N. K. Bose and C. R. Rao, eds., *Handbook of Statistics, Vol. 10*

16

Subspace Methods for Directions-of-Arrival Estimation

A. Paulraj, B. Ottersten, R. Roy, A. Swindlehurst, G. Xu and T. Kailath

1. Introduction

In many practical signal processing problems, the objective is to estimate from noisy measurements a set of *constant* parameters upon which the underlying true signals depend. For example, estimating the directions-of-arrival (DOAs) of impinging wavefronts given the set of signals received at an antenna array is important in fields such as radar, sonar, electronic surveillance and seismic exploration. High resolution frequency estimation is important in numerous applications including Doppler radar and system identification. The quantities to be estimated are parameters (e.g., DOAs of plane waves, cisoid frequencies) upon which the observations depend, and these parameters are assumed to be *constant* over the observation interval which is long enough to collect sufficient data to ensure parameter estimates of the desired accuracy.

There have been several approaches to such problems including the so-called Capon's maximum likelihood (ML) method [12] and Burg's maximum entropy (ME) method [10]. Though often successful and widely used, these methods have certain fundamental limitations (esp. bias and sensitivity in parameter estimates), largely because they use an incorrect model (e.g., AR rather than special ARMA) of the measurements. Pisarenko [37] was one of the first to exploit the structure of the data model, doing so in the context of estimation of parameters of cisoids in additive noise using a covariance approach. Schmidt [46] and independently Bienvenu and Kopp [5] were the first to do this in the case of sensor arrays of arbitrary form. Schmidt, in particular, accomplished this by first deriving a complete geometric solution in the absence of noise, then cleverly extending the geometric concepts to obtain a *reasonable* approximate solution in the presence of noise. The resulting algorithm was called MUSIC (*mu*ltiple *si*gnal *c*lassification) and has been widely studied since its inception. The geometric concepts upon which MUSIC is founded form the basis for a much broader class of *subspace-based* algorithms and is the focus on this discussion.

Since the pioneering work of Schmidt, several new techniques based on the subspace approach have been developed. Notably, the *ESPRIT* technique (estimation of signal parameters via rotational invariance techniques) proposed by Paulraj, Roy and Kailath [40, 41, 44]. More recently, new results in multi-dimensional techniques such as the classical maximum likelihood method and weighted-subspace-fitting (WSF) of Ottersten and Kailath [68] have attracted attention due to their potentially superior performance. In fact, an optimal choice of weighting can be shown to yield estimates that achieve the Cramér–Rao lower bound on the error variance [34]. In our presentation we shall exploit the unifying subspace-fitting framework of [68] to provide a methodology for analyzing a variety of subspace algorithms.

We have attempted to adhere to standard notational conventions. Lower case boldface italic characters will generally refer to vectors. Capital boldface italic characters will generally refer to matrices. For either real or complex-valued matrices, $(\cdot)^*$ will be used to denote the Hermitian conjugate (or complex-conjugate transpose) operation. Eigenvalues of square Hermitian matrices are assumed to be ordered in decreasing magnitude, as are the singular values of non-square matrices.

For the purposes of developing our concepts we have chosen a simple single parameter (i.e., azimuth-only) direction-of-arrival (DOA) estimation problem. Extension to more general multi-parameter problems is usually straightforward. We begin by introducing in the next section the *low-rank* data model and describe its geometric properties. In Section 3, the subspace fitting framework for signal parameter estimation is developed. In Section 4, the asymptotic properties of the estimators are established. Inclusion of sensor array errors in asymptotic analyses is important in real-world considerations and is discussed in Section 5. In the next section, the issue of computational complexity and its reduction via exploitation of some of the low rank properties inherent in the data model is addressed through the fast subspace decomposition (FSD) algorithm and its computational advantages described. We end with a brief summary of the chapter in the final section.

2. The subspace-based data model and applications

There are a number of important parameter estimation problems in which the following model of the measurements is satisfactory;

$$\boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{s}(t) + \boldsymbol{n}(t)\,, \tag{1}$$

where $\boldsymbol{x}(t) \in \mathbb{C}^M$ is a single observation or snapshot of the data at time t, $\boldsymbol{A} \in \mathbb{C}^{M\times d}$ is a time-invariant matrix that depends on a set of parameters $\boldsymbol{\eta} \in \mathbb{R}^q$, $\boldsymbol{s}(t) \in \mathbb{C}^d$ represents some underlying signal waveform, and $\boldsymbol{n}(t) \in \mathbb{C}^M$ represents additive noise. The three most important features are that the

matrix $A(\eta)$ is time-invariant over the observation interval, the model is bilinear in A and s, and the noise is additive. When there are more measurements made than signals present (i.e., $d < M$) in these problems, the signal component of $x(t)$ $(A(\eta)s(t))$ is confined to (at most) a d-dimensional subspace of $\mathbb{C}^M$, referred to as the *signal subspace*. Since the noise is typically assumed to possess energy in all dimensions of the observation space, equation (1) is often referred to as a *low-rank signal* in *full-rank noise* data model. Herein it is referred to as a subspace-based data model.

This data model admits an appealing geometric interpretation when $d < M$ and provides insight into the sensor array processing problem. The measurements are vectors in a complex space of dimension equal to the number of elements in the array (M). In the absence of noise, these measurements are confined to a subspace (the *signal subspace*) of the measurement space whose dimension (d') is at most equal to the number of signals present and it spans either the entire or some restriction of the column space of A (itself a subspace of the measurement space). If any of the impinging signals are perfectly correlated, i.e., one signal is simply a complex scalar multiple of another, the span of the *signal subspace* (d') will be less than d. If there is sufficient *excitation*, i.e., no signals are *perfectly correlated*, the signal subspace is d-dimensional.

With this geometric picture, solutions to the parameter estimation problem are readily manifest as subspace intersection problems. Because of the many applications for which the subspace-based data model is appropriate, numerous *subspace-based* techniques have been developed to exploit it. The approach taken by subspace based methods is in general a three-step procedure.

(1) Construct a suitable (unique) parametrization $A(\eta)$ of the measurement distribution vectors (columns of $A(\eta)$) for all parameter values η of interest.

(2) Obtain an estimate $\hat{A}$ of the signal subspace from observations in the measurement space.

(3) Find a parameter estimate $\hat{\eta}$ so that $A(\hat{\eta})$ best matches $\hat{A}$ in some sense.

The requirement in step (1) is simply one of *identifiability*, i.e., given the matrix $A(\eta)$, it should be possible to uniquely identify the parameter vector η associated with it. The problem of estimating the low-rank signal subspace in step (2) is generally performed in an unconstrained manner and can be viewed as a factor analysis problem. These two steps are common to all subspace-based algorithms. It is primarily the implementation of step (3) that has attracted the most research attention.

There are several important signal processing problems that possess the data model of (1) and for which subspace-based approaches are appropriate. For the purposes of illustration, we focus on the directions-of-arrival estimation problem in sensor arrays in this chapter. Subspace methods have more recently been applied to several other applications including estimation of the parameters of exponentials in noise (often referred to as *harmonic retrieval*), state-space system identification, estimating position and alignment of straight lines in images (see, e.g., the recent reference [1]).

2.1. Directions-of-arrival estimation problem

The goal of many radar, sonar, and geophysical signal processing systems is to take data collected by a spatially distributed array of sensors and extract information about the sources of energy received by the sensor array. This information is typically in the form of parameters that describe the number of distinct energy sources, their location (e.g., azimuth and elevation angles-of-arrival, range) their temporal characteristics (e.g., wavelength, frequency, amplitude) or other important quantities (polarization, doppler shift, etc.).

A simple example of this type of application is depicted in Figure 1, where two signals are shown impinging on an array of three receivers. The patterns associated with each receiver indicate their relative directional sensitivity. For simplicity, the receivers are assumed to be located in the same plane, and the signal sources are in far field and hence the wavefronts are assumed to be planar (unless the signal sources are close to the array, in which case the wavefronts would have some curvature). In the case depicted, the parameters of interest are simply the directions-of-arrival (DOAs) of the wavefronts.

The sensors and receivers are basically transducers that convert the received energy to electrical signals. Typically, the signals output by the sensors are first appropriately conditioned, simultaneously sampled (*snapshot*), digitized and then processed to obtain estimates of the signal parameters.

Signal and noise subspaces

In this discussion, we will focus on the so-called *narrowband* data model for the sensor array problem. This model inherently assumes that as the signal wavefronts propagate across the array, the *envelop* of the signal remains

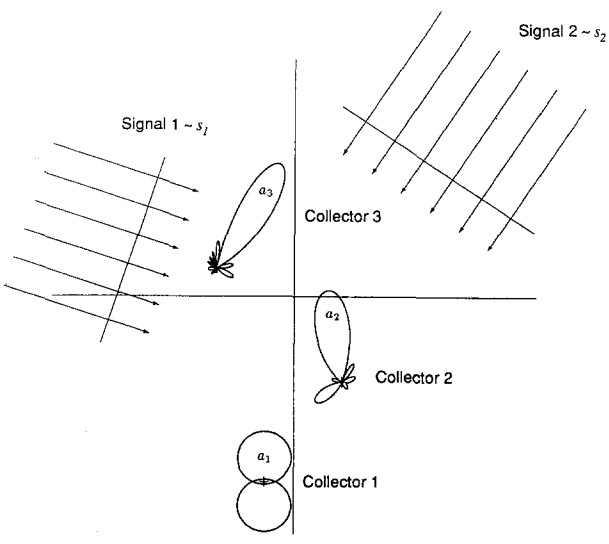

Fig. 1. Illustration of a simple source location estimation problem.

essentially unchanged. The term *narrowband* is used here since the assumption of a slowly varying signal envelop is most often satisfied when either the signals or sensor elements have a bandwidth that is small relative to the center frequency of operation. However, this assumption can also be satisfied by wideband signals provided that the frequency response of the array is approximately flat over the bandwidth of the signals, and provided that the propagation time across the array is small compared with the reciprocal bandwidths.

Under the narrowband assumption, the complex (in-phase and quadrature) output of an M-element array $\boldsymbol{x}(t) \in \mathbb{C}^M$ due to d sources may be described by the following equation:

$$\boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{s}(t) + \boldsymbol{n}(t)\,, \tag{2}$$

where $\boldsymbol{s}(t) \in \mathbb{C}^d$ represents the signal amplitude and phase at time t, $\boldsymbol{n}(t) \in \mathbb{C}^M$ is additive noise, and $\boldsymbol{A}(\boldsymbol{\eta}) \in \mathbb{C}^{M \times d}$ is the matrix of array response vectors and parameterized by the vector $\boldsymbol{\eta} = [\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d]^{\mathrm{T}}$ that is a collection of all the unknown spatial parameters,

$$\boldsymbol{A}(\boldsymbol{\eta}) = [\boldsymbol{a}(\boldsymbol{\eta}_1) \cdots \boldsymbol{a}(\boldsymbol{\eta}_d)]\,. \tag{3}$$

The array response $\boldsymbol{a}(\boldsymbol{\eta}_k)$ for the kth source is a function of the parameters contained in the subvector $\boldsymbol{\eta}_k$, which might include for example the location of the source in some coordinate system, the signal carrier frequency, polarization angles, etc. If there are $p < M$ elements in $\boldsymbol{\eta}_k$, then $\boldsymbol{a}(\boldsymbol{\eta}_k)$ will in general trace out a p-dimensional surface in $\mathbb{C}^M$ as $\boldsymbol{\eta}_k$ is varied over the parameter space. This surface is referred to as the *array manifold*, and is denoted mathematically as

$$\mathcal{A} = \{\boldsymbol{a}(\boldsymbol{\eta}_k)\colon\ \boldsymbol{\eta}_k \in \boldsymbol{\Theta}\}\,, \tag{4}$$

where $\boldsymbol{\Theta}$ denotes the set of all possible values of the parameter vector. For example, if $p = 1$ (e.g., if $\boldsymbol{\eta}_k = \theta_k$, the signal DOA), then $\mathcal{A}$ is a one-dimensional *rope* winding through $\mathbb{C}^M$ as illustrated in Figure 2.

An *unambiguous* array manifold $\mathcal{A}$ is defined to be one for which any collection of $d \leqslant M$ distinct vectors from $\mathcal{A}$ form a linearly independent set. For example, an element from the manifold of a uniform linear array of identical sensors is proportional to

$$\boldsymbol{a}(\theta_k) = \begin{bmatrix} 1 \\ \mathrm{e}^{-\mathrm{j}2\pi\delta \sin\theta_k/\lambda} \\ \mathrm{e}^{-\mathrm{j}4\pi\delta \sin\theta_k/\lambda} \\ \vdots \\ \mathrm{e}^{-\mathrm{j}2(M-1)\pi\delta \sin\theta_k/\lambda} \end{bmatrix}, \tag{5}$$

where λ is the wavelength of the signal and δ is the distance in wavelengths

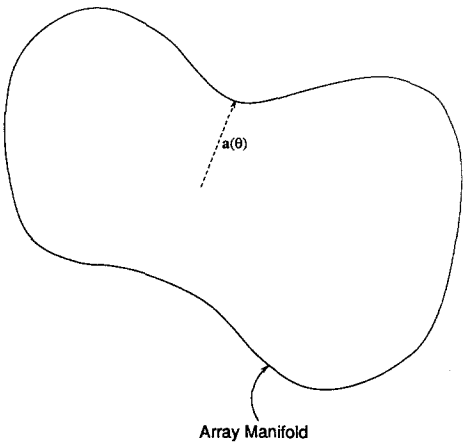

Fig. 2. A one-dimensional array manifold.

separating the sensors. It is quite easy to show that such an array is unambiguous for $\theta \in [0, \pi]$ only if $\delta < \frac{1}{2}\lambda$. For more widely spaced sensors, there will be pairs of angles θ_1 and θ_2 for which $\boldsymbol{a}(\theta_1) = \boldsymbol{a}(\theta_2)$. In such cases, the array response for a signal from θ_1 is indistinguishable from that for a signal from θ_2. Clearly, whether or not an array is ambiguous has some important implications for the identifiability of the parameters in $\boldsymbol{\eta}$. Note that the requirement of half-wavelength interelement spacing is entirely analogous to the well-known Nyquist sampling theorem for time series. In this case, the signal of interest must be sampled at twice its highest *spatial* frequency in order for it to be uniquely identified.

If $\mathcal{A}$ is unambiguous and $M > d$, the matrix $\boldsymbol{A}(\boldsymbol{\eta})$ defined above will be full-rank d. If, in addition, the matrices

$$\boldsymbol{X} = [\boldsymbol{x}(1) \quad \cdots \quad \boldsymbol{x}(N)],$$
$$\boldsymbol{S} = [\boldsymbol{s}(1) \quad \cdots \quad \boldsymbol{s}(N)]$$

are defined for $N > d$ vector observations, or *snapshots*, from the array, then

$$\boldsymbol{X} = \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{S}, \tag{6}$$

and it is seen that each snapshot from the array is just a linear combination of the columns of $\boldsymbol{A}(\boldsymbol{\eta})$. In other words, each observation is constrained to lie in the d-dimensional subspace of $\mathbb{C}^M$ defined by the d columns of $\boldsymbol{A}(\boldsymbol{\eta})$. Furthermore, if $\boldsymbol{S}$ is itself full-rank d, then $\text{span}(\boldsymbol{A}) = \text{span}(\boldsymbol{A}(\boldsymbol{\eta}))$, and the observations will *fill-out* this low-rank subspace.

Figure 3 illustrates this idea for the special case of a single parameter per signal source ($p = 1$), two sources ($d = 2$), and four snapshots ($N = 4$). Each of

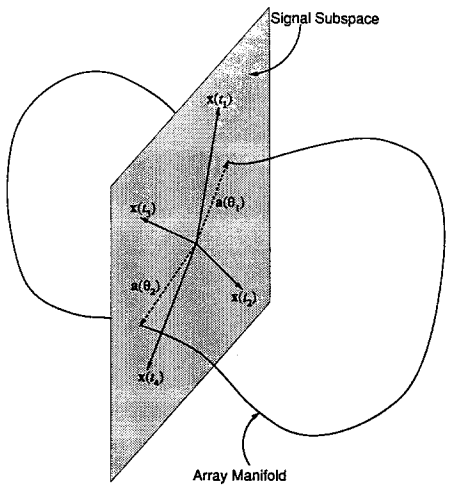

Fig. 3. A geometric view of the narrowband DOA estimation problem.

the two sources has associated with it a response vector $\boldsymbol{a}(\theta_k)$ from the array manifold, and the four snapshots $\boldsymbol{x}(t_1), \ldots, \boldsymbol{x}(t_4)$ lie in the two-dimensional subspace spanned by these vectors. The specific positions of these vectors depend on the values of the signal waveforms at each time instant. As indicated in the figure, the term *signal subspace* is used to denote the subspace spanned by the columns of $\boldsymbol{A}(\boldsymbol{\eta})$, denoted $\mathcal{R}\{\boldsymbol{A}(\boldsymbol{\eta})\}$. Note that the array manifold intersects the signal subspace at only two points, each corresponding to a response vector of one of the signals. A third intersection at some $\boldsymbol{a}(\theta_3)$ would imply that the manifold was not unambiguous, since $\boldsymbol{a}(\theta_3)$ could be written as a linear combination of $\boldsymbol{a}(\theta_1)$ and $\boldsymbol{a}(\theta_2)$. In other words, it would be impossible to distinguish a signal subspace generated by signals with parameters θ_1, θ_2, and θ_3 from one generated by any pair of parameters, e.g., θ_1 and θ_2, alone. Thus, ambiguities in the array manifold can lead to a loss of parameter identifiability.

It is also possible for the signal subspace to have dimension smaller than d if the matrix of signal samples $\boldsymbol{S}$ has a rank less than d. This situation will arise, for example, if one or more of the signals is equal to a linear combination of the others. Such signals are referred to as *coherent*, or fully-correlated signals, and occur most frequently in the sensor array problem when the *multipath* phenomenon is present. As its name implies, multipath results when a given signal is received at the array from several different directions or paths due to reflections from various objects in the propagation environment. It is also possible there are fewer snapshots than sources present, in which case the signal subspace dimension cannot exceed the number of snapshots. In either case, the dimension of the signal subspace is less than the number of sources present. This does not mean, however, that estimates of the number of sources and the associated parameters cannot be obtained. For example, it can be

shown [8, 78] that for one-dimensional parameter vectors (i.e., one parameter per source) the signal parameters are still identifiable if $\mathscr{A}$ is unambiguous and $M > 2d - d'$, where $d' = \text{rank}\ (\boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{S})$. The identifiability condition, which is again geometrically obvious, is that the signal subspace be spanned by a *unique* set of d vectors from the array manifold.

Additive noise

Noise may enter into the array measurement either externally through the sensor, or internally through the receiver electronics. The internal component is typically due to thermal noise and quantization effects within the receiver. On the other hand, external noise can include random background radiation and clutter, in addition to any type of interference that elicits an array response significantly different from any element of the assumed manifold (e.g., near-field, wideband, or distributed emitters under the assumption of far-field, narrowband point sources). In either case, it is often assumed that the noise causes an additive, zero-mean disturbance to the receiver output,

$$\boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{s}(t) + \boldsymbol{n}(t)\,, \tag{7}$$

which for multiple snapshots is written as

$$\boldsymbol{X} = \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{S} + \boldsymbol{N}\,, \tag{8}$$

using an obvious notation. In addition, the noise is also typically assumed to be a complex stationary circular Gaussian random process. When deriving maximum likelihood estimation formulas based on equation (7), the noise is further assumed to be uncorrelated from snapshot to snapshot.

Of perhaps more importance than the temporal properties of the noise are its *spatial* properties. In particular, most algorithms for this problem require that the *spatial covariance* of the noise be known. This quantity will be denoted by the matrix $\sigma^2\boldsymbol{\Sigma}$, and is defined to be

$$\sigma^2\boldsymbol{\Sigma} = \mathscr{E}\{\boldsymbol{n}(t)\boldsymbol{n}^*(t)\}\,, \tag{9}$$

where $\boldsymbol{\Sigma}$ is normalized so that $\det(\boldsymbol{\Sigma}) = 1$ and σ^2 represents the noise *power*. Often the simplifying assumption of spatial whiteness (i.e., $\boldsymbol{\Sigma} = \boldsymbol{I}$) is made.

In practical applications, the noise is usually full-rank, and consequently any collection of $N \geq M$ samples of $\boldsymbol{n}(t)$ will with probability one span all of $\mathbb{C}^M$. Although the signal component in equation (7) is confined to a d'-dimensional subspace of $\mathbb{C}^M$, the full-rank nature of the noise implies that $\boldsymbol{x}(t)$ will be a full-rank signal as well. Comparison of equations (7) and (1) reveals that the model for narrowband sensor array data has the desired structure required for exploitation by subspace-based algorithms.

As noted earlier, in the absence of noise, a collection of $N \geq d$ data vectors determine the signal subspace and hence the solution can be found as intersections of the array manifold and this subspace. In the presence of noise,

the approach is to estimate this signal subspace from the data and then determine parameters $\boldsymbol{\eta}$ such that the corresponding array manifold vectors best fit this estimate. Indeed, dealing effectively with noise is the key issue in subspace algorithms and is described in some detail in this chapter.

Subspace estimation in presence of noise

Though techniques are available that proceed directly from the data vectors (e.g., singular value decompositions or SVDs), it is conceptually easier to explain the subspace estimation based on the spatial array covariance matrix. Assuming that noise is spatially white, the *array covariance* matrix $\boldsymbol{R}_{XX}$ has the following form

$$\boldsymbol{R}_{XX} = \boldsymbol{E}[\boldsymbol{X}\boldsymbol{X}^*] = \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{R}_{SS}\boldsymbol{A}^*(\boldsymbol{\eta}) + \sigma^2\boldsymbol{I}\,, \tag{10}$$

where $\boldsymbol{R}_{SS} = \boldsymbol{E}[\boldsymbol{S}\boldsymbol{S}^*]$ is the *signal covariance* matrix and $\boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{R}_{SS}\boldsymbol{A}^*(\boldsymbol{\eta})$ is a rank-d' matrix and $d' < M$. The eigendecomposition of $\boldsymbol{R}_{XX}$ is given by

$$\boldsymbol{R}_{XX} = \sum_{k=1}^{M} \lambda_k \boldsymbol{e}_k \boldsymbol{e}_k^*\,, \tag{11}$$

where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_M$ are the eigenvalues and the $\{\boldsymbol{e}_k\}$ are the corresponding eigenvectors of $\boldsymbol{R}_{XX}$. For $\boldsymbol{R}_{XX}$ with the structure given in equation (10), i.e., a rank-d' matrix plus a scaled identity, it is not difficult to see that the $M - d'$ smaller eigenvalues are repeated, viz., $\lambda_{d'+1} = \cdots = \lambda_M = \sigma^2$. Define $\boldsymbol{E}_S = [\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_{d'}]$ and $\boldsymbol{E}_N = [\boldsymbol{e}_{d'+1}, \boldsymbol{e}_{d'+2}, \ldots, \boldsymbol{e}_M]$. It is easy to see that the span of the columns of $\boldsymbol{E}_S$ is the *signal subspace*. Its orthogonal complement is spanned by $\boldsymbol{E}_N$ and is referred to as the *noise subspace*. Note that the signal subspace is contained in a space spanned by $\boldsymbol{A}(\boldsymbol{\eta})$, i.e.,

$$\mathcal{R}\{\boldsymbol{E}_S\} \subseteq \mathcal{R}\{\boldsymbol{A}(\boldsymbol{\eta})\}\,. \tag{12}$$

This observation is the basis of most subspace-based estimation techniques. The spaces are equal if and only if $d = d'$ which requires that $\boldsymbol{R}_{SS}$ is full-rank and the array has no ambiguities in effect. The d' largest eigenvalues λ_i, $1 \leqslant i \leqslant d$, are often referred to as the *signal eigenvalues*, while $\lambda_i = \sigma^2$, $d' + 1 \leqslant i \leqslant M$ are referred to as the *noise eigenvalues*. It is important to note the fundamental asymmetry between the signal and noise subspaces with respect to $\mathcal{R}\{\boldsymbol{A}(\boldsymbol{\eta})\}$. It is entirely possible for the noise subspace to have a non-vanishing intersection with $\mathcal{R}\{\boldsymbol{A}(\boldsymbol{\eta})\}$, whereas the signal subspace is entirely confined to $\mathcal{R}\{\boldsymbol{A}(\boldsymbol{\eta})\}$.

In practice, the ideal covariance $\boldsymbol{R}_{XX}$ is rarely available. A consistent estimate may be obtained by forming the sample covariance, $\hat{\boldsymbol{R}}_{XX}$, based on N *snapshots*, i.e.,

$$\hat{\boldsymbol{R}}_{XX} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}(t)\boldsymbol{x}^*(t)\,. \tag{13}$$

Let $(\hat{\lambda}_k, \hat{\boldsymbol{e}}_k)$, $1 \leq k \leq M$ be the eigenpairs of $\hat{\boldsymbol{R}}_{XX}$. The collection of eigenvectors corresponding to the largest eigenvalues, $\mathcal{R}\{\hat{\boldsymbol{E}}_S\} = \mathcal{R}\{\hat{\boldsymbol{e}}_1, \ldots, \hat{\boldsymbol{e}}_{d'}\}$, forms a consistent estimate of the signal subspace. The dimension of the signal subspace may be estimated based on the eigenvalue distribution. There are several well-known detection schemes available such as likelihood ratio, MDL, and AIC tests [2, 73]. As is well known, the eigendecomposition described above requires $O(M^3)$ operations and is often the major computational burden. In Section 6, a computationally more efficient scheme is presented.

3. Subspace methods for parameters estimation

This section discusses subspace methods for parameter estimation by using a subspace fitting approach. It also relates these subspace fitting ideas to several other techniques that have been applied to the sensor array problem. We begin by considering the classical maximum likelihood estimator formulated based on a stochastic modeling of the emitter signals. This technique, while having excellent performance is usually not computationally efficient, thus providing a motivation for the more computationally efficient subspace methods.

3.1. Maximum likelihood estimation

In many parameter estimation problems, a natural systematic procedure is the maximum likelihood method. Such techniques have the intuitively appealing interpretation of finding the set of parameter values that make the observations made the ones most likely to have occurred given the model of the measurements. In the sensor array parameter estimation problem, two models differing only in assumptions concerning the underlying signals have been proposed. These models have led to the development of two different ML procedures. The *deterministic* or least-squares ML method assumes deterministic signals and has been studied in, for example, [7, 17, 88, 55]. In Section 3.2, it is shown that this technique may be interpreted as a subspace fitting method and can be analyzed in that framework.

The *stochastic* ML technique, herein referred to as SML, is based on stochastic modeling of the emitter signals [3]. As described in the previous section, the data model under consideration is

$$\boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{s}(t) + \boldsymbol{n}(t)\,. \tag{14}$$

Under the assumption that the emitter signals and noise are i.i.d. Gaussian random waveforms, the array output constitutes a stationary, temporally white, complex Gaussian random process with covariance matrix $\boldsymbol{R}_{XX}$, given by

$$\boldsymbol{R}_{XX} = \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{R}_{SS}\boldsymbol{A}^*(\boldsymbol{\eta}) + \sigma^2\boldsymbol{I}\,. \tag{15}$$

The normalized negative log-likelihood function of the observations $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)$, has the following form,

$$l(\boldsymbol{\xi}) = M \log \pi + \log|\boldsymbol{R}_{XX}(\boldsymbol{\xi})| + \mathrm{Tr}\{\boldsymbol{R}_{XX}^{-1}(\boldsymbol{\xi})\hat{\boldsymbol{R}}_{XX}\}\,, \tag{16}$$

where $\boldsymbol{R}_{XX} = E\{\boldsymbol{x}\boldsymbol{x}^*\} = \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{R}_{SS}\boldsymbol{A}^*(\boldsymbol{\eta}) + \sigma^2\boldsymbol{I}$, $\hat{\boldsymbol{R}}_{XX}$ is the sample covariance, and $\boldsymbol{\xi}$ represents the unknown parameters of the array covariance matrix. The SML estimate of $\boldsymbol{\xi}$ is obtained by minimizing $l(\boldsymbol{\xi})$. Herein, the unknown parameters are assumed to be $\boldsymbol{\eta}$, $\boldsymbol{R}_{SS}$, and σ^2.

The SML method requires a nonlinear $(d^2 + d + 1)$-dimensional optimization for one-parameter manifolds. Although some methods have been reported for performing this search [17, 6], it is often unacceptably expensive.

As noted in [9], the log likelihood function equation (16) can be separated and thus the dimension of the optimization can be reduced. The following concentrated negative log likelihood function is obtained (omitting $\boldsymbol{\eta}$-independent terms),

$$J(\boldsymbol{\eta}) = \log|\boldsymbol{A}(\boldsymbol{\eta})\hat{\boldsymbol{R}}_{SS}(\boldsymbol{\eta})\boldsymbol{A}(\boldsymbol{\eta})^* + \hat{\sigma}^2(\boldsymbol{\eta})\boldsymbol{I}|\,, \tag{17}$$

$$\hat{\boldsymbol{R}}_{SS}(\boldsymbol{\eta}) = \boldsymbol{A}^{\dagger}(\boldsymbol{\eta})(\hat{\boldsymbol{R}}_{XX} - \hat{\sigma}^2(\boldsymbol{\eta})\boldsymbol{I})\boldsymbol{A}^{\dagger *}(\boldsymbol{\eta})\,, \tag{18}$$

$$\hat{\sigma}^2(\boldsymbol{\eta}) = \frac{1}{m-d}\,\mathrm{Tr}\{\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\eta})}\hat{\boldsymbol{R}}_{XX}\}\,, \tag{19}$$

where $\boldsymbol{A}^{\dagger} = (\boldsymbol{A}^*\boldsymbol{A})^{-1}\boldsymbol{A}^*$, $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{A}^{\dagger}$, and $\boldsymbol{P}^{\perp}_{\boldsymbol{A}} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}$. The criterion function, $J(\boldsymbol{\eta})$ has a rather complicated structure. However, by rewriting the cost function the computational burden can be reduced; in [35] a Gauss–Newton search technique requiring $O(m^2d)$ operations per iteration is presented. If the global minimum is found, excellent estimates are to be expected since the method is asymptotically efficient.

Since the SML cost function may have many local extrema, accurate initial estimates are required. There has been a great interest in computationally efficient suboptimal alternatives to the SML technique. Among these, the subspace-based techniques have attracted most attention. Subspace methods use an elegant geometric interpretation of the sensor array problem and provide very good performance. Though these methods are, in general, suboptimal techniques, it has recently been shown [34] that if the parameter estimation problem is formulated properly, subspace techniques yield *asymptotically efficient* estimates. This is further discussed in Section 4.

3.2. Subspace fitting framework

This section formulates different signal parameter estimation methods in a common subspace fitting framework, and displays the algebraic relations between the algorithms. The framework can be used for designing common

numerical algorithms and for obtaining new methods. For example, extensions of the *ESPRIT* algorithm arise in a natural way from the subspace fitting formulation. The subspace fitting formulation of the deterministic ML method is first discussed, and the other algorithms are related to this formulation.

The subspace fitting approach was first described by Schmidt [47] and formalized by Cadzow [11] and can be posed as

$$[\hat{\boldsymbol{A}}, \hat{\boldsymbol{T}}] = \arg\min_{\boldsymbol{A},\boldsymbol{T}} \|\boldsymbol{M} - \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{T}\|_F^2 , \tag{20}$$

where $\boldsymbol{M}$ is an $m \times q$ matrix that is obtained from the data, the $m \times p$ matrix $\boldsymbol{A}$ is parameterized by $\boldsymbol{\eta}$, and $\boldsymbol{T}$ is any $p \times q$ matrix. The estimate of the parameter vector $\boldsymbol{\eta}$ is the argument of $\hat{\boldsymbol{A}}$. The matrix $\boldsymbol{M}$ and its dimension, as well as the parametrization (and dimension) of $\boldsymbol{A}$, can be chosen in different ways leading to different estimates. This is the key observation in describing the several methods as solutions to variations of the same basic problem.

Note that the subspace fitting problem is separable in $\boldsymbol{A}$ and $\boldsymbol{T}$ [19]. By substituting the pseudo-inverse solution $\hat{\boldsymbol{T}} = \boldsymbol{A}^{\dagger}\boldsymbol{M}$ back into (20), we obtain the following equivalent problem

$$\hat{\boldsymbol{A}} = \arg\max_{\boldsymbol{A}} \operatorname{Tr}\{\boldsymbol{P}_{\boldsymbol{A}}\boldsymbol{M}\boldsymbol{M}^*\} , \tag{21}$$

where $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}^*\boldsymbol{A})^{-1}\boldsymbol{A}^*$ is the projection matrix that projects onto the column space of $\boldsymbol{A}$.

Table 1 lists some different choices of $\boldsymbol{M}$ and the set containing $\boldsymbol{A}$ for the basic subspace fitting problem. This section attempts to explain how these choices lead to the different methods.

Deterministic maximum likelihood (DML)

Maximizing the likelihood function of the data matrix $\boldsymbol{X}_{\mathrm{N}}$ is easily seen to be equivalent to the following problem [70],

$$\min_{\boldsymbol{\eta},\boldsymbol{S}_{\mathrm{N}}} \operatorname{Tr}\{(\boldsymbol{X}_{\mathrm{N}} - \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{S}_{\mathrm{N}})^*(\boldsymbol{X}_{\mathrm{N}} - \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{S}_{\mathrm{N}})\} = \min_{\boldsymbol{\eta},\boldsymbol{S}_{\mathrm{N}}} \|\boldsymbol{X}_{\mathrm{N}} - \boldsymbol{A}(\boldsymbol{\eta})\boldsymbol{S}_{\mathrm{N}}\|_F^2 , \tag{22}$$

where we have assumed that the signal $\boldsymbol{S}_{\mathrm{N}}$ is deterministic but unknown. Thus

Table 1
Subspace fitting methods

Choice of $\boldsymbol{M}$	Constraint on $\boldsymbol{A}$		
	$\boldsymbol{A} \in \mathscr{A}^d$	$\boldsymbol{A} \in \mathscr{A}$	$\boldsymbol{A} \in \mathscr{E}$
$\boldsymbol{M}\boldsymbol{M}^* = \hat{\boldsymbol{R}}_{XX}$	Det-ML	Beamforming	ML-ESPRIT
$\boldsymbol{M} = \hat{\boldsymbol{E}}_{\mathrm{S}}$	MD-MUSIC	MUSIC	TLS-ESPRIT
$\boldsymbol{M} = \hat{\boldsymbol{E}}_{\mathrm{S}}\boldsymbol{W}_{\mathrm{opt}}^{1/2}$	WSF	Weighted MUSIC	Weighted ESPRIT

choosing $\boldsymbol{M} = \boldsymbol{X}_{\mathrm{N}}$ and $\boldsymbol{A} \in \mathscr{A}^d$ in (20) leads immediately to the DML method. The deterministic ML estimate of the emitter waveforms, $\boldsymbol{S}_{\mathrm{N}}$, is obtained from (22) by

$$\hat{\boldsymbol{S}}_{\mathrm{N}} = \boldsymbol{A}^{\dagger}(\boldsymbol{\theta})\boldsymbol{X}_{\mathrm{N}} . \tag{23}$$

By substituting (23) into (22) and normalizing by N, the following familiar optimization problem is obtained

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \mathrm{Tr}\{\boldsymbol{P}_{A(\boldsymbol{\eta})}\hat{\boldsymbol{R}}_{XX}\} , \tag{24}$$

where $\hat{\boldsymbol{R}}_{XX}$ is the sample covariance matrix.

Unfortunately, the resulting constrained optimization problem is too computationally expensive in many practical applications. Given the formulation in (20), it is natural to look for suboptimal methods by performing a one-dimensional search as opposed to a full d-dimensional search of the parameter space.

Beamforming

The traditional beamforming method, or 'delay-and-sum method', for a sensor array attempts to maximize the power of a weighted sum of the outputs of the elements of the array. The array is steered to all directions in the range of interest. The normalized output power $P(\boldsymbol{\eta}_j)$, as a function of direction $\boldsymbol{\eta}_j$, is given by

$$P(\boldsymbol{\eta}_j) = \frac{1}{N}\sum_{t=1}^{N} \frac{|\boldsymbol{a}^*(\boldsymbol{\eta}_j)\boldsymbol{x}(t)|^2}{|\boldsymbol{a}(\boldsymbol{\eta}_j)|^2} = \frac{\boldsymbol{a}^*(\boldsymbol{\eta}_j)\hat{\boldsymbol{R}}_{XX}\boldsymbol{a}(\boldsymbol{\eta}_j)}{\boldsymbol{a}^*(\boldsymbol{\eta}_j)\boldsymbol{a}(\boldsymbol{\eta}_j)} = \mathrm{Tr}\{\boldsymbol{P}_a(\boldsymbol{\eta}_j)\hat{\boldsymbol{R}}_{XX}\} , \tag{25}$$

where $\boldsymbol{P}_a(\boldsymbol{\eta}_j) = \boldsymbol{a}(\boldsymbol{a}^*\boldsymbol{a})^{-1}\boldsymbol{a}^*$. The directions with maximum power are taken as the estimates of the DOAs.

The beamforming method is also immediately obtained from the basic subspace fitting problem by comparing (21) and (25). Set $p = 1$ in (21), i.e., take $\boldsymbol{A} \in \mathscr{A}$ and choose the d largest (separated) maxima of the criterion function as the estimates of the directions of arrivals. When only one signal is present, beamforming is equivalent to deterministic ML, and for well separated sources, the estimates are almost the same. The aperture of the array is the limiting factor, not the amount of data when trying to resolve closely spaced emitters.

MUSIC

The MUSIC algorithm [47] is based on a geometric formulation of the sensor array problem and has superior resolution capabilities compared to classical beamforming methods. Assumed that the signal covariance matrix $\boldsymbol{R}_{SS}$, has full

rank, i.e., that $d' = d$. The array manifold vectors corresponding to the true signal parameters $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d$ then span the signal subspace, $\mathcal{R}(\boldsymbol{E}_S)$, and are thus orthogonal to the noise subspace, i.e.,

$$\boldsymbol{a}^*(\boldsymbol{\eta}_j)\boldsymbol{E}_N = \boldsymbol{0}, \quad j = 1, \ldots, d. \tag{26}$$

In practice, an estimate of the noise subspace is obtained by selecting the eigenvectors corresponding to the $m - d$ smallest eigenvalues of $\hat{\boldsymbol{R}}_{XX}$. A scalar measure of the distance between the array manifold and the estimated noise subspace is formed and its inverse,

$$P_{MU}(\boldsymbol{\eta}_j) = \frac{\boldsymbol{a}^*(\boldsymbol{\eta}_j)\boldsymbol{a}(\boldsymbol{\eta}_j)}{\boldsymbol{a}^*(\boldsymbol{\eta}_j)\hat{\boldsymbol{E}}_N\hat{\boldsymbol{E}}_N^*\boldsymbol{a}(\boldsymbol{\eta}_j)} = \frac{\boldsymbol{a}^*(\boldsymbol{\eta}_j)\boldsymbol{a}(\boldsymbol{\eta}_j)}{\boldsymbol{a}^*(\boldsymbol{\eta}_j)(\boldsymbol{I} - \hat{\boldsymbol{E}}_S\hat{\boldsymbol{E}}_S^*)\boldsymbol{a}(\boldsymbol{\eta}_j)}, \tag{27}$$

is searched for peaks. The parameter estimates are determined by the location of the d largest peaks of this so called *MUSIC spectrum*. Maximizing P_{MU} is equivalent to maximizing $1 - 1/P_{MU}$. Thus the MUSIC estimate can be written

$$\hat{\boldsymbol{\eta}}_j = \arg\max_{\boldsymbol{\eta}_j \in \boldsymbol{\Theta}} \frac{\boldsymbol{a}^*(\boldsymbol{\eta}_j)\hat{\boldsymbol{E}}_S\hat{\boldsymbol{E}}_S^*\boldsymbol{a}(\boldsymbol{\eta}_j)}{\boldsymbol{a}^*(\boldsymbol{\eta}_j)\boldsymbol{a}(\boldsymbol{\eta}_j)} = \arg\max_{\boldsymbol{a} \in \mathcal{A}} \mathrm{Tr}\{\boldsymbol{P}_{\boldsymbol{a}}\hat{\boldsymbol{E}}_S\hat{\boldsymbol{E}}_S^*\}. \tag{28}$$

The above corresponds to (21) with $\boldsymbol{M} = \hat{\boldsymbol{E}}_S$ and $\boldsymbol{A} \in \mathcal{A}$.

The MUSIC and beamforming methods are thus obtained by doing a 'one-signal-search', i.e., by taking $\boldsymbol{a} \in \mathcal{A}$ and searching for the d best fits of the array manifold with the matrix $\boldsymbol{M}$. In beamforming, $\boldsymbol{M}$ is chosen as a Hermitian square root of $\hat{\boldsymbol{R}}_{XX}$ and in MUSIC, $\boldsymbol{M} = \hat{\boldsymbol{E}}_S$, i.e., only the eigenvectors of $\hat{\boldsymbol{R}}_{XX}$ corresponding to the d largest eigenvalues are used.

MD-MUSIC

One major disadvantage of applying a one-dimensional search to the sensor array problem is the resulting inability to handle coherent emitter signals. If $d' < d$, the true steering vectors cannot all be orthogonal to the noise subspace. Different schemes for coping with this problem have been discussed in the literature, e.g., spatial smoothing [71, 48, 50]. These methods require specific array structures. In general, to handle the coherence problem the search procedure has to be multidimensional. The array manifold is simultaneously searched for d vectors that span a space *containing* the d', $d' \leq d$ largest eigenvectors of $\boldsymbol{R}_{XX}$.

Adopting the subspace fitting formulation the method can be posed as

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{A} \in \mathcal{A}^d, \boldsymbol{T}} \|\hat{\boldsymbol{E}}_S - \boldsymbol{A}\boldsymbol{T}\|_F^2 = \arg\max_{\boldsymbol{A} \in \mathcal{A}^d} \mathrm{Tr}\{\boldsymbol{P}_{\boldsymbol{A}(\boldsymbol{\eta})}\hat{\boldsymbol{E}}_S\hat{\boldsymbol{E}}_S^*\}. \tag{29}$$

The above is easily recognized as (21) with $\boldsymbol{M} = \hat{\boldsymbol{E}}_S$, and $\boldsymbol{A} \in \mathcal{A}^d$. The method suffers, just like ML, from a costly multidimensional optimization. It is thus not clear why one would choose this MD-MUSIC method in favor of DML. It

will be shown in the next section that by giving MD-MUSIC a specific subspace weighting it will have the same asymptotic properties as DML. We also present another weighting that makes the resulting optimally *weighted subspace fitting* (WSF) method outperform DML. It will now be demonstrated that the MD-MUSIC formulation has an elegant solution for an important special choice of array geometry and parametrization.

ESPRIT

The *ESPRIT* algorithm [40, 41, 44] assumes a specific array geometry and is thus not as general as the algorithms discussed up to now. By exploiting this array structure, a very interesting solution to array problem is obtained that does not require specific knowledge of the array manifold. No search is involved as in the previous methods, thus significantly reducing the computational and storage requirements.

Assume that the array is composed of two identical subarrays, each containing $\frac{1}{2}m$ elements. The subarrays are displaced from each other by a known displacement vector $\boldsymbol{\Delta}$. The output of the array is modeled as

$$\boldsymbol{x}(t) = \begin{bmatrix} \bar{\boldsymbol{A}} \\ \bar{\boldsymbol{A}}\boldsymbol{\Phi} \end{bmatrix} \boldsymbol{s}(t) + \begin{bmatrix} \boldsymbol{n}_1(t) \\ \boldsymbol{n}_2(t) \end{bmatrix}, \tag{30}$$

where the $\frac{1}{2}m \times d$ matrix $\bar{\boldsymbol{A}}$ contains the common array manifold vectors of the two subarrays. The *ESPRIT* algorithm does not exploit the entire array manifold. The only knowledge that is used (and consequently the only knowledge that is required) is the displacement structure of the array. This limits the number of resolvable sources to $\frac{1}{2}m$, thus $d \leq \frac{1}{2}m$ is assumed for *ESPRIT*. The propagation between the two subarrays is described by the diagonal matrix $\boldsymbol{\Phi} = \operatorname{diag}[e^{j\omega\tau_1}, \ldots, e^{j\omega\tau_d}]$ where τ_k is the time delay in the propagation of the k-th emitter signal between the two subarrays and ω is the center frequency of the emitters. The time delay is related to the angle of arrival by $\tau_k = |\boldsymbol{\Delta}| \sin \theta_k / c$, where c is the speed of propagation. Thus, estimates of the DOAs can readily be obtained from estimates of the time delays τ_k.

Assume that the emitter covariance is full rank, i.e., $d' = d$, the subspace fitting formulation of *ESPRIT* is then

$$\min_{\boldsymbol{A} \in \mathscr{E}, \boldsymbol{T}} \|\hat{\boldsymbol{E}}_S - \boldsymbol{A}\boldsymbol{T}\|_F^2, \tag{31}$$

where the set $\mathscr{E}$ is defined by

$$\mathscr{E} = \left\{ \boldsymbol{Z} \mid \boldsymbol{Z} = \begin{bmatrix} \bar{\boldsymbol{A}} \\ \bar{\boldsymbol{A}}\boldsymbol{\Phi} \end{bmatrix}, \bar{\boldsymbol{A}} \in \mathbb{C}^{m/2 \times d}, \boldsymbol{\Phi} = \operatorname{diag}[\phi_1, \ldots, \phi_d], \phi_i \in \mathbb{C} \right\}, \tag{32}$$

see [33]. Minimizing (31) can be done in a very efficient manner [45]. A more detailed discussion of the identifiability conditions and a generalized version of *ESPRIT* is given in [33] and also [34].

Weighted subspace fitting (WSF)

Since the columns of $\hat{E}_S$ are each perturbed in different ways, it is reasonable to expect that a weighted subspace fitting method can be superior to the unweighted method.

The weighted subspace fitting can be written as

$$[\hat{A}, \hat{T}] = \arg\min_{A,T} \|MW^{1/2} - A(\eta)T\|_F^2 , \tag{33}$$

where W is a positive definite weighting matrix. This again can be reduced to equivalent problem

$$\hat{\eta} = \arg\max_{\eta} \operatorname{Tr}\{P_{A(\eta)}\hat{E}_S W \hat{E}_S^*\} . \tag{34}$$

It has been shown that the optimal weighting matrix is given by $W_{\text{opt}} = \tilde{\Lambda}^2 \Lambda_S^{-1}$ (see [69]). As is discussed in the next section, this weighting is optimal in the sense it gives the lowest asymptotic estimation error variance of any unbiased estimator, i.e., it reaches the stochastic CRB.

The weighted versions of MUSIC and *ESPRIT* can also be formulated as shown in Table 1. The ML-*ESPRIT* proposed in [45] can be formulated by replacing $\hat{E}_S$ in (31) by any Hermitian square root of $\hat{R}_{XX}$. This completes Table 1.

4. Asymptotic properties

It is often of great interest to be able to predict the accuracy of an estimator. Valuable insight into the estimation problem may be gained and an algorithm can be related to various performance bounds. In sharp contrast, extensive simulation studies are time consuming and give limited information about the behavior of an algorithm. This section investigates the asymptotic properties of the multidimensional estimation techniques discussed in the previous section. Thus, the discussion is restricted to $A \in \mathcal{A}^d$ and $A \in \mathcal{E}$.

In the subspace fitting (SSF) framework the data (M in (20)) are represented in two ways. Either M is chosen as a Hermitian square root of the covariance matrix $\hat{R}_{XX}$, or as a weighted estimated of the signal subspace matrix $\hat{E}_S W^{1/2}$. However, the following observation from [67], shows that it is sufficient to considered the case of a weighted subspace for all algorithms considered herein.

LEMMA 4.1. *The deterministic ML method has the same asymptotic distribution as the following estimator,*

$$\hat{\eta} = \arg\min \|\hat{E}_S \tilde{\Lambda}^{1/2} - AT\|_F^2 = \arg\max_{\eta} \operatorname{Tr}\{P_A \hat{E}_S \tilde{\Lambda} \hat{E}_S^*\} , \tag{35}$$

where $\tilde{\Lambda} = \Lambda_S - \sigma^2 I = \tilde{\Lambda}^{1/2}\tilde{\Lambda}^{*1/2}$.

An important consequence of Lemma 4.1 is that all methods within the subspace fitting framework asymptotically maximize the following criterion function:

$$V(\boldsymbol{\eta}) = \mathrm{Tr}\{\boldsymbol{P}_{\boldsymbol{A}(\boldsymbol{\eta})}\hat{\boldsymbol{E}}_S \boldsymbol{W}\hat{\boldsymbol{E}}_S^*\}\,, \tag{36}$$

where $\boldsymbol{W}$ is a $d' \times d'$ weighting matrix. We will require that the parametrization of $\boldsymbol{A}$ is identifiable but may otherwise be arbitrary. This implies that the analysis holds for $\boldsymbol{A} \in \mathscr{A}^d$, all the different ESPRIT formulations, and the generalized (multiple invariance) versions of the ESPRIT algorithm [61]. We will refer to this class of methods as *multidimensional subspace fitting* techniques.

4.1. *Consistency*

Asymptotic consistency of estimates is an important property. Also, asymptotic distribution of estimation error can only be established for consistent estimators.

The multidimensional subspace fitting estimate $\hat{\boldsymbol{\eta}}$ is given by

$$\hat{\boldsymbol{\eta}} = \arg\max V(\boldsymbol{\eta})\,, \tag{37}$$

$$V(\boldsymbol{\eta}) = \mathrm{Tr}\{\boldsymbol{P}_{\boldsymbol{A}(\boldsymbol{\eta})}\hat{\boldsymbol{E}}_S \boldsymbol{W}\hat{\boldsymbol{E}}_S^*\}\,. \tag{38}$$

We will restrict the weighting matrix $\boldsymbol{W}$ in (38) to be Hermitian and positive definite. Under certain regularity conditions for $\boldsymbol{A}(\boldsymbol{\eta})$, the criterion function, $V(\boldsymbol{\eta})$, converges uniformly to $\bar{V}(\boldsymbol{\eta}) = \lim_{N\to\infty} E[V(\boldsymbol{\eta})]$ (see [34]) since the signal eigenvectors converge to their true values. Consequently, $\hat{\boldsymbol{\eta}}$ converges to the maximizing argument of $\bar{V}(\boldsymbol{\eta})$. We have,

$$\begin{aligned}\bar{V}(\boldsymbol{\eta}) &= \mathrm{Tr}\{\boldsymbol{P}_{\boldsymbol{A}(\boldsymbol{\eta})}\boldsymbol{E}_S \boldsymbol{W}\boldsymbol{E}_S^*\} \\ &= \mathrm{Tr}\{\boldsymbol{E}_S \boldsymbol{W}\boldsymbol{E}_S^*\} - \mathrm{Tr}\{\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\eta})}\boldsymbol{E}_S \boldsymbol{W}\boldsymbol{E}_S^*\} \\ &\leqslant \mathrm{Tr}\{\boldsymbol{E}_S \boldsymbol{W}\boldsymbol{E}_S^*\} = \bar{V}(\boldsymbol{\eta}_0)\,.\end{aligned} \tag{39}$$

In the inequality, we have used the fact that the signal subspace is a subset of the space spanned by $\boldsymbol{A}(\boldsymbol{\eta}_0)$. As discussed previously, the parametrization $\boldsymbol{A} \in \mathscr{A}^d$ is identifiable if $d < (M + d')/2$. For the case $\boldsymbol{A} \in \mathscr{C}$, introduce

$$\boldsymbol{\Phi} = \mathrm{diag}[\rho_1\, e^{j2\pi|\Delta|\sin\theta_1}, \ldots, \rho_d\, e^{j2\pi|\Delta|\sin\theta_d}]\,, \tag{40}$$

$$\bar{\boldsymbol{A}} = \mathrm{Re}\{\bar{\boldsymbol{A}}\} + j\,\mathrm{Im}\{\bar{\boldsymbol{A}}\} = \bar{\boldsymbol{A}}_R + j\bar{\boldsymbol{A}}_I\,. \tag{41}$$

Define the $2md$-dimensional parameter vector $\boldsymbol{\eta}$ as

$$\boldsymbol{\eta} = [\mathrm{vec}(\tilde{\boldsymbol{I}}\bar{\boldsymbol{A}}_R^T, \mathrm{vec}(\tilde{\boldsymbol{I}}\bar{\boldsymbol{A}}_I^T, \rho_1, \ldots, \rho_d, \theta_1, \ldots, \theta_d]^T\,, \tag{42}$$

where $\tilde{I} = [\mathbf{0}_{(m-1)\times 1}\ I]$ is an $(m-1)\times m$ matrix and vec(A) denotes a vector obtained by stacking the columns of A. In [33] it is shown that identifiability is then guaranteed if $d' = d$ and $d \leq \frac{1}{2}M$. Thus the inequality (39) shows that $\hat{\eta}$ is a consistent estimate of η_0.

4.2. Asymptotic distribution

Having established consistency of the SSF method, we now proceed to the asymptotic distribution of the estimate. Since $\hat{\eta}$ maximizes V given by (36), we have ([24], p. 240)

$$0 = V'(\hat{\eta}) = V'(\eta_0) + V''(\eta_\xi)(\hat{\eta} - \eta_0)\,, \tag{43}$$

where V' denotes the gradient and V'' is the Hessian. The vector η_ξ lies between η_0 and $\hat{\eta}$, so $\eta_\xi \to \eta_0$ as $N \to \infty$. Denote the limiting second derivative of V

$$\bar{V}''(\eta) = \lim_{N\to\infty} V''(\eta)\,. \tag{44}$$

Under sufficient regularity conditions, [67], we have for large N,

$$(\hat{\eta} - \eta_0) \simeq -\{\bar{V}''(\eta_0)\}^{-1} V'(\eta_0)\,, \tag{45}$$

provided $\bar{V}''(\eta_0)$ is invertible. Let V_i denote the i-th component of the gradient $V'(\eta_0)$. For large N the gradient satisfies

$$V_i \simeq 2\mathrm{Re}\left\{ \sum_{k=1}^{d'} w_k^* E_S^* A^{\dagger *} A_\eta^* P_A^\perp \hat{e}_k \right\}, \tag{46}$$

where w_k is the k-th column of W and the derivative of the projection matrix is used. Using the asymptotic distribution of the estimated signal eigenvectors, it is possible to show, that the gradient, $V'(\eta_0)$, is asymptotically normal with zero-mean and covariance matrix Q, given by

$$Q = \lim_{N\to\infty} N\,\mathrm{E}[V'(\eta_0)V'(\eta_0)^{\mathrm{T}}]\,. \tag{47}$$

Equations (45) and (47) imply that the asymptotic distribution of the estimation error is given by

$$\sqrt{N}(\hat{\eta} - \eta_0) \sim \mathrm{As\,N}(0, C)\,, \tag{48}$$

where

$$C = (\bar{V}'')^{-1} Q (\bar{V}'')^{-1}\,. \tag{49}$$

Both matrices are evaluated in $\boldsymbol{\eta}_0$. The following result, [67], provides the asymptotic distribution of the multidimensional SSF signal parameter estimates.

THEOREM 4.1. *Consider the subspace fitting method*, (37)–(38). *The asymptotic distribution of the estimation error is given by* (48)–(49) *where the* (i, j)*-th elements of the matrices are given by*

$$(\bar{V}'')_{ij} = -2\mathrm{Re}\{\boldsymbol{A}_j^* \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{A}_i \boldsymbol{A}^{\dagger} \boldsymbol{E}_S \boldsymbol{W} \boldsymbol{E}_S^* \boldsymbol{A}^{\dagger *}\}\,, \tag{50}$$

$$(\boldsymbol{Q})_{ij} = 2\sigma^2 \mathrm{Re}\{\boldsymbol{A}_j^* \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{A}_i \boldsymbol{A}^{\dagger} \boldsymbol{E}_S \boldsymbol{W} \boldsymbol{\Lambda}_S \tilde{\boldsymbol{\Lambda}}^{-2} \boldsymbol{W} \boldsymbol{E}_S^* \boldsymbol{A}^{\dagger *}\}\,, \tag{51}$$

where $\boldsymbol{W}$ *is any Hermitian positive definite weighting matrix and all expressions above are evaluated at the true parameter values. For the special case* $\boldsymbol{A} \in \mathscr{A}^d$ *and there is only one parameter,* θ*, associated with each signal, the matrices of the asymptotic distribution have a simple matrix form*

$$\bar{V}'' = -2\mathrm{Re}\{(\boldsymbol{D}^* \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{D}) \odot (\boldsymbol{A}^{\dagger} \boldsymbol{E}_S \boldsymbol{W} \boldsymbol{E}_S^* \boldsymbol{A}^{\dagger *})^{\mathrm{T}}\}\,, \tag{52}$$

$$\boldsymbol{Q} = 2\sigma^2 \mathrm{Re}\{(\boldsymbol{D}^* \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{D}) \odot (\boldsymbol{A}^{\dagger} \boldsymbol{E}_S \boldsymbol{W} \boldsymbol{\Lambda}_S \tilde{\boldsymbol{\Lambda}}^{-2} \boldsymbol{W} \boldsymbol{E}_S^* \boldsymbol{A}^{\dagger *})^{\mathrm{T}}\}\,, \tag{53}$$

where

$$\boldsymbol{D} = \left[\frac{\partial \boldsymbol{a}}{\partial \theta}\bigg|_{\theta=\theta_1}, \ldots, \frac{\partial \boldsymbol{a}}{\partial \theta}\bigg|_{\theta=\theta_d} \right]. \tag{54}$$

The symbol $\odot$ *denotes element-wise multiplication.*

Notice that Theorem 4.1 holds for arbitrary signal correlation (including full coherence). Below, the covariance expression is examined for some special choices of the weighting matrix. For convenience, $\boldsymbol{A} \in \mathscr{A}^d$ and $\boldsymbol{\eta}_k = \theta_k$ is assumed.

COROLLARY 4.1.1. (i) *The asymptotic covariance of the unweighted MD-MUSIC method* ($\boldsymbol{W} = \boldsymbol{I}$) *is given by* (49) *with* ($\boldsymbol{S}$ *is full rank*)

$$\bar{V}'' = -2\mathrm{Re}\{(\boldsymbol{D}^* \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{D}) \odot (\boldsymbol{A}^* \boldsymbol{A})^{-\mathrm{T}}\}\,, \tag{55}$$

$$\begin{aligned} \boldsymbol{Q} = 2\sigma^2 \mathrm{Re}\{&(\boldsymbol{D}^* \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{D}) \odot [(\boldsymbol{A}^* \boldsymbol{A} \boldsymbol{S} \boldsymbol{A}^* \boldsymbol{A})^{-1} \\ &+ \sigma^2 (\boldsymbol{A}^* \boldsymbol{A} \boldsymbol{R}_{SS} \boldsymbol{A}^* \boldsymbol{A} \boldsymbol{R}_{SS} \boldsymbol{A}^* \boldsymbol{A})^{-1}]^{\mathrm{T}}\}\,. \end{aligned} \tag{56}$$

(ii) *The asymptotic covariance of the deterministic ML method* ($\boldsymbol{W} = \tilde{\boldsymbol{\Lambda}}$) *is*

$$\boldsymbol{C} = \mathrm{CRB}_{\mathrm{DET}} + (\bar{V}'')^{-1} \tilde{\boldsymbol{Q}} (\bar{V}'')^{-1}\,, \tag{57}$$

where

$$\bar{V}'' = -2\mathrm{Re}\{(\boldsymbol{D}^*\boldsymbol{P}_A^{\perp}\boldsymbol{D}) \odot \boldsymbol{R}_{SS}^{\mathrm{T}}\}\,, \tag{58}$$

$$\tilde{\boldsymbol{Q}} = 2\sigma^4\mathrm{Re}\{(\boldsymbol{D}^*\boldsymbol{P}_A^{\perp}\boldsymbol{D}) \odot (\boldsymbol{A}^*\boldsymbol{A})^{-\mathrm{T}}\}\,, \tag{59}$$

$$\mathrm{CRB}_{\mathrm{DET}} = \tfrac{1}{2}\sigma^2[\mathrm{Re}\{(\boldsymbol{D}^*\boldsymbol{P}_A^{\perp}\boldsymbol{D}) \odot \boldsymbol{R}_{SS}^{\mathrm{T}}\}]^{-1}\,. \tag{60}$$

The $\mathrm{CRB}_{\mathrm{DET}}$ *is the asymptotic deterministic Cramér–Rao lower bound as derived in* [57, 14].

(iii) *The asymptotic covariance of the weighted subspace fitting method with weights* $\boldsymbol{W} = \tilde{\Lambda}^2\Lambda_S^{-1}$ *is*

$$\boldsymbol{C} = \tfrac{1}{2}\sigma^2[\mathrm{Re}\{(\boldsymbol{D}^*\boldsymbol{P}_A^{\perp}\boldsymbol{D}) \odot (\boldsymbol{S}\boldsymbol{A}^*\boldsymbol{R}\boldsymbol{x}\boldsymbol{x}^{-1}\boldsymbol{A}\boldsymbol{S})^{\mathrm{T}}\}]^{-1}\,. \tag{61}$$

The derivation of these expressions can be found in [67]. Note that the estimates of the deterministic ML method do not achieve the deterministic CRB. The DML technique does not provide consistent estimates of all model parameters and therefore efficiency is lost. In the deterministic model, the signal waveforms themselves are regarded as unknown parameters. As a result the dimension of the parameter vector grows without bound with increasing N and consistent estimation of the signal waveforms is impossible.

The expressions above are well suited for numerical evaluation and the theoretical standard deviation of the signal parameter estimates can be used to accurately predict algorithm performance [31] and to compare with results for other methods. The ESPRIT algorithm has been displayed here in its simplest form. A more general version, and an analysis thereof can be found in [33].

The expression for the asymptotic DML covariance in (57) is consistent with the expressions derived in [32] and [56].

4.3. Relation to maximum likelihood

In this section, the relation between the estimates obtained with the SML technique described in Section 3.1 and the SSF estimates of the previous section is examined. As in the previous analysis, this is initially done under the assumption of Gaussian signal waveforms.

The $\mathrm{CRB}_{\mathrm{STO}}$ is a lower bound on the estimation error variance for *any* unbiased estimator. An estimator that (asymptotically) achieves the $\mathrm{CRB}_{\mathrm{STO}}$ is said to be (asymptotically) *efficient*. The CRB based on the Gaussian signal model is discussed in [3] and is easily derived from the negative log likelihood function in (16). The (i, j)-th element of the inverse of the CRB for a single observation ($N = 1$) is given by

$$\{\mathrm{CRB}_{\mathrm{STO}}^{-1}\}_{ij} = \mathrm{E}\left[\frac{\partial^2 l(\boldsymbol{\xi}_0)}{\partial \boldsymbol{\xi}_j\, \partial \boldsymbol{\xi}_i}\right] = \mathrm{Tr}\{\boldsymbol{R}^{-1}\boldsymbol{R}_i\boldsymbol{R}^{-1}\boldsymbol{R}_j\}\,. \tag{62}$$

Although the above formula can easily be numerically evaluated, it does not give much insight into the behavior of the CRB as a function of the scenario dependent parameters.

Provided the likelihood function is *sufficiently regular*, the ML estimate is asymptotically Gaussian with covariance equal to the CRB [23]. The CRB gives the covariance of the entire parameter vector $\boldsymbol{\xi}$ whereas the SSF method only provides estimates of the signal parameters $\boldsymbol{\eta}$. In order to compare the asymptotic properties of ML and SSF, an expression for the asymptotic distribution of the signal parameters, $\boldsymbol{\eta}$ is required.

Extracting the covariance of $\hat{\boldsymbol{\eta}}$ from the CRB matrix directly seems like a formidable task. The following result, [29, 59], provides the asymptotic distribution of the ML signal parameter estimates.

THEOREM 4.2. *Let $\hat{\boldsymbol{\eta}}$ be the minimizing argument of* (17). *The normalized estimation error $\sqrt{N}\tilde{\boldsymbol{\eta}} = \sqrt{N}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ has a limiting Gaussian distribution with zero-mean and the (i, j)-th element of the covariance is given by*

$$\tfrac{1}{2}\sigma^2[\mathrm{Re}\{\boldsymbol{A}_j^* \boldsymbol{P}_A^{\perp} \boldsymbol{A}_i \boldsymbol{R}_{SS} \boldsymbol{A}^* \boldsymbol{R}_{XX}^{-1} \boldsymbol{A} \boldsymbol{R}_{SS}\}]^{-1} . \tag{63}$$

Since the SML estimates are asymptotically efficient, the above result provides a compact expression for the Cramér–Rao bound on the signal parameters. More important, the following result is obtained by comparing the expression above with (61).

THEOREM 4.3. *The estimate of the signal parameter vector $\boldsymbol{\eta}$ obtained with the optimally weighted subspace fitting method has the same asymptotic distribution as the SML estimate and is thus asymptotically efficient.*

As discussed previously the multidimensional SSF techniques require a multidimensional non-linear optimization for computing the signal parameter estimates. However, for the special case $\boldsymbol{A} \in \mathscr{C}$ the global optimum may be found in a very efficient manner using the ESPRIT algorithm. For the case $\boldsymbol{A} \in \mathscr{A}^d$ a Newton-type algorithm is proposed in [69] for performing the numerical search of the cost function. This method requires $O(Md^2)$ operations per iteration and is thus an order of magnitude computationally less expensive compared to iterations of the SML technique.

4.4. Asymptotic robustness

The asymptotic results presented in the previous section are based on the assumption that the signal waveforms $\boldsymbol{s}(t)$ and the additive noise $\boldsymbol{n}(t)$ have Gaussian distributions. This section we comment on the applicability of the previous results for the more general case of arbitrary emitter signals. In the following, the noise is still assumed to be Gaussian. It is found that the asymptotic distribution derived for Gaussian signals is preserved in the non-Gaussian case. This interesting property is known as *asymptotic robustness*.

Assume that the emitter signals, $\{s(t)\}$, are arbitrary deterministic (i.e., fixed), second-order ergodic sequence of vectors. The signal 'covariance' is still denoted $\boldsymbol{R}_{SS}$, but it is defined by the following deterministic limit

$$\boldsymbol{R}_{SS} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} s(t)s^*(t) \,. \tag{64}$$

Consistency of the estimates immediately generalizes to the non-Gaussian case. For all methods discussed, consistency depends only on the convergence of the sample covariance matrix and if the signals are second-order ergodic, $\hat{\boldsymbol{R}}_{XX}$ can easily be shown to converge with probability 1 to $\boldsymbol{R}_{XX}$.

The asymptotic distribution of subspace fitting estimates is based on the distribution of the signal eigenvectors of the sample covariance matrix. The case of complex observations with a deterministic, second-order ergodic signal in additive Gaussian noise is treated in [34]. There it is also shown that the estimates obtained through the subspacing fitting criterion (36), are asymptotically robust. We state this as a theorem.

THEOREM 4.4. *Let the emitter signals be second-order ergodic sequences. The SSF parameter estimates obtained in* (36) *have the same asymptotic properties as in the case of Gaussian signal waveforms, i.e.*

$$\sqrt{N}\tilde{\boldsymbol{\eta}} \in \text{As N}(0, \boldsymbol{C}_{\text{SML}}) \,, \tag{65}$$

where $\boldsymbol{C}$ *is given by* (49).

Note that this applies to the subspace fitting criterion (36) for any choice of weighting matrix $\boldsymbol{W}$ and also for an arbitrary (identifiable) parametrization of $\boldsymbol{A}$. Hence, one can conclude that the deterministic ML method, MD-MUSIC, WSF, and ESPRIT are all asymptotically robust estimation techniques.

4.5. Discussion

The asymptotic efficiency and robustness properties of the stochastic ML and WSF methods have some interesting implications on the modeling aspects of the sensor array problem. There are two competing models for the sensor array problem, namely *deterministic* versus *stochastic* modeling of the emitter signals. These lead to different ML criteria and CRBs. If the emitter signals are Gaussian, it is clear that the stochastic ML method is optimal. If, on the other hand, the signals have an unknown non-Gaussian distribution or are modeled as deterministic, the relation between the deterministic ML criterion and the Gaussian based ML method is unclear. It is also of interest to compare the CRB for the Gaussian case with the corresponding bounds for other signal distributions. The following corollary to Theorems 4.2, 4.3, and 4.4 establishes these relations.

COROLLARY 4.4.1. *Let* C_{WSF} *and* C_{DET} *denote the asymptotic covariances of* $\sqrt{N}\hat{\boldsymbol{\theta}}$, *for the WSF and deterministic ML estimates respectively. Furthermore, let* $\mathrm{CRB}_{\mathrm{STO}}$ *and* $\mathrm{CRB}_{\mathrm{ARB}}$ *be the stochastic CRB and the CRB for any other signal distribution or deterministic signals respectively. The following inequalities then hold*

$$C_{\mathrm{DET}} \geq C_{\mathrm{WSF}} = \mathrm{CRB}_{\mathrm{STO}} \geq \mathrm{CRB}_{\mathrm{ARB}} . \tag{66}$$

Proof of the first inequality is evident if we apply the deterministic ML method to Gaussian signals. The DML estimator provides asymptotically unbiased parameter estimates. Theorem 4.3 and the CRB inequality then imply that the first inequality in (66) holds. This, together with the asymptotic robustness of the DML and WSF methods implies that the first inequality holds for arbitrary signals.

To prove the second inequality, note that $\mathrm{CRB}_{\mathrm{STO}}$ only depends on second-order properties of the signals. Apply the WSF method to non-Gaussian signals, since the WSF estimates are asymptotically unbiased and the asymptotic covariance of the estimates still equals $C_{\mathrm{WSF}} = \mathrm{CRB}_{\mathrm{STO}}$, this result follows from the Cramér–Rao inequality.

The corollary above is a more general version of a similar theorem from [30]. The first inequality provides strong justification for the stochastic model being appropriate for the sensor array problem. The second inequality says that the case of Gaussian signal waveforms can be interpreted as a *worst case*. If the signal waveforms have a known non-Gaussian distribution, it is usually possible to achieve better performance than the SML and WSF methods. This would require the maximization of the appropriate likelihood function, which can be very complicated. Notice also that the resulting method cannot be asymptotically robust.

It should be noted that the results here refer to asymptotic in *the number of snapshots N*. Although there is a strong connection between DOA estimation in sensor arrays and cisoid estimation in time series, the asymptotic statements are different. The dual statement to having a large number of data in time series, is that *the number of sensors M* is large. For large M, the deterministic ML method is efficient and the deterministic model is appropriate.

Although the stochastic ML and WSF techniques have identical asymptotic properties, the finite sample behavior may be different. There are other aspects which also need to be considered when comparing the methods, for example computational complexity and global minimization properties. These issues require further study.

It should be noted that there is a class of multidimensional array processing techniques referred to as MODE, [57, 59, 63, 62], which are quite similar to the SSF techniques. The MODE technique is generally formulated in terms of the estimated noise subspace

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} V_{\mathrm{NSF}}(\boldsymbol{\theta}) , \tag{67}$$

where

$$V_{\mathrm{NSF}}(\boldsymbol{\theta}) = \|\hat{\boldsymbol{E}}_{\mathrm{N}}^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{\theta})\|_{U}^{2}$$
$$= \mathrm{Tr}\{\hat{\boldsymbol{E}}_{\mathrm{N}}^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{U}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{\theta})\hat{\boldsymbol{E}}_{\mathrm{N}}\}$$
$$= \mathrm{Tr}\{\boldsymbol{U}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{\theta})\hat{\boldsymbol{E}}_{\mathrm{N}}\hat{\boldsymbol{E}}_{\mathrm{N}}^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{\theta})\}\,, \tag{68}$$

where $\boldsymbol{U} \geqslant \boldsymbol{0}$ is a $d \times d$ weighting matrix. This technique has been extensively analyzed and the optimal weighting derived [57, 59, 63, 62]. When the optimal weighting is applied, the MODE technique provides asymptotically efficient estimates provided $d' = d$. This weighting however depends on $\boldsymbol{\eta}_0$ and a two-step procedure must be applied as described in [63]. The connection between MODE and WSF is discussed in [35].

All methods discussed herein assume knowledge of the number of emitter signals, d. In many practical situations, the number of emitters is not known and must be estimated in addition to the parameters. If no coherent sources are present, well-established techniques such as Rissanen's minimum description length (MDL) and Akaike's information criterion (AIC) can be employed to estimate the number of emitters. In the presence of coherent sources, $d' < d$ and alternative methods are required. In [77] an MDL based approach to the detection problem is proposed. An alternative detection method based on the WSF criterion is discussed in [69]. Therein, the distribution of the WSF cost function at the global minimum is derived and based on this, a hypothesis testing scheme is proposed.

5. A performance analysis for model errors

In the above asymptotic analysis, it has been assumed that the array manifold and the spatial covariance of the noise were both known exactly. Unfortunately, these assumptions are never satisfied in practice. Due to changes in weather, the surrounding environment, and antenna location, the response of the array may be significantly different than when it was last calibrated. Furthermore, the calibration measurements themselves are subject to gain and phase errors. For the case of analytically calibrated arrays of nominally identical, identically oriented elements, errors result since the elements are not really identical and their locations are not precisely known. Depending on the degree to which the actual antenna response differs from its nominal value, algorithm performance may be significantly degraded.

The requirement of known noise statistics is also difficult to satisfy in practice, since the surrounding environment and orientation of the array may be time-varying. In addition, one is often unable to account for the effect of unmodeled 'noise' phenomena such as distributed sources, reverberation, noise due to the antenna platform, and undesirable channel crosstalk. Measurement of the noise statistics is complicated by the fact that there are often signals-of-

interest observed along with the noise and interference. Because of these difficulties, it is often assumed by default that the noise field is isotropic, that it is independent from channel to channel, and that its power in each channel is equal. When the SNR is high, deviations of the noise from these assumptions are not critical since they contribute little to the array covariance. However, at low SNR, the degradation may be severe.

In this section, we briefly discuss the sensitivity of subspace based algorithms to the assumptions of known array response and noise statistics. This work has been previously reported in [60, 51, 64, 52, 65, 54, 53,]. Others have also investigated this issue. Friedlander [15, 16] has considered the sensitivity of MUSIC and deterministic ML for the special case of array errors, and has developed a methodology for determining both the DOA estimation error and resolution threshold of the algorithms. Wong et al., have also considered the performance of MUSIC under the assumption of random sensor motion [76]. The performance of least-squares (LS)-ESPRIT and TAM (among other algorithms) has been studied by Li and Vaccaro for both calibration [27, 25] and noise modeling errors [26]. Soon and Huang [49] have also recently considered the sensitivity of LS-ESPRIT to random sensor errors. Several authors have also investigated DOA estimation algorithms which attempt to mitigate the effects of the antenna and noise model errors described above [38, 39, 42, 18, 75, 72, 74].

A general error model

To isolate the effects of model errors on the DOA estimates, it will be assumed that the finite sample effects due to additive noise are negligible and that an exact measurement of the perturbed covariance $\hat{\boldsymbol{R}}_{XX}$ is available. A very general model for $\hat{\boldsymbol{R}}_{XX}$ is the following:

$$\hat{\boldsymbol{R}}_{XX} = (\boldsymbol{I} + \boldsymbol{\Delta})[(\boldsymbol{A} + \tilde{\boldsymbol{A}})\boldsymbol{R}_{SS}(\boldsymbol{A} + \tilde{\boldsymbol{A}})^* + \sigma^2(\boldsymbol{I} + \tilde{\boldsymbol{\Sigma}})](\boldsymbol{I} + \boldsymbol{\Delta})^* , \tag{69}$$

where the matrices $\boldsymbol{\Delta}$, $\tilde{\boldsymbol{A}}$, and $\tilde{\boldsymbol{\Sigma}}$ are the result of various types of model perturbations. The matrix $\boldsymbol{\Delta}$ contains errors that affect both the signal and noise components of the data. Such errors include gain imbalances in the receiver electronics, channel crosstalk, and mutual coupling effects. The matrix $\tilde{\boldsymbol{A}}$ represents the error in the nominal array response, and incorporates the effects of imprecisely known sensor locations, perturbations in the antenna amplitude and phase patterns, and signal-only mutual coupling. Finally, deviations from the assumption of spatially white noise are denoted by the Hermitian matrix $\tilde{\boldsymbol{\Sigma}}$.

In the asymptotic analysis of the previous section, the distribution of the signal subspace eigenvectors was obtained by exploiting the fact that the sample covariance matrix is Wishart distributed. Here we would like to quantify how the presence of $\boldsymbol{\Delta}$, $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{\Sigma}}$ affect the span of the signal and noise subspaces. Suppose $\boldsymbol{E}_S$ and $\boldsymbol{E}_N$ span the true signal and noise subspaces, and let $\hat{\boldsymbol{E}}_S = \boldsymbol{E}_S + \tilde{\boldsymbol{E}}_S$ and $\hat{\boldsymbol{E}}_N = \boldsymbol{E}_N + \tilde{\boldsymbol{E}}_N$ represent the perturbed spanning sets. It is

shown in [54, 53, 65] that, to first order, the model errors lead to the following subspace perturbations:

$$\hat{\boldsymbol{E}}_{\mathrm{N}}\hat{\boldsymbol{E}}_{\mathrm{N}}^{*}\boldsymbol{A}(\boldsymbol{\theta}_0) = \boldsymbol{E}_{\mathrm{N}}\tilde{\boldsymbol{E}}_{\mathrm{N}}^{*}\boldsymbol{A}(\boldsymbol{\theta}_0) = -\boldsymbol{E}_{\mathrm{N}}\boldsymbol{E}_{\mathrm{N}}^{*}\boldsymbol{\Xi}\,, \tag{70}$$

$$\boldsymbol{P}_{A}^{\perp}\tilde{\boldsymbol{E}}_{\mathrm{S}} = \boldsymbol{P}_{A}^{\perp}\boldsymbol{\Xi}\boldsymbol{T}\,, \tag{71}$$

$$\boldsymbol{\Xi} = \tilde{\boldsymbol{A}} + \boldsymbol{\Delta}\boldsymbol{A} + \sigma^2(\tilde{\boldsymbol{\Sigma}} + \boldsymbol{\Delta}^{*} + \boldsymbol{\Delta})\boldsymbol{A}\boldsymbol{T}\tilde{\boldsymbol{\Lambda}}^{-1}\boldsymbol{T}^{\dagger}\,. \tag{72}$$

In the discussion that follows we consider the special case where $\boldsymbol{\Xi} = \tilde{\boldsymbol{A}}$, and show how an analysis of model errors may be useful.

Models for array perturbations

A perturbed sensor array is typically assumed to be composed of uniformly spaced identical elements, i.e., each sensor is assumed to have identical response, the signal conditioning electronics (e.g., filter gain and phase responsed, etc.) are assumed to perform identically, and the analog-to-digital (A/D) converters are assumed to be synchronized. However, the sensors are not identical (their response patterns are different), and their positions are not uniform. In addition, the filter and AGC characteristics will not be uniform from receiver to receiver, the A/D converters will not be exactly in phase, and there may be uncalibrated or non-uniform mutual coupling present[1]. All of these factors combine in varying degrees to produce the array perturbation $\tilde{\boldsymbol{A}}$.

There are a variety of models that could be used to describe $\tilde{\boldsymbol{A}}$. In practice, the response of a given sensor is typically known to within some tolerance in gain and phase that accounts for variations in the construction of the sensor and the conditions under which it is to operate. This tolerance may be specified as limits above and below some nominal response, or as an expected deviation around the nominal. Consequently, one might assume that $\tilde{\boldsymbol{A}}$ is specified in probabilistic terms (e.g., the mean and variance of the elements of $\tilde{\boldsymbol{A}}$ are assumed known), and that the sensor array is just one realization from the probability space of arrays specified by $\boldsymbol{A}$ and the distribution of $\tilde{\boldsymbol{A}}$.

A particularly simple model of this type that has been widely used [15, 25, 87, 20] is to assume that the columns of $\tilde{\boldsymbol{A}}$ are zero-mean, white, circular, and stationary:

$$\mathscr{E}\{\tilde{\boldsymbol{a}}(\theta_p)\tilde{\boldsymbol{a}}^{*}(\theta_p)\} = v^2\boldsymbol{B}\,, \tag{73}$$

$$\mathscr{E}\{\tilde{\boldsymbol{a}}(\theta_p)\tilde{\boldsymbol{a}}^{*}(\theta_q)\} = 0\,, \quad p \neq q\,, \tag{74}$$

$$\mathscr{E}\{\tilde{\boldsymbol{a}}(\theta_p)\tilde{\boldsymbol{a}}^{\mathrm{T}}(\theta_q)\} = 0 \quad \forall p, q\,. \tag{75}$$

[1] Mutual coupling occurs when a nominally passive collector acts as a transmitter and reradiates some of its received energy.

If the errors are independent from sensor-to-sensor, $\boldsymbol{B}$ is clearly diagonal. Off-diagonal terms indicate sensor-to-sensor correlations that result, for example, if there are uncalibrated mutual coupling effects, or if some sensors tend to perturb uniformly (such as identical or adjacent elements).

In [54, 53, 65], expressions are derived for the DOA estimation error covariance of several popular subspace based methods assuming the error model of (73)–(75). In particular, this is done for a weighted version of MUSIC:

$$\hat{\theta} = \arg\min_{\theta} \frac{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{E}}_{\mathrm{N}}\boldsymbol{W}_{\mathrm{MU}}\hat{\boldsymbol{E}}_{\mathrm{N}}^*\boldsymbol{a}(\theta)}{\boldsymbol{a}^*(\theta)\boldsymbol{a}(\theta)} , \tag{76}$$

and for row- and column-weighted subspace fitting algorithms:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\min_{\theta,T} \|\boldsymbol{W}_{\mathrm{R}}(\hat{\boldsymbol{E}}_{\mathrm{S}}\boldsymbol{W}_{\mathrm{C}}^{1/2} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{T})\|_F^2 \\ &= \arg\min_{\theta} \mathrm{Tr}(\boldsymbol{P}_A^{\perp}\boldsymbol{W}_{\mathrm{R}}\hat{\boldsymbol{E}}_{\mathrm{S}}\boldsymbol{W}_{\mathrm{C}}\hat{\boldsymbol{E}}_{\mathrm{S}}^*\boldsymbol{W}_{\mathrm{R}}) ,\end{aligned} \tag{77}$$

where $\boldsymbol{W}_{\mathrm{R}}$ and $\boldsymbol{W}_{\mathrm{C}}$ are respectively the row and column weightings.

An important outcome of the analysis in [54, 53, 65] was that, for the error model of (73)–(75), the following weightings lead to minimum variance DOA estimates when finite sample effects are neglected:

$$\begin{aligned}\boldsymbol{W}_{\mathrm{MU}} &= (\hat{\boldsymbol{E}}_{\mathrm{N}}^*\boldsymbol{B}\hat{\boldsymbol{E}}_{\mathrm{N}})^{-1} , \\ \boldsymbol{W}_{\mathrm{R}} &= \boldsymbol{B}^{-1/2} , \\ \boldsymbol{W}_{\mathrm{C}} &= (\boldsymbol{T}^*\boldsymbol{T})^{-1} ,\end{aligned} \tag{78}$$

where $\boldsymbol{T}$ is an estimate of the matrix satisfying

$$\boldsymbol{E}_{\mathrm{S}} = \boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{T} .$$

Note that these are quite different weightings than those obtained in the asymptotic analysis of WSF and MUSIC [58]. The simulation results that follow illustrate the advantage of these optimal weightings in cases where array errors are the limiting factor in estimation performance.

Simulation examples

In each of these examples, finite sample effects are neglected by computing in each trial the exact perturbed covariance $\hat{\boldsymbol{R}}_{XX}$. A total of 1000 trials were conducted for each example, with $\hat{\boldsymbol{R}}_{XX}$ generated for each trial using the error-free covariance $\boldsymbol{R}_{XX}$ and the distribution of the perturbation under consideration. The sample RMS error of the DOA estimates was then calculated and compared to that predicted by the corresponding theoretical expressions. The nominal gain of all sensors was assumed to be unity in the direction

of the impinging signals, and it was assumed that the number of emitters d had been correctly determined.

To begin with, consider the array of Figure 4 which is composed of a seven element ULA and two additional elements separated from either end of the ULA by six wavelengths. Suppose that the array is subject to the type of gain and phase perturbation described by the model of (73)–(75), and further suppose that the error covariance $\boldsymbol{B}$ is given by

$$\boldsymbol{B} = \text{diag}\{[10^{-2}, 10^{-4}, 10^{-4}, \ldots, 10^{-4}, 10^{-2}]\} . \tag{79}$$

In other words, the standard deviation of the array perturbation for the end elements is ten times greater than for the elements of the ULA. This particular choice of $\boldsymbol{B}$ corresponds to a standard deviation of about 1% in gain and 0.6° in phase for the ULA, and 10% in gain and 5.7° in phase for the end elements. For this array, a relevant system design problem would be how to appropriately trade-off (a) the performance improvement that results from using the large aperture provided by the end elements, and (b) the performance degradation caused by the unreliable calibration information for these elements.

The array of Figure 4 was simulated with the perturbation described by (79). Two uncorrelated emitters were assumed, each of power 0 dB relative to the additive noise. The first emitter was fixed at 0° broadside, while the second was varied from 2° to 50° over several experiments. Figure 5 gives a plot of the RMS error of the DOA estimates versus the separation of the two sources. The dotted curve and the symbols ○ and × represent the predicted and measured performance of MUSIC and root-MUSIC when using only the 7 element ULA and ignoring the end elements (no MUSIC result is shown for the case where the second source was at 2° since the algorithm failed to resolve the two sources in over half the trials). The solid curve and the symbol ★ denote the predicted and measured performance of MUSIC for the full array (including the end elements) without weighting, while the dashed line and + denote the same for the optimally weighted version of MUSIC.

When the sources are closely spaced, the smaller ULA does not provide enough aperture for MUSIC to accurately estimate the DOAs. As the second source is moved away from broadside, the performance of the ULA improves until at $\theta_2 \simeq 8°$ it does as well as the full unweighted array. For DOAs beyond 8°, using the information from the unweighted end elements actually degrades algorithm performance relative to just ignoring them. However, the lowest estimation error is achieved using the full optimally weighted array. In fact, the

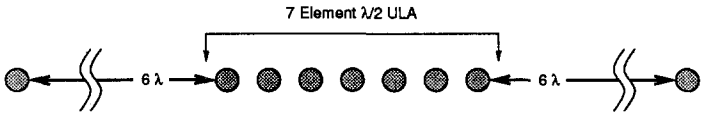

Fig. 4. Non-uniform linear array.

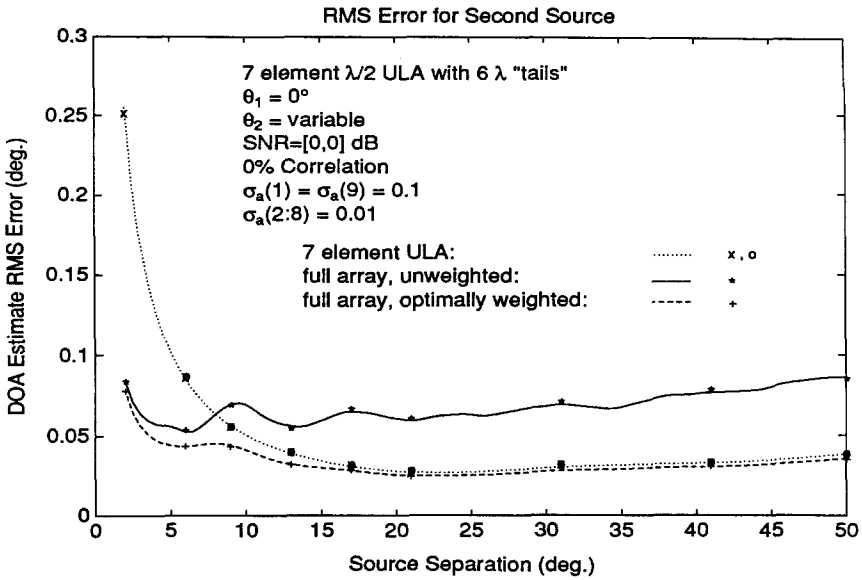

Fig. 5. Performance of MUSIC and root-music vs. source separation for non-uniform array.

CRB for this case is virtually equivalent to the performance predicted for weighted MUSIC. Thus, in this case at least, the weighted MUSIC algorithm appears to be efficient. In all cases, the measured RMS errors are in excellent agreement with those predicted by theory.

As another example, a 10 element uniform linear array (ULA) with $\frac{1}{2}\lambda$ interelement spacing was simulated, with two emitters at 0° broadside and 7°. The SNR for the 7° source was 20 dB, while that of the broadside source was varied between 0 and 12 dB, and the two sources were 90% correlated with 0° correlation phase. The covariance $\boldsymbol{B}$ of (73) was assumed to be of the form $v^2\boldsymbol{I}$, with $v = 0.01$. This corresponds to a -40 dB gain error and a 0.57° phase error standard deviation. Figure 6 shows a plot of the standard deviation of the 0° source estimates for MD-MUSIC, deterministic ML, WSF, and WSF using the weighting of (78) and an estimate of $\boldsymbol{T}$ obtained from initial conditions generated using ESPRIT. The latter algorithm is referred to as RSF to distinguish it from WSF. The connected lines indicate the theoretical predictions and the symbols represent the results of the simulations. Note the improvement achieved by the optimal RSF weighting $\boldsymbol{W}_C = (\hat{\boldsymbol{T}}^*\hat{\boldsymbol{T}})^{-1}$ relative to WSF and deterministic ML (in this case, $\boldsymbol{W}_R = \boldsymbol{I}$). Though not shown on the plot, the one-dimensional MUSIC algorithm has virtually identical performance to that of MD-MUSIC and RSF.

Figure 7 shows the predicted finite sample performance of the algorithms for $N = 250$ snapshots under the assumption of no array errors. Note that the relative performance of the algorithms is reversed compared to Figure 6; in this case WSF and deterministic ML outperform RSF and MD-MUSIC (only two curves seem to be present since the DOA error is virtually identical for WSF

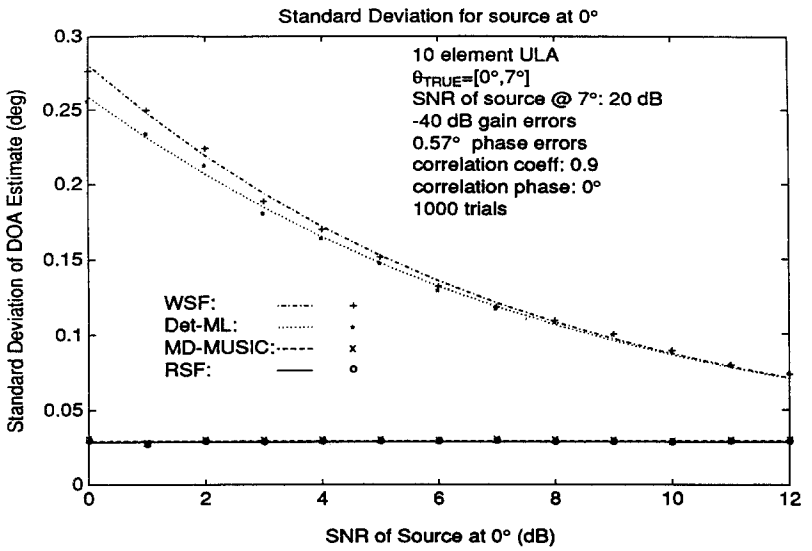

Fig. 6. Actual and predicted DOA errors vs. SNR.

and ML, and also for RSF and MD-MUSIC). This phenomenon has been observed in other cases as well, i.e., in scenarios involving array errors where RSF has the greatest advantage over WSF, it has relatively poor finite sample performance. This illustrates the need for study of combined weightings that optimally account for both sources of error.

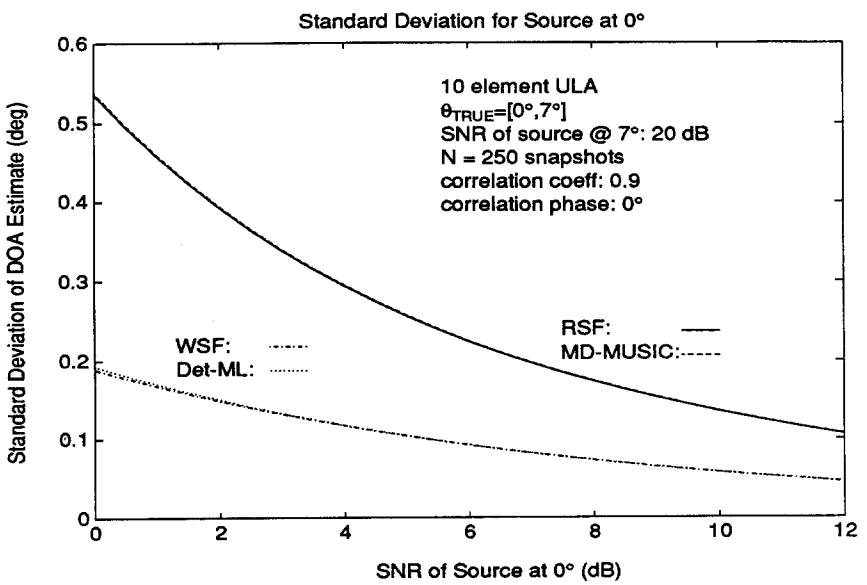

Fig. 7. Predicted DOA errors vs. SNR, finite sample effects only.

6. Fast subspace decomposition (FSD)

This section is concerned with a fast and parallelizable subspace decomposition approach for estimating signal subspaces. In Section 3, the conventional signal subspace decomposition technique employing an eigendecomposition was described. Though the eigendecomposition technique is widely used, the signal eigenvectors form only *one* orthonormal basis for the signal subspace. Furthermore, standard eigendecomposition algorithms do not exploit the (low rank plus a shift) matrix structure of $\boldsymbol{R}_{XX}$ to speed up computation. More importantly, computing the eigendecomposition involves at least $O(M^3)$ multiplications for an $M \times M$ covariance matrix, and in many of its more stable instantiations, involves global communication and significant amounts of local storage, properties that make VLSI parallel implementation difficult. Therefore, most subspace-based algorithms for very large sensor arrays are generally only proposed for off-line data analysis instead of *real-time* implementations.

Several papers have been published in the last five years in an attempt to reduce the computational load for the signal subspace decomposition. These include the power method [66], QR factorization [43] and the subspace iteration method [13], among others. These methods do reduce computational complexity in friendly situations, e.g., when the SNR is very high and the signal subspace dimension is known a priori. They will suffer from slow convergence in more difficult scenarios such as low SNR and unknown d'. None of the above algorithms exploit the fact that the covariance matrix associated with subspace-based algorithms is close to a low-rank (rank d') matrix plus a shift,

$$\boldsymbol{R}_{XX} = \boldsymbol{A}\boldsymbol{R}_{SS}\boldsymbol{A}^* + \sigma^2\boldsymbol{I}\,. \tag{80}$$

This matrix structure is also the basis for the signal subspace techniques. That is to say, if $\boldsymbol{R}_{XX}$ is not of the form defined in (80), the signal subspace algorithms will not yield *consistent* estimates.

Recently, Xu and Kailath [81, 83, 85] exploited such matrix structures and developed a class of $O(M^2d')$ fast subspace decomposition (FSD) techniques for ideal and sample covariance matrices. For ideal covariance matrices, FSD can *exactly* determine the signal subspace $\mathcal{R}\{\boldsymbol{A}\}$ in a *finite* number of steps. This results looks a little surprising, since conventional signal subspace decomposition relies on an eigendecomposition that requires an *infinite* number of iterations for an *exact* solution. Though most traditional detection schemes for determining d' require knowledge of all the eigenvalues, the FSD approach suggests a detection scheme that does not rely on the eigenvalues and that can be carried out in parallel with the process of finding the signal subspace.

For sample covariance matrices, a different version of the FSD algorithm including different detection schemes [83, 80, 85] has been developed to handle the finite-data effect. Numerical and statistical analyses have been appropriately combined to establish the *strong consistency* of the new detection schemes

and the *asymptotic equivalence* of the signal subspace estimates obtained from FSD and a more costly eigendecomposition.

It is worth mentioning that the above FSD algorithms can be modified to proceed directly from the data matrix. The computational complexity is $O(NMd')$ flops, which is even smaller than the cost of forming the covariance matrix, i.e., $O(NM^2)$, where N is the number of data samples. It can be shown that all the asymptotic analysis results also apply to this FSD algorithm. However, in this chapter, we shall only present the FSD algorithms for covariance matrices and FSD based on data matrices can be found in [84, 82, 85].

6.1. Ideal covariance case

An important theorem

In general the *exact* eigendecomposition of an $M \times M$ Hermitian matrix requires an infinite number of steps (if $M > 4$), though in practice satisfactory *approximations* can be calculated in $O(M^3)$ flops. Now, let us consider the special *structured* matrix defined in (80).

$$\boldsymbol{R}_{XX} = \boldsymbol{A}\boldsymbol{R}_{SS}\boldsymbol{A}^* + \sigma^2\boldsymbol{I}\,. \tag{81}$$

We first introduce the important linear algebraic concept of *Krylov* subspaces, perhaps familiar to some readers via the well-known Lanczos algorithm (see, e.g., [36]).

DEFINITION 6.1. For a Hermitian matrix $\boldsymbol{R}_{XX}$ and a vector $\boldsymbol{f}$, Krylov matrices $\boldsymbol{K}^m(\boldsymbol{R}_{XX}, \boldsymbol{f})$ and Krylov subspaces $\mathcal{K}^m(\boldsymbol{R}_{XX}, \boldsymbol{f})$ are defined by

$$\boldsymbol{K}^m(\boldsymbol{R}_{XX}, \boldsymbol{f}) = [\boldsymbol{f}, \boldsymbol{R}_{XX}\boldsymbol{f}, \ldots, \boldsymbol{R}_{XX}^{m-1}\boldsymbol{f}]\,, \tag{82}$$

$$\mathcal{K}^m(\boldsymbol{R}_{XX}, \boldsymbol{f}) = \mathcal{R}\{\boldsymbol{K}^m(\boldsymbol{R}_{XX}, \boldsymbol{f})\}\,, \quad m = 1, 2, \ldots, M\,. \tag{83}$$

As we shall see, the Krylov subspace allows us to exploit the special structure (81). This will become clear from the following lemmas, proved in [81, 79, 85].

LEMMA 6.1. *For any scalar ρ and any Hermitian matrix $\boldsymbol{R}_{XX}$, $\mathcal{K}^m(\boldsymbol{R}_{XX}, \boldsymbol{f}) = \mathcal{K}^m(\boldsymbol{R}_{XX} - \rho\boldsymbol{I}, \boldsymbol{f})$.*

REMARK. Letting $\rho = \sigma^2$, we can easily see that $\mathcal{K}^m(\boldsymbol{R}_{XX}, \boldsymbol{f}) = \mathcal{K}^m(\boldsymbol{R}_{XX} - \sigma^2\boldsymbol{I}, \boldsymbol{f}) = \mathcal{K}^m(\boldsymbol{A}\boldsymbol{R}_{SS}\boldsymbol{A}^*, \boldsymbol{f})$. In other words, though the structured matrix $\boldsymbol{R}_{XX}$ is of full rank for $\sigma^2 \neq 0$, it acts like a rank d' matrix in its Krylov subspaces. This fact is crucial for the successful exploitation of the underlying matrix structure to achieve fast computation.

LEMMA 6.2. *Let $\boldsymbol{R}_{XX}$ be the structured Hermitian matrix* (81) *and suppose it has $d' < M-1$ distinct eigenvalues and one repeated eigenvalue with multiplicity $M-d'$. Assume that we can find an initial $M \times 1$ vector $\boldsymbol{f}$ that is not orthogonal to each of the d' signal eigenvectors and also not orthogonal to the noise subspace. For the Krylov subspaces of $\boldsymbol{R}_{XX}$, it holds that*

(1) $\mathcal{K}^{d'+1}(\boldsymbol{R}_{XX}, \boldsymbol{f}) = \mathcal{R}\{\boldsymbol{E}_S, \boldsymbol{f}\} = \mathcal{R}\{\boldsymbol{A}, \boldsymbol{f}\}$,

(2) $\mathcal{K}^{m}(\boldsymbol{R}_{XX}, \boldsymbol{f}) = \mathcal{K}^{d'+1}(\boldsymbol{R}_{XX}, \boldsymbol{f})$, $m \geq d'+1$.

REMARK. The condition on $\boldsymbol{f}$ will not be hard to satisfy. If we pick $\boldsymbol{f}$ randomly in the M-dimensional space, the probability will be zero that we shall get a *degenerate* $\boldsymbol{f}$, the vector that does not satisfy the stated orthogonality conditions.

These two lemmas lead us to our main result.

THEOREM 6.1. *With $\boldsymbol{R}_{XX}$, $\boldsymbol{f}$ defined in Lemma* 6.2, *the signal subspace, i.e., $\mathcal{R}\{\boldsymbol{A}\}$, can be exactly computed in a* finite *number of steps, namely* $O(M^2(d'+1))$ *flops.*

PROOF. According to Lemma 6.2, we know that $\mathcal{K}^{d'+1}(\boldsymbol{R}_{XX}, \boldsymbol{f}) = \mathcal{R}\{\boldsymbol{E}_S, \boldsymbol{f}\}$. Then $\operatorname{rank}\{\mathcal{K}^{d'+1}(\boldsymbol{R}_{XX}, \boldsymbol{f})\} = \operatorname{rank}\{\boldsymbol{E}_S, \boldsymbol{f}\} = d'+1$. Since $\operatorname{rank}(\boldsymbol{K}^{d'+1}(\boldsymbol{R}_{XX})) = \dim(\mathcal{K}^{d'+1}(\boldsymbol{R}_{XX}, \boldsymbol{f})) = d'+1 < M$, there is at least one non-trivial vector $\boldsymbol{q}$ that is orthogonal to $\mathcal{K}^{d'+1}(\boldsymbol{R}_{XX}, \boldsymbol{f}) = \mathcal{R}\{\boldsymbol{f}, \boldsymbol{A}\}$. Hence, $\boldsymbol{q}^*\boldsymbol{A} = 0$, $\boldsymbol{q}^*\boldsymbol{f} = 0$, and

$$\boldsymbol{q}^*\boldsymbol{R}_{XX} = \boldsymbol{q}^*\boldsymbol{A}\boldsymbol{R}_{SS}\boldsymbol{A}^* + \sigma^2\boldsymbol{q}^* = \sigma^2\boldsymbol{q}^* . \tag{84}$$

Obviously, $\boldsymbol{q}$ is an eigenvector associated with the repeated eigenvalues σ^2. Therefore, with $\boldsymbol{q}$, we can easily obtain σ^2. With σ^2 found, any d' linearly independent columns in $\boldsymbol{R}_{XX} - \sigma^2\boldsymbol{I} = \boldsymbol{A}\boldsymbol{R}_{SS}\boldsymbol{A}^*$ can form a basis for the signal subspace $\mathcal{R}\{\boldsymbol{A}\}$.

To determine $\mathcal{K}^{d'+1}(\boldsymbol{R}_{XX}, \boldsymbol{f})$ requires $d'+1$ matrix–vector products, which take around $O(M^2d')$ multiplications. $\boldsymbol{q}$ can be determined via a Gram–Schmidt orthogonalization, which takes about $O(Md')$ flops if we have already obtained an orthogonal basis for $\mathcal{K}^{d'+1}(\boldsymbol{R}_{XX}, \boldsymbol{f})$. Therefore, the overall computational cost for finding the signal subspace is $O(M^2d')$ flops. More accurate operational counts will be given below. □

The Lanczos algorithm

As will be obvious in the following, determination of d' requires an orthogonal basis $\boldsymbol{Q}_j = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_j]$ of $\mathcal{K}^m(\boldsymbol{R}_{XX}, \boldsymbol{f})$. Of course, this can be done by the Gram–Schmidt orthogonalization procedure. Here, we introduce a more efficient way of finding an orthonormal basis $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_j$. We start with a vector $\boldsymbol{f} = \boldsymbol{q}_1$ and calculate $\boldsymbol{R}_{XX}\boldsymbol{q}_1$, which is then orthogonalized against $\boldsymbol{q}_1$. Let $\boldsymbol{q}_2$ be the *normalized* orthogonal component. We then calculate $\boldsymbol{R}_{XX}\boldsymbol{q}_2$ and orthogonalize it against $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ to yield $\boldsymbol{q}_3$. $\cdots$ At the j-th step, we must

orthogonalize $\boldsymbol{R}_{XX}\boldsymbol{q}_{j-1}$ against $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_{j-1}$ to obtain $\boldsymbol{q}_j$. In fact, however, we do not need to do so. Lanczos [21] showed that $\boldsymbol{R}_{XX}\boldsymbol{q}_j$ is already orthogonal to $\boldsymbol{q}_i$, $i < j-1$, i.e.,

$$\boldsymbol{q}_i^* \boldsymbol{R}_{XX} \boldsymbol{q}_j = 0\,, \quad i < j-1\,. \tag{85}$$

Therefore, to obtain $\boldsymbol{q}_{j+1}$, we only need to orthogonalize $\boldsymbol{R}_{XX}\boldsymbol{q}_j$ against $\boldsymbol{q}_{j-1}$ and $\boldsymbol{q}_j$. Furthermore, it is easy to show that $\mathcal{R}\{\boldsymbol{Q}_j\} = \mathcal{R}\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_j\} = \mathcal{K}^j(\boldsymbol{R}_{XX}, \boldsymbol{f})$.

If we define

$$\alpha_j = \boldsymbol{q}_j^* \boldsymbol{R}_{XX} \boldsymbol{q}_j \qquad \beta_{j-1} = \boldsymbol{q}_{j-1}^* \boldsymbol{R}_{XX} \boldsymbol{q}_j\,, \tag{86}$$

then it turns out that

$$\boldsymbol{R}_{XX}\boldsymbol{q}_j = \boldsymbol{q}_{j-1}\beta_{j-1} + \boldsymbol{q}_j\alpha_j + \boldsymbol{q}_{j+1}\beta_j\,, \tag{87}$$

$$\beta_j = \|\boldsymbol{R}_{XX}\boldsymbol{q}_j - \boldsymbol{q}_j\alpha_j - \boldsymbol{q}_{j-1}\alpha_{j-1}\|\,. \tag{88}$$

We can also see that

$$\boldsymbol{Q}_j^* \boldsymbol{R}_{XX} \boldsymbol{Q}_j = \boldsymbol{T}_j = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \beta^2 & \ddots & \ddots & \\ & & \ddots & \alpha_{j-1} & \beta_{j-1} \\ & & & \beta_{j-1} & \alpha_j \end{bmatrix}, \tag{89}$$

which is a tridiagonal matrix.

The proof of the above results can be found, for example, in [36, Section 12-6].

Determination of signal subspace dimension

In this section, we shall show how to determine d'. One obvious approach is to check the multiplicity of the smallest eigenvalues among all the M eigenvalues. If we do not have the knowledge of eigenvalues, the following lemma will show how to find d' in the process of creating the Krylov subspace.

LEMMA 6.3. *For the same $\boldsymbol{R}_{XX}$ and $\boldsymbol{f}$ as defined in Lemma 6.2, if the Lanczos algorithm is used, then β_j defined in* (86) *is strictly positive for $j \leq d'$ and zero for $j = d'+1$, i.e.,*

$$\beta_j > 0\,, \quad j \leq d' \qquad \beta_j = 0\,, \quad j = d'+1\,. \tag{90}$$

The proof of Lemma 6.3 can be found in [85]. With Lemma 6.3, we can determine d' by checking whether β_j is zero or not, i.e., $\beta_j \stackrel{?}{=} 0$.

The algorithm for signal subspace decomposition
In the proof of Theorem 6.1, we have already provided a method for computing σ^2, after $\beta_j = 0$ is detected. In fact, there is an even simpler approach justified by the following lemma, which is proved in [85].

LEMMA 6.4. *With* $\boldsymbol{T}_{d'+1}$ *as in* (89) *and* $\boldsymbol{R}_{XX}$, f *as in Lemma* 6.2, *we maintain that*

$$\sigma^2 = \frac{\mathrm{Tr}(\boldsymbol{R}_{XX}) - \sum_{i=1}^{d'+1} \alpha_i}{M - d' - 1}. \tag{91}$$

The above results can be summarized as follows.

FSD ALGORITHM FOR IDEAL COVARIANCE MATRICES.
(1) Let $\alpha = 0$.
(2) Carry out the Lanczos algorithm.
(3) Update α and assign the result to α: $\alpha \leftarrow \alpha + \alpha_j$.
(4) If $\beta_j > 0$, return to Step (2) and continue the Lanczos iteration. Otherwise, $d' \overset{\text{set}}{=} j - 1$ and go to Step (5).
(5) Compute $(\mathrm{Tr}(\boldsymbol{R}_{XX}) - \alpha)/(M - d' - 1)$ and store the result in σ^2:

$$\sigma^2 \leftarrow (\mathrm{Tr}(\boldsymbol{R}_{XX}) - \alpha)/(M - d' - 1).$$

(6) Compute $\tilde{\boldsymbol{R}} = \boldsymbol{R}_{XX} - \sigma^2 \boldsymbol{I}$.
(7) Conduct a Gram–Schmidt orthogonalization to find d' orthonormal basis vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{d'}$ from all the columns of $\tilde{\boldsymbol{R}}$.

As shown in [85], the computational complexity of the above algorithm is $M^2(d'+1) + M(d'^2 + 7d' + 5)$ flops.

6.2. *Sample covariance case*

Summary of the FSD algorithm
In the above section, we laid the theoretical ground for exploiting the common matrix structure associated with subspace based algorithms and proposed a *finite-step* algorithm to determine the signal subspace from $\boldsymbol{R}_{XX}$. In this section, we propose a different version of FSD [83, 80, 85] to estimate the signal subspace and its dimension from a sample covariance matrix, an estimate of the ideal covariance based on a finite number (N) of snapshots. More importantly, a combination of statistical and numerical analyses is presented, showing that the FSD estimates of the signal subspace dimension are *strongly consistent* and that its signal subspace estimation *asymptotically equivalent* to those relying on a more costly eigendecomposition.

Certainly, the above finite-step FSD algorithm relies on $\boldsymbol{R}_{XX}$ having *exactly* the structure (81). However, this will not be true in practice where we usually only have a *sample* covariance matrix $\hat{\boldsymbol{R}}_{XX}$ which is an $O(N^{-1/2})$ estimate of $\boldsymbol{R}_{XX}$ based on N data samples,

$$\boldsymbol{R}_{XX} - \hat{\boldsymbol{R}}_{XX} = O(N^{-1/2}) . \tag{92}$$

It will be useful to note that (see [2])

$$\boldsymbol{e}_k - \hat{\boldsymbol{e}}_k = O(N^{-1/2}) , \quad k = 1, 2, \ldots, d' , \tag{93}$$

$$\lambda_k - \hat{\lambda}_k = O(N^{-1/2}) , \quad k = 1, \ldots, M . \tag{94}$$

Therefore,

$$\mathcal{R}\{\boldsymbol{A}\} = \mathcal{R}\{\boldsymbol{e}_k\}_{k=1}^{d'} = \mathcal{R}\{\hat{\boldsymbol{e}}_k\}_{k=1}^{d'} + O(N^{-1/2}) . \tag{95}$$

Of course the computation of $\{\hat{\lambda}_k, \hat{\boldsymbol{e}}_k\}$ is of order $O(M^3)$ flops; so this is not done in the extension of the FSD. Instead, we use certain so-called Rayleigh–Ritz (RR) values $\{\theta_k^{(m)}\}$ and RR vectors $\{\boldsymbol{y}_k^{(m)}\}$ that can be obtained in a manner to be described later at the m-th step of the Lanczos recursion. The point is that the RR values and vectors that we shall use can be found with only $O(d^3)$ flops and they have the following important asymptotic properties: for $m > d'$,

$$\theta_k^{(m)} - \hat{\lambda}_k = O(N^{-(m-d')}) , \quad k = 1, 2, \ldots, d' , \tag{96}$$

$$\boldsymbol{y}_k^{(m)} - \hat{\boldsymbol{e}}_k = O(N^{-(m-d')/2}) , \quad k = 1, 2, \ldots, d' . \tag{97}$$

Hence, once $m \geq d' + 2$,

$$\lim_{N\to\infty} \sqrt{N}\{(\boldsymbol{y}_k^{(m)} - \boldsymbol{e}_k) - (\hat{\boldsymbol{e}}_k - \boldsymbol{e}_k)\} = 0 , \quad k = 1, 2, \ldots, d' . \tag{98}$$

In other words, $\mathcal{R}\{\boldsymbol{y}_k^{(m)}\}_{k=1}^{d'}$ and $\mathcal{R}\{\hat{\boldsymbol{e}}_k\}_{k=1}^{d'}$ are *asymptotically equivalent* estimates of the *true* signal subspace $\mathcal{R}\{\boldsymbol{A}\} = \mathcal{R}\{\boldsymbol{e}_k\}_{k=1}^{d'}$.

Unless the signal subspace dimension d' is known a priori, the first step will be to estimate d'. Conventional methods use all the eigenvalues of $\hat{\boldsymbol{R}}_{XX}$ to carry out the likelihood ratio test [22] or to form certain information theoretic criteria, e.g., AIC, MDL [73, 70]. The FSD detection scheme is based on a sequential set of hypothesis tests involving the RR values at each Lanczos recursion or Lanczos step. It goes as follows.

Assume we are at the m-th Lanczos recursion where we have found $\{\alpha_i, \beta_i\}_{i=1}^{m}$ and $\boldsymbol{Q}_m = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_m]$ as an orthonormal basis for $\mathcal{K}^m(\hat{\boldsymbol{R}}_{XX}, \boldsymbol{f})$. It turns out [36] that

$$\boldsymbol{Q}_m^* \hat{\boldsymbol{R}}_{XX} \boldsymbol{Q}_m \triangleq \boldsymbol{T}_m , \tag{99}$$

where $\boldsymbol{T}_m$ is an $m \times m$ *real* and *tridiagonal* matrix. Find the eigenvalues $\theta_k^{(m)}$ of $\boldsymbol{T}_m$. (The notation $\theta_k^{(m)}$ is used because it can be shown that the eigenvalues of $\boldsymbol{T}_m$ are exactly the RR values associated with $\mathcal{K}^m(\hat{\boldsymbol{R}}_{XX}, \boldsymbol{f})$.) Now, form the test statistic $\varphi_{\hat{d}}$.

$$\varphi_{\hat{d}} = N(M-\hat{d}) \log \left(\frac{\sqrt{\dfrac{1}{M-\hat{d}} \left(\|\hat{\boldsymbol{R}}_{XX}\|^2 - \sum_{k=1}^{\hat{d}} \theta_k^{(m)2} \right)}}{\dfrac{1}{M-\hat{d}} \left(\mathrm{Tr}(\hat{\boldsymbol{R}}_{XX}) - \sum_{k=1}^{\hat{d}} \theta_k^{(m)} \right)} \right),$$

$$\hat{d} = 0, 1, \ldots, m-2 . \tag{100}$$

Under the (null) hypothesis that the signal subspace is $\hat{d}$-dimensional, we will show that $\varphi_{\hat{d}}$ is asymptotically chi-squared (χ^2) distributed if $\hat{d} = d'$ and $\varphi_{\hat{d}} \overset{N\to\infty}{\longrightarrow} \infty$ with probability if $\hat{d} < d'$. Then, the first $\hat{d}$ such that $\varphi_{\hat{d}} \leqslant \gamma_{\hat{d}}$ is taken as the estimate of the signal subspace dimension d', where $\gamma_{\hat{d}}$ is a pre-calculated threshold based on the tail area of a certain χ^2 distribution. It will be shown (see Section 6.1) that this estimate is *strongly consistent*. Once d' is estimated, we compute the eigenvectors $\{\boldsymbol{s}_k^{(m)}\}_{k=1}^m$ of $\boldsymbol{T}_m$ and select d' of them corresponding to the d' largest eigenvalues of $\boldsymbol{T}_m$. We shall relabel them as $\boldsymbol{s}_1^{(m)}, \ldots, \boldsymbol{s}_{d'}^{(m)}$ and then form the vectors $\boldsymbol{Q}_m \boldsymbol{s}_k^{(m)} = \boldsymbol{y}_k^{(m)}$. (This notation is used because it can be shown that the $\{\boldsymbol{y}_k^{(m)}\}_{k=1}^{d'}$ are exactly the RR vectors associated with $\mathcal{K}^m(\hat{\boldsymbol{R}}_{XX}, \boldsymbol{f})$.) The signal subspace estimate is $\mathcal{R}\{\boldsymbol{y}_k^{(m)}\}_{k=1}^{d'}$. As will be shown in Section 6.1, this signal subspace estimate is asymptotically equivalent to the principal eigenspace $\mathcal{R}\{\hat{\boldsymbol{e}}_k\}_{k=1}^{d'}$ of $\hat{\boldsymbol{R}}_{XX}$. We have completed a *brief* explanation of our FSD algorithm for sample covariance matrices. More rigorous justification will be presented in the following sections.

A combination of numerical and statistical analyses

Let us first give a formal definition of the aforementioned Rayleigh–Ritz (RR) values and vectors.

DEFINITION 6.2. For an m-dimensional subspace $\mathcal{S}^m$, the Rayleigh–Ritz (RR) values $\theta_i^{(m)}$ and vectors $\boldsymbol{y}_i^{(m)}$ of a Hermitian matrix $\hat{\boldsymbol{R}}_{XX}$ are defined such that

$$\hat{\boldsymbol{R}}_{XX} \boldsymbol{y}_i^{(m)} - \theta_i^{(m)} \boldsymbol{y}_i^{(m)} \perp \mathcal{S}^m . \tag{101}$$

LEMMA 6.5. *Let* $(\theta_i^{(m)}, \boldsymbol{y}_i^{(m)})$, $i = 1, \ldots, m$ *be the RR values and vectors of the subspace* $\mathcal{S}^m$ *and* $\boldsymbol{Q}_m = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_m]$ *be an orthonormal basis of the same subspace. If* $(\alpha_i, \boldsymbol{s}_i)$ *be the eigenpairs of the* $m \times m$ *matrix* $\boldsymbol{Q}_m^* \hat{\boldsymbol{R}}_{XX} \boldsymbol{Q}_m$, $i = 1, 2, \ldots, m$, *then*

$$\theta_i^{(m)} = \alpha_i , \tag{102}$$

$$\boldsymbol{y}_i^{(m)} = \boldsymbol{Q}_m \boldsymbol{s}_i . \tag{103}$$

The proofs of Lemmas 6.5 can be found in [36]. The above lemma show us a method of finding RR values and vectors. The RR values and vectors from $\mathcal{K}^m(\hat{\boldsymbol{R}}_{XX}, \boldsymbol{f})$ are used to approximate the desired eigenvalues and eigenvectors. With the Lanczos basis, the problem of finding the *desired* eigenvalues and eigenvectors of $M \times M$ regular Hermitian matrix turns into a problem of computing the eigendecomposition of a small $m \times m$ *real* and *tridiagonal* matrix. This is one important and appealing property of the Lanczos method.

THEOREM 6.2. *Let* $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_M$ *be the eigenvalues and* $\hat{\boldsymbol{e}}_1, \ldots, \hat{\boldsymbol{e}}_M$ *be the eigenvectors of the sample covariance matrix* $\hat{\boldsymbol{R}}_{XX}$ *obtained from N data samples with a Gaussian distribution* $\mathrm{N}(0, \boldsymbol{R}_{XX})$, *where* $\hat{\boldsymbol{R}}_{XX}$ *is the structured covariance (rank-d' matrix* $+ \sigma^2\boldsymbol{I}$) *defined in* (81). Let $\lambda_1 > \cdots > \lambda_{d'} > \lambda_{d'+1} = \cdots = \lambda_M = \sigma^2$ *and* $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_M$ *be the eigenvalues and eigenvectors of the ideal covariance matrix* $\boldsymbol{R}_{XX}$. $\theta_1^{(m)} \geq \theta_2^{(m)} \geq \cdots \geq \theta_m^{(m)}$ *and* $\boldsymbol{y}_1^{(m)}, \ldots, \boldsymbol{y}_m^{(m)}$ *denote the RR values and their associated RR vectors derived from* $\mathcal{K}^m(\boldsymbol{R}_{XX}, \boldsymbol{f})$. *If* $\boldsymbol{f}$ *is chosen such that* $\boldsymbol{f}^*\hat{\boldsymbol{e}}_i \neq 0$, *where* $1 \leq i \leq d'$, *then, for* $k = 1, 2, \ldots, d'$, *we claim*:

(1) *If* $m \geq d' + 1$, *the RR values* $\theta_k^{(m)}$ *approximate their corresponding eigenvalues* $\hat{\lambda}_k$ *to order* $\mathrm{O}(N^{-(m-d')})$, *and the RR vectors* $\boldsymbol{y}_k^{(m)}$ *approximate their corresponding eigenvectors* $\hat{\boldsymbol{e}}_k$ *to order* $\mathrm{O}(N^{-(m-d')/2})$, *i.e.*,

$$\theta_k^{(m)} = \hat{\lambda}_k + \mathrm{O}(N^{-(m-d')}), \tag{104}$$

$$\boldsymbol{y}_k^{(m)} = \hat{\boldsymbol{e}}_k + \mathrm{O}(N^{-(m-d)/2}). \tag{105}$$

(2) *If* $m \geq d' + 1$, $\theta_k^{(m)}$ *and* λ_k *are asymptotically equivalent estimates of* λ_k. *If* $m \geq d' + 2$, $\boldsymbol{y}_k^{(m)}$ *and* $\hat{\boldsymbol{e}}_k$ *are also asymptotically equivalent estimates of* $\boldsymbol{e}_k$.

PROOF. The proof of (1) in this theorem involves quite a few deep numerical linear algebra results and is also complicated. Hence, it is not presented here and interested readers are referred to [80, 85].

With (1), it is quite easy to prove (2). Knowing that $\boldsymbol{y}_k^{(m)} - \hat{\boldsymbol{e}}_k = \mathrm{o}(N^{-1/2})$ when $m \geq d' + 2$, we can similarly establish the *asymptotic equivalence* between $\boldsymbol{y}_k^{(m)}$ and $\hat{\boldsymbol{e}}_k$ for estimating $\boldsymbol{e}_k$. The above proof can be very simply illustrated by Figure 8, where ε is the coefficient of the most significant $\mathrm{O}(N^{-1/2})$ error term. □

Estimation of the signal subspace dimension

For ideal covariance matrices, we only need to check $\beta_m \overset{?}{=} 0$ to determine d'. In the sample covariance case, β_m is almost surely nonzero, when $m = d' + 1$. Nonetheless, by Theorem 6.2, we know that the d' larger RR values are asymptotically equivalent to the signal eigenvalues. This property can be *properly* used to estimate d'. In the following, we first form a statistics $\varphi_{\hat{d}}$, based on which the statistical hypothesis is tested to determine d'. The limiting distribution of $\varphi_{\hat{d}}$ is then derived and a new detection scheme relying on $\varphi_{\hat{d}}$ is

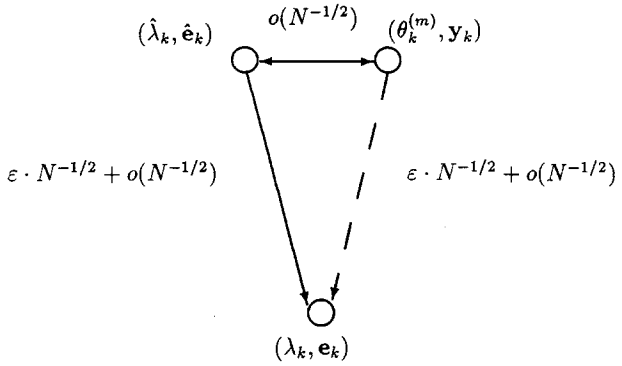

Fig. 8. Illustration of triangular relationship among $(\lambda_k, \boldsymbol{e}_k)$, $(\hat{\lambda}_k, \hat{\boldsymbol{e}}_k)$ and $(\theta_k^{(m)}, \boldsymbol{y}_k^{(m)})$.

designed. Finally, we show that the proposed FSD detection scheme is *strongly consistent*.

A new detection scheme

By (94), we know $\hat{\lambda}_k$, $d' + 1 \leq k \leq M$ cluster around $\sigma^2 (= \lambda_k)$ in a small region with size $O(N^{-1/2})$. In essence, the test of $\hat{d} \stackrel{?}{=} d'$ is performed by checking whether $\hat{\lambda}_{\hat{d}+1}, \ldots, \hat{\lambda}_M$ are close to one another. One frequently used measure of the difference among $\hat{\lambda}_{\hat{d}+1}, \ldots, \hat{\lambda}_M$ is the ratio of the geometric mean and arithmetic mean of these quantities, i.e.,

$$L(\hat{\lambda}_{\hat{d}+1}, \ldots, \hat{\lambda}_M) = \frac{\left(\prod_{k=\hat{d}+1}^{M} \hat{\lambda}_k \right)^{1/(M-\hat{d})}}{\frac{1}{M-\hat{d}} \sum_{k=\hat{d}+1}^{M} \hat{\lambda}_k} \tag{106}$$

which is also used in most conventional detection schemes, e.g., likelihood ratio test [22, 4], AIC, MDL [73, 70]. Unfortunately, at each Lanczos step, say the m-th step, we never have these eigenvalues, i.e., $\hat{\lambda}_{\hat{d}+1}, \ldots, \hat{\lambda}_M$. In fact, only m Ritz–Rayleigh (RR) values and other intermediate quantities are accessible. Nevertheless, as will be shown below, we can still carry out the likelihood ratio test at each Lanczos step without knowing the eigenvalues of $\hat{\boldsymbol{R}}_{XX}$.

By Theorem 6.2, we know that the d' larger RR values from the m-th Lanczos step ($m \geq d' + 1$) can replace the signal eigenvalues without affecting the asymptotic properties; more specifically, $\hat{\lambda}_k = \theta_k^{(m)} + O(N^{-(m-d')})$, for $1 \leq k \leq d'$. Since

$$\mathrm{Tr}(\hat{\boldsymbol{R}}_{XX}) = \sum_{k=1}^{M} \hat{\lambda}_k \quad \det(\hat{\boldsymbol{R}}_{XX}) = \prod_{k=1}^{M} \hat{\lambda}_k \tag{107}$$

thus,

$$\sum_{k=d'+1}^{M} \hat{\lambda}_k = \mathrm{Tr}(\hat{\boldsymbol{R}}_{XX}) - \sum_{k=1}^{d'} \hat{\lambda}_k = \mathrm{Tr}(\hat{\boldsymbol{R}}_{XX}) - \sum_{k=1}^{d'} \theta_k^{(m)} + \mathrm{O}(N^{-(m-d')}) , \tag{108}$$

$$\prod_{k=d'+1}^{M} \hat{\lambda}_k = \frac{\det(\hat{\boldsymbol{R}}_{XX})}{\prod_{k=1}^{d'} \hat{\lambda}_k} = \frac{\det(\hat{\boldsymbol{R}}_{XX})}{\prod_{k=1}^{d'} \theta_k^{(m)}} + \mathrm{O}(N^{-(m-d')}) . \tag{109}$$

Therefore, $\Sigma_{k=d'+1}^{M} \hat{\lambda}_k$, $\Pi_{k=d'+1}^{M} \hat{\lambda}_k$ and eventually their ratio (106) can be equivalently evaluated by the RR values, $\mathrm{Tr}(\hat{\boldsymbol{R}}_{XX})$, and $\det(\hat{\boldsymbol{R}}_{XX})$. Nevertheless, since the computation of $\det(\hat{\boldsymbol{R}}_{XX})$ generally requires $\mathrm{O}(M^3)$ flops, it is rather computationally intensive. Herein, we will form another statistic which can be evaluated by no more than $\mathrm{O}(M^2)$ flops. The trick is to replace the geometric mean of $M - d'$ smaller eigenvalues by their quadratic mean, viz., $\sqrt{(1/M - d') \Sigma_{k=d'+1}^{M} \hat{\lambda}_k^2}$. Knowing that $\|\hat{\boldsymbol{R}}_{XX}\|^2 = \Sigma_{k=1}^{M} \hat{\lambda}_k^2$, we can use the same trick to obtain

$$\sum_{k=d'+1}^{M} \hat{\lambda}_k^2 = \|\hat{\boldsymbol{R}}_{XX}\|^2 - \sum_{k=1}^{d'} \theta_k^{(m)^2} + \mathrm{O}(N^{-(m-d')}) . \tag{110}$$

With these in mind, we can form the statistics $\varphi_{\hat{d}}$ for estimating d' as follows:

$$\varphi_{\hat{d}} = N(M - \hat{d}) \log \left(\frac{\sqrt{\dfrac{1}{M - \hat{d}} \left(\|\hat{\boldsymbol{R}}_{XX}\|^2 - \sum_{k=1}^{\hat{d}} \theta_k^{(m)^2} \right)}}{\dfrac{1}{M - \hat{d}} \left(\mathrm{Tr}(\hat{\boldsymbol{R}}_{XX}) - \sum_{k=1}^{\hat{d}} \theta_k^{(m)} \right)} \right) . \tag{111}$$

The following theorem gives the limiting distribution of $\varphi_{\hat{d}}$ under the null hypothesis H_0: $\hat{d} = d'$.

THEOREM 6.3. *For $m > d' + 1$, the statistic $\varphi_{\hat{d}}$ defined in* (111) *is asymptotically chi-squared (χ^2) distributed with $\frac{1}{2}(M - d')(M - d' + 1) - 1$ degrees of freedom under H_0: $\hat{d} = d'$, if $\hat{\boldsymbol{R}}_{XX}$ is real. When $\hat{\boldsymbol{R}}_{XX}$ is complex, $2\varphi_{\hat{d}}$ is asymptotically χ^2 with $(M - d')^2 - 1$ degrees of freedom.*

The proof of Theorem 6.3 is given in [80, 85]. The proposed detection scheme for the $\varphi_{\hat{d}}$ at the m-th Lanczos step is outlined below. $c(N)$ is a predetermined sequence of numbers. The choice of $c(N)$ is discussed in the next section.

FSD DETECTION SCHEME.

(1) Set $\hat{d} = 1$.

(2) Set null hypothesis H_0: $d = \hat{d}$.

(3) Properly choose a threshold $\gamma_{\hat{d}}$ based on the tail area of the χ^2 distribution with $\frac{1}{2}(M - \hat{d})(M - \hat{d} + 1) - 1$ (for a real $\hat{\boldsymbol{R}}_{XX}$) or $(M - \hat{d})^2 - 1$ (for a complex $\hat{\boldsymbol{R}}_{XX}$) degrees of freedom.

(4) Evaluate the test statistic $\varphi_{\hat{d}}$.

(5) If $\varphi_{\hat{d}} \leq \gamma_{\hat{d}} c(N)$ (for a real $\hat{\boldsymbol{R}}_{XX}$) or $2\varphi_{\hat{d}} \leq \gamma_{\hat{d}} c(N)$ (for a complex $\hat{\boldsymbol{R}}_{XX}$), accept H_0 and stop.

(6) If $\varphi_{\hat{d}} > \gamma_{\hat{d}} c(N)$, reject H_0; if $\hat{d} < m - 2$, $\hat{d} := \hat{d} + 1$, return to Step (2). Otherwise, $m := m + 1$, continue the $(m+1)$-th step.

Strong consistency

This section concerns how to select $c(N)$ for the detection scheme to yield a *strongly consistent* estimate of the signal subspace dimension d'. The estimate of d' is called *strongly consistent*, if the correct estimate occurs with probability one as the number of data goes to infinity. The consistency proof requires the following preliminary results.

LEMMA 6.6. *For the sample covariance matrix* $\hat{\boldsymbol{R}}_{XX}$ *and its eigenvalues* $\hat{\lambda}_k$, *the following relations hold with probability* 1,

$$\hat{\boldsymbol{R}}_{XX} - \boldsymbol{R}_{XX} = \mathrm{O}(\sqrt{\log\log N/N})\,, \tag{112}$$

$$\hat{\lambda}_k - \lambda_k = \mathrm{O}(\sqrt{\log\log N/N})\,, \quad k = 1, \ldots, M\,. \tag{113}$$

The proof of (112) can be found in [28], while (113) is proved in [86].

Let $k = d' + 1$ or M, by (113), we have $\hat{\lambda}_k - \sigma^2 = \mathrm{O}(\sqrt{\log\log N/N})$. Hence,

$$\varepsilon = \hat{\lambda}_{d'+1} - \hat{\lambda}_M = (\hat{\lambda}_{d'+1} - \sigma^2) - (\hat{\lambda}_M - \sigma^2) = \mathrm{O}(\sqrt{\log\log N/N}) \tag{114}$$

with probability 1.

Now let us give the following theorem concerning strong consistency of the detection scheme.

THEOREM 6.4. *For the statistic* $\varphi_{\hat{d}}$ *defined in* (111), *the detection scheme of the previous section is strongly consistent if* $c(N)$ *satisfies the following conditions*:

$$\lim_{N\to\infty} \frac{c(N)}{N} = 0\,, \qquad \lim_{N\to\infty} \frac{c(N)}{\log\log N} = \infty\,. \tag{115}$$

The proof of this theorem can be found in [80, 85]. In fact, the sequence $c(N)$ which satisfies the conditions in (115) is not difficult to find. For example, $\sqrt{\log N}$ is a good candidate. The unique advantage of the FSD detection

scheme is that it does not require the knowledge of the eigenvalues but the intermediate quantities of the $(\hat{d}+2)^{th}$ Lanczos step, i.e., $\theta_k^{(m)}$, $k=1,\ldots,\hat{d}+1$. Therefore, the detection scheme can be carried out in parallel with the process of finding the signal subspace. Also *only* RR values (not RR vectors) need to be computed at each intermediate Lanczos step. According to [36, Section 8–15], the eigenvalues (no eigenvectors) of an $m \times m$ tridiagonal matrix can be found in $9m^2$ flops (instead of $10m^3$ flops for a complete eigendecomposition). Therefore, the detection only requires $O(m^2)$ flops at the m-th Lanczos recursion or $O(d^3)$ flops in total plus $\frac{1}{2}M^2$ flops for computing $\|\hat{\boldsymbol{R}}_{XX}\|^2$.

BASIC FSD ALGORITHM. With the new *strongly consistent* detection scheme, our FSD algorithm is complete. The basic FSD algorithm is briefly described below.

(1) Properly select a nondegenerate f, set $m=1$ and $\hat{d}=1$.

(2) Carry out the m-th Lanczos recursion.

(3) Compute the RR values $\theta_i^{(m)}$, $i=1,2,\ldots,m$.

(4) Evaluate $\varphi_{\hat{d}}$, for $\hat{d}=1,\ldots,m-1$. If $\varphi_{\hat{d}} \leqslant \gamma_{\hat{d}} c(N)$, set $d'=\hat{d}$ (accept H_0), go to Step (5). Otherwise, $m := m+1$ and return to Step (2).

(5) Compute the d' principal RR vectors $\boldsymbol{y}_k^{(m)}$ associated with $\mathscr{K}^m(\hat{\boldsymbol{R}}_{XX}, \boldsymbol{f})$. The final signal subspace estimate is $\mathscr{R}\{\boldsymbol{y}_1^{(m)},\ldots,\boldsymbol{y}_{d'}^{(m)}\}$. Stop.

Numerical concerns

One practical issue regarding the FSD algorithm is possible numerical instability associated with the Lanczos method. Nevertheless, we can always remedy the numerical problem by adopting a selective reorthogonalization [36], Section 13-8, or a complete reorthogonalization. Since $d' \ll M$, the extra computational complexity is no more than $O(Md^2)$, which is not so significant. Another remedy to this problem is to replace the Lanczos algorithm by Householder reflections and Givens rotations, which are much more numerically robust. In this case, the computational cost is still of the same order $(O(M^2d'))$, but the coefficient is larger.

6.3. Summary

In this section, we first showed the possibility of achieving the signal subspace decomposition in a *finite* number of steps by properly exploiting the *structure* of ideal covariance matrices. The proposed algorithm can achieve nearly an order of computational reduction over eigendecomposition methods.

We also proposed a new $O(M^2d')$ fast subspace decomposition algorithm for *sample* covariance matrices, including a *new simple* detection scheme for estimating the signal subspace dimension, which can be carried out in parallel with the estimation of the signal subspace. Rigorous statistical and numerical analyses have been conducted, which assure the *strong consistency* of our new detection scheme and the *asymptotic equivalence* between the FSD estimates and those based on the more costly $O(M^3)$ eigendecompositions.

7. Concluding remarks

In this chapter we presented an approach to solving a class of nonlinear parameter estimation problems that fit a model characterized by 'low rank signal and full rank noise'. Our method first estimated this low rank signal subspace from the full rank observations (by exploiting certain structure of the noise contributions) and thereafter the problem of estimating the unknown parameters reduced to determining parameter values that induced a subspace that best fitted the estimated signal subspace.

This core non-linear parameter estimation problem underlies a surprising number of problems ranging from antenna array processing, system identification, extraction of exponentials in noise, high resolution range estimation, and several others. Our discussion of this parameter estimation problem was mainly in the context of antenna array processing since it is here that subspace techniques have been traditionally developed. However, a number of problems in general signal processing as well as other applications in engineering/sciences admit a similar formulation. The application of subspace techniques is slowly penetrating new areas and our hope is that our exposition in this chapter will encourage further exploration of this approach and its application to new problems.

Some comments on the application of subspace methods to antenna array signal processing are needed to put in perspective the material covered in this chapter.

(1) Array signal processing has been a particularly fertile area of research and has been approached from a number of different view points with varying degrees of success. Subspace methods (MUSIC, ESPRIT, etc.) have been only one, but particularly successful, paradigm for array signal processing. The success of the subspace methods arises largely from imposing a simple underlying model on the signal and noise and then usefully exploiting the properties of this model.

(2) The assumptions on the data model may not always be practical or convenient. For example an important restriction of the subspace method is that the number of signals be less than the number of sensors (assuming no coherent sources). This restriction is not usually needed in the other array processing methods. Another restriction is that the noise field be uncorrelated and equi-power at each sensor (equivalently the noise field covariance is completely known to within a scalar). Again such restrictions are not necessary in many other array signal processing methods. This assumption on the noise field characterization is inconvenient or impractical in many real scenarios. Several extensions of subspace methods allowing them to relax some of these restrictions have been proposed, and further extensions remains of great interest.

(3) Signal processing addresses a large number of problems including the direction-of-arrival estimation, signal waveform estimation (sometimes referred to as signal copy), signal covariance estimation etc. In our discussion we have restricted ourselves to the DOA estimation problem as it is the key to solving

the other problems. However subspace methods also provide a nice solution to the other related problems and the reader is referred to the extensive literature in the field.

In summary, this chapter introduced subspace methods in the context of array signal processing. We began with a description of the underlying data model and presented an unified approach to this non-linear parameter estimation problem via subspace fitting techniques. We then discussed some of the statistical properties of these methods and followed this up with an analysis of their sensitivity to modeling errors. Determining the signal subspace (and its rank) is a key underlying problem and in the last section we described certain fast techniques for subspace estimation and analyzed their statistical properties.

References

[1] Aghajan, H. A. and T. Kailath (1992). A subspace fitting approach to super resolution multi-line fitting and straight edge detection. In: *Proc. IEEE ICASSP*. San Francisco, CA, March, III, 121–124.

[2] Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34**, 122–148.

[3] Bangs, W. J. (1971). Array processing with generalized beamformers. Ph.D. thesis, Yale University, New Haven, CT.

[4] Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *J. Roy. Statist. Soc. Ser. B* **16**, 296–298.

[5] Bienvenu, G. and L. Kopp (1979). Principle de la goniometrie passive adaptive. In: *Proc. 7'eme Colloque GRESIT*. Nice, France, 106/1–106/10.

[6] Böhme, J. F. and D. Kraus (1988). On least squares methods for direction of arrival estimation in the presence of unknown noise fields. In: *Proc. ICASSP 88*. New York, 2833–2836.

[7] Bresler, Y. and A. Macovski (1986). Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Trans. Acoust. Speech Signal Process.* **34**(5), 1081–1089.

[8] Bresler, Y. and A. Macovski (1986). On the number of signals resolvable by a uniform linear array. *IEEE Trans. Acoust. Speech Signal Process.* **34**(6), 1361–1375.

[9] Böhme, J. F. (1986). Estimation of spectral parameters of correlated signals in wavefields. *Signal Process.* **10**, 329–337.

[10] Burg, J. P. (1975). Maximum entropy spectral analysis. Ph.D. thesis, Stanford University, Stanford, CA.

[11] Cadzow, J. A. (1988). A high resolution direction-of-arrival algorithm for narrowband coherent and incoherent sources. *IEEE Trans. Acoust. Speech Signal Process.* **36**, 965–979.

[12] Capon, J. (1969). High resolution frequency wave number spectrum analysis. *Proc. IEEE* **57**, 1408–1418.

[13] Comon, P. and G. Golub (1990). Tracking a few extreme singular values and vectors in signal processing. *Proc. IEEE* **78**(8), 1327–1343.

[14] Clergeot, H., S. Tressens and A. Ouamri (1989). Performance of high resolution frequencies estimation methods compared to the Cramér–Rao bounds. *IEEE Trans. Acoust. Speech Signal Process.* **37**(11), 1703–1720.

[15] Friedlander, B. (1990). A sensitivity analysis of the MUSIC algorithm. *IEEE Trans. Acoust. Speech Signal Process.* **38**(10), 1740–1751.

[16] Friedlander, B. (1990). Sensitivity of the maximum likelihood direction finding algorithm. *IEEE Trans. Aerospace Electron. Systems* **26**(6), 953–968.

[17] Feder, M. and E. Weinstein (1988). Parameter estimation of superimposed signals using the EM algorithm. *IEEE Trans. Acoust. Speech Signal Process.* **36**, 477–489.
[18] Friedlander, B. and A. J. Weiss (1988). Eigenstructure methods for direction finding with sensor gain and phase uncertainties. In: *Proc. ICASSP*, 2681–2684.
[19] Golub, G. H. and A. Pereyra (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal.* **10**, 413–432.
[20] Kuruc, A. R. (1989). Lower bounds on multiple-source direction finding in the presence of direction-dependent antenna-array-calibration errors. Technical Report TR-799, MIT Lincoln Laboratory.
[21] Lanczos, C. (1950). An iterative method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Standards, Sect. B* **45**, 225–280.
[22] Lawley, D. N. (1956). Tests of significance of the latent roots of the covariance and correlation matrices. *Biometrica* **43**, 128–136.
[23] Lehmann, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.
[24] Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.
[25] Li, F. and R. Vaccaro (1990). Statistical comparison of subspace based DOA estimation algorithms in the presence of sensor errors. In: *Proc. 5th ASSP Spectral Estimation Workshop*, Rochester, NY, October, 327–331.
[26] Li, F. and R. Vaccaro (1991). Performance degradation of DOA estimators due to unknown noise fields. In: *Proc. IEEE ICASSP*, Toronto, Canada.
[27] Li, F., R. Vaccaro and D. Tufts (1990). Unified performance analysis of subspace-based estimation algorithms. In: *Proc. IEEE ICASSP*. Vol. 5, Albuquerque, NM, 2575–2578.
[28] Mardia, K. V., J. T. Kent and J. M. Bibby (1979). *Multivariate Analysis*. Academic Press, New York.
[29] Ottersten, B., T. Kailath, M. Viberg and B. Wahlberg (1989). Stochastic maximum likelihood estimation in sensor arrays by weighted subspace fitting. In: *Proc. 23rd Asilomar Conf. on Signals, Systems, and Computers*. San Jose, CA, November, 599–603.
[30] Ottersten, B. and L. Ljung (1989). Asymptotic results for sensor array processing. In: *Proc. ICASSP 89*, Glasgow, Scotland, May, 2266–2269.
[31] Ottersten, B. (1989). Parametric subspace fitting methods for array signal processing. Ph.D. thesis, Stanford University, Stanford, CA.
[32] Ottersten, B. and M. Viberg (1988). Asymptotic results for multidimensional sensor array processing. In: *Proc. 22nd Asilomar Conf. on Signals, Systems, and Computers*, Monterey, CA, November, 833–837.
[33] Ottersten, B., M. Viberg and T. Kailath (1991). Performance analysis of the total least squares *ESPRIT* algorithm. *IEEE Trans. Signal Process.* **39**(5), 1122–1135.
[34] Ottersten, B., M. Viberg and T. Kailath (1992). Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data. *IEEE Trans. Signal Process.* **40**(3), 590–600.
[35] Ottersten, B., M. Viberg, P. Stoica and A. Nehorai (1993). Exact and large sample ML techniques for parameter estimation and detection in array processing. In: S. Haykin, ed., *Radar Array Signal Processing*. Springer, Wien, New York. To appear.
[36] Parlett, B. N. (1980). *The Symmetric Eigenvalue Problem*. Prentice-Hall, Englewood Cliffs, NJ.
[37] Pisarenko, V. F. (1973). The retrieval of harmonics from a covariance function. *Geophys. J. Roy. Astronom. Soc.* **33**, 347–366.
[38] A. Paulraj, and T. Kailath (1985). Direction-of-arrival estimation by eigenstructure methods with unknown sensor gain and phase. In: *Proc. IEEE ICASSP*. Tampa, FL, March, 17.7.1–17.7.4.
[39] Paulraj, A. and T. Kailath (1986). Eigenstructure methods for direction of arrival estimation in the presence of unknown noise fields. *IEEE Trans. Acoust. Speech Signal Process.* **34**(1), 13–20.
[40] Paulraj, A., R. Roy and T. Kailath (1985). Estimation of signal parameters via rotational invariance techniques – *ESPRIT*. In: *Proc. 19th Asilomar Conf. on Circuits, Systems and Computers*. San Jose, CA, November, 83–89.

[41] Paulraj, A., R. Roy and T. Kailath (1986). A subspace rotation approach to signal parameter estimation. *Proc. IEEE* **74**(7), 1044–1045.

[42] Paulraj, A., V. U. Reddy, T. J. Shan and T. Kailath (1986). A subspace approach to determine sensor gain and phase with applications to array processing. In: *Proc. 30th SPIE International Technical Symposium, Advanced Algorithms and Architectures for Signal Processing*. San Diego, CA, August.

[43] Reilly, J. P. (1987). A real-time high-resolution technique for angle-of-arrival estimation. *Proc. IEEE* **75**(12), 1692–1694.

[44] Roy, R. and T. Kailath (1989). *ESPRIT* – Estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **37**(7), 984–995.

[45] Roy, R. H. (1987). *ESPRIT* – Estimation of signal parameters via rotational invariance techniques. Ph.D. thesis, Stanford University, Stanford, CA.

[46] Schmidt, R. O. (1979). Multiple emitter location and signal parameter estimation. In: *Proc. RADC Spectrum Estimation Workshop*. Griffiss AFB, NY, 243–258.

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

[48] Shan, T. J., M. Wax and T. Kailath (1985). On spatial smoothing for direction of arrival estimation of coherent signals. *IEEE Trans. Acoust. Speech Signal Process.* **33**(4), 806–811.

[49] Soon, V. C. and Y. F. Huang (1992). An analysis of *ESPRIT* under random sensor uncertainties. *IEEE Trans. Acoust. Speech Signal Process.* **40**(9), 2353–2358.

[50] Shan, T. J. (1986). Array processing for coherent sources. Ph.D. thesis. Stanford University, Stanford, CA.

[51] Swindlehurst, A. and T. Kailath (1990). An analysis of subspace fitting algorithms in the presence of sensor errors. In: *Proc. IEEE ICASSP*. Vol. 5, Albuquerque, NM, 2647–2650.

[52] Swindlehurst, A. and T. Kailath (1990). On the sensitivity of the *ESPRIT* algorithm to non-identical subarrays. *Sādhanā Acad. Proc. Engrg. Sci.* **15**(3), 197–212.

[53] Swindlehurst, A. and T. Kailath (1990). A performance analysis of subspace-based methods in the presence of model errors – Part 2: Multidimensional algorithms. *IEEE Trans. Signal Process.* To be published.

[54] Swindlehurst, A. and T. Kailath (1992). A performance analysis of subspace-based methods in the presence of model errors – Part 1: The MUSIC algorithm. *IEEE Trans. Signal Process.* **40**(7), 1758–1774.

[55] Starer, D. and A. Nehorai (1988). Maximum likelihood estimation of exponential signals in noise using a newton algorithm. In: *Proc. 4th ASSP Workshop on Spectrum Estimation and Modeling*. Minneapolis, MN, August 240–245.

[56] Stoica, P. and A. Nehorai (1988). MUSIC, Maximum likelihood and Cramér–Rao bound: Further results and comparisons. Technical Report 8819, Yale University, New Haven.

[57] Stoica, P. and A. Nehorai (1989). MUSIC, maximum likelihood and Cramér–Rao bound. *IEEE Trans. Acoust. Speech Signal Process.* **37**, 720–741.

[58] Stoica, P. and A. Nehorai (1990). MUSIC, maximum likelihood, and Cramér–Rao bound: Further results and comparisons. *IEEE Trans. Acoust. Speech Signal Process.* **38**(12), 2140–2150.

[59] Stoica, P. and A. Nehorai (1990). Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Trans. Acoust. Speech Signal Process.* **38**, 1783–1795.

[60] Swindlehurst, A., B. Ottersten and T. Kailath (1989). An analysis of MUSIC and root-MUSIC in the presence of sensor perturbations. In: *Proc. 23rd Asilomar Conf. on Signals, Systems, and Computers*. Asilomar, CA, November, 930–934.

[61] Swindlehurst, A. L., B. Ottersten, R. Roy and T. Kailath (1992). Multiple invariance *ESPRIT*. *IEEE Trans. Signal Process.* **40**(4), 867–881.

[62] Stoica, P. and K. Sharman (1990). Maximum likelihood methods for direction-of-arrival estimation. *IEEE Trans. Acoust. Speech Signal Process.* **38**, 1132–1143.

[63] Stoica, P. and K. Sharman (1990). A novel eigenanalysis method for direction estimation. *Proc. IEE*-F, **137**(1), 19–26.

[64] Swindlehurst, A. (1990). Robust algorithms for direction-finding in the presence of model errors. In: *Proc. 5th ASSP Workshop on Spectral Estimation and Modeling*. Rochester, NY, October, 362–366.

[65] Swindlehurst, A. (1991). Applications of subspace fitting to estimation and identification. Ph.D. thesis, Stanford University, Stanford, CA.

[66] Tufts, D.W. and C.D. Melissinos (1986). Simple, effective computation of principal eigenvectors and their eigenvalues and application to high-resolution estimation of frequencies. *IEEE Trans. Acoust. Speech Signal Process.* **5**(34), 1046–1052.

[67] Viberg, M. and B. Ottersten (1991). Sensor array processing based on subspace fitting. *IEEE Trans. Signal Process.* **39**(5), 1110–1121.

[68] Viberg, M., B. Ottersten and T. Kailath (1989). Direction of arrival estimation and detection using weighted subspace fitting. In: *Proc. 23rd Asilomar Conf. on Signals, Systems, and Computers*, San Jose, CA, November, 604–608.

[69] Viberg, M., B. Ottersten and T. Kailath (1991). Detection and estimation in sensor arrays using weighted subspace fitting. *IEEE Trans. Signal Process.* **39**(11), 2436–2449.

[70] Wax, M. (1985). Detection and estimation of superimposed signals. Ph.D. thesis, Stanford University, Stanford, CA.

[71] Widrow, B., K. M. Duvall, R. P. Gooch and W. C. Newman (1982). Signal cancellation phenomena in adaptive antennas: Causes and cures. *IEEE Trans. Antennas and Propagation* **30**, 469–478.

[72] Weiss, A. J. and B. Friedlander (1988). Direction finding in the presence of mutual coupling. In: *Proc. 22nd Asilomar Conf. on Signals, Systems, and Computers*. 598–602.

[73] Wax, M. and T. Kailath (1985). Detection of signals by information theoretic criteria. *IEEE Trans. Acoust. Speech Signal Process.* **33**(2), 387–392.

[74] Wahlberg, B., B. Ottersten and M. Viberg (1991). Robust signal parameter estimation in the presence of array perturbations. In: *Proc. IEEE ICASSP.* Toronto, Canada, 3277–3280.

[75] Weiss, A. J., A. S. Willsky and B. C. Levy (1988). Eigenstructure approach for array processing with unknown intensity coefficients. *IEEE Trans. Acoust. Speech Signal Process.* **36**(10), 1613–1617.

[76] Wong, K. M., R. S. Walker and G. Niezgoda (1988). Effects of random sensor motion on bearing estimation by the MUSIC algorithm. *Proc. IEE-F* **135**(3), 233–250.

[77] Wax, and I. Ziskind (1989). Detection of the number of coherent signals by the mdl principle. *IEEE Trans. Acoust. Speech Signal Process.* **37**(8), 1190–1196.

[78] Wax, M. and I. Ziskind (1989). On unique localization of multiple sources by passive sensor arrays. *IEEE Trans. Acoust. Speech Signal Process.* **37**(7), 996–1000.

[79] Xu, G. and T. Kailath (1990). Fast signal subspace decomposition – Part I: Ideal covariance matrices. Submitted to *IEEE Trans. Acoust. Speech Signal Process.*

[80] Xu, G. and T. Kailath (1990). Fast signal subspace decomposition – Part II: Sample covariance matrices. Submitted to *IEEE Trans. Acoust. Speech Signal Process.*

[81] Xu, G. and T. Kailath (1990). Fast signal subspace decomposition without eigendecomposition. In: *Proc. 24th Asilomar Conf. on Signals Systems, and Computers*. Pacific Grove, CA, November. 966–970.

[82] Xu, G. and T. Kailath (1990). Strong consistency of FSD detection schemes. In: *Proc. 24th Asilomar Conf. on Signals Systems, and Computers*. Pacific Grove, CA, November. 267–271.

[83] Xu, G. and T. Kailath (1991). A fast algorithm for signal subspace decomposition and its performance analysis. In: *Proc. ICASSP.* Vol. 5, Toronto, Canada, May, 3069–3072.

[84] Xu, G. and T. Kailath (1991). Fast subspace decomposition of data matrices. In: *Proc. 25th Asilomar Conf. on Signals Systems, and Computers*. Pacific Grove, CA, November 1991.

[85] Xu, G. (1991). Fast subspace decomposition and its applications. Ph.D. thesis, Stanford University, Stanford, CA.

[86] Zhao, L. C., P. R. Krishnaiah and Z. D. Bai (1986). On the detection of the number of signals in the presence of white noise. *J. Multivariate Anal.* **20**(1), 1–25.

[87] Zhu, J. X. and H. Wang (1988). Effects of sensor position and pattern perturbations on CRLB for direction finding of multiple narrowband sources. In: *Proc. 4th ASSP Workshop on Spectral Estimation and Modeling*. Minneapolis, MN, August, 98–102.

[88] Ziskind, I. and M. Wax (1988). Maximum likelihood localization of multiple sources by alternating projection. *IEEE Trans. Acoust. Speech Signal Process.* **36**, 1553–1560.