mathbf{e}_l)^T\} = (1 - \delta_{kl}) \frac{-\lambda_k \lambda_l}{N(\lambda_k - \lambda_l)^2} \mathbf{e}_l \mathbf{e}_k^T + o(N^{-1}) . \quad (\text{B.3})$$

□

It is convenient to introduce a set of transformed variables

$$\hat{\mathbf{u}}_k = \mathbf{Z}^H \hat{\mathbf{e}}_k, \quad k = 1, \dots, d' \quad (\text{B.4})$$

where \mathbf{Z} is an $m \times (m - d)$ full rank, square root factor of $\mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta}_0)$, i.e.,

$$\mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta}_0) = \mathbf{Z}\mathbf{Z}^H. \quad (\text{B.5})$$

Note that $\mathbf{Z}^H \mathbf{Z} = \mathbf{I}$, since $\mathbf{P}_{\mathbf{A}}^\perp$ is an idempotent matrix. Clearly, $\hat{\mathbf{u}}_k$, $k = 1, \dots, d'$ have a limiting normal distribution. From (B.1)–(B.3), the first and second moments are given by

$$\mathbb{E}\{\hat{\mathbf{u}}_k\} = \mathbf{0} \quad (\text{B.6})$$

$$\mathbb{E}\{\hat{\mathbf{u}}_k \hat{\mathbf{u}}_l^H\} = \delta_{kl} \frac{\sigma^2 \lambda_k}{N(\lambda_k - \sigma^2)^2} \mathbf{I} \quad (\text{B.7})$$

$$\mathbb{E}\{\hat{\mathbf{u}}_k \hat{\mathbf{u}}_l^T\} = \mathbf{0} . \quad (\text{B.8})$$

From (4.85), we can rewrite the normalized WSF cost function as

$$\begin{aligned}
\frac{2N}{\hat{\sigma}^2} V_{WSF}(\boldsymbol{\theta}_0) &= \frac{2N}{\hat{\sigma}^2} \text{Tr}\{\mathbf{Z}\mathbf{Z}^H \hat{\mathbf{E}}_s \hat{\mathbf{W}}_{opt} \hat{\mathbf{E}}_s^H\} \\
&= \frac{2N}{\hat{\sigma}^2} \sum_{k=1}^{d'} \hat{w}_k \hat{\mathbf{u}}_k^H \hat{\mathbf{u}}_k \\
&= \frac{2N}{\hat{\sigma}^2} \hat{\mathbf{u}}^H \hat{\mathbf{u}} \\
&= \frac{2N}{\hat{\sigma}^2} (\bar{\mathbf{u}}^T \bar{\mathbf{u}} + \tilde{\mathbf{u}}^T \tilde{\mathbf{u}}), \tag{B.9}
\end{aligned}$$

where $\hat{w}_k = (\hat{\lambda}_k - \hat{\sigma}^2)^2 / \hat{\lambda}_k$, $\hat{\mathbf{u}} = [\sqrt{\hat{w}_1} \hat{\mathbf{u}}_1^T \cdots \sqrt{\hat{w}_{d'}} \hat{\mathbf{u}}_{d'}^T]^T$, $\bar{\mathbf{u}} = \text{Re}(\hat{\mathbf{u}})$, and $\tilde{\mathbf{u}} = \text{Im}(\hat{\mathbf{u}})$. Since $\hat{\mathbf{u}}$ is of order $O_p(N^{-1/2})$ for large N , we can replace the consistent estimates of σ^2 and λ_k above with the true quantities without changing the asymptotic properties. The assertion on the asymptotic distribution of $V_{WSF}(\boldsymbol{\theta}_0)$ then follows from the fact that (B.9) is a sum of $2d'(m-d)$ squared random variables that are asymptotically independent and $N(0,1)$.

To determine the distribution of the cost function evaluated at the parameter estimate, consider the Taylor series expansions

$$V(\hat{\boldsymbol{\theta}}) = V(\boldsymbol{\theta}_0) + V'^T(\boldsymbol{\theta}_0) \tilde{\boldsymbol{\theta}} + \frac{1}{2} \tilde{\boldsymbol{\theta}}^T V''(\boldsymbol{\theta}_0) \tilde{\boldsymbol{\theta}} + o(|\tilde{\boldsymbol{\theta}}|^2) \tag{B.10}$$

$$V'(\hat{\boldsymbol{\theta}}) = \mathbf{0} = V'(\boldsymbol{\theta}_0) + V''(\boldsymbol{\theta}_0) \tilde{\boldsymbol{\theta}} + o(|\tilde{\boldsymbol{\theta}}|) \tag{B.11}$$

(where the subscript has been dropped from $V_{WSF}(\boldsymbol{\theta})$ for notational convenience). Here, $\tilde{\boldsymbol{\theta}}$ denotes the estimation error, $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$. The regularity assumptions on the array response vectors guarantee that $V''(\boldsymbol{\theta}_0)$ converges (w.p.1) as $N \rightarrow \infty$. The limiting Hessian, denoted \mathbf{H} , is assumed to be non-singular. Equations (B.10)–(B.11) can be written

$$V(\hat{\boldsymbol{\theta}}) = V(\boldsymbol{\theta}_0) + V'^T(\boldsymbol{\theta}_0) \tilde{\boldsymbol{\theta}} + \frac{1}{2} \tilde{\boldsymbol{\theta}}^T \mathbf{H} \tilde{\boldsymbol{\theta}} + o(|\tilde{\boldsymbol{\theta}}|^2), \tag{B.12}$$

$$V'(\hat{\boldsymbol{\theta}}) = \mathbf{0} = V'(\boldsymbol{\theta}_0) + \mathbf{H} \tilde{\boldsymbol{\theta}} + o(|\tilde{\boldsymbol{\theta}}|). \tag{B.13}$$

The “in-probability” rate of convergence of the estimation error is $O_p(N^{-1/2})$. Thus, inserting (B.13) into (B.12) yields

$$V(\hat{\boldsymbol{\theta}}) \simeq V(\boldsymbol{\theta}_0) - V'^T(\boldsymbol{\theta}_0) \mathbf{H}^{-1} V'(\boldsymbol{\theta}_0) + \frac{1}{2} V'^T(\boldsymbol{\theta}_0) \mathbf{H}^{-1} V'(\boldsymbol{\theta}_0)$$

$$= V(\boldsymbol{\theta}_0) - \frac{1}{2}V'^T(\boldsymbol{\theta}_0)\mathbf{H}^{-1}V'(\boldsymbol{\theta}_0), \quad (\text{B.14})$$

where the approximation is of order $o_p(N^{-1})$. Consider the second term of (B.14). The i^{th} component of the gradient is given in (4.93). Since $\mathbf{P}_{\mathbf{A}}^\perp(\boldsymbol{\theta}_0)\hat{\mathbf{E}}_s$ is of order $O(N^{-1/2})$, we can write

$$\begin{aligned} \frac{\partial}{\partial \theta_i} V(\boldsymbol{\theta}_0) &= -2\text{Re} \left\{ \sum_{k=1}^{d'} \hat{w}_k \mathbf{e}_k^H \mathbf{A}^\dagger \mathbf{A}_i^H \mathbf{P}_{\mathbf{A}}^\perp \hat{\mathbf{e}}_k \right\} + o_p(N^{-1/2}) \\ &= -2\text{Re} \left\{ \sum_{k=1}^{d'} \hat{w}_k \mathbf{e}_k^H \mathbf{A}^\dagger \mathbf{A}_i^H \mathbf{Z} \hat{\mathbf{u}}_k \right\} \\ &= \text{Re} \{ \mathbf{T}_i \hat{\mathbf{u}} \} + o_p(N^{-1/2}), \end{aligned} \quad (\text{B.15})$$

where

$$\mathbf{T}_i = \left[-2\sqrt{w_1} \mathbf{e}_1^H \mathbf{A}^\dagger \mathbf{A}_i^H \mathbf{Z}, \dots, -2\sqrt{w_{d'}} \mathbf{e}_{d'}^H \mathbf{A}^\dagger \mathbf{A}_i^H \mathbf{Z} \right]. \quad (\text{B.16})$$

Thus, $V'(\boldsymbol{\theta}_0)$ can be written as

$$V'(\boldsymbol{\theta}_0) = \text{Re} \{ \mathbf{T} \hat{\mathbf{u}} \} + o_p(N^{-1/2}) = \bar{\mathbf{T}} \bar{\mathbf{u}} - \tilde{\mathbf{T}} \tilde{\mathbf{u}} + o_p(N^{-1/2}), \quad (\text{B.17})$$

where $\mathbf{T} = [\mathbf{T}_1^T \dots \mathbf{T}_{pd}^T]^T$, $\bar{\mathbf{T}} = \text{Re}(\mathbf{T})$, and $\tilde{\mathbf{T}} = \text{Im}(\mathbf{T})$. Using (B.9) and (B.17), we can rewrite Equation (B.14) as

$$V(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \bar{\mathbf{u}}^T & \tilde{\mathbf{u}}^T \end{bmatrix} \left(\mathbf{I} - \frac{1}{2} \begin{bmatrix} \bar{\mathbf{T}}^T \\ -\tilde{\mathbf{T}}^T \end{bmatrix} \mathbf{H}^{-1} \begin{bmatrix} \bar{\mathbf{T}} & -\tilde{\mathbf{T}} \end{bmatrix} \right) \begin{bmatrix} \bar{\mathbf{u}} \\ \tilde{\mathbf{u}} \end{bmatrix} + o_p(N^{-1}). \quad (\text{B.18})$$

The matrices \mathbf{T} and \mathbf{H} are related using $\sigma^2 \mathbf{H} = \mathbf{Q}$ (cf. the proof of Corollary 4.1a) together with (B.17)

$$\begin{aligned} \sigma^2 \mathbf{H} &= \lim_{N \rightarrow \infty} N \text{E} \{ V'(\boldsymbol{\theta}_0) V'(\boldsymbol{\theta}_0)^T \} \\ &= \lim_{N \rightarrow \infty} N \text{E} \{ (\bar{\mathbf{T}} \bar{\mathbf{u}} - \tilde{\mathbf{T}} \tilde{\mathbf{u}}) (\bar{\mathbf{T}} \bar{\mathbf{u}} - \tilde{\mathbf{T}} \tilde{\mathbf{u}})^T \} \\ &= \lim_{N \rightarrow \infty} N (\bar{\mathbf{T}} \text{E} \{ \bar{\mathbf{u}} \bar{\mathbf{u}}^T \} \bar{\mathbf{T}}^T - \bar{\mathbf{T}} \text{E} \{ \bar{\mathbf{u}} \tilde{\mathbf{u}}^T \} \tilde{\mathbf{T}}^T \\ &\quad - \tilde{\mathbf{T}} \text{E} \{ \tilde{\mathbf{u}} \bar{\mathbf{u}}^T \} \bar{\mathbf{T}}^T + \tilde{\mathbf{T}} \text{E} \{ \tilde{\mathbf{u}} \tilde{\mathbf{u}}^T \} \tilde{\mathbf{T}}^T) \\ &= \frac{\sigma^2}{2} (\bar{\mathbf{T}} \bar{\mathbf{T}}^T + \tilde{\mathbf{T}} \tilde{\mathbf{T}}^T), \end{aligned} \quad (\text{B.19})$$

where we have used that $E\{\tilde{\mathbf{u}}\tilde{\mathbf{u}}^T\} = o(N^{-1})$ and $E\{\bar{\mathbf{u}}\bar{\mathbf{u}}^T\} = E\{\tilde{\mathbf{u}}\tilde{\mathbf{u}}^T\} = \sigma^2/2N + o(N^{-1})$.

Substitute (B.19) in (B.18) to obtain

$$\begin{aligned}
V(\hat{\boldsymbol{\theta}}) &= \begin{bmatrix} \bar{\mathbf{u}}^T & \tilde{\mathbf{u}}^T \end{bmatrix} \left(\mathbf{I} - \begin{bmatrix} \bar{\mathbf{T}}^T \\ -\tilde{\mathbf{T}}^T \end{bmatrix} (\bar{\mathbf{T}}\bar{\mathbf{T}}^T + \tilde{\mathbf{T}}\tilde{\mathbf{T}}^T)^{-1} \begin{bmatrix} \bar{\mathbf{T}} & -\tilde{\mathbf{T}} \end{bmatrix} \right) \begin{bmatrix} \bar{\mathbf{u}} \\ \tilde{\mathbf{u}} \end{bmatrix} + o_p(N^{-1}) \\
&= \begin{bmatrix} \bar{\mathbf{u}}^T & \tilde{\mathbf{u}}^T \end{bmatrix} (\mathbf{I} - \mathbf{X}) \begin{bmatrix} \bar{\mathbf{u}} \\ \tilde{\mathbf{u}} \end{bmatrix} + o_p(N^{-1}). \tag{B.20}
\end{aligned}$$

We recognize the matrix \mathbf{X} above as a pd -dimensional orthogonal projection in $\mathbb{R}^{2d'(m-d)}$. Thus $\mathbf{I} - \mathbf{X}$ is a $(2d'(m-d) - pd)$ -dimensional projection matrix. Hence, $(2N/\hat{\sigma}^2)V_{WSF}(\hat{\boldsymbol{\theta}})$ can be written as a sum of the square of $2d'(m-d) - pd$ random variables that are asymptotically independent and $N(0,1)$. The result follows.

References

- [1] S. P. Applebaum, “Adaptive Arrays”, Technical Report SPL TR 66-1, Syracuse University Research Corporation, August 1966.
- [2] B. Widrow, P.E. Mantez, L.J. Griffiths, and B.B. Goode, “Adaptive Antenna Systems”, *Proc. IEEE*, **55**:2143–2159, Dec. 1967.
- [3] F.C. Schwegge, “Sensor Array Data Processing for Multiple Signal Sources”, *IEEE Trans. on IT*, **IT-14**:294–305, 1968.
- [4] V.H. MacDonald and P.M. Schultheiss, “Optimum Passive Bearing Estimation in a Spatially Incoherent Noise Environment”, *J. Acoust. Soc. Am.*, **46**(1):37–43, 1969.
- [5] J. Capon, “High Resolution Frequency Wave Number Spectrum Analysis”, *Proc. IEEE*, **57**:1408–1418, 1969.
- [6] R.A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays*, Wiley, 1980.
- [7] S. Haykin, editor, *Array Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
- [8] B.D. Van Veen and K.M. Buckley, “Beamforming: A Versatile Approach to Spatial Filtering”, *IEEE ASSP Magazine*, pages 4–24, April 1988.
- [9] J. P. Burg, “Maximum Entropy Spectral Analysis”, In *Proceedings of the 37th Annual International SEG Meeting*, Oklahoma City, OK., 1967.
- [10] D.W. Tufts and R. Kumaresan, “Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood”, *Proc. IEEE*, **70**:975–989, Sept. 1982.
- [11] R. Kumaresan and D.W. Tufts, “Estimating the Angles of Arrival of Multiple Plane Waves”, *IEEE Trans. Aerosp. Electron. Syst.*, **AES-19**:134–139, Jan. 1983.
- [12] R.O. Schmidt, “Multiple Emitter Location and Signal Parameter Estimation”, In *Proc. RADC Spectrum Estimation Workshop*, pages 243–258, Rome, NY, 1979.

- [13] G. Bienvenu and L. Kopp, “Principle de la goniometrie passive adaptive”, In *Proc. 7^{eme} Colloque GRESIT*, pages 106/1–106/10, Nice, France, 1979.
- [14] G. Bienvenu and L. Kopp, “Adaptivity to Background Noise Spatial Coherence for High Resolution Passive Methods”, In *Proc. IEEE ICASSP*, pages 307–310, Denver, CO., 1980.
- [15] N.L. Owsley, “Data Adaptive Orthonormalization”, In *Proc. ICASSP 78*, pages 109–112, Tulsa, OK, 1978.
- [16] V.F. Pisarenko, “The retrieval of harmonics from a covariance function”, *Geophys. J. Roy. Astron. Soc.*, **33**:347–366, 1973.
- [17] G. Bienvenu and L. Kopp, “Optimality of High Resolution Array Processing”, *IEEE Trans. ASSP*, **ASSP-31**:1235–1248, Oct. 1983.
- [18] A. Paulraj, R. Roy, and T. Kailath, “A Subspace Rotation Approach to Signal Parameter Estimation”, *Proceedings of the IEEE*, pages 1044–1045, July 1986.
- [19] S.Y. Kung, C.K. Lo, and R. Foka, “A Toeplitz Approximation Approach to Coherent Source Direction Finding”, In *Proc. ICASSP 86*, 1986.
- [20] S.J. Orfanidis, “A Reduced MUSIC Algorithm”, In *Proc. IEEE ASSP 3rd Workshop Spectrum Est. Modeling*, pages 165–167, Boston, MA, Nov. 1986.
- [21] W. J. Bangs, *Array Processing with Generalized Beamformers*, PhD thesis, Yale University, New Haven, CT, 1971.
- [22] M. Wax, *Detection and Estimation of Superimposed Signals*, PhD thesis, Stanford Univ., Stanford, CA, March 1985.
- [23] J.F. Böhme, “Estimation of Source Parameters By Maximum Likelihood and Non-linear Regression”, In *Proc. ICASSP 84*, pages 7.3.1–7.3.4, 1984.
- [24] J.F. Böhme, “Estimation of Spectral Parameters of Correlated Signals in Wavefields”, *Signal Processing*, **10**:329–337, 1986.

- [25] P. Stoica and A. Nehorai, “MUSIC, Maximum Likelihood and Cramér-Rao Bound”, In *Proc. ICASSP 88 Conf*, pages 2296–2299, New York, April 1988.
- [26] B. Ottersten and L. Ljung, “Asymptotic Results for Sensor Array Processing”, In *Proc. ICASSP 89*, pages 2266–2269, Glasgow, Scotland, May 1989.
- [27] B. Ottersten and M. Viberg, “Analysis of Subspace Fitting Based Methods for Sensor Array Processing”, In *Proc. ICASSP 89*, pages 2807–2810, Glasgow, Scotland, May 1989.
- [28] R. O. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*, PhD thesis, Stanford Univ., Stanford, CA, Nov. 1981.
- [29] B. Ottersten, B. Wahlberg, M. Viberg, and T. Kailath, “Stochastic Maximum Likelihood Estimation in Sensor Arrays by Weighted Subspace Fitting”, In *Proc. 23rd Asilomar Conf. Sig., Syst., Comput.*, pages 599–603, Monterey, CA, Nov. 1989.
- [30] P. Stoica and A. Nehorai, “MUSIC, Maximum Likelihood and Cramér-Rao Bound”, *IEEE Trans. ASSP*, **ASSP-37**:720–741, May 1989.
- [31] 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.
- [32] P. Stoica and A. Nehorai, “Performance Study of Conditional and Unconditional Direction-of-Arrival Estimation”, *IEEE Trans. ASSP*, **ASSP-38**:1783–1795, October 1990.
- [33] K. Sharman, T.S. Durrani, M. Wax, and T. Kailath, “Asymptotic Performance of Eigenstructure Spectral Analysis Methods”, In *Proc. ICASSP 84 Conf*, pages 45.5.1–45.5.4, San Diego, CA, March 1984.
- [34] D.J. Jeffries and D.R. Farrier, “Asymptotic Results for Eigenvector Methods”, *IEE Proceedings F*, **132**(7):589–594, June 1985.

- [35] M. Kaveh and A. J. Barabell, “The Statistical Performance of the MUSIC and the Minimum-Norm Algorithms in Resolving Plane Waves in Noise”, *IEEE Trans. ASSP*, **ASSP-34**:331–341, April 1986.
- [36] B. Porat and B. Friedlander, “Analysis of the Asymptotic Relative Efficiency of the MUSIC Algorithm”, *IEEE Trans. ASSP*, **ASSP-36**:532–544, April 1988.
- [37] B. D. Rao and K. V. S. Hari, Performance analysis of root-music, *IEEE Trans. ASSP*, **ASSP-37**(12):1939–1949, Dec. 1989.
- [38] B. D. Rao and K. V. S. Hari, Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise, *IEEE Trans. ASSP*, **ASSP-37**(12):1990–1995, Dec. 1989.
- [39] P. Stoica and A. Nehorai, “MUSIC, Maximum Likelihood and Cramér-Rao Bound: Further Results and Comparisons”, *IEEE Trans. ASSP*, **ASSP-38**:2140–2150, December 1990.
- [40] P. Stoica and A. Nehorai, “Performance Comparison of Subspace Rotation and MUSIC Methods for Direction Estimation”, *IEEE Trans. ASSP*, Feb. 1991.
- [41] P. Stoica and K. Sharman, “A Novel Eigenanalysis Method for Direction Estimation”, *Proc. IEE*, **F**:19–26, Feb 1990.
- [42] P. Stoica and K. Sharman, “Maximum Likelihood Methods for Direction-of-Arrival Estimation”, *IEEE Trans. ASSP*, **ASSP-38**:1132–1143, July 1990.
- [43] D. Starer, *Algorithms for Polynomial-Based Signal Processing*, PhD thesis, Yale University, New Haven, CT, May 1990.
- [44] M. Viberg and B. Ottersten, “Sensor Array Processing Based on Subspace Fitting”, *IEEE Trans. ASSP*, **ASSP-39**, May 1991.
- [45] B. Ottersten, M. Viberg, and T. Kailath, “Performance Analysis of the Total Least Squares ESPRIT Algorithm”, *IEEE Trans. on ASSP*, **ASSP-39**, May 1991.

- [46] Y. Bresler, “Maximum Likelihood Estimation of Linearly Structured Covariance with Application to Antenna Array Processing”, In *Proc. 4th ASSP Workshop on Spectrum Estimation and Modeling*, pages 172–175, Minneapolis, MN, Aug. 1988.
- [47] B. Ottersten, R. Roy, and T. Kailath, “Signal Waveform Estimation in Sensor Array Processing”, In *Proc. 23rd Asilomar Conf. Sig., Syst., Comput.*, Nov. 1989.
- [48] M. Wax and I. Ziskind, “On Unique Localization of Multiple Sources by Passive Sensor Arrays”, *IEEE Trans. on ASSP*, **ASSP-37**(7):996–1000, July 1989.
- [49] A. Nehorai, D. Starer, and P. Stoica, “Direction-of-Arrival Estimation in Applications with Multipath and Few Snapshots”, *Circuits, Systems, and Signal Processing*, 1991, to appear.
- [50] K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979.
- [51] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd edition, John Wiley & Sons, New York, 1984.
- [52] N.R. Goodman, “Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)”, *Annals Math. Stat.*, **Vol. 34**:152–176, 1963.
- [53] A. G. Jaffer, “Maximum Likelihood Direction Finding of Stochastic Sources: A Separable Solution”, In *Proc. ICASSP 88*, volume 5, pages 2893–2896, New York, New York, April 1988.
- [54] G. Golub and V. Pereyra, “The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems Whose Variables Separate”, *SIAM J. Num. Anal.*, **10**:413–432, 1973.
- [55] S.D. Silvey, *Statistical Inference*, Penguin Books, London, 1970.
- [56] E.L. Lehmann, *Theory of Point Estimation*, John Wiley & Sons, New York, 1983.

- [57] A.J. Weiss and E. Weinstein, “Lower Bounds in Parameter Estimation – Summary of Results”, In *Proc. ICASSP 86*, pages 569–572, Tokyo, Japan, 1986.
- [58] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Ellis Horwood Ltd, Chichester, UK, 1981.
- [59] A.J. Weiss and B. Friedlander, “On the Cramér-Rao Bound for Direction Finding of Correlated Signals”, Technical report, Signal Processing Technology, Ltd., Palo Alto, CA, July 1990.
- [60] B. Ottersten, M. Viberg, and T. Kailath, “Asymptotic Robustness of Sensor Array Processing Methods”, In *Proc. ICASSP 90 Conf*, pages 2635–2638, Albuquerque, NM, April 1990.
- [61] T. Söderström and P. Stoica, *System Identification*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [62] R. H. Roy, *ESPRIT, Estimation of Signal Parameters via Rotational Invariance Techniques*, PhD thesis, Stanford Univ., Stanford, CA, Aug. 1987.
- [63] J.A. Cadzow, “A High Resolution Direction-of-Arrival Algorithm for Narrow-Band Coherent and Incoherent Sources”, *IEEE Trans. ASSP*, **ASSP-36**:965–979, July 1988.
- [64] M. Viberg, B. Ottersten, and T. Kailath, “Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting”, *IEEE Trans. ASSP*, **ASSP-39**, Nov. 1991.
- [65] T.J. Shan, M. Wax, and T. Kailath, “On Spatial Smoothing for Direction-of-Arrival Estimation of Coherent Signals”, *IEEE Trans. ASSP*, **ASSP-33**:806–811, Aug. 1985.
- [66] M. Feder and E. Weinstein, “Parameter Estimation of Superimposed Signals Using the EM Algorithm”, *IEEE Trans. ASSP*, **ASSP-36**:477–489, April 1988.

- [67] M.I. Miller and D.R. Fuhrmann, “Maximum-Likelihood Narrow-Band Direction Finding and the EM Algorithm”, *IEEE Trans. ASSP*, **ASSP-38**(9):1560–1577, Sept. 1990.
- [68] I. Ziskind and M. Wax, “Maximum Likelihood Localization of Multiple Sources by Alternating Projection”, *IEEE Trans. on ASSP*, **ASSP-36**:1553–1560, Oct. 1988.
- [69] Y. Bresler and A. Macovski, “Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise”, *IEEE Trans. ASSP*, **ASSP-34**:1081–1089, Oct. 1986.
- [70] K. Sharman, “Maximum Likelihood Estimation by Simulated Annealing”, In *Proc. ICASSP 88 Conf*, pages 2741–2744, New York, April 1988.
- [71] D. Goryn and M. Kaveh, “Neural Networks for Narrowband and Wideband Direction Finding”, In *Proc. ICASSP 88*, volume 4, pages 2164–2167, New York, New York, April 1988.
- [72] K. Sharman and G.D. McClurkin, “Genetic Algorithms for Maximum Likelihood Parameter Estimation”, In *Proc. ICASSP 89 Conf*, pages 2716–2719, Glasgow, Scotland, May 1989.
- [73] M. Wax and T. Kailath, “Optimal Localization of Multiple Sources by Passive Arrays”, *IEEE Trans. on ASSP*, **ASSP-31**:1210–1218, Oct. 1983.
- [74] J.F. Böhme and D. Kraus, “On Least Squares Methods for Direction of Arrival Estimation in the Presence of Unknown Noise Fields”, In *Proc. ICASSP 88*, pages 2833–2836, New York, N.Y., 1988.
- [75] D. Starer and A. Nehorai, “Newton Algorithms for Conditional and Unconditional Maximum Likelihood Estimation of the Parameters of Exponential Signals in Noise”, *IEEE Trans. ASSP*, 1991, to appear.
- [76] P.E. Gill, W. Murray, and M.H. Wright, *Practical Optimization*, Academic Press, London, 1981.

- [77] J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, NJ., 1983.
- [78] P.E. Gill, G.H. Golub, W. Murray, and M.A. Saunders, “Methods for Modifying Matrix Factorizations”, *Math. Comp.*, **28**(126):505–535, 1974.
- [79] M. Wax and T. Kailath, “Detection of Signals by Information Theoretic Criteria”, *IEEE Trans. on ASSP*, **ASSP-33**(2):387–392, April 1985.
- [80] L.C. Zhao, P.R. Krishnaiah, and Z.D. Bai, “On Detection of the Number of Signals in Presence of White Noise”, *J. of Multivariate Analysis*, **20**:1:1–25, 1986.
- [81] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 2nd edition, Johns Hopkins University Press, Baltimore, MD., 1989.
- [82] G. Xu and T. Kailath, “Fast Signal Subspace Decomposition Without Eigendecomposition”, In *Proc. 24th Asilomar Conf. Sig., Syst., Comput.*, Monterey, CA, November 1990.
- [83] I. Karasalo, “Estimating the Covariance Matrix by Signal Subspace Averaging ”, *IEEE Trans. ASSP*, **ASSP-34**(1):8–12, February 1986.
- [84] R. Schreiber, “Implementation of Adaptive Array Algorithms”, *IEEE Trans. ASSP*, **ASSP-34**:1038–1045, Oct. 1986.
- [85] P. Comon, “Fast Updating of A Low-Rank Approximation to a Varying Hermitean Matrix”, In *Proc. 23rd Asilomar Conf. Sig., Syst., Comput.*, pages 358–362, Monterey, CA, Nov. 1989.
- [86] A.H. Abdallah and Y.H. Hu, “Parallel VLSI Computing Array Implementation for Signal Subspace Updating Algorithm”, *IEEE Trans. ASSP*, **ASSP-37**:742–748, May 1989.
- [87] R. Roy and T. Kailath, “ESPRIT – Estimation of Signal Parameters via Rotational Invariance Techniques”, *IEEE Trans. on ASSP*, **ASSP-37**(7):984–995, July 1989.

- [88] J. Rissanen, “Modeling by Shortest Data Description”, *Automatica*, **14**:465–471, 1978.
- [89] M. Wax and I. Ziskind, “Detection of the Number of Coherent Signals by the MDL Principle”, *IEEE Trans. on ASSP*, **ASSP-37**(8):1190–1196, Aug. 1989.
- [90] P. Stoica, R.L. Moses, B. Friedlander, and T. Söderström, “Maximum Likelihood Estimation of the Parameters of Multiple Sinusoids from Noisy Measurements”, *IEEE Trans. ASSP*, **ASSP-37**:378–392, March 1989.
- [91] T.W. Anderson, “Asymptotic Theory for Principal Component Analysis”, *Ann. Math. Statist.*, **34**:122–148, 1963.
- [92] R. P. Gupta, “Asymptotic Theory for Principal Component Analysis in the Complex Case”, *Journal of the Indian Stat. Assoc.*, **Vol. 3**:97–106, 1965.