Analysis of Joint Angle-Frequency Estimation Using ESPRIT

Aweke N. Lemma, Alle-Jan van der Veen, *Senior Member, IEEE*, and Ed F. Deprettere, *Fellow, IEEE*

***Abstract*—High-resolution parameter estimation techniques have recently been applied to jointly estimate multiple signal parameters. In this work, we consider the problem of determining the directions and center frequencies of a number of narrowband sources in a band of interest. We present a joint angle-frequency estimation method, based on the multidimensional ESPRIT algorithm. A perturbation error analysis gives bounds on the parameter estimates and provides optimal values for the temporal and spatial smoothing parameters. The analysis is shown to be consistent with simulation results.**

***Index Terms*—Joint diagonalization, joint parameter estimation, multidimensional ESPRIT, multiresolution ESPRIT, shift-invariance.**

I. Introduction

In MANY practical signal processing problems, it is desired to estimate from measurements a set of parameters upon which the received signals depend. Optimal techniques based on maximum likelihood are often applicable but might be computationally prohibitive. Algebraic techniques based on a batch of data have an edge in terms of computational complexity. Such techniques make specific use of certain algebraic structures present in the data matrix.

A prime example of an algebraic technique is the ESPRIT algorithm [1]. Since its formal derivation in 1985, ESPRIT has been used for direction-of-arrival (DOA) estimation, harmonic analysis, frequency estimation, delay estimation, and combinations thereof. In essence, the algorithm makes use of the shift invariance structure present in the array response vector $\boldsymbol{a}(\theta)$, where $\theta = e^{j\mu}$, and μ is a phase shift to be estimated. In narrowband DOA estimation, the phase shift is due to the difference in arrival times of the wavefront at the elements of an antenna array. For a uniform linear array (ULA), it is well known that $\boldsymbol{a}(\theta) = [1 \quad \theta \quad \theta^2 \quad \cdots]^T$ and $\mu = 2\pi\Delta\sin(\alpha)$, where Δ is the distance between the elements (in wavelengths), and α is the angle of arrival measured with respect to the normal of the array axis. A similar situation occurs in frequency estimation where we have $\mu = -2\pi fT$. Here, T is the sampling period, and f is the frequency to be estimated.

When ESPRIT is used to estimate multiple signal parameters, such as angle and frequency, one may solve the problem in one of the following two ways. In the first approach, the individual signal parameters are estimated independently, and only then (using some matching algorithm) are the parameters that belong to the same signal grouped together. Apart from the computational overhead, this also leads to a numerically less robust set of problems as it does not exploit the relation between the individual estimation problems.

A second method is to combine the individual estimation problems into a single joint parameter estimation problem. In our context, joint parameter estimation is discussed in a number of papers, including joint azimuth and elevation angle estimation [2], joint frequency and 2-D angle estimation [3], [4], and joint angle and delay estimation [5]. Basically, these methods rely on the fact that each parameter is estimated from a certain eigenvalue problem, where all eigenvalue problems share the same eigenvectors (which are related to the beamforming vectors). This allows the posing of the problem as a joint diagonalization problem of a collection of data matrices. The prime advantage of joint estimation is that the individual parameters are paired for free and show a better robustness to signal and parameter disturbances.

In the literature, a number of ESPRIT-based joint angle and frequency estimation methods have been proposed. In particular, Zoltowski *et al.* [3] discuss this problem in the context of radar applications. Because of ambitious goals, however, their solutions are very much directed by engineering considerations, which incur a certain sacrifice in elegance and clarity. Haardt *et al.* [4] discuss the problem in the context of mobile communications for space division multiple access (SDMA) applications. Their method is based on Unitary-ESPRIT, which involves a certain Cayley transformation of the data to real-valued matrices. This provides a computationally efficient solution scheme but might lead to numerical inaccuracies, particularly when the eigenvalues are close to $\pm\pi$. A similar but simpler algorithm called joint angle-frequency estimation (JAFE) has been proposed by us in [6].

The objective of the paper is to give a comprehensive error analysis of the JAFE algorithm. The algorithm is a function of certain stacking (or smoothing) parameters, and the error analysis provides us with the optimal choices for these parameters.

Manuscript received March 25, 1999; revised November 1, 2002. The work of A. N. Lemma was supported by TNO-FEL, The Hague, The Netherlands. The publication of this work was delayed due to administrative problems. The associate editor coordinating the review of this paper and approving it for publication was Dr. Maria Joao Rendas.

A. N. Lemma was with Delft University of Technology, Department of Information Technology, Delft, The Netherlands. He is now with Philips Digital Systems Labs Eindhoven, Eindhoven, The Netherlands (e-mail: Aweke.Lemma@Philips.com).

A.-J. van der Veen is with the Department of Information Technology, Delft University of Technology, Delft, The Netherlands (e-mail: Allejan@cobalt.et.tudelft.nl).

E. F. Deprettere was with the Department of Information Technology, Delft University of Technology, Delft, The Netherlands. He is now with the Department of Mathematics and Natural Sciences, LIACS, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands (e-mail: Edd@wi.leidenuniv.nl).

Digital Object Identifier 10.1109/TSP.2003.810306

A. Outline

We begin our discussion by describing the data model and the parameter estimation problem. In Section II, we extend the ESPRIT algorithm to JAFE. Section III looks into different data extension and processing techniques that influence the robustness and performance of the algorithm. Following this, in Sections IV–VII, we present a performance analysis of the JAFE algorithm. Finally, after deriving Cramér–Rao lower bounds on the parameter estimation errors (Section VIII), we present simulation results that illustrate the various aspects of this work (Section IX).

B. Notation

Throughout, row and column vectors are denoted by lowercase bold-faced letters and matrices with uppercase bold-faced letters. For any positive integer p, $\boldsymbol{I}_p$ denotes a $p\times p$ identity matrix. We suppress the index when this does not lead to confusion. Superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transposition and Hermitian transposition, respectively. Complex conjugation by itself is denoted by conj$(\cdot)$. $E\{\cdot\}$ denotes mathematical expectation.

A vector constructed from a sequence of entries, such as $\boldsymbol{\phi} = [\phi_0 \;\; \phi_1 \;\; \cdots \;\; \phi_N]$, or a sequence of function values, such as $\boldsymbol{\phi} = [\phi(0)\; \phi(1) \;\cdots\; \phi(N)]^T$, may be written as

$$\boldsymbol{\phi} = \left[\{\phi_i\}_{i=0}^N\right], \quad \text{or} \quad \boldsymbol{\phi} = \left[\{\phi(n)\}_{n=0}^N\right]$$

respectively. Similarly, a diagonal matrix Φ with the above diagonal entries may be expressed as

$$\boldsymbol{\Phi} = \text{diag}\{\phi_i\}_{i=0}^N \quad \text{or} \quad \boldsymbol{\Phi} = \text{diag}\{\phi(n)\}_{n=0}^N,$$

respectively. For two matrices $\boldsymbol{A} \in \mathbb{C}^{m,\,n}$ and $\boldsymbol{B} \in \mathbb{C}^{p,\,q}$, the Kroneker product $\boldsymbol{A}\otimes\boldsymbol{B} \in \mathbb{C}^{mp,\,nq}$ is defined as

$$\boldsymbol{A}\otimes\boldsymbol{B} = \begin{bmatrix} a_{1,\,1}\boldsymbol{B} & a_{1,\,2}\boldsymbol{B} & \cdots & a_{1,\,n}\boldsymbol{B} \\ a_{2,\,1}\boldsymbol{B} & a_{2,\,2}\boldsymbol{B} & \cdots & a_{2,\,n}\boldsymbol{B} \\ \vdots & \vdots & & \vdots \\ a_{m,\,1}\boldsymbol{B} & a_{m,\,2}\boldsymbol{B} & \cdots & a_{m,\,n}\boldsymbol{B} \end{bmatrix}.$$

Unless stated otherwise, in the paper, $\|\cdot\|$ is a Frobenius norm operator.

II. Joint Angle and Frequency Estimation

Suppose that we have an antenna array, observe a frequency band of interest, and want to separate and identify the directions and carrier frequencies of all sources that are present. For frequency estimation to be meaningful, we assume that the sources are sufficiently narrowband, typically with different carrier frequencies, but the spectra might be partly overlapping. The objective is to estimate the parameters and to construct a beamformer to separate the sources based on differences in angles or carrier frequencies. We will assume that the sample rates are much higher than the data rates of each source and that multipath is negligible.

A. Model

Suppose that there are d sources of interest, with complex baseband representations $s_i(t)$, for $i = 1, \ldots, d$. Let the band of interest have a center frequency f_c, and suppose that the ith source has a carrier frequency of $f_c + f_i$. After demodulation to IF, the signal due to the ith source is $e^{j2\pi f_i t}s_i(t)$, and the signal received at the kth antenna $k = 1, \ldots, M$ is

$$x_k(t) = \sum_{i=1}^{d} a_k(\theta_i)e^{j2\pi f_i t}b_i s_i(t) + w_k(t)$$

where θ_i is the parameterization of the DOA of the ith signal, with respect to a common phase reference, $a_k(\theta)$ is the antenna response of the kth antenna to a signal from direction θ, $b_i \in \mathbb{R}^+$ is the amplitude of the ith signal, and $w_k(t)$ is noise. It is natural to stack the antenna outputs into a single vector $\boldsymbol{x}(t)$.

Further suppose that the narrowband signals have a bandwidth of less than $1/T$ so that they can be sampled with a period T to satisfy the Nyquist rate. We normalize to $T = 1$. Let us say that the bandwidth of the band to be scanned is an integer number P times larger: After demodulation to IF, we have to sample at a rate P [obviously we require $-(P/2) \le f_i < P/2$ to prevent aliasing]. The data sample at the receiver is

$$\boldsymbol{x}\left(\frac{n}{P}\right) = \sum_{i=1}^{d} \boldsymbol{a}(\theta_i)b_i \exp\left(j\frac{2\pi}{P}f_i n\right) s_i\left(\frac{n}{P}\right) + \boldsymbol{w}\left(\frac{n}{P}\right)$$

where $\boldsymbol{a}(\theta_i)$ is the array response vector of the ith source, and $\boldsymbol{w}(n/P) \in \mathbb{C}^{M,\,1}$ is the noise vector collecting the samples of the noise terms at the output of each antenna element. In matrix form, this can be written as

$$\boldsymbol{x}\left(\frac{n}{P}\right) = \boldsymbol{A}\boldsymbol{B}\boldsymbol{\Phi}^n\boldsymbol{s}\left(\frac{n}{P}\right) + \boldsymbol{w}\left(\frac{n}{P}\right) \tag{1}$$

where $\boldsymbol{\Phi} = \text{diag}\{\phi_i\}_{i=1}^d$, $\phi_i = e^{j(2\pi/P)f_i}$, $\boldsymbol{B} = \text{diag}\{b_i\}_{i=1}^d$ is a signal gain matrix, $\boldsymbol{A}$ is an $M \times d$ matrix collecting the d steering vectors, and the vector $\boldsymbol{s}(t)$ is a stack of the d signals, where each signal has a unit amplitude. In the remainder of the paper, unless it is necessary to write it explicitly, the diagonal matrix $\boldsymbol{B}$ in the data model is absorbed by $\boldsymbol{s}(t)$, in which case, the amplitude of the ith signal is equal to b_i instead of 1. Assume that we have collected N samples of the array output $\boldsymbol{x}(t)$ at a rate P into the $M \times N$ data matrix $\boldsymbol{X}$, i.e.,

$$\boldsymbol{X} = \boldsymbol{A}\left[\boldsymbol{s}(0) \quad \boldsymbol{\Phi}\boldsymbol{s}\left(\frac{1}{P}\right) \quad \cdots \quad \boldsymbol{\Phi}^{N-1}\boldsymbol{s}\left(\frac{N-1}{P}\right)\right] + \boldsymbol{W} \in \mathbb{C}^{M,\,N} \tag{2}$$

where $\boldsymbol{W} \in \mathbb{C}^{M,\,N}$ is a matrix collecting N samples of the $M \times 1$ array noise vector.

B. Temporal Smoothing

In this section, we consider a data stacking technique (referred to as temporal smoothing) that adds structure to the data model for the implementation of the JAFE algorithm. Apart from this, temporal smoothing introduces an interesting feature. That is, the data matrix in (2) is rank deficient when two or more signals have the same DOA. This is because the array steering vectors corresponding to signals with the same DOAs are identical, and therefore, the rank of the data matrix will be less than d, where d is the number of signals. It will be shown in this section that, under a certain condition, temporal smoothing restores

the rank of the data matrix. An m-*factor* temporally smoothed data matrix is constructed by stacking m temporally shifted versions of the original data matrix. This results in the following $mM \times N - m + 1$ matrix [*viz.* (2)]:

$$\boldsymbol{X}_m = \begin{bmatrix} \boldsymbol{A}\left[\boldsymbol{s}(0)\ \boldsymbol{\Phi}\boldsymbol{s}\left(\frac{1}{P}\right)\ \cdots\ \boldsymbol{\Phi}^{N-m}\boldsymbol{s}\left(\frac{N-m}{P}\right)\right] \\ \boldsymbol{A}\boldsymbol{\Phi}\left[\boldsymbol{s}\left(\frac{1}{P}\right)\ \boldsymbol{\Phi}\boldsymbol{s}\left(\frac{2}{P}\right)\ \cdots\right] \\ \vdots \\ \boldsymbol{A}\boldsymbol{\Phi}^{m-1}\left[\boldsymbol{s}\left(\frac{m-1}{P}\right)\ \boldsymbol{\Phi}\boldsymbol{s}\left(\frac{m}{P}\right)\ \cdots\right] \end{bmatrix} + \boldsymbol{W}_m \quad (3)$$

where $\boldsymbol{W}_m$ represents the noise term constructed from $\boldsymbol{W}$ in a similar way as $\boldsymbol{X}_m$ is obtained from $\boldsymbol{X}$. Assume that the signals are narrow band, i.e.,

$$\boldsymbol{s}(t) \approx \boldsymbol{s}\left(t + \frac{1}{P}\right) \approx \cdots \approx \boldsymbol{s}\left(t + \frac{m-1}{P}\right).$$

In this case, all the block rows in the right-hand term of ((3) are approximately equal, which means that $\boldsymbol{X}_m$ has the factorization

$$\begin{aligned} \boldsymbol{X}_m &\approx \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{A}\boldsymbol{\Phi} \\ \vdots \\ \boldsymbol{A}\boldsymbol{\Phi}^{m-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{s}(0) & \boldsymbol{\Phi}\boldsymbol{s}\left(\frac{1}{P}\right) & \cdots \end{bmatrix} + \boldsymbol{W}_m \\ &=: \boldsymbol{A}_m\boldsymbol{F}_s + \boldsymbol{W}_m \in \mathbb{C}^{mM,\,N-m+1} \end{aligned} \quad (4)$$

where $\boldsymbol{A}_m$, throughout the sequel, which is referred to as the extended array steering matrix, is given by

$$\boldsymbol{A}_m = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{A}\boldsymbol{\Phi} \\ \vdots \\ \boldsymbol{A}\boldsymbol{\Phi}^{m-1} \end{bmatrix} \in \mathbb{C}^{mM,\,d} \quad (5)$$

and

$$\boldsymbol{F}_s = \begin{bmatrix} \boldsymbol{s}(0) & \cdots & \boldsymbol{\Phi}^{N-m}\boldsymbol{s}\left(\frac{N-m}{P}\right) \end{bmatrix} \in \mathbb{C}^{d,\,N-m+1} \quad (6)$$

is a matrix collecting $N - m + 1$ samples of the d sources.

Theorem II.1: Consider an M element antenna array impinged by $d < M$ narrowband far-field signals. Assume that all the signals have distinct (different) center frequencies. Suppose that the signals are divided into r groups, such that the signals from each group have the same DOA. Let p_i, for $i = 1, \ldots, r$, represent the number of sources in the ith group. Then, the m-factor temporally smoothed data matrix $\boldsymbol{X}_m$ of (4) is full rank d if and only if $m \geq \max_i p_i$.

Proof: See Appendix A ∎

The above theorem shows that with m-factor temporally smoothed data, the ESPRIT algorithm can resolve up to m signals having the same DOA. Apart from rank restoration, it enriches the structure of the data matrix, resulting in some interesting properties. In Section II-C, these properties are exploited for joint angle and frequency estimation. For now, it suffices to note that temporal smoothing preserves the shift invariance structure needed for the DOA estimation. That is, the extended array steering matrix $\boldsymbol{A}_m$ has the required shift invariance structure, and the DOAs are estimated in the usual way.

C. Estimation Algorithm

At this point, we have obtained a model with much the same structure as in the classical ESPRIT algorithm but with $\boldsymbol{A}$ replaced by $\boldsymbol{A}_m$. The estimation of the parameters and the construction of the beamformer can now follow the same strategy as well. First, note that the rank of $\boldsymbol{X}$ is only d since this is the number of rows of $\boldsymbol{F}_s$. We compute the SVD of $\boldsymbol{X}$, i.e., $\boldsymbol{X} =: \boldsymbol{U}_s\boldsymbol{\Sigma}_s\boldsymbol{V}_s$, where $\boldsymbol{U}_s$ has d columns, spanning the column space of $\boldsymbol{X}$. Thus, for some nonsingular $d \times d$ matrix $\boldsymbol{T}$

$$\boldsymbol{U}_s = \boldsymbol{A}_m\boldsymbol{T}^{-1}.$$

We begin the estimation of the parameters by defining two types of selection matrices: a pair to select submatrices for estimating $\boldsymbol{\Phi}$ and a pair for estimating $\boldsymbol{\Theta} = \text{diag}\{\theta_i\}_{i=1}^{d}$:

$$\begin{cases} \boldsymbol{J}^x(\phi) := [\boldsymbol{I}_{m-1}\ \ \boldsymbol{0}_1] \otimes \boldsymbol{I}_M \\ \boldsymbol{J}^y(\phi) := [\boldsymbol{0}_1\ \ \boldsymbol{I}_{m-1}] \otimes \boldsymbol{I}_M \end{cases} \quad (7)$$

$$\begin{cases} \boldsymbol{J}^x(\theta) := \boldsymbol{I}_m \otimes [\boldsymbol{I}_{M-1}\ \ \boldsymbol{0}_1] \\ \boldsymbol{J}^y(\theta) := \boldsymbol{I}_m \otimes [\boldsymbol{0}_1\ \ \boldsymbol{I}_{M-1}]. \end{cases} \quad (8)$$

To estimate $\boldsymbol{\Phi}$, we take submatrices consisting of the first and the last $M(m-1)$ rows of $\boldsymbol{U}_s$, respectively, i.e.,

$$\boldsymbol{U}_{x,\phi} = \boldsymbol{J}^x(\phi)\boldsymbol{U}_s, \qquad \boldsymbol{U}_{y,\phi} = \boldsymbol{J}^y(\phi)\boldsymbol{U}_s$$

whereas to estimate $\boldsymbol{\Theta}$, we stack, for each of the m blocks, its first and last $M-1$ rows, respectively

$$\boldsymbol{U}_{x,\theta} = \boldsymbol{J}^x(\theta)\boldsymbol{U}_s, \qquad \boldsymbol{U}_{y,\theta} = \boldsymbol{J}^y(\theta)\boldsymbol{U}_s.$$

These data matrices have the structures

$$\begin{cases} \boldsymbol{U}_{x,\phi} = \boldsymbol{A}'\boldsymbol{T}^{-1} \\ \boldsymbol{U}_{y,\phi} = \boldsymbol{A}'\boldsymbol{\Phi}\boldsymbol{T}^{-1} \end{cases} \quad \begin{cases} \boldsymbol{U}_{x,\theta} = \boldsymbol{A}''\boldsymbol{T}^{-1} \\ \boldsymbol{U}_{y,\theta} = \boldsymbol{A}''\boldsymbol{\Theta}\boldsymbol{T}^{-1} \end{cases} \quad (9)$$

where $\boldsymbol{A}'$ and $\boldsymbol{A}''$ are both submatrices of $\boldsymbol{A}_m$. If dimensions are such that these are low-rank factorizations, then

$$\begin{aligned} \boldsymbol{E}_\phi &:= \boldsymbol{U}_{x,\phi}^{\dagger}\boldsymbol{U}_{y,\phi} = \boldsymbol{T}\boldsymbol{\Phi}\boldsymbol{T}^{-1} \\ \boldsymbol{E}_\theta &:= \boldsymbol{U}_{x,\theta}^{\dagger}\boldsymbol{U}_{y,\theta} = \boldsymbol{T}\boldsymbol{\Theta}\boldsymbol{T}^{-1}. \end{aligned} \quad (10)$$

It is seen that the data matrices $\boldsymbol{E}_\phi$ and $\boldsymbol{E}_\theta$ are jointly diagonalizable by the same matrix $\boldsymbol{T}$. There are several algorithms to compute this joint diagonalization, e.g., by means of QZ iteration [7], [8] or Jacobi iterations [2]. For this to work, it is necessary that each submatrix in (9) has at least d rows. After $\boldsymbol{T}$ has been found, we also have estimates of $\{(\theta_i, \phi_i)\}$ for each of the d sources. This provides us with angle and frequency estimates:

$$\alpha_i = \text{asin}\left(\frac{\arg(\theta_i)}{2\pi\Delta}\right), \qquad f_i = \arg(\phi_i)\frac{P}{2\pi}.$$

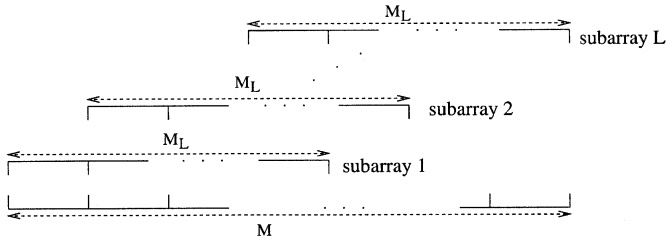


Fig. 1. Spatial smoothing.

III. Factors Affecting the Performance of the JAFE Algorithm

The data matrix given in (3) is the basic JAFE data model. In this section, we consider some additional data manipulation techniques and processing stages that have some influence on the performance of the algorithm.

A. Spatial Smoothing

As discussed above, temporal smoothing enables us to estimate the underlying parameters correctly even if the DOAs of more than one signal are the same. Employing a similar technique in the spatial domain, coherent signals can be separated. This is called spatial smoothing [9]–[11]. In spatial smoothing, an array of M sensors is subdivided into L subarrays. The number of elements in a subarray depends on the way the division is made. For instance, in ULA, allowing a maximum overlap[1] as in Fig. 1, the number of elements per subarray is $M_L = M - L + 1$.

For $l = 1, \ldots, L$, let the $M_L \times M$ matrix $\boldsymbol{J}_l$ be a selection matrix that selects part of the $M \times N$ data matrix $\boldsymbol{X}$ that correspond to the lth subarray. Then, a spatially smoothed $M_L \times LN$ data matrix $\boldsymbol{X}_L$ is constructed as

$$\boldsymbol{X}_L = [\,\boldsymbol{J}_1\boldsymbol{X} \quad \boldsymbol{J}_2\boldsymbol{X} \quad \cdots \quad \boldsymbol{J}_L\boldsymbol{X}\,] \in \mathbb{C}^{M_L,\,LM}. \tag{11}$$

Using the structure of $\boldsymbol{X}$ in (2), we can re-express (11) as

$$\boldsymbol{X}_L = [\,\boldsymbol{J}_1\boldsymbol{A} \quad \cdots \quad \boldsymbol{J}_L\boldsymbol{A}\,] \begin{bmatrix} \boldsymbol{F}_s & & \\ & \ddots & \\ & & \boldsymbol{F}_s \end{bmatrix} + \boldsymbol{W}_L$$

where $\boldsymbol{W}_L$ is a noise term that has also been shuffled in a similar way as $\boldsymbol{X}_L$. Let $\boldsymbol{A}'$ contain the rows of $\boldsymbol{A}$ that correspond to the first subarray; then, from the shift-invariance property, we have the following relation for $k = 1, \ldots, L$:

$$\boldsymbol{J}_k\boldsymbol{A} = \boldsymbol{J}_1\boldsymbol{A}\boldsymbol{\Theta}^{k-1} =: \boldsymbol{A}'\boldsymbol{\Theta}^{k-1}.$$

Using these properties, $\boldsymbol{X}_L$ can be written in a compact form as

$$\begin{aligned}\boldsymbol{X}_L &= \boldsymbol{A}'\left[\boldsymbol{F}_s\ \boldsymbol{\Theta}\boldsymbol{F}_s\ \cdots\ \boldsymbol{\Theta}^{L-1}\boldsymbol{F}_s\right] + \boldsymbol{W}_L \\ &=: \boldsymbol{A}'\boldsymbol{F}_L + \boldsymbol{W}_L \in \mathbb{C}^{M_L,\,LN}.\end{aligned} \tag{12}$$

Theorem III.1: Consider an M element antenna array impinged by d narrowband far-field signals. Assume that all the signals have distinct (different) DOAs. Suppose that the signals are divided into r groups, such that the signals from each group have the same center frequencies. Let, for $i = 1, \ldots, r$, q_i represent the number of sources in the ith group. Then, the L-factor spatially smoothed data matrix $\boldsymbol{X}_L$, with $M_L > d$, (4) is full rank d if and only if $L \geq \max_i q_i$.

Proof: Consider the L-factor spatially smoothed data matrix of (12). As we have assumed that all the sources have different DOAs, the rank of $\boldsymbol{A}'$ is d. Thus, since $\boldsymbol{F}_L$ has only d rows, it is sufficient to show that these are linearly independent. The proof is similar to that given for Theorem II.1. First, note that $\boldsymbol{F}_L^T$ has the same structure as $\boldsymbol{A}_m$, with $\boldsymbol{\Theta}$ playing the role of $\boldsymbol{\Phi}$. Thus, with the same argument, it follows that $\boldsymbol{F}_L$ is full rank if

$$L \geq \max_i q_i. \qquad \blacksquare$$

B. Forward-Backward Averaging

Another way of extending the data matrix is termed as forward-backward averaging [12]–[15]. It uses the fact that the eigenvalues (θ_i, ϕ_i) lie on a unit circle and that the structure of $\boldsymbol{A}$ is centro-symmetric.[2] A forward-backward averaged data matrix $\boldsymbol{X}_{fb}$ is constructed from the data $\boldsymbol{X}$ given in (2) as

$$\boldsymbol{X}_{fb} = [\,\boldsymbol{X} \quad \text{conj}(\boldsymbol{\Pi}\boldsymbol{X})\,] \in \mathbb{C}^{M,\,2N} \tag{13}$$

where $\boldsymbol{\Pi}$ is an anti-diagonal exchange matrix that reverses the ordering of the rows of $\boldsymbol{X}$. It can be shown [5], [18] that if the centro-symmetric property is satisfied, the forward-backward averaged data $\boldsymbol{X}_{fb}$ has the required shift-invariant structure. We can, therefore, apply ESPRIT to solve for the underlying parameters. Note that with this data extension, the number of available temporal samples per antenna element has essentially doubled from N to $2N$, which gives a significant improvement in accuracy. It also provides some protection against loss of rank in the case of coherent sources, i.e., even if $L = 1$ (see above), we can tolerate coherent signals with multiplicity 2.

C. Spatio-Temporally Smoothed and Forward–Backward Averaged Data Model

In this section, we derive a generalized data model that incorporates the above three data extension procedures. We start with the temporally smoothed data $\boldsymbol{X}_m$ given in (3). Let M_L be the number of antenna elements in the subarrays of the spatially smoothed data, and let, for $l = 1, \ldots, L$, the selection matrix $\boldsymbol{J}_l \in \mathbb{R}^{mM_L,\,mM}$ select part of the data matrix $\boldsymbol{X}_m$ that corresponds to the lth subarray. Then, an (m, L) factor spatio-temporally smoothed data matrix $\boldsymbol{X}_{m,L}$ is constructed as

$$\boldsymbol{X}_{m,L} = [\,\boldsymbol{J}_1\boldsymbol{X}_m \quad \cdots \quad \boldsymbol{J}_l\boldsymbol{X}_m\,] \in \mathbb{C}^{mM_L,\,L(N-m+1)}. \tag{14}$$

Using the structure of $\boldsymbol{X}_m$ from (3), this can be factored as

$$\boldsymbol{X}_{m,L} = [\,\boldsymbol{J}_1\boldsymbol{A}_m \cdots \boldsymbol{J}_l\boldsymbol{A}_m\,] \begin{bmatrix} \boldsymbol{F}_s & & \\ & \ddots & \\ & & \boldsymbol{F}_s \end{bmatrix} + \boldsymbol{W}_{m,L}$$

[1]Note that a maximum overlap of subarrays is obtained by shifting a selected window over a single antenna as in Fig. 1

[2]An antenna array is said to be centro-symmetric if the element locations of the array are symmetric with respect to the centroid and the complex radiation characteristics of paired elements are the same (*viz.* [16]–[18]).

where $\boldsymbol{W}_{m,L}$ is a noise term that has also been shuffled in a similar way as $\boldsymbol{X}_{m,L}$. Let $\boldsymbol{A}'_m = \boldsymbol{J}_1\boldsymbol{A}_m \in \mathbb{C}^{mM_L,d}$. Then, from the shift invariance structure of $\boldsymbol{A}_m \in \mathbb{C}^{mM,d}$, it follows that for $k = 1, \ldots, L$

$$\boldsymbol{J}_k\boldsymbol{A}_m = \boldsymbol{J}_1\boldsymbol{A}_m\boldsymbol{\Theta}^{k-1} =: \boldsymbol{A}'_m\boldsymbol{\Theta}^{k-1}.$$

Thus, $\boldsymbol{X}_{m,L}$ can be written in a compact form as

$$\begin{aligned}\boldsymbol{X}_{m,L} &= \boldsymbol{A}'_m[\,\boldsymbol{F}_s \quad \boldsymbol{\Theta}\boldsymbol{F}_s \quad \cdots \quad \boldsymbol{\Theta}^{L-1}\boldsymbol{F}_s\,] + \boldsymbol{W}_{m,L}\\ &=: \boldsymbol{A}'_m\boldsymbol{F}_L + \boldsymbol{W}_{m,L}.\end{aligned} \tag{15}$$

Finally, performing forward-backward averaging on the above (spatio-temporally smoothed) data, we get the $mM_L \times 2L(N-m+1)$ data matrix

$$\boldsymbol{X}_{m,L,fb} = [\,\boldsymbol{X}_{m,L} \quad \text{conj}(\boldsymbol{\Pi}\boldsymbol{X}_{m,L})\,] \tag{16}$$

where $\boldsymbol{\Pi}$ is an exchange matrix that reverses the ordering of the rows of $\boldsymbol{X}_{m,L}$. All of the above data models contain the shift invariance properties needed by the JAFE algorithm (*viz.* [5] and [18]). Thus, the angle-frequency pairs may be estimated in the usual way (by considering shift invariance pairs).

D. Identifiability

The extended data $\boldsymbol{X}_{m,L,fb}$ of (16) is the generalized data model we want to work with. It incorporates three processes:

1) temporal smoothing;
2) spatial smoothing
3) forward-backward averaging.

To derive identifiability conditions we assume that initially, a total of N samples per antenna element are present. Thus, after temporal smoothing, spatial smoothing, and forward-backward averaging, the extended data matrix $\boldsymbol{X}_{m,L,fb}$ has the dimensions $mM_L \times 2L(N-m+1)$. Let $\boldsymbol{U}_s$ be a full rank $mM_L \times d$ matrix that spans the column space of $\boldsymbol{X}_{m,L,fb}$.

Condition 1: To correctly estimate $\boldsymbol{U}_s$, $\boldsymbol{X}_{m,L,fb}$ must have at least d rows and d columns.

Once $\boldsymbol{U}_s$ is determined, the next step in JAFE is to construct submatrices with the required shift invariance properties using selection matrices. To this end, let the four selection matrices $\boldsymbol{J}_x(\phi)$, $\boldsymbol{J}_y(\phi)$, $\boldsymbol{J}_x(\theta)$, and $\boldsymbol{J}_y(\theta)$ be such that

$$\begin{cases}\boldsymbol{U}_{x,\phi} = \boldsymbol{J}_x(\phi)\boldsymbol{U}_s\\ \boldsymbol{U}_{y,\phi} = \boldsymbol{J}_y(\phi)\boldsymbol{U}_s\end{cases} \qquad \begin{cases}\boldsymbol{U}_{x,\theta} = \boldsymbol{J}_x(\theta)\boldsymbol{U}_s\\ \boldsymbol{U}_{y,\theta} = \boldsymbol{J}_y(\theta)\boldsymbol{U}_s\end{cases} \tag{17}$$

form shift invariant pairs.

Condition 2: To estimate the DOAs and frequencies properly, these matrices must have at least d rows.

The actual number of rows in these matrices depend on the way the selection matrices are defined. For a ULA for instance, with the subarrays chosen as shown in Fig. 1, the selection matrices are given by

$$\begin{cases}\boldsymbol{J}_x(\theta) = \boldsymbol{I}_m \otimes [\boldsymbol{I}_{M_L-1} \;\; \boldsymbol{0}_1]\\ \boldsymbol{J}_y(\theta) = \boldsymbol{I}_m \otimes [\boldsymbol{0}_1 \;\; \boldsymbol{I}_{M_L-1}]\\ \boldsymbol{J}_x(\phi) = \boldsymbol{I}_m \otimes [\boldsymbol{I}_{M_L-1} \;\; \boldsymbol{0}_1]\\ \boldsymbol{J}_y(\phi) = \boldsymbol{I}_m \otimes [\boldsymbol{0}_1 \;\; \boldsymbol{I}_{M_L-1}].\end{cases} \tag{18}$$

Putting these into (17) and noting that $M_L = M - L + 1$, it follows that $\boldsymbol{U}_{x,\phi}$ and $\boldsymbol{U}_{y,\phi}$ are both $(m-1)(M-L+1) \times d$ matrices, whereas $\boldsymbol{U}_{x,\theta}$ and $\boldsymbol{U}_{y,\theta}$ are both $m(M-L) \times d$ matrices. Thus, for ULA, combining conditions 1 and 2, we get the following identifiability criteria:

$$\begin{aligned}&\text{a)} \quad d \le m(M-L)\\ &\text{b)} \quad d \le (m-1)(M-L+1)\\ &\text{c)} \quad d \le 2L(N-m+1).\end{aligned} \tag{19}$$

Given the number of sensors M and the number of snapshots N, we want to find the pair (m, L) that maximizes the number of signals that can be identified. In analogy to a similar problem considered in [5], we obtain as the solution to this maximization problem:

$$\text{if } N \ge M + \frac{1}{\sqrt{2}} \qquad \begin{cases}d_{\max} = M(N+1)(2-\sqrt{2})^2\\ m_o = (N+1)(2-\sqrt{2})\\ L_o = M(\sqrt{2}-1)\end{cases} \tag{20}$$

$$\text{if } N \le M - \frac{1}{\sqrt{2}} \qquad \begin{cases}d_{\max} = N(M+1)(2-\sqrt{2})^2\\ m_o = N(2-\sqrt{2})+1\\ L_o = (M+1)(\sqrt{2}-1).\end{cases} \tag{21}$$

The first set of equations corresponds to a region where conditions a) and c) are satisfied with equality and the second set corresponds to a region where conditions b) and c) are satisfied with equality. The actual maxima are slightly smaller because m and L can take integer values only.

For identifiability in addition to the above conditions, the submatrices $\boldsymbol{U}_{x\phi}$, $\boldsymbol{U}_{y\phi}$, $\boldsymbol{U}_{x\phi}$, and $\boldsymbol{U}_{y\phi}$ in (9) must also be full rank d. If the impinging wavefronts have distinct frequencies and DOAs, the Vandermonde structures of the matrices ensures that this is the case. Under conditions where there are multiple DOAs or multiple center frequencies, the matrices may still be full ranked if, in addition to (20) and (21), the following are also satisfied (*viz.* [5]):

$$\begin{cases}m \ge p\\ N \ge \frac{3}{2}p\end{cases}$$

and

$$\begin{cases}L \ge \frac{1}{2}q\\ M \ge \frac{3}{2}q\end{cases}$$

where p and q are the multiplicity of the DOAs and center frequencies, respectively. These inequalities are derived by considering the results of Theorems II.1 and III.1.

E. Whitening as the JAFE Processing Stage

The spatio-temporal smoothing procedure introduces correlation between the noise terms in the different rows of the data matrix. In many cases, this correlation causes degradation as it tends to reduce the degree of averaging that could have been

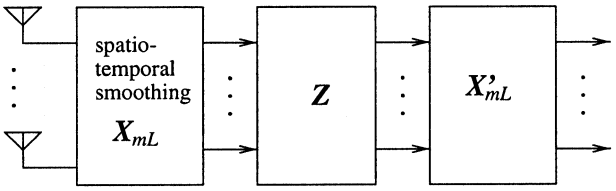


Fig. 2. Whitening the spatio-temporally smoothed data.

obtained had the noise been white. In this context, the JAFE algorithm can be preceded with a whitening filter, as shown in Fig. 2. Consider the noise part of spatio-temporally smoothed data matrix given in (14). Let the singular value decomposition of the noise covariance matrix $\boldsymbol{R}_w = \boldsymbol{W}_{m,L}\boldsymbol{W}_{m,L}^H$ be given by

$$\boldsymbol{R}_w = \boldsymbol{U}_n\boldsymbol{\Sigma}_n^2\boldsymbol{U}_n^H.$$

Then, the whitened data matrix is derived as (*viz*. [19])

$$\boldsymbol{X}'_{m,L} = \boldsymbol{U}_n\boldsymbol{\Sigma}_n^{-1}\boldsymbol{U}_n^H\boldsymbol{X}_{m,L}. \tag{22}$$

Thus, in Fig. 2, the transformation matrix $\boldsymbol{Z}$ is equal to $\boldsymbol{U}_n\boldsymbol{\Sigma}_n^{-1}\boldsymbol{U}_n^H$. Let the SVD of $\boldsymbol{X}'_{m,L}$ be given by

$$\boldsymbol{X}'_{m,L} = \boldsymbol{U}'\boldsymbol{\Sigma}'\boldsymbol{V}'^H$$

and let $\boldsymbol{U}'_s$ be the d dominant columns of $\boldsymbol{U}$ corresponding to the d largest singular values; then, the JAFE algorithm is implemented in the usual way by considering shift invariance properties of $\boldsymbol{U}_s = \boldsymbol{U}_n\boldsymbol{\Sigma}_n^{-1}\boldsymbol{U}_n^H\boldsymbol{U}'_s$. In Section IX, we present simulation results comparing the performances of the JAFE algorithm implemented with and without whitening.

F. Multiresolution ESPRIT Algorithm

Recall that the data sampling rate used in constructing the data matrix of (1) is P times the Nyquist rate of the baseband signals. Since P can be quite large, it would be very expensive to construct a full data matrix of all samples. In fact, it is sufficient to subsample: Collect 2 subsequent samples at a rate P followed by m samples at a rate P/k, where k is an integer greater than one. This leads to the so-called multiresolution ESPRIT (MR-ESPRIT) [6], [20], [21] based JAFE algorithm. In the MR-based JAFE algorithm, multiple spatio-temporal sampling rates are used to improve the parameter estimation accuracy without raising the estimation complexity. In the above example, for instance, if the size of the data matrix is preserved, the MR approach provides a k-times accuracy improvement. On the other hand, if we perform downsampling on the original number of samples (effectively reducing the number of samples by the factor k), we obtain a significant reduction in complexity for the same estimation accuracy. In the simulation results of Section IX, we give quantitative analysis of this effect. See the above papers for further understanding.

IV. Performance Analysis

As described in Section II-C, the JAFE algorithm involves three main steps, namely

1) singular value decomposition (SVD) of the data matrix;
2) diagonalization of a set of eigenvalue decomposition (EVD) problems;
3) transformation of the eigenvalues into signal parameters.

The first step, which is equivalent to finding the EVD of the data covariance matrix, is well studied in the literature [22]–[26] for the case of white Gaussian noise contaminated data model. In our case, however, since some data stacking techniques have been employed, the noise is no longer white. Thus, in this section, we will first derive the eigenvalue estimation error for the JAFE data model and show how this can be applied to derive errors on shift invariance parameters. In Section V, we make use of the results of this section to derive more specific error expressions for the parameterized DOA and frequency estimations. Similar analyses, in the context of white Gaussian noise, have been presented in [27] and [28]. The results obtained here could be seen as the generalization of these results.

When we assume that prewhitening has been applied to the data before the application of the JAFE algorithm, as discussed in Section III-E, the analysis reduces to the forms similar to those described in [27] and [28]. However, the results in [27] are derived considering a ULA only. Their final result does not give explicit relations between the parameter estimation errors and the noise. In [28], the results of [27] are extended to more general array geometries, and the analysis there is fairly complete. However, the results are derived for DOA estimation only, and they consider a data model without any extension or stacking. Here, we give derivations for both angle and frequency estimations, and we also show how the different data extension procedures affect the estimation performances.

A. Eigenvectors of the Data Covariance Matrix

The following theorem, whose proof is given in Appendix-B, gives the eigenvectors estimation errors for the JAFE data model given in (15).

Theorem IV.1: Consider an M-element antenna array impinged by d far-field narrowband signals. Let the (m, L) factor spatio-temporally smoothed data matrix $\boldsymbol{X}_{m,L} \in \mathbb{C}^{mM_L \times L(N-m)}$ be as given in (14), $\boldsymbol{R}_{m,L} = (1/N)\boldsymbol{X}_{m,L}\boldsymbol{X}_{m,L}^H$ be the finite sample data covariance matrix, and $\boldsymbol{U} \in \mathbb{C}^{mM_L, mM_L}$ and $\boldsymbol{\Sigma} \in \mathbb{C}^{mM_L, mM_L}$ be such that the eigenvalue decomposition of $\overline{\boldsymbol{R}}_{m,L} = E\{\boldsymbol{R}_{m,L}\}$ is given by

$$\overline{\boldsymbol{R}}_{m,L} = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^H.$$

Let $\boldsymbol{U}_s$ be the first d columns of $\boldsymbol{U}$. In Section II-C, it has been shown that for some invertible matrix $\boldsymbol{T}$

$$\boldsymbol{U}_s = \boldsymbol{A}_m\boldsymbol{T}^{-1}. \tag{23}$$

Now, let $\boldsymbol{u}_j$ be the jth eigenvector of $\overline{\boldsymbol{R}}_{m,L}$, where $j \leq d$. Moreover, let $\Delta\boldsymbol{u}_j$ represent a noise-caused perturbation on $\boldsymbol{u}_j$, $\sigma_k^2 = \overline{\sigma}_k^2 + \sigma_n^2$ be the kth eigenvalue of $\overline{\boldsymbol{R}}_{m,L}$, where $\overline{\sigma}_i^2$ is the kth noise free eigenvalue, and let σ_n^2 is the noise contribution. Let $\boldsymbol{\Phi} = \text{diag}\{\phi_i\}_{i=1}^d$, and let $\boldsymbol{\Theta} = \text{diag}\{\theta_i\}_{i=1}^d$ be the parameterizations of the center frequencies and DOAs of the d signals. Moreover, let

- $\boldsymbol{Z}^x$ be a Toeplitz matrix with all the elements equal to zero, except for those unity valued entries on the xth parallel to the main diagonal;
- $m_o := \min(m, N-m)$ and $L_o := \min(M_L, L)$
- $\boldsymbol{Q} := \boldsymbol{T}^{-1}$, $\boldsymbol{q}_x^H$ and $\boldsymbol{t}_x$ be the xth row and column of the matrices $\boldsymbol{Q}$ and $\boldsymbol{T}$, respectively. (Note that $\boldsymbol{q}_x^H \boldsymbol{t}_y = 0$, for $x \neq y$ and $\boldsymbol{q}_x^H \boldsymbol{t}_y = 1$ for $x = y$.)
- $\Upsilon(x, y, v, w) := \sum_{h=-m_o}^{m_o} \sum_{r=-L_o}^{L_o} \boldsymbol{u}_x^H \boldsymbol{Z}^{hM_L+r} \boldsymbol{u}_y \boldsymbol{q}_v^H \boldsymbol{\Phi}^h \boldsymbol{\Theta}^r \boldsymbol{t}_w$;
- $\Omega(x, y, v, w)^H := \sum_{h=-m_o}^{m_o} \sum_{r=-L_o}^{L_o} \boldsymbol{u}_x^H \boldsymbol{Z}^{-hM_L-r} \boldsymbol{u}_y \boldsymbol{q}_v^H \boldsymbol{\Phi}^{-h} \boldsymbol{\Theta}^{-r} \boldsymbol{t}_w$;
- $\eta(x, y, v, w) := \sum_{h=-m_o}^{m_o} \sum_{r=-L_o}^{L_o} \boldsymbol{u}_x^H \boldsymbol{z}^{hM_L+r} \boldsymbol{u}_y \boldsymbol{u}_v^H \boldsymbol{z}^{-hM_L-r} \boldsymbol{u}_w$.

Assuming that $\overline{\boldsymbol{R}}_{m,L}$ has distinct eigenvalues, the covariance of the eigenvector estimation error $E\{\Delta \boldsymbol{u}_j \Delta \boldsymbol{u}_p^H\}$ is given by

$$\begin{aligned} &E\{\Delta \boldsymbol{u}_v \Delta \boldsymbol{u}_w^H\}_{m,L} \\ &\quad = \frac{\sigma_n^2}{L(N-m+1)} \\ &\qquad \times \sum_{\substack{j=1 \\ j\neq v}}^{mM_L} \sum_{\substack{i=1 \\ i\neq w}}^{mM_L} \frac{\Omega(j, i, v, w)\overline{\sigma}_v^2 + \eta(j, i, v, w)\sigma_n^2}{(\sigma_v^2 - \sigma_j^2)(\sigma_w^2 - \sigma_k^2)} \boldsymbol{u}_j \boldsymbol{u}_k^H \\ &\qquad + \sum_{\substack{j=1 \\ j\neq v}}^{d} \sum_{\substack{i=1 \\ i\neq w}}^{d} \frac{\Upsilon(v, w, j, i)\overline{\sigma}_j^2}{(\sigma_v^2 - \sigma_j^2)(\sigma_w^2 - \sigma_k^2)} \boldsymbol{u}_j \boldsymbol{u}_k^H. \end{aligned} \tag{24}$$

Lemma IV.1: For the whitened spatio-temporally smoothed data matrix, the covariance of the eigenvector estimation error reduces to $E\{\Delta \boldsymbol{u}_v \Delta \boldsymbol{u}_w^H\} = 0$ for $w \neq v$, and

$$\begin{aligned} &E\{\Delta \boldsymbol{u}_v \Delta \boldsymbol{u}_v^H\}_{m,L} \\ &\quad = \frac{1}{L(N-m+1)} \\ &\qquad \times \left(\sum_{\substack{j=1 \\ j\neq v}}^{d} \frac{\sigma_v^2\sigma_j^2 - \overline{\sigma}_v^2\overline{\sigma}_j^2}{(\sigma_v^2 - \sigma_j^2)^2} \boldsymbol{u}_j \boldsymbol{u}_j^H + \sigma_n^2 \sigma_v^2 \sum_{j=d+1}^{mM_L} \frac{\boldsymbol{u}_j \boldsymbol{u}_j^H}{(\sigma_v^2 - \sigma_j^2)^2} \right) \end{aligned} \tag{25}$$

for $w = v$.

Proof: See Appendix C. ∎

B. Shift Invariance Parameters

Let $\boldsymbol{X}_{m,L} \in \mathbb{C}^{mM_L, L(N-m)}$ and $\boldsymbol{U}_s = \boldsymbol{A}_m \boldsymbol{T}^{-1} \in \mathbb{C}^{mM_L, d}$ be as defined in Theorem IV.1. In Section II-B it has been shown that the angle and frequency estimation is attained by considering the dual shift invariance structure present in $\boldsymbol{U}_s$. In this section, we take a close look at the behavior of this computation. To this end, as before, define two selection matrices $\boldsymbol{J}_x$ and $\boldsymbol{J}_y$, such that the two full column ranked matrices $\boldsymbol{U}_x$ and $\boldsymbol{U}_y$ defined as

$$\begin{aligned} \boldsymbol{U}_x &= \boldsymbol{J}_x \boldsymbol{U}_s \\ \boldsymbol{U}_y &= \boldsymbol{J}_y \boldsymbol{U}_s \end{aligned} \tag{26}$$

are such that the $d \times d$ matrix $\boldsymbol{E} = \boldsymbol{U}_x^{\dagger} \boldsymbol{U}_y$ has the eigenvalue decomposition

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix} = \boldsymbol{QET} \tag{27}$$

$$= \begin{bmatrix} \boldsymbol{q}_1^H \\ \vdots \\ \boldsymbol{q}_d^H \end{bmatrix} \boldsymbol{E} [\boldsymbol{t}_1 \quad \cdots \quad \boldsymbol{t}_d] \tag{28}$$

where $\boldsymbol{Q} = \boldsymbol{T}^{-1}$, and $\boldsymbol{T}$ is as in (23). In the above model, coinciding eigenvalues are allowed, as long as the eigenvectors (the corresponding columns of $\boldsymbol{T}$) are linearly independent. In fact, we require that the eigenvectors be sufficiently distinct[3] such that after small perturbation they remain linearly independent. This assumption is needed because $\boldsymbol{T}^{-1}$ appears in the derivations, and thus, we want $\boldsymbol{T}$ to be invertible. From (28), it follows that

$$\lambda_i = \boldsymbol{q}_i^H \boldsymbol{E} \boldsymbol{t}_i \tag{29}$$

where $\boldsymbol{q}_i$ and $\boldsymbol{t}_i$ are the left and right eigenvectors of $\boldsymbol{E}$, respectively, i.e., $\boldsymbol{q}_i^H \boldsymbol{E} = \lambda_i \boldsymbol{q}_i^H$ and $\boldsymbol{E}\boldsymbol{t}_i = \lambda_i \boldsymbol{t}_i$. Moreover, for $i, j = 1, 2, \ldots, d$, $\boldsymbol{q}_i$ and $\boldsymbol{t}_j$ satisfy

$$\boldsymbol{q}_i^H \boldsymbol{t}_j = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

Let $\Delta\lambda_i$, $\Delta \boldsymbol{q}_i$, $\Delta \boldsymbol{t}_i$ and $\Delta \boldsymbol{E}$ represent noise caused perturbations on λ_i, $\boldsymbol{q}_i$, $\boldsymbol{t}_i$ and $\boldsymbol{E}$, respectively. We assume that the eigenvectors are sufficiently separated and remain distinct after these small variations. Thus, under noisy situations, (29) may be rewritten as

$$\lambda_i + \Delta\lambda_i = (\boldsymbol{q}_i^H + \Delta \boldsymbol{q}_i^H)(\boldsymbol{E} + \Delta \boldsymbol{E})(\boldsymbol{t}_i + \Delta \boldsymbol{t}_i).$$

Taking only the linear terms in the above equation, and after some rearrangement of the terms, we obtain

$$\begin{aligned} \Delta\lambda_i &\approx \boldsymbol{q}_i^H \Delta \boldsymbol{E} \boldsymbol{t}_i + \Delta \boldsymbol{q}_i^H \boldsymbol{E} \boldsymbol{t}_i + \boldsymbol{q}_i^H \boldsymbol{E} \Delta \boldsymbol{t}_i \\ &= \boldsymbol{q}_i^H \Delta \boldsymbol{E} \boldsymbol{t}_i + \Delta \boldsymbol{q}_i^H \lambda_i \boldsymbol{t}_i + \lambda_i \boldsymbol{q}_i^H \Delta \boldsymbol{t}_i \\ &= \boldsymbol{q}_i^H \Delta \boldsymbol{E} \boldsymbol{t}_i + \lambda_i (\Delta \boldsymbol{q}_i^H \boldsymbol{t}_i + \boldsymbol{q}_i^H \Delta \boldsymbol{t}_i). \end{aligned} \tag{31}$$

Since $\boldsymbol{Q} = \boldsymbol{T}^{-1}$, the noise terms $\Delta \boldsymbol{q}_i$ and $\Delta \boldsymbol{t}_i$ are not independent. Their relation is derived by noting that (30) is valid under noisy conditions as well, i.e.,

$$(\boldsymbol{q}_i^H + \Delta \boldsymbol{q}_i^H)(\boldsymbol{t}_j + \Delta \boldsymbol{t}_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases}$$

Taking the first-order terms only, this simplifies to

$$\begin{aligned} \boldsymbol{q}_i^H \boldsymbol{t}_i + \boldsymbol{q}_i^H \Delta \boldsymbol{t}_i + \Delta \boldsymbol{q}_i^H \boldsymbol{t}_i &\approx 1 \\ \boldsymbol{q}_i^H \Delta \boldsymbol{t}_i + \Delta \boldsymbol{q}_i^H \boldsymbol{t}_i &\approx 0. \end{aligned}$$

[3]Note that the eigensubspaces belonging to such a multiple eigenvalue do not have unique eigenvectors. Thus, an orthogonal basis of eigenvectors in that subspace can be chosen.

This means that the second term in (31) is approximately zero, and the first-order approximations of the errors on the eigenvalues are given by (*viz.* [27])

$$\Delta\lambda_i \approx \boldsymbol{q}_i^H \Delta\boldsymbol{E}\boldsymbol{t}_i. \tag{32}$$

Let $\Delta\boldsymbol{U}_x$ represent the noise perturbation on $\boldsymbol{U}_x$ and similarly for $\Delta\boldsymbol{U}_y$ and $\boldsymbol{U}_y$; then, an expression for $\Delta\boldsymbol{E}$ is derived by noting that

$$\boldsymbol{E} + \Delta\boldsymbol{E} = (\boldsymbol{U}_x + \Delta\boldsymbol{U}_x)^\dagger (\boldsymbol{U}_y + \Delta\boldsymbol{U}_y). \tag{33}$$

If the perturbation is small enough, the first term in the above equation can be approximated (up to first order) as

$$(\boldsymbol{U}_x + \Delta\boldsymbol{U}_x)^\dagger \approx (\boldsymbol{I} - \boldsymbol{U}_x^\dagger \Delta\boldsymbol{U}_x)\boldsymbol{U}_x^\dagger.$$

Putting this into (33) and taking the linear terms only, we get the following approximation for $\Delta\boldsymbol{E}$

$$\Delta\boldsymbol{E} \approx \boldsymbol{U}_x^\dagger (\Delta\boldsymbol{U}_y - \Delta\boldsymbol{U}_x\boldsymbol{E}).$$

Using this relation, and noting that $\boldsymbol{E}\boldsymbol{t}_i = \lambda_i\boldsymbol{t}_i$, $\Delta\boldsymbol{U}_x = \boldsymbol{J}_x\Delta\boldsymbol{U}_s$ and $\Delta\boldsymbol{U}_y = \boldsymbol{J}_y\Delta\boldsymbol{U}_s$, the expression for $\Delta\lambda_i$ in (32) becomes

$$\Delta\lambda_i \approx \boldsymbol{q}_i^H \boldsymbol{U}_x^\dagger (\boldsymbol{J}_y - \lambda_i\boldsymbol{J}_x)\Delta\boldsymbol{U}_s\boldsymbol{t}_i.$$

Note that $\boldsymbol{U}_s$ and the selection matrices have dimensions that are functions of the spatio-temporal smoothing factors m and L. In the following, these dependencies are made explicit using indexed references. Thus, putting $\boldsymbol{r}_{m,L}^H(\lambda_i) = \boldsymbol{q}_i^H\boldsymbol{U}_x^\dagger(\boldsymbol{J}_y - \lambda_i\boldsymbol{J}_x)$, the error $\Delta\lambda_i$ may then be written as $\Delta\lambda_i \approx \boldsymbol{r}_{m,L}^H(\lambda_i)\Delta\boldsymbol{U}_s\boldsymbol{t}_i$ and its mean square value as

$$\begin{aligned}\sigma_{m,L}(\lambda_i) &:= E\{|\Delta\lambda_i|^2\} \\ &\approx \boldsymbol{r}_{m,L}^H(\lambda_i) E\{\Delta\boldsymbol{U}_s\boldsymbol{t}_i\boldsymbol{t}_i^H\Delta\boldsymbol{U}_s^H\}\boldsymbol{r}_{m,L}(\lambda_i).\end{aligned}$$

Let $\boldsymbol{u}_j$ and $\Delta\boldsymbol{u}_j$ be such that $\boldsymbol{U}_s = [\boldsymbol{u}_1 \ \cdots \ \boldsymbol{u}_d]$ and $\Delta\boldsymbol{U}_s = [\Delta\boldsymbol{u}_1 \ \cdots \ \Delta\boldsymbol{u}_d]$. With these definitions, $\sigma_{m,L}(\lambda_i)$ is given as

$$\sigma_{m,L}(\lambda_i) = \boldsymbol{r}_{m,L}^H(\lambda_i) E\left\{\sum_{j=1}^{d}\sum_{k=1}^{d}\Delta\boldsymbol{u}_j t_{ji} t_{ki}^H \Delta\boldsymbol{u}_k^H\right\}\boldsymbol{r}_{m,L}(\lambda_i)$$

where t_{ji} is the jth entry of the column vector $\boldsymbol{t}_i$ defined earlier. Noting that t_{ji} is noise independent, the above can be rewritten as

$$\sigma_{m,L}(\lambda_i) = \boldsymbol{r}_{m,L}^H(\lambda_i)\left(\sum_{j=1}^{d}\sum_{k=1}^{d} t_{ji}t_{ki}^H E\{\Delta\boldsymbol{u}_j\Delta\boldsymbol{u}_k^H\}\right)\boldsymbol{r}_{m,L}(\lambda_i). \tag{34}$$

Note that a similar term to $\boldsymbol{r}_{m,L}(\lambda)$ has been derived in [28]. Later, we will make use of some interesting properties of this term derived in [28] to get simplified expressions for $\sigma_{m,L}(\lambda_i)$.

One further simplification of (34) is conveniently obtained for a whitened data matrix. In the remainder of this section, we will assume that a whitening has been applied to the data matrix prior to the application of the algorithm. In this case, the eigenvector estimation error $E\{\Delta\boldsymbol{u}_j\Delta\boldsymbol{u}_k^H\}$ is as given in (25). Let $\zeta_{j,k}$ be defined as

$$\zeta_{j,k} := \frac{\sigma_j^2\sigma_k^2 - \overline{\sigma}_j^2\overline{\sigma}_k^2}{(\sigma_j^2 - \sigma_k^2)^2}$$

and then, putting (25) into (34), it follows that

$$\begin{aligned}&\sigma_{m,L}(\lambda_i) \\ &\quad\approx \boldsymbol{r}_{m,L}^H(\lambda_i)\sum_{j=1}^{d}|t_{ji}|^2\frac{1}{L(N-m+1)} \\ &\qquad\times\left(\sum_{\substack{k=1\\k\neq j}}^{d}\zeta_{j,k}(\boldsymbol{u}_k\boldsymbol{u}_k^H) + (\sigma_j^2\sigma_n^2)\sum_{k=d+1}^{mM_L}\frac{\boldsymbol{u}_k\boldsymbol{u}_k^H}{(\sigma_j^2-\sigma_n^2)^2}\right) \\ &\qquad\times\boldsymbol{r}_{m,L}(\lambda_i).\end{aligned} \tag{35}$$

Let the left eigenvectors of the data covariance matrix be partitioned into $\boldsymbol{U}_s$ and $\boldsymbol{U}_n$, such that $\boldsymbol{U}_s$ spans the signal subspace, and $\boldsymbol{U}_n$ spans the noise subspace. Let the (punctured) diagonal matrix $\boldsymbol{\Gamma}_j$, with a zero at the jjth position, be defined as

$$\boldsymbol{\Gamma}_j = \begin{bmatrix}\operatorname{diag}\{\zeta_{j,k}\}_{k=1}^{j-1} & & \\ & 0 & \\ & & \operatorname{diag}\{\zeta_{j,k}\}_{k=j+1}^{d}\end{bmatrix}.$$

Then, we can rewrite (35) in a simplified way as

$$\begin{aligned}\sigma_{m,L}(\lambda_i) \approx \boldsymbol{r}_{m,L}^H(\lambda_i)\Bigg(&\sum_{j=1}^{d}|t_{ji}|^2\frac{1}{L(N-m+1)} \\ &\times\left[\boldsymbol{U}_s\boldsymbol{\Gamma}_j\boldsymbol{U}_s^H + \frac{\sigma_j^2\sigma_n^2}{(\sigma_j^2-\sigma_n^2)^2}\boldsymbol{U}_n\boldsymbol{U}_n^H\right]\Bigg)\boldsymbol{r}_{m,L}(\lambda_i).\end{aligned} \tag{36}$$

Note that in the above discussions, we have made no assumption on the array geometry. The geometry information is contained in $\boldsymbol{r}_{m,L}^H(\lambda_i) = \boldsymbol{q}_i^H\boldsymbol{U}_x^\dagger(\boldsymbol{J}_y - \lambda_i\boldsymbol{J}_x)$, which is referred to as the array geometry parameter. Let $\boldsymbol{A}_x = \boldsymbol{J}_x\boldsymbol{A}$, $\boldsymbol{A}_y = \boldsymbol{J}_y\boldsymbol{A}$, and $\boldsymbol{w}_i^H$ be the ith row of $\boldsymbol{A}_x^\dagger$; then, noting that $\boldsymbol{U}_x = \boldsymbol{A}_x\boldsymbol{T}^{-1}$, where $\boldsymbol{T}$ is as given in (27), the array geometry parameter may be expressed as

$$\begin{aligned}\boldsymbol{r}_{m,L}^H(\lambda_i) &= \boldsymbol{q}_i^H\boldsymbol{T}\boldsymbol{A}_x^\dagger(\boldsymbol{J}_y - \lambda_i\boldsymbol{J}_x) \\ &= \boldsymbol{w}_i^H(\boldsymbol{J}_y - \lambda_i\boldsymbol{J}_x).\end{aligned} \tag{37}$$

A further simplification of (36) is obtained by noting $\boldsymbol{A}_y = \boldsymbol{A}_x\boldsymbol{\Lambda}$ and the following fact (*viz.* [28]):

$$\begin{aligned}\boldsymbol{r}_{m,L}^H(\lambda_i)\boldsymbol{U}_s &= \boldsymbol{r}_{m,L}^H(\lambda_i)\boldsymbol{A}\boldsymbol{T}^{-1} \\ &\approx \boldsymbol{q}_i^H\boldsymbol{T}\boldsymbol{A}_x^\dagger(\boldsymbol{J}_y - \lambda_i\boldsymbol{J}_x)\boldsymbol{A}\boldsymbol{T}^{-1} \\ &= \boldsymbol{q}_i^H\boldsymbol{T}\boldsymbol{A}_x^\dagger(\boldsymbol{A}_y - \lambda_i\boldsymbol{A}_x)\boldsymbol{T}^{-1} \\ &= \boldsymbol{q}_i^H\boldsymbol{T}\boldsymbol{A}_x^\dagger(\boldsymbol{A}_x\boldsymbol{\Lambda} - \lambda_i\boldsymbol{A}_x)\boldsymbol{T}^{-1} \\ &= \boldsymbol{q}_i^H\boldsymbol{T}\boldsymbol{A}_x^\dagger\boldsymbol{A}_x(\boldsymbol{\Lambda} - \lambda_i\boldsymbol{I})\boldsymbol{T}^{-1} \\ &= \boldsymbol{q}_i^H\boldsymbol{T}(\boldsymbol{\Lambda} - \lambda_i\boldsymbol{I})\boldsymbol{T}^{-1} = [0\cdots 0].\end{aligned}$$

From this, it follows that the first term in (36) vanishes, and with the parameter η_i defined as

$$\frac{1}{\eta_i} = \sum_{j=1}^{d} |t_{ji}|^2 \frac{\sigma_j^2}{(\sigma_j^2 - \sigma_n^2)^2} \tag{38}$$

we get the following expression for $\sigma_{m,L}(\lambda_i)$:

$$\sigma_{m,L}(\lambda_i) \approx \frac{\sigma_n^2/\eta_i}{L(N-m+1)} \boldsymbol{r}_{m,L}^H(\lambda_i) \boldsymbol{U}_n \boldsymbol{U}_n^H \boldsymbol{r}_{m,L}(\lambda_i). \tag{39}$$

Let $\rho_{m,L}(\lambda_i)$ be defined as

$$\rho_{m,L}(\lambda_i) = \boldsymbol{r}_{m,L}^H(\lambda_i) \boldsymbol{r}_{m,L}(\lambda_i). \tag{40}$$

Then, noting that $\boldsymbol{r}_{m,L}^H(\lambda_i)\boldsymbol{U}_n\boldsymbol{U}_n^H = \boldsymbol{r}_{m,L}^H(\lambda_i)(\boldsymbol{I} - \boldsymbol{U}_s\boldsymbol{U}_s^H) = \boldsymbol{r}_{m,L}^H(\lambda_i)$, (39) simplifies to

$$\sigma_{m,L}(\lambda_i) \approx \frac{1}{L(N-m+1)} \frac{\sigma_n^2}{\eta_i} \rho_{m,L}(\lambda_i). \tag{41}$$

Theorem IV.2: Suppose that an M-element antenna array that is split into L subarrays each with M_L elements, is impinged by $d < M_L$ narrowband signals. Let $\boldsymbol{X} \in \mathbb{C}^{mM_L, L(N-m)}$ be an (m, L)-factor spatio-temporally smoothed data matrix collected at the output of the array, σ_n^2 be the power of the noise, and $\boldsymbol{U}_s \in \mathbb{C}^{mM_L, d}$ be the d dominant eigenvectors of the data covariance matrix. Now, let $\boldsymbol{U}_x$ and $\boldsymbol{U}_y$ be two submatrices of $\boldsymbol{U}_s$ such that, for some nonsingular $d \times d$ matrix $\boldsymbol{T}$ and a $d \times d$ nonsingular diagonal matrix $\boldsymbol{\Lambda}$, $\boldsymbol{U}_x^{\dagger}\boldsymbol{U}_y = \boldsymbol{T}^{-1}\boldsymbol{\Lambda}\boldsymbol{T}$. Let λ_i be the ith entry of $\boldsymbol{\Lambda}$ and $\sigma_{m,L}(\lambda_i)$ be the variance of the estimation error on λ_i. Then, the dependence of $\sigma_{m,L}(\lambda_i)$ on the array geometry and the spatio-temporal smoothing factors is completely described by the factor

$$G(m, L) := \frac{1}{L(N-m)} \rho_{m,L}(\lambda_i).$$

Proof: For the condition stated in the theorem, it has already been shown that $\sigma_{m,L}(\lambda_i)$ is given by (41). Thus, for the proof, it is sufficient to show that η_i defined in (38) is independent of the array geometry and the factors m and L. To this end, let the $Mm \times d$ matrix $\boldsymbol{A}_m$ be the extended array steering matrix associated with a data matrix with a temporal smoothing factor of m, $\boldsymbol{U}_s$ be a unitary matrix that spans the column space of the extended data matrix, and $\boldsymbol{\Sigma}_s$ be a diagonal matrix that contains the d largest singular values of the extended data matrix. Then, there exists a nonsingular $d \times d$ matrix $\boldsymbol{T}$ such that the data covariance matrix $\boldsymbol{R}_x = (1/L(N-m+1))\boldsymbol{X}\boldsymbol{X}^H$ may be expressed as

$$\begin{aligned} \boldsymbol{U}_s\boldsymbol{\Sigma}_s^2\boldsymbol{U}_s^H &= \boldsymbol{R}_x = \boldsymbol{A}_m\boldsymbol{R}_s\boldsymbol{A}_m^H + \sigma_n^2\boldsymbol{I} \\ &= \boldsymbol{U}_s\boldsymbol{T}\boldsymbol{R}_s\boldsymbol{T}^H\boldsymbol{U}_s^H + \sigma_n^2\boldsymbol{I} \end{aligned}$$

where $\boldsymbol{R}_s$ is the signal covariance matrix. Solving for $\boldsymbol{R}_s^{-1}$, we obtain

$$\boldsymbol{R}_s^{-1} = \boldsymbol{T}^H(\boldsymbol{\Sigma}_s^2 - \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{T}$$

and thus, with t_{ji} equal to the jith entry of $\boldsymbol{T}$

$$\{\boldsymbol{R}_s^{-1}\}_{ii} = \sum_{j=1}^{d} |t_{ji}|^2 \frac{1}{\sigma_j^2 - \sigma_n^2},$$

where σ_j is the jjth entry of the diagonal matrix $\boldsymbol{\Sigma}_s$. Note that the signal covariance matrix (the left-hand expression) is independent of the array geometry. This means that the right-hand summation and, therefore, η_i is also independent of the array geometry and of m. In a similar way, one can show that η_i is also independent of the spatial smoothing factor L. Note that for good SNR, $\sigma_j \gg \sigma_n$ for all $j \leq d$, in which case, $\eta_i \approx 1/\{\boldsymbol{R}_s^{-1}\}_{ii}$ and, hence, is independent of m and L. ■

This result is useful when we consider the effects of m and L on the estimation errors because the dependency of $\sigma_{m,L}(\lambda_i)$ on these parameters is completely described by the less complex factor $\rho_{m,L}(\lambda_i)$. To emphasize the fact that the geometry information is fully described by $\rho_{m,L}(\lambda_i)$, in the sequel, it will be referred to as the geometric factor. Moreover, since η_i is dependent purely on the signal covariance matrix and on the SNR, it is referred to as the signal factor, and the ratio $\widehat{\mathrm{SNR}} = \eta_i/\sigma_n^2$ is termed as the effective SNR.

V. Parameterized DOA and Frequency Estimation

In the foregoing discussions, we have made no reference to the parameter to be estimated. The analysis up to now, therefore, applies for both the parameterized DOA and frequency estimations alike. The distinction comes in the way the selection matrices are defined. In the following, we derive more specific results by separately considering the parameterized DOA and the parameterized frequency estimations.

A. Parameterized DOA Estimation

In line with the discussions in Section IV-B, the parameterized DOA estimation error $\sigma_{m,L}(\theta_i)$ is obtained from (41) by replacing λ_i with θ_i and $\rho_{m,L}(\lambda_i)$ with $\rho_{m,L}(\theta_i)$:

$$\sigma_{m,L}(\theta_i) \approx \frac{1}{L(N-m+1)} \frac{\sigma_n^2}{\eta_i} \rho_{m,L}(\theta_i). \tag{42}$$

Here, θ_i is the ith parameterized DOA defined in Section II-A, and $\rho_{m,L}(\theta_i)$ is the corresponding array factor constructed using the selection matrices $\boldsymbol{J}_x(\theta)$ and $\boldsymbol{J}_y(\theta)$. In this section, we present the analysis where[4] $m = 2$ and $L = 1$, in which case, the column span of the data matrix is given by

$$\boldsymbol{A}_m = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{A}\boldsymbol{\Phi} \end{bmatrix}. \tag{43}$$

Here, $\boldsymbol{A}$ represents the array steering matrix. For a ULA and $(m, L) = (2, 1)$, the selection matrices $\boldsymbol{J}_x(\theta)$ and $\boldsymbol{J}_y(\theta)$ select the first and, respectively, the last $M-1$ rows from each of the two block entries of $\boldsymbol{A}_m$. Thus

$$\boldsymbol{J}_y(\theta) - \theta_i\boldsymbol{J}_x(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \boldsymbol{\Upsilon}(\theta_i)$$

[4]Note that for JAFE, the minimum possible value of m is 2, and that of L is 1. Behaviors corresponding to larger m and L values are considered in Section IV.

where

$$\boldsymbol{\Upsilon}(\theta_i) = \begin{bmatrix} -\theta_i & 1 & & & \\ & -\theta_i & 1 & & \\ & & \ddots & \ddots & \\ & & & -\theta_i & 1 \end{bmatrix}. \tag{44}$$

Let $\boldsymbol{A}_x$ denote the first $M-1$ rows of $\boldsymbol{A}$ and $\boldsymbol{A}_{m,x}$ be defined as

$$\boldsymbol{A}_{m,x} = \begin{bmatrix} \boldsymbol{A}_x \\ \boldsymbol{A}_x\boldsymbol{\Phi} \end{bmatrix}.$$

It follows that $\boldsymbol{W}_x = (1/\sqrt{2})[\boldsymbol{A}_x^{\dagger}\boldsymbol{\Phi}^{-1}\boldsymbol{A}_x^{\dagger}]$ is a left inverse of $\boldsymbol{A}_{m,x}$, and if we let $\boldsymbol{w}_i^H = [w_{i,1}\cdots w_{i,M-1}]$ represent the ith row of $\boldsymbol{A}_x^{\dagger}$, then the ith row of $\boldsymbol{W}_x$ is given by $\boldsymbol{w}_{x,i}^H = (1/\sqrt{2})[\boldsymbol{w}_i^H \;\; \phi_i^{-1} \;\; \boldsymbol{w}_i^H]$. Thus, using (37) and (40), it follows that

$$\rho_{2,1}(\theta_i) = \tfrac{1}{2}\left(|w_{i,1}|^2 + |w_{i,M-1}|^2 + \sum_{k=2}^{M-1} |w_{i,k-1} - \theta_i w_{i,k}|^2\right). \tag{45}$$

For a single source, $w_{1k} = (1/(M-1))\theta_1^{-k}$, and thus, $\rho_{2,1}(\theta_1) = 1/(M-1)^2$. Moreover, $\widehat{\text{SNR}}_1 = \eta_1/\sigma_n^2 = \text{SNR}_1$. Combining these and using (42), we get the following approximation for the estimation error:

$$\sigma_{2,1}(\theta_1) \approx \frac{1}{(M-1)^2(N-1)}\,\frac{1}{\text{SNR}_1}. \tag{46}$$

If we have started by setting $m = 1$, we get the expression

$$\sigma_{2,1}(\theta_1) \approx \frac{2}{(M-1)^2 N}\,\frac{1}{\text{SNR}_1} \tag{47}$$

which agrees with those described in [27] and [28] for a single source scenario. It is seen that the DOA estimation error is proportional to the inverse of the square of the number of antennas. This means that the algorithm, for large M, fails to achieve the Cramér–Rao Lower Bound (CRB)[5] (i.e., it is inefficient). However, in Section VI, it will be shown that by choosing an appropriate value for L, the DOA estimation error can be made to decay in proportion to M^{-3}.

B. Parameterized Frequency Estimation

For the frequency estimation, with reference to the discussions in Section IV-B, the parameterized frequency estimation error $\sigma_{m,L}(\phi_i)$ is obtained from (41) by replacing λ_i with ϕ_i and the geometric factor $\rho_{m,L}(\lambda_i)$ with $\rho_{m,L}(\phi_i)$:

$$\sigma_{m,L}(\phi_i) \approx \frac{1}{L(N-m+1)}\,\frac{\sigma_n^2}{\eta_i}\,\rho_{m,L}(\phi_i) \tag{48}$$

where $\rho_{m,L}(\phi_i)$ is defined using the selection matrices $\boldsymbol{J}_x(\phi)$ and $\boldsymbol{J}_y(\phi)$. For a ULA, these selection matrices are given in (7). In the following, we give a performance analysis for the case $m = 2$ and $L = 1$. The effects of other (m, L) values is considered in Section VI. For $(m, L) = (2, 1)$, the column span of the data matrix is given by (43). For the parameterized frequency estimation, the selection matrix $\boldsymbol{J}_x(\phi)$ selects the first M rows of $\boldsymbol{A}_m$ (which is $\boldsymbol{A}$) and $\boldsymbol{J}_y(\phi)$ its last M rows. Thus

$$\boldsymbol{J}_y(\phi) - \phi_i\boldsymbol{J}_x(\phi) = [-\phi_i\boldsymbol{I} \quad \boldsymbol{I}].$$

Let $\boldsymbol{w}_i^H$ be the ith row of $\boldsymbol{A}^{\dagger}$; then, using (37), it follows that $\boldsymbol{r}_{2,1}^H(\phi_i) = [-\phi_i\boldsymbol{w}_i^H \;\; \boldsymbol{w}_i^H]$,[6] and hence

$$\rho_{2,1}(\phi_i) = \boldsymbol{r}_{2,1}^H(\phi_i)\boldsymbol{r}_{2,1}(\phi_i) = 2|\boldsymbol{w}_i|^2$$

and

$$\sigma_{2,1}(\phi_i) \approx \frac{2|\boldsymbol{w}_i|^2}{N-1}\,\frac{\sigma_n^2}{\eta_i}.$$

For a single source, $|\boldsymbol{w}_1|^2 = 1/M$, and $\widehat{\text{SNR}}_1 = \eta_1/\sigma_n^2 \approx \text{SNR}_1$; thus

$$\sigma_{2,1}(\phi_1) \approx \frac{2}{M(N-1)}\,\frac{1}{\text{SNR}_1}.$$

It is seen that the estimation error, for $m = 2$, decays only in linear proportion with N. This is, of course, an extremely poor result. However, it is important to note that for $m = 2$, the effective number of temporal samples used in actual phase computation is 2. This means that by choosing larger m values, the performance can be improved significantly. In fact, in the following section, it will be shown that by choosing an appropriate value for m, the frequency estimation error can be made to decay in proportion to N^{-3}.

VI. Effects of Data Extensions

The performance analysis outlined in the above section considers the data model of (14), with a spatial smoothing factor $L = 1$ and a temporal smoothing factor $m = 2$. In this section, we give analysis of how the data extension procedures affect the estimation performances. Moreover, we derive the optimum values of L and m (L_o, m_o) that minimize the angle, frequency, and joint estimation errors.

Consider an antenna array with an arbitrary geometry. Let, for an (m, L) spatio-temporally smoothed data matrix, the pairs $(\boldsymbol{J}_x(\phi), \boldsymbol{J}_y(\phi))$ and $(\boldsymbol{J}_x(\theta), \boldsymbol{J}_y(\theta))$ be the selection matrices that produce the shift invariance pairs for the parameterized frequency and DOA estimations, respectively; then

$$\begin{cases} \boldsymbol{J}_x(\phi) := [\boldsymbol{I}_{m-1} \;\; \boldsymbol{0}_1] \otimes \boldsymbol{I}_{M_L} \\ \boldsymbol{J}_y(\phi) := [\boldsymbol{0}_1 \;\; \boldsymbol{I}_{m-1}] \otimes \boldsymbol{I}_{M_L} \end{cases} \tag{49}$$

$$\begin{cases} \boldsymbol{J}_x(\theta) := \boldsymbol{I}_m \otimes \boldsymbol{J}_{1,x}(\theta) \\ \boldsymbol{J}_x(\theta) := \boldsymbol{I}_m \otimes \boldsymbol{J}_{1,y}(\theta). \end{cases} \tag{50}$$

where $\boldsymbol{J}_{1,x}(\theta)$ and $\boldsymbol{J}_{1,y}(\theta)$ are the selection matrices for the case $(m, L) = (1, 1)$ and are array geometry dependent. For a ULA, they are given by

$$\boldsymbol{J}_{1,x}(\theta) = [\boldsymbol{I}_{M-1} \;\; \boldsymbol{0}_1]$$
$$\boldsymbol{J}_{1,y}(\theta) = [\boldsymbol{0}_1 \;\; \boldsymbol{I}_{M-1}].$$

[5] In CRB, the angle estimation error is proportional to M^{-3}

[6] Note that in contrast to $\boldsymbol{J}_x(\theta)$ and $\boldsymbol{J}_y(\theta)$, the structures of the selection matrices $\boldsymbol{J}_x(\phi)$ and $\boldsymbol{J}_y(\phi)$ are independent of array geometry, and thus, this result applies to an arbitrary array geometry.

Now, let $\boldsymbol{U}_s$ be a unitary matrix that spans the column space of the extended data matrix; then, for some invertible matrix $\boldsymbol{T}$, the matrices $\boldsymbol{A}_\phi := \boldsymbol{J}_x(\phi)\boldsymbol{U}_s\boldsymbol{T}^{-1}$ and $\boldsymbol{A}_\theta := \boldsymbol{J}_x(\theta)\boldsymbol{U}_s\boldsymbol{T}^{-1}$ are given by

$$\boldsymbol{A}_\phi = \begin{bmatrix} \boldsymbol{A}(L) \\ \boldsymbol{A}(L)\boldsymbol{\Phi} \\ \vdots \\ \boldsymbol{A}(L)\boldsymbol{\Phi}^{m-2} \end{bmatrix} \quad \text{and} \quad \boldsymbol{A}_\theta = \begin{bmatrix} \boldsymbol{A}'(L) \\ \boldsymbol{A}'(L)\boldsymbol{\Phi} \\ \vdots \\ \boldsymbol{A}'(L)\boldsymbol{\Phi}^{m-1} \end{bmatrix}$$

respectively, where $\boldsymbol{A}(L)$ is the array steering matrix corresponding to the L-factor spatially smoothed data matrix, and $\boldsymbol{A}'(L)$ is a submatrix of $\boldsymbol{A}(L)$, whose row dimension depends on how the spatial smoothing is performed.

Let $\boldsymbol{w}_{\phi_i}^H$ and $\boldsymbol{w}_{\theta_i}^H$ be the ith rows of $\boldsymbol{W}_\phi = \boldsymbol{A}_\phi^\dagger$ and $\boldsymbol{W}_\theta = \boldsymbol{A}_\theta^\dagger$, respectively, and let the bi-diagonal matrix $\boldsymbol{\Upsilon}(\phi_i)$ be defined as

$$\boldsymbol{\Upsilon}(\phi_i) = \begin{bmatrix} -\phi_i & 1 & & & \\ & -\phi_i & 1 & & \\ & & \ddots & \ddots & \\ & & & -\phi_i & 1 \end{bmatrix}. \tag{51}$$

Then, using the selection matrices defined in (49) and (50), it can be shown that

$$\begin{aligned} \boldsymbol{r}_{m,L}^H(\theta_i) &= \boldsymbol{w}_{\theta_i}^H [\boldsymbol{I}_m \otimes \boldsymbol{\Upsilon}(\theta_i)] \\ \boldsymbol{r}_{m,L}^H(\phi_i) &= \boldsymbol{w}_{\phi_i}^H [\boldsymbol{\Upsilon}(\phi_i) \otimes \boldsymbol{I}_{M_L}]. \end{aligned} \tag{52}$$

Note that although $\boldsymbol{\Upsilon}(\theta_i)$ is array geometry-dependent, $\boldsymbol{\Upsilon}(\phi_i)$ is independent of geometry. It is seen that for ULA, these matrices have the same structure [compare (51) with (44)]. Now, let $\boldsymbol{W}(L) = \boldsymbol{A}^\dagger(L)$ and $\boldsymbol{W}'(L) = \boldsymbol{A}'^\dagger(L)$; then, it is seen that the two matrices

$$\boldsymbol{W}_m = \frac{1}{m-1}[\boldsymbol{W}(L) \quad \boldsymbol{\Phi}^{-1}\boldsymbol{W}(L) \quad \cdots \quad \boldsymbol{\Phi}^{2-m}\boldsymbol{W}(L)]$$

and

$$\boldsymbol{W}'_m = \frac{1}{m}[\boldsymbol{W}'(L) \quad \boldsymbol{\Phi}^{-1}\boldsymbol{W}'(L) \quad \cdots \quad \boldsymbol{\Phi}^{1-m}\boldsymbol{W}'(L)]$$

are left inverses of $\boldsymbol{A}_\phi$ and $\boldsymbol{A}_\theta$, respectively, i.e.,

$$\boldsymbol{W}_m\boldsymbol{A}_\phi = \boldsymbol{I} \quad \text{and} \quad \boldsymbol{W}'_m\boldsymbol{A}_\theta = \boldsymbol{I}.$$

It is well known, however, that the Moore–Penrose inverse (the pseudo-inverse) gives unique left inverses of these matrices with minimum (Frobenius) norms. Thus, with $\boldsymbol{W}_\phi = \boldsymbol{A}_\phi^\dagger$ and $\boldsymbol{W}_\theta = \boldsymbol{A}_\theta^\dagger$

$$\|\boldsymbol{W}_\phi\|^2 \le \|\boldsymbol{W}_m\|^2 = \frac{\|\boldsymbol{W}(L)\|^2}{m-1}$$

and

$$\|\boldsymbol{W}_\theta\|^2 \le \|\boldsymbol{W}'_m\|^2 = \frac{\|\boldsymbol{W}'(L)\|^2}{m}.$$

If the underlying sources are sufficiently separated, $\boldsymbol{W}_\phi \approx \boldsymbol{W}_m$ and $\boldsymbol{W}_\theta \approx \boldsymbol{W}'_m$, and the above inequalities are tight. When the sources are close to each other, on the other hand, while the original array steering matrix (the steering matrix with $m = 1$) is near singular, the extended steering matrix may be well conditioned. This means that for closely separated sources, the above inequalities may become loose, and the given approximations might be too pessimistic. As a rule of thumb, in the case of a ULA for instance, we say two sources with DOAs α_1 and α_2 are sufficiently separated if

$$|\sin\alpha_1 - \sin\alpha_2| \ge \frac{1}{M\Delta}$$

where Δ is antenna spacing measured in fractions of signal wavelength. Note that when this condition is satisfied, the peaks of the FFT of the columns of the array steering matrix $\boldsymbol{A}$ are at least $2\pi/M$ rad apart, which is equal to the BW of the M-point FFT bins. Thus, this states that if the peaks of the FFT of $\boldsymbol{A}$ are separated at least by an amount equal to the BW of the bins of the M-point FFT, then the sources are said to be sufficiently separated. Considering the fact that the FFT matrix is a unitary matrix, this is a justifiable assumption. In the following, we will assume that the sources under consideration are sufficiently separated and that the above inequalities are tight. Under this condition, we may write

$$\begin{aligned} \boldsymbol{w}_{\phi_i}^H &\approx \frac{1}{m-1}[\boldsymbol{w}_i^H(L) \quad \phi_i^{-1}\boldsymbol{w}_i^H(L) \quad \cdots \quad \phi_i^{2-m}\boldsymbol{w}_i^H(L)] \\ \boldsymbol{w}_{\theta_i}^H &\approx \frac{1}{m}[\boldsymbol{w}_i'^H(L) \quad \phi_i^{-1}\boldsymbol{w}_i'^H(L) \quad \cdots \quad \phi_i^{1-m}\boldsymbol{w}_i'^H(L)] \end{aligned}$$

where $\boldsymbol{w}_i^H(L)$ and $\boldsymbol{w}_i'^H(L)$ are the ith rows of $\boldsymbol{W}(L)$ and $\boldsymbol{W}'(L)$, and $\boldsymbol{w}_{\phi_i}^H$ and $\boldsymbol{w}_{\theta_i}^H$ are the ith rows of $\boldsymbol{W}_\phi$ and $\boldsymbol{W}_\theta$, respectively. Putting these into (37) and using (40), it follows that

$$\begin{aligned} \rho_{2,L}(\theta_i) &= \|\boldsymbol{w}_i'^H(L)\boldsymbol{\Upsilon}(\theta_i)\|^2 \\ \rho_{2,L}(\phi_i) &= 2\|\boldsymbol{w}_i(L)\|^2 \end{aligned}$$

and

$$\begin{aligned} \rho_{m,L}(\theta_i) &= \frac{2}{m}\rho_{2,L}(\theta_i) \\ \rho_{m,L}(\phi_i) &\approx \frac{1}{(m-1)^2}\rho_{2,L}(\phi_i). \end{aligned} \tag{53}$$

The expressions for the parameterized angle and frequency estimation errors are obtained by replacing these into (42) and (48), respectively. Thus, with

$$\begin{cases} G_{\theta_i}(m) = \dfrac{2}{m(N-m+1)} \\ G_{\phi_i}(m) = \dfrac{1}{(m-1)^2(N-m+1)} \end{cases} \tag{54}$$

and

$$\begin{cases} G_{\theta_i}(L) = \dfrac{1}{L}\rho_{2,L}(\theta_i) \\ G_{\phi_i}(L) = \dfrac{1}{L}\rho_{2,L}(\phi_i) \end{cases} \tag{55}$$

it follows that

$$\begin{aligned} E_{m,L}\{|\Delta\theta_i|^2\} &\approx \frac{\sigma_n^2}{\eta}G_{\theta_i}(m)G_{\theta_i}(L) \\ E_{m,L}\{|\Delta\phi_i|^2\} &\approx \frac{\sigma_n^2}{\eta}G_{\phi_i}(m)G_{\phi_i}(L). \end{aligned} \tag{56}$$

The above relations show that the estimation errors are separable functions in m and L. This means that the optimum values of the spatio-temporal smoothing factors (m_o, L_o) are independent of each other. In the following, we will compute these values, considering each separately.

A. Optimum Temporal Smoothing Factor

From (56), it is seen that the optimum values of m for the parameterized DOA and frequency estimations are obtained by minimizing $G_\theta(m)$ and $G_\phi(m)$, respectively.[7] Denoting by $m_o(\theta)$ and $m_o(\phi)$ the optimum temporal smoothing factors for the parameterized DOA and frequency estimations, respectively, we have

$$m_o(\theta) = \frac{N}{2}, \qquad \text{for } \theta, \text{ and}$$
$$m_o(\phi) = \frac{2N+1}{3}, \quad \text{for } \phi.$$

The above value of $m_o(\phi)$ agrees with a similar result reported in [29] for the harmonic retrieval problem. The corresponding variances of the parameterized DOA and frequency estimation are then

$$E_{m_o,\,L}\{|\Delta\theta|^2\} \approx \frac{\sigma_n^2}{\eta}\,\frac{4}{N^2}\,G_\theta(L)$$
$$E_{m_o,\,L}\{|\Delta\phi|^2\} \approx \frac{\sigma_n^2}{\eta}\,\frac{27}{4(N-1)^3}\,G_\phi(L).$$

For JAFE, it makes more sense to look for an optimum m that minimizes the joint estimation error. To this end, we define a joint estimation error as the geometric mean of the variances of the angle and frequency estimation errors:

$$E_{m,\,L}(\theta,\,\phi) \triangleq \sqrt{E_{m,\,L}\{|\Delta\theta|^2\}E_{m,\,L}\{|\Delta\phi|^2\}} \qquad (57)$$

which is equivalent to the arithmetic mean on the logarithmic scale. This definition has twofold advantages: First, it alleviates the scaling problems associated with the arithmetic mean, and second, it preserves the separability of the error function in the variables m and L. Minimizing $E_{m,\,L}(\theta,\,\phi)$ with respect to m and after some elaboration (approximation), we obtain

$$m_o \approx \frac{3N+2}{5}.$$

It is seen that m_o for the joint estimation is approximately equal to the average of its values obtained considering the angle and frequency estimations separately.

B. Optimum Spatial Smoothing Factor

Referring to (56), it follows that minimizing the estimation errors with respect to L is equivalent to minimizing $G_\theta(L)$ and $G_\phi(L)$. To this end, first, we need to derive explicit expressions for these functions. Let M_L and M'_L be the number of rows in $\boldsymbol{A}(L)$ and $\boldsymbol{A}'(L)$, respectively; then, for sufficiently independent sources, the following approximations are valid:

$$G_\theta(L) \approx \frac{1}{LM'^2_L} \quad \text{and} \quad G_\phi(L) \approx \frac{1}{LM_L}.$$

For a single source, these relations are exact. For more sources, however, the above approximation is valid only if the steering vectors are sufficiently independent. This is always (asymptotically) satisfied for large M_L and M'_L values.

From their definitions, it is clear that the values of M_L and M'_L depend on the way the spatial smoothing is performed. For a ULA, for instance, if we assume a maximum overlap of subarrays as described in Section III-A, $M_L = M - L + 1$, and $M'_L = M - L$. Putting these into the above expressions and minimizing $G_\theta(L)$ and $G_\phi(L)$ with respect to L, we obtain

$$L_o = \begin{cases} \dfrac{M}{3}, & \text{for } \theta, \text{ and} \\ \dfrac{M+1}{2}, & \text{for } \phi. \end{cases}$$

The corresponding variances of the parameterized DOA and frequency estimation are then

$$E_{m,\,L_o}\{|\Delta\theta|^2\} \approx \frac{\sigma_n^2}{\eta}\,\frac{27}{4M^3}\,G_\theta(m)$$
$$E_{m,\,L_o}\{|\Delta\phi|^2\} \approx \frac{\sigma_n^2}{\eta}\,\frac{4}{M^2-1}\,G_\phi(m).$$

For JAFE, defining a joint estimation error as in (57), we obtain

$$L_o \approx \frac{2M}{5}.$$

As in the case of m_o, this value of L_o is approximately equal to the average of the optimum values obtained considering the angle and frequency estimations separately.

C. Forward-Backward Averaging

Forward-backward averaging [5], [18] is equivalent to doubling the number of temporal samples, with the rest of the data parameters remaining unchanged. Thus, for both θ and ϕ, this data extension provides a factor-of-2 improvement in the estimation accuracies. The important aspect of forward-backward averaging is that the resulting data matrix can be transformed into a real matrix of the same size [16]. This provides a substantial reduction in complexity [18] while improving the accuracy.

VII. Errors in Computing the Actual Parameters

The final step in the JAFE algorithm is to transform the parameterized DOAs and the parameterized frequencies into their actual values in radians and hertz, respectively. In this section, we consider how these transformations affect the error behaviors.

[7] To simplify notations, in the remainder of this section, we omit the signal number index i, For instance, we write θ instead of θ_i.

A. Errors in Computing the DOAs

The relation between the parameterized and the physical DOAs is defined by the array geometry. In the case of the ULA, for instance, this relation is given by

$$\alpha = \operatorname{asin}\left(\frac{\arg\theta}{2\pi\Delta}\right).$$

Assuming the real and imaginary parts of θ are affected by statistically independent equal variance noises, the mean square errors on the angle estimates can be computed as (first-order approximations) (*viz*. [27], [30]–[33])

$$\sigma_{m,L}(\alpha) := E\{\Delta\alpha^2\}_{m,L} = \frac{\sigma_{m,L}(\theta)}{2(2\pi\Delta\cos\alpha)^2}. \quad (58)$$

The final expressions for $\sigma_{m,L}(\alpha)$ are obtained by replacing $\sigma_{m,L}(\theta)$ with the expressions from (42).

B. Errors in Computing the Frequencies

As recalled from Section II-A, the actual signal frequency is computed from the parameterized frequency ϕ using the transformation

$$f = \frac{P}{2\pi}\arg\phi.$$

The first-order approximation of the perturbation Δf on f is

$$\Delta f \approx \operatorname{Re}\left(\frac{\partial f}{\partial\phi}\Delta\phi\right) = \frac{P}{2\pi}\operatorname{Re}(\Delta\phi),$$

and thus, the variance of the frequency estimation error is

$$\sigma_{m,L}(f) = \frac{1}{2}\frac{P^2}{4\pi^2}\sigma_{m,L}(\phi). \quad (59)$$

The final expression for $\sigma_{m,L}(f)$ is obtained by replacing $\sigma_{m,L}(\phi)$ with the expression from (48).

VIII. Cramér–Rao Lower Bound

Putting a lower bound for any estimator proves to be extremely useful. It provides a benchmark against which we can compare the performance of any unbiased estimator. Moreover, it tells us the impossibility of finding an unbiased estimator whose variance is less than the bound. One such bound is the CRB [34]. In this section, we derive the CRB for the JAFE algorithm.

Let us assume that an $M \times 1$ deterministic signal vector $\boldsymbol{s}(k;\ \boldsymbol{\mu})$ with unknown parameter vector

$$\boldsymbol{\mu} = [\,\mu_1 \quad \mu_2 \quad \cdots \quad \mu_q\,]$$

is observed in additive noise

$$\boldsymbol{x}(k) = \boldsymbol{s}(k;\ \boldsymbol{\mu}) + \boldsymbol{\eta}(k) = \begin{bmatrix} s_1(k;\ \boldsymbol{\mu}) \\ s_2(k;\ \boldsymbol{\mu}) \\ \cdots \\ s_M(k;\boldsymbol{\mu}) \end{bmatrix} + \boldsymbol{\eta}(k) \in \mathbb{C}^M \quad (60)$$

where $\boldsymbol{\eta}(k)$ is an $M \times 1$ noise vector. Here, the dependence of the signal $\boldsymbol{s}(k)$ on $\boldsymbol{\mu}$ is made explicit by writing $\boldsymbol{s}(k;\boldsymbol{\mu})$. Assume that $\boldsymbol{\eta}(k)$ is a white Gaussian noise (WGN) with variance σ^2 and that we have collected N time samples of the signal $\boldsymbol{x}(k)$. Then, the log likelihood function of the signal (defined as the logarithm of the probability density function) is given by

$$\mathcal{L}(\boldsymbol{x};\ \boldsymbol{\mu}) = -\frac{MN}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}[\boldsymbol{x}(k) - \boldsymbol{s}(k;\ \boldsymbol{\mu})]^H[\boldsymbol{x}(k) - \boldsymbol{s}(k;\ \boldsymbol{\mu})]. \quad (61)$$

Let the gradient of the signal vector $\boldsymbol{s}(k;\boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$ be denoted by $\boldsymbol{D}_k(\boldsymbol{\mu})$, i.e.,

$$\boldsymbol{D}_k(\boldsymbol{\mu}) = \left[\begin{array}{cccc} \dfrac{\partial\boldsymbol{s}(k;\ \boldsymbol{\mu})}{\partial\mu_1} & \dfrac{\partial\boldsymbol{s}(k;\ \boldsymbol{\mu})}{\partial\mu_2} & \cdots & \dfrac{\partial\boldsymbol{s}(k;\ \boldsymbol{\mu})}{\partial\mu_q} \end{array}\right]$$

then the so-called Fisher information matrix is given by

$$\boldsymbol{I}(\boldsymbol{\mu}) = \frac{1}{\sigma^2}\operatorname{Re}\left(\sum_{k=1}^{N}\boldsymbol{D}_k^H(\boldsymbol{\mu})\boldsymbol{D}_k(\boldsymbol{\mu})\right). \quad (62)$$

The CRB for estimating the ith parameter μ_i is obtained from the inverse of the Fisher information matrix (*viz*. [34]) as

$$\operatorname{CRB}(\mu_i) = [\boldsymbol{I}^{-1}(\boldsymbol{\mu})]_{ii}.$$

A. CRB for the JAFE Data Model

Consider a simplified version of the JAFE data model, in which the modulating signals are set to have $BW = 0$

$$\boldsymbol{x}(k) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{B}\boldsymbol{\phi}^k + \boldsymbol{\eta}(k) =: \boldsymbol{s}(k;\ \boldsymbol{\mu}) + \boldsymbol{\eta}(k) \quad (63)$$

where we have the following.

- $\boldsymbol{A}(\boldsymbol{\theta}) = [\,\boldsymbol{a}(\theta_1) \quad \boldsymbol{a}(\theta_2) \quad \boldsymbol{a}(\theta_d)\,] \in \mathbb{C}^{M,d}$, and $\boldsymbol{a}(\theta)$ is the array response vector for a signal with a parameterized DOA of θ.
- $\boldsymbol{\theta} = [\theta_1 \cdots \theta_d]^T$ is a vector containing the parameterized DOA of the d signals.
- $\boldsymbol{\phi}^k = [\,\phi_1^k \quad \phi_2^k \quad \cdots \quad \phi_d^k\,]^T$, where ϕ_i is the parameterized frequency of the ith signal.
- $\boldsymbol{B} = \operatorname{diag}\{\beta_i\}_{i=1}^d$ is a diagonal gain matrix, where $\beta_i \in \mathbb{R}^+$ is the amplitude of the ith signal as received by the antenna array.
- $\boldsymbol{\eta}(k)$ is an $M \times 1$ white Gaussian noise vector.

Let $\boldsymbol{\beta} = [\beta_1 \ \cdots \ \beta_d]^T$ be a vector containing the channel gains; then, the conditioning parameters that affect the signal likelihood function are collected into the $3d \times 1$ vector $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = [\,\boldsymbol{\theta}^T \quad \boldsymbol{\phi}^T \quad \boldsymbol{\beta}^T\,].$$

Define

$$\boldsymbol{D}_\theta = \left[\begin{array}{ccc} \dfrac{\partial\boldsymbol{a}}{\partial\theta_1}(\theta_1) & \cdots & \dfrac{\partial\boldsymbol{a}}{\partial\theta_d}(\theta_d) \end{array}\right]$$
$$\boldsymbol{\Phi}^k = \operatorname{diag}\{\boldsymbol{\phi}^k\}.$$

Evaluating the derivative of $\boldsymbol{s}(k;\ \boldsymbol{\mu})$ with respect to each parameter, we get the following:

$$\frac{\partial \boldsymbol{s}(k;\ \boldsymbol{\mu})}{\partial \boldsymbol{\theta}} = \boldsymbol{D}_\theta \boldsymbol{B}\boldsymbol{\Phi}^k =: \boldsymbol{D}_k(\boldsymbol{\theta}),$$
$$\frac{\partial \boldsymbol{s}(k;\ \boldsymbol{\mu})}{\partial \boldsymbol{\phi}} = k\boldsymbol{A}\boldsymbol{B}\boldsymbol{\Phi}^{k-1} =: \boldsymbol{D}_k(\boldsymbol{\phi}),$$
$$\frac{\partial \boldsymbol{s}(k;\ \boldsymbol{\mu})}{\partial \boldsymbol{\beta}} = \boldsymbol{A}\boldsymbol{\Phi}^k =: \boldsymbol{D}_k(\boldsymbol{\beta}).$$

Let $\boldsymbol{I}_k(\boldsymbol{\mu})$ be defined as

$$\boldsymbol{I}_k(\boldsymbol{\mu}) = \frac{1}{\sigma^2}\,\mathrm{Re}\begin{bmatrix} \boldsymbol{D}_k(\boldsymbol{\theta})^H \\ \boldsymbol{D}_k(\boldsymbol{\phi})^H \\ \boldsymbol{D}_k(\boldsymbol{\beta})^H \end{bmatrix}\begin{bmatrix} \boldsymbol{D}_k(\boldsymbol{\theta})^H \\ \boldsymbol{D}_k(\boldsymbol{\phi})^H \\ \boldsymbol{D}_k(\boldsymbol{\beta})^H \end{bmatrix}^H. \tag{64}$$

Then, the Fisher information matrix is

$$\boldsymbol{I}(\boldsymbol{\mu}) = \sum_{k=1}^{N} \boldsymbol{I}_k(\boldsymbol{\mu}) = \frac{1}{\sigma^2}\,\mathrm{Re}\begin{bmatrix} \boldsymbol{\Delta} & \boldsymbol{P}^H & \boldsymbol{Q}^H \\ \boldsymbol{P} & \boldsymbol{\Gamma} & \boldsymbol{R}^H \\ \boldsymbol{Q} & \boldsymbol{R} & \boldsymbol{\Lambda} \end{bmatrix} \tag{65}$$

where N is the number of time samples, and

$$\boldsymbol{\Delta} = \sum_{k=1}^{N} \boldsymbol{\Phi}^{-k}\boldsymbol{B}\boldsymbol{D}_\alpha^H \boldsymbol{D}_\alpha \boldsymbol{B}\boldsymbol{\Phi}^k$$
$$\boldsymbol{\Gamma} = \sum_{k=1}^{N} k^2\boldsymbol{\Phi}^{1-k}\boldsymbol{B}\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{B}\boldsymbol{\Phi}^{k-1}$$
$$\boldsymbol{\Lambda} = \sum_{k=1}^{N} \boldsymbol{\Phi}^{-k}\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{\Phi}^k$$
$$\boldsymbol{P} = \sum_{k=1}^{N} k\boldsymbol{\Phi}^{-k}\boldsymbol{B}\boldsymbol{D}_\theta^H\boldsymbol{A}\boldsymbol{B}\boldsymbol{\Phi}^{k-1}$$
$$\boldsymbol{Q} = \sum_{k=1}^{N} \boldsymbol{\Phi}^{-k}\boldsymbol{B}\boldsymbol{D}_\theta^H\boldsymbol{A}\boldsymbol{\Phi}^k$$
$$\boldsymbol{R} = \sum_{k=1}^{N} k\boldsymbol{\Phi}^{1-k}\boldsymbol{B}\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{\Phi}^k.$$

IX. Simulation Examples

In this simulation example, we consider a four-element ULA with baseline separation of $\Delta = 1/2$. We assume that two far field, equal power signals s_1 and s_2 are impinging on the antenna array. The DOA and center frequency of s_1 are $\alpha_1 = 10^\circ$ and $f_1 = 2$ MHz, and those of s_2 are $\alpha_2 = 55^\circ$ and $f_2 = 5$ MHz, respectively. The source signals are narrowband (25 kHz) amplitude-modulated signals. The data is sampled at a rate of 20 MHz, and the processing is done over $N = 32$ time samples. All simulation results are based on 100 Monte Carlo runs. The behaviors are summarized in Figs. 5–8. In Fig. 6 and 8, while keeping the rest of the parameters fixed at their original values, the DOA and center frequency of s_2 are varied to generate behaviors as functions of angular and frequency separations, respectively.

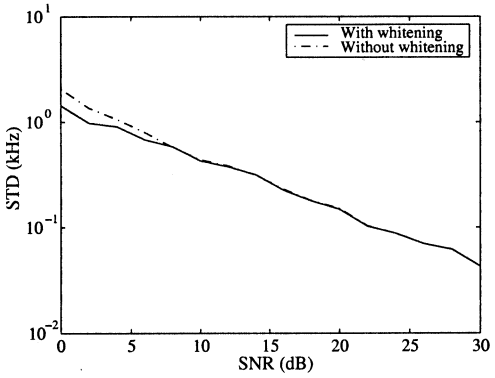


Fig. 3. Effect of whitening on the frequency estimation error.

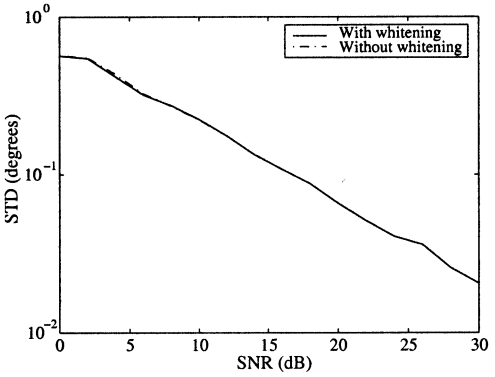


Fig. 4. Effect of whitening on the DOA estimation error.

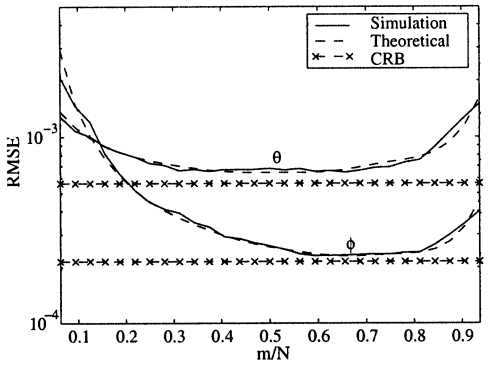


Fig. 5. Parameterized DOA and frequency estimation errors as functions of temporal smoothing factor m at SNR $= 30$ dB and $L = 2$. It is seen that the theoretical behaviors perfectly agree with the simulation results.

First, results comparing the performances of the JAFE algorithm implemented with and without a pre-whitening filter (*viz.* Section III-E) are shown in Figs. 3 and 4 (For clarity, only behaviors corresponding to the first source are shown. Similar situations are observed for the second source as well.). In the simulation, the temporal and spatial smoothing factors are chosen to be $m = 3$ and $L = 2$, respectively. From the results, one sees that the whitening has very minor effect on the DOA estimation error. On the other hand, an appreciable performance improvement is observed in the frequency estimation, particularly at the low SNR region.

Fig. 5 shows how temporal smoothing affects the parameter estimation errors. From the plots, it is seen that the DOA estimation error is minimum for $m = N/2$ and that the fre-

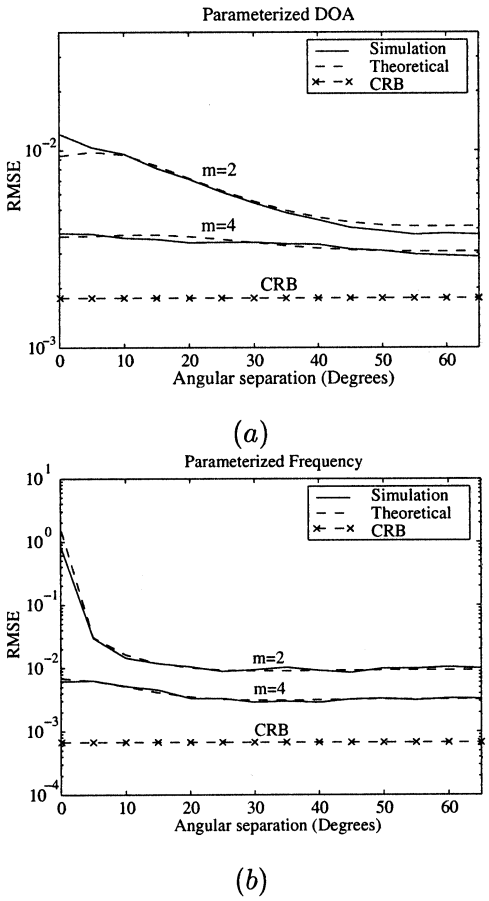


Fig. 6. Behavior of (a) the parameterized DOA and (b) the parameterized frequency estimation errors as functions of angular separation. Note the improvement obtained via temporal smoothing, particularly at small angular separations. (SNR $= 20$ dB).

quency estimation error is minimum for $m = 2N/3$, as predicted in Section VI. It is seen that when[8] $m/N \to 2/N$, the frequency/DOA estimation errors increase sharply and that for the algorithm to attain the CRB, m should be close to its optimum value. Choosing large m, however, increases the computational complexity. This means that one has to find a compromise between complexity and accuracy. In Fig. 6, it is shown that, apart from improving the estimation accuracy, temporal smoothing also provides robustness against rank loss when there exist multiple signals with the same DOA. This is in agreement with the identifiability conditions discussed in Section III-D.

The effect of spatial smoothing on the estimation errors is summarized in Fig. 7. The simulation was run using a ULA with $M = 16$ elements, $N = 16$, $m = m_o = 10$, and SNR $= 20$ dB. The DOAs and the center frequencies of the two sources under consideration are the same as before. As predicted in Section VI, the parameterized DOA and frequency estimation errors are minimum for $L = M/3$ and $L = (M+1)/2$, respectively. Moreover, in Fig. 8 it is seen that, apart from performance improvement, spatial smoothing achieves rank restoration when several signals have the same center frequencies. In Fig. 8(a), a seemingly unexpected behavior is seen. That is, for large frequency separation, the DOA estimation error increased when L was changed from 1 to 2. This should not be surprising because for $M = 4$ and $L = 2$, the ratio $(M-1)^2/L(M-L)^2 > 1$ (cf. Section IV), and thus, $E_2 > E_1$, where E_1 and E_2 are estimation errors corresponding to $L = 1$ and $L = 2$, respectively.

As stated above, the JAFE algorithm approaches the CRB only when m and L are close to their optimum values. This is, in most cases, computationally prohibitive. To alleviate this problem, we can use the MR-ESPRIT algorithm described in

[8]Note that the minimum possible value for m is 2.

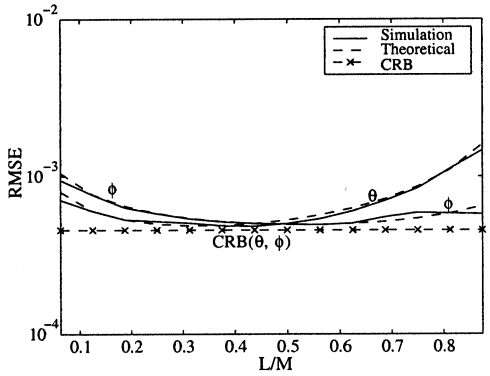


Fig. 7. Parameterized DOA and frequency estimation errors as functions of spatial smoothing factor L at SNR $= 20$ dB and $m = 2$. It is seen that the theoretical behaviors perfectly agree with the simulation results.

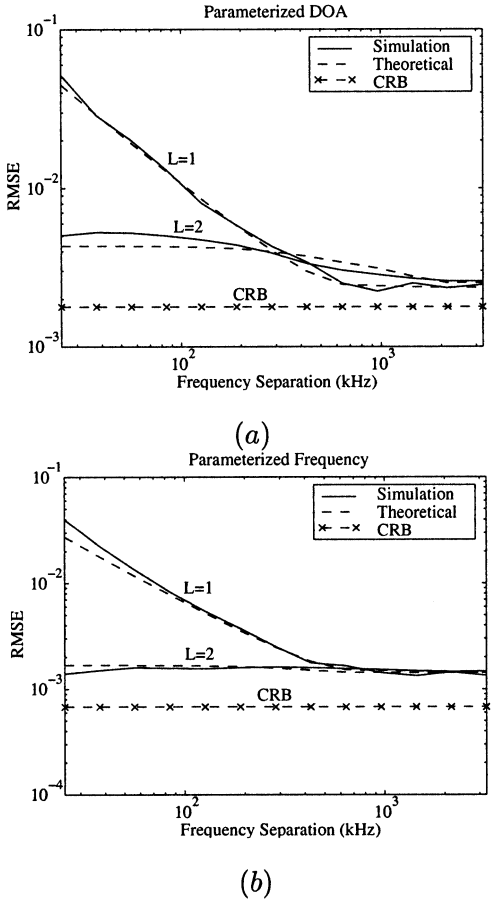


Fig. 8. Behavior of (a) the parameterized DOA and (b) the parameterized frequency estimation errors as functions of frequency separation. Note the superior performance of the spatially smoothed data approach at small frequency separations. (SNR $= 20$ dB.)

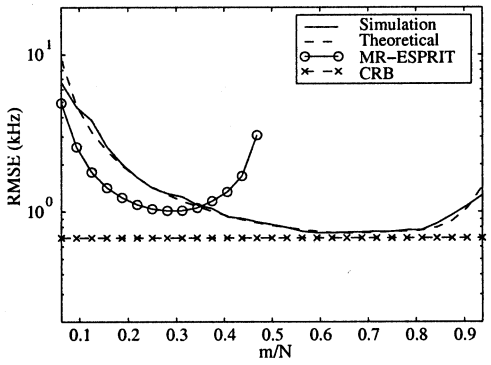


Fig. 9. MR-ESPRIT-based JAFE improves the performance of the frequency estimation in the small m region. (SNR $=$ 30 dB).

Section III-F in the JAFE context. With this approach, it is possible to improve the estimation accuracy significantly in the small m and L regions without raising the computational complexity. For instance, the effect of MR temporal sampling on the frequency estimation is summarized in Fig. 9,[9] where results obtained via the MR-ESPRIT, with a resolution gain factor of 2 are compared against those of the direct estimation method. The superiority of the MR-based approach is obvious, particularly at small m values.

Let k represent the resolution gain factor; then, in [6], [20], and [21], it has been shown that the MR-ESPRIT gives a factor k improvement in accuracy assuming the size of the data matrix is preserved. If the size of the data matrix is reduced as is the case here, on the other hand, the gained accuracy will be less than k. More precisely, if we let $\varepsilon(m)$ and $\varepsilon_{mr}(m)$ represent the variances of the direct and the multiresolution approaches at a temporal smoothing factor of m, respectively, the above two effects can be combined to obtain the relation

$$\varepsilon_{mr}(m) = k\frac{N_s/k - m}{N_s - m}\,\varepsilon(km) \qquad m < N_s/k - d$$

where N_s is the original number of samples. For $N_s/k \gg m$, this simplifies to

$$\varepsilon_{mr}(m) \approx \varepsilon(km).$$

If N_s is comparable to m, on the other hand, the MR-ESPRIT never achieves the best performance obtainable with the direct estimation approach. This is clearly seen in the plots of Fig. 9.

X. Conclusions

In this paper, we have presented an analysis of the ESPRIT-based JAFE algorithm. Using a simple perturbation model, we were able to derive analytical expressions for the estimation errors and for the optimum values of the spatial and temporal smoothing factors.

We have discussed two sets of optimum values for the spatial and temporal smoothing factors: the first set maximizing the number of identifiable sources and the second set minimizing the joint parameter estimation error. It has been seen that these optimum values are different, and thus, one cannot satisfy simultaneous optimality in both identifiability and accuracy.

Moreover, it has been shown that the JAFE algorithm achieves the CRB only when the spatio-temporal smoothing factors are close to their optimum values. However, choosing the smoothing factors close to their optimum values is computationally too expensive. We have shown that in this case, the multiresolution concept can elegantly be used in the JAFE context to solve this problem with an acceptable complexity.

Finally, it should be noted that the performance analysis presented here is independent of the joint diagonalization technique that may have been used in solving the joint matrix pencil problem described in Section II. The analysis assumes that an optimum joint diagonalization of the matrix pencil problem has been achieved. The actual performance is therefore dependent on the quality of the joint diagonalization procedure employed. Although there are several accounts in the literature on this problem, it is an open research topic to find a reliable optimal solution. One way to evaluate the quality of a joint diagonalization method is to compare it against the expected performance derived in this work.

[9]With an MR spatial sampling, a similar behavior is exhibited in DOA estimation.

Appendix
Proofs of Theorems

A. Proof of Theorem II.1

Since all the sources are assumed to have distinct frequencies, $\boldsymbol{F}_s$ in (4) has a full row rank. Thus, for the proof, it is sufficient to show that $\boldsymbol{A}_m$ is full column ranked matrix. To this end, for $i = 1, \ldots, r$, let θ_i represent the DOA of the signals from the ith group, and for $j = 1, \ldots, p_i$, let $\phi_{i,j}$ be the center frequency of the jth signal from the ith group. Moreover, let $\boldsymbol{A}_i$ and $\boldsymbol{\Phi}_i$ be defined as

$$\boldsymbol{A}_i = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \theta_i & \theta_i & \cdots & \theta_i \\ \vdots & \vdots & & \vdots \\ \theta_i^{M-1} & \theta_i^{M-1} & \cdots & \theta_i^{M-1} \end{bmatrix} \in \mathbb{C}^{M \times Pi}$$

and $\boldsymbol{\Phi}_i = \operatorname{diag}\{\phi_{i,j}\}_{j=1}^{p_i}$, then the extended steering matrix $\boldsymbol{A}_m$ can be expressed as

$$\boldsymbol{A}_m = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots & \boldsymbol{A}_r \\ \boldsymbol{A}_1\boldsymbol{\Phi}_1 & \boldsymbol{A}_2\boldsymbol{\Phi}_2 & \cdots & \boldsymbol{A}_r\boldsymbol{\Phi}_r \\ \vdots & \vdots & & \vdots \\ \boldsymbol{A}_1\boldsymbol{\Phi}_1^{m-1} & \boldsymbol{A}_2\boldsymbol{\Phi}_2^{m-1} & \cdots & \boldsymbol{A}_r\boldsymbol{\Phi}_r^{m-1} \end{bmatrix}.$$

To prove the theorem, it is sufficient to show that for a given d-vector $\boldsymbol{x}$

$$\boldsymbol{A}_m\boldsymbol{x} = 0 \Leftrightarrow \boldsymbol{x} = 0.$$

Let $\boldsymbol{x}$ be partitioned into r sub-vectors $\boldsymbol{x}_1, \ldots \boldsymbol{x}_r$ with dimensions such that

$$\boldsymbol{A}_m\boldsymbol{x} = \begin{bmatrix} \boldsymbol{A}_1\boldsymbol{x}_1 + \boldsymbol{A}_2\boldsymbol{x}_2 + \cdots + \boldsymbol{A}_r\boldsymbol{x}_r \\ \boldsymbol{A}_1\boldsymbol{\Phi}_1\boldsymbol{x}_1 + \boldsymbol{A}_2\boldsymbol{\Phi}_2\boldsymbol{x}_2 + \cdots + \boldsymbol{A}_r\boldsymbol{\Phi}_r\boldsymbol{x}_r \\ \vdots \\ \boldsymbol{A}_1\boldsymbol{\Phi}_1^{m-1}\boldsymbol{x}_1 + \cdots + \boldsymbol{A}_r\boldsymbol{\Phi}_r^{m-1}\boldsymbol{x}_r \end{bmatrix} = 0. \quad (66)$$

If $\boldsymbol{A}_m$ is full column rank, the above will be satisfied if and only if $\boldsymbol{x}_1 = 0, \boldsymbol{x}_2 = 0, \ldots, \boldsymbol{x}_r = 0$. For a p-vector $\boldsymbol{v} = [v_1, \ldots, v_p]$, let the function $S(\boldsymbol{v})$ be given by

$$S(\boldsymbol{v}) = \sum_{k=1}^{p} v_i$$

and let $\boldsymbol{A}_\theta$ be defined as

$$\boldsymbol{A}_\theta = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \theta_1 & \theta_2 & \cdots & \theta_r \\ \vdots & \vdots & & \vdots \\ \theta_1^{M-1} & \theta_2^{M-1} & \cdots & \theta_r^{M-1} \end{bmatrix}.$$

Since the θ_i are different and that we have assumed $M \geq d \geq r$, from the Vandermonde structure, it follows that $\boldsymbol{A}_\theta$ has a full column rank of r. Now, using the above definitions, (66) can equivalently be expressed as

$$\boldsymbol{A}_\theta \begin{bmatrix} S(\boldsymbol{x}_1) \\ S(\boldsymbol{x}_2) \\ \vdots \\ S(\boldsymbol{x}_r) \end{bmatrix} = \boldsymbol{A}_\theta \begin{bmatrix} S(\boldsymbol{\Phi}_1 \boldsymbol{x}_1) \\ S(\boldsymbol{\Phi}_2 \boldsymbol{x}_2) \\ \vdots \\ S(\boldsymbol{\Phi}_r \boldsymbol{x}_r) \end{bmatrix} = \cdots$$

$$= \boldsymbol{A}_\theta \begin{bmatrix} S\boldsymbol{\Phi}_1^{m-1}(\boldsymbol{x}_1) \\ S(\boldsymbol{\Phi}_2^{m-1} \boldsymbol{x}_2) \\ \vdots \\ S\boldsymbol{\Phi}_r^{m-1}(\boldsymbol{x}_r) \end{bmatrix} = 0.$$

Because $\boldsymbol{A}_\theta$ is a full rank matrix, it follows that $\boldsymbol{A}_m$ is also full rank if and only if there does not exists an $\boldsymbol{x}_i \neq 0$ such that

$$S(\boldsymbol{x}_i) = S(\boldsymbol{\Phi}_i \boldsymbol{x}_i) = \cdots = S(\boldsymbol{\Phi}_i^{m-1} \boldsymbol{x}_i) = 0 \quad i = 1, \ldots, r. \tag{67}$$

Let

$$\boldsymbol{A}_{\phi_i} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \phi_{i,1} & \phi_{i,2} & \cdots & \phi_{i,p_i} \\ \vdots & \vdots & & \vdots \\ \phi_{i,1}^{m-1} & \phi_{i,2}^{m-1} & \cdots & \phi_{i,p_i}^{m-1} \end{bmatrix}$$

then the conditions in (67) can be combined into the single expression

$$\boldsymbol{A}_{\phi_i} \boldsymbol{x}_i = 0, \qquad i = 1 \cdots, r. \tag{68}$$

Thus, $\boldsymbol{A}_m$ is full rank if and only if there does not exists an $\boldsymbol{x}_i \neq 0$ that satisfies (68), or, in other words, $\boldsymbol{A}_m$ is full rank if and only if $\boldsymbol{A}_{\phi_i}$ is a full rank matrix. Observe that $\boldsymbol{A}_{\phi_i}$ has a Vandermonde structure. Since all the $\phi_{i,j}$ are assumed to be distinct, it will have full rank provided that $m \geq p_i$. From this, it follows that the data matrix $\boldsymbol{X}_m$ is full rank if and only if, for $i = 1, \ldots, r$, all the $\boldsymbol{A}_{\phi_i}$ are all full rank matrices, which is satisfied if

$$m \geq \max_i p_i.$$

This concludes the proof.

B. Proof of Theorem IV-1

Consider the Gaussian, circulant noise-contaminated data model $\boldsymbol{X}_{m,L}$ given in (14). Let $\boldsymbol{Y}_{m,L} = \boldsymbol{X}_{m,L} - \boldsymbol{W}_{m,L}$ be the noise-free data

$$\boldsymbol{R}_{xx} := \frac{1}{L(N-m)} \boldsymbol{X}_{m,L} \boldsymbol{X}_{m,L}^H$$

$$\boldsymbol{R}_{yy} := \frac{1}{L(N-m)} \boldsymbol{Y}_{m,L} \boldsymbol{Y}_{m,L}^H$$

$$\boldsymbol{R}_{ww} := \frac{1}{L(N-m)} \boldsymbol{W}_{m,L} \boldsymbol{W}_{m,L}^H.$$

In the following, we use a perturbation analysis to derive the covariance of the errors on the eigenvectors of $\boldsymbol{R}_{xx}$. In the derivation, we will make the assumption that the noise-free data $\boldsymbol{Y}_{m,L}$ is deterministic, and hence, $E\{\boldsymbol{R}_{yy}\} = \boldsymbol{R}_{y,y}$. Now, let $\overline{\boldsymbol{R}}_{xx} = E\{\boldsymbol{R}_{xx}\}$ and

$$\boldsymbol{V} = \boldsymbol{R}_{xx} - \overline{\boldsymbol{R}}_{xx}.$$

$\boldsymbol{V}$ is a perturbation on $\overline{\boldsymbol{R}}_{xx}$. From its definition, it is seen that $\boldsymbol{V}$ is Hermitian and $E\{\boldsymbol{V}\} = 0$. Let $\overline{\boldsymbol{R}}_{xx} = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^H$ be the EVD of $\overline{\boldsymbol{R}}_{xx}$, $\boldsymbol{u}_v$ be the vth column of $\boldsymbol{U}$ (the vth eigenvector of $\overline{\boldsymbol{R}}_{xx}$), and $\Delta\boldsymbol{u}_v$ represent the perturbation on $\boldsymbol{u}_v$ due to $\boldsymbol{V}$. In [35], it has been sown that for the above model, the first-order approximation of $\Delta\boldsymbol{u}_v$, $v \leq d$ (d is the number of signals in the channel) is given by

$$\Delta\boldsymbol{u}_v = \sum_{\substack{i=1 \\ i \neq w}}^{mM_L} \frac{\boldsymbol{u}_j^H \boldsymbol{V} \boldsymbol{u}_v}{\sigma_v^2 - \sigma_j^2} \boldsymbol{u}_j.$$

From this, it follows that

$$E\left\{\Delta\boldsymbol{u}_v \Delta\boldsymbol{u}_w^H\right\} = \sum_{\substack{j=1 \\ j \neq v}}^{mM_L} \sum_{\substack{i=1 \\ i \neq w}}^{mM_L} \frac{E\left\{\boldsymbol{u}_j^H \boldsymbol{V} \boldsymbol{u}_v \boldsymbol{u}_w^H \boldsymbol{V} \boldsymbol{u}_i\right\}}{(\sigma_v^2 - \sigma_j^2)(\sigma_w^2 - \sigma_i^2)} \boldsymbol{u}_j \boldsymbol{u}_i^H. \tag{69}$$

Putting $\boldsymbol{V} = \boldsymbol{R}_{xx} - \overline{\boldsymbol{R}}_{xx}$ into the above and noting that for $x \neq y$, $\boldsymbol{u}_x^H \overline{\boldsymbol{R}}_{xx} \boldsymbol{u}_y = 0$, the above can be rewritten as

$$E\left\{\Delta\boldsymbol{u}_v \Delta\boldsymbol{u}_w^H\right\} = \sum_{\substack{j=1 \\ j \neq v}}^{mM_L} \sum_{\substack{i=1 \\ i \neq w}}^{mM_L} \frac{E\left\{\boldsymbol{u}_j^H \boldsymbol{R}_{xx} \boldsymbol{u}_v \boldsymbol{u}_w^H \boldsymbol{R}_{xx} \boldsymbol{u}_i\right\}}{(\sigma_v^2 - \sigma_j^2)(\sigma_w^2 - \sigma_i^2)} \boldsymbol{u}_j \boldsymbol{u}_i^H. \tag{70}$$

Let $\boldsymbol{x}_k$ be the kth column of $\boldsymbol{X}_{m,L}$; then, we have

$$\boldsymbol{R}_{xx} = \frac{1}{L(N-m+1)} \sum_{k=1}^{L(N-m+1)} \boldsymbol{x}_k \boldsymbol{x}_k^H.$$

Thus, putting this in place of $\boldsymbol{R}_{xx}$ in (70), the numerator term is expressed as

$$E\left\{\boldsymbol{u}_j^H \boldsymbol{R}_{xx} \boldsymbol{u}_v \boldsymbol{u}_w^H \boldsymbol{R}_{xx} \boldsymbol{u}_i\right\} = \frac{1}{L^2(N-m+1)^2} \cdot \sum_{k,n=1}^{L(N-m+1)} E\left\{\boldsymbol{u}_j^H \boldsymbol{x}_k \boldsymbol{x}_k^H \boldsymbol{u}_v \boldsymbol{u}_w^H \boldsymbol{x}_n \boldsymbol{x}_n^H \boldsymbol{u}_i\right\}. \tag{71}$$

It is well known that for nonzero mean Gaussian random variables z_1, z_2, z_3 and z_4

$$E\{z_1 z_2 z_3 z_4\} = E\{z_1 z_2\} E\{z_3 z_4\} + E\{z_1 z_3\} E\{z_2 z_4\} \cdot E\{z_1 z_4\} E\{z_2 z_3\} - 2E\{z_1\} E\{z_2\} E\{z_3\} E\{z_4\}.$$

Using this property and noting $E\{\boldsymbol{x}_k\} = \boldsymbol{y}_k$, we get

$$\begin{aligned}\sum_{k,\,n} & E\left\{\boldsymbol{u}_j^H\boldsymbol{x}_k\boldsymbol{x}_k^H\boldsymbol{u}_v\boldsymbol{u}_w^H\boldsymbol{x}_n\boldsymbol{x}_n^H\boldsymbol{u}_i\right\}\\ &= \sum_{k,\,n} E\left\{\boldsymbol{u}_j^H\boldsymbol{x}_k\boldsymbol{x}_k^H\boldsymbol{u}_v\right\}E\{\boldsymbol{u}_w^H\boldsymbol{x}_n\boldsymbol{x}_n^H\boldsymbol{u}_i\}\\ &\quad+\sum_{k,\,n} E\left\{\boldsymbol{u}_j^H\boldsymbol{x}_k\boldsymbol{u}_w^H\boldsymbol{x}_n\right\}E\left\{\boldsymbol{x}_k^H\boldsymbol{u}_v\boldsymbol{x}_n^H\boldsymbol{u}_i\right\}\\ &\quad+\sum_{k,\,n} E\left\{\boldsymbol{u}_j^H\boldsymbol{x}_k\boldsymbol{x}_n\boldsymbol{u}_i\right\}E\left\{\boldsymbol{u}_w^H\boldsymbol{x}_n\boldsymbol{x}_k\boldsymbol{u}_v\right\}\\ &\quad-2\sum_{k,\,n}\boldsymbol{u}_j^H\boldsymbol{y}_k\boldsymbol{y}_k\boldsymbol{u}_v\boldsymbol{u}_w^H\boldsymbol{y}_n\boldsymbol{y}_n\boldsymbol{u}_i. \qquad (72)\end{aligned}$$

The first term may be written as $\boldsymbol{u}_j^H E\left\{\sum_k \boldsymbol{x}_k\boldsymbol{x}_k^H\right\}\boldsymbol{u}_v\boldsymbol{u}_w^H E\left\{\sum_n \boldsymbol{x}_n\boldsymbol{x}_n^H\right\}\boldsymbol{u}_i$. However, $E\{\sum_k \boldsymbol{x}_k\boldsymbol{x}_k^H\} = L(N-m+1)\overline{\boldsymbol{R}}_{xx}$. Thus, the first term in (72) can be written more compactly as $L^2(N-m+1)^2\boldsymbol{u}_j^H\overline{\boldsymbol{R}}_{xx}\boldsymbol{u}_v\boldsymbol{u}_w^H\overline{\boldsymbol{R}}_{xx}\boldsymbol{u}_i$. Since $j \neq v$ and $i \neq w$, this term reduces to zero.

Consider the second term. Let $\boldsymbol{w}_k$ be the kth noise vector (which is assumed to be Gaussian and circulant) such that $\boldsymbol{x}_k = \boldsymbol{y}_k + \boldsymbol{w}_k$. Putting this into the second term, and noting that all the odd number moments in the noise are zero and, for all $k,\ n$, $E\{\boldsymbol{w}_k\boldsymbol{w}_n^T\} = 0$, we obtain

$$\begin{aligned}\sum_{k,\,n} & E\left\{\boldsymbol{u}_j^H\boldsymbol{x}_k\boldsymbol{u}_w^H\boldsymbol{x}_n\right\}E\left\{\boldsymbol{x}_k^H\boldsymbol{u}_v\boldsymbol{x}_n^H\boldsymbol{u}_i\right\}\\ &= \sum_{k,\,n}\boldsymbol{u}_j^H\boldsymbol{y}_k\boldsymbol{y}_k\boldsymbol{u}_v\boldsymbol{u}_w^H\boldsymbol{y}_n\boldsymbol{y}_n\boldsymbol{u}_i. \qquad (73)\end{aligned}$$

Rearranging the terms, the above can be compactly written as

$$L^2(N-m+1)^2\boldsymbol{u}_j^H\overline{\boldsymbol{R}}_{yy}\boldsymbol{u}_v\boldsymbol{u}_w^H\overline{\boldsymbol{R}}_{yy}\boldsymbol{u}_i. \qquad (74)$$

For $j \neq v$ or $i \neq w$, this vanishes. Note that the fourth term in (72) also reduces to (74). Thus, it also vanishes. This means that the only remaining term in (72) is the third one; hence

$$\begin{aligned}\sum_{k,\,n} & E\left\{\boldsymbol{u}_j^H\boldsymbol{x}_k\boldsymbol{x}_k^H\boldsymbol{u}_v\boldsymbol{u}_w^H\boldsymbol{x}_n\boldsymbol{x}_n^H\boldsymbol{u}_i\right\}\\ &= \sum_{k,\,n} E\left\{\boldsymbol{u}_j^H\boldsymbol{x}_k\boldsymbol{x}_n\boldsymbol{u}_i\right\}E\left\{\boldsymbol{u}_w^H\boldsymbol{x}_n\boldsymbol{x}_k\boldsymbol{u}_v\right\}.\end{aligned}$$

Putting $\boldsymbol{x}_k = \boldsymbol{y}_k + \boldsymbol{w}_k$ in the above, we get 16 terms. Of these, all the odd order moments in the noise are zero, and since we have assumed circulant noise, $E\{\boldsymbol{w}_k\boldsymbol{w}_n^T\} = 0$ for all $n,\ k$. After eliminating all these terms, we obtain

$$\begin{aligned}\sum_{k,\,n} & E\left\{\boldsymbol{u}_j^H\boldsymbol{x}_k\boldsymbol{x}_n\boldsymbol{u}_i\right\}E\left\{\boldsymbol{u}_w^H\boldsymbol{x}_n\boldsymbol{x}_k\boldsymbol{u}_v\right\}\\ &= \sum_{k,\,n}\boldsymbol{u}_j^H\boldsymbol{y}_k\boldsymbol{y}_n^H\boldsymbol{u}_i\boldsymbol{u}_w^H\boldsymbol{y}_n\boldsymbol{y}_k^H\boldsymbol{u}_v\\ &\quad+\sum_{k,\,n}\boldsymbol{u}_j^H\boldsymbol{y}_k\boldsymbol{y}_n^H\boldsymbol{u}_i E\left\{\boldsymbol{u}_w^H\boldsymbol{w}_n\boldsymbol{w}_k^H\boldsymbol{u}_v\right\}\\ &\quad+\sum_{k,\,n} E\left\{\boldsymbol{u}_j^H\boldsymbol{w}_k\boldsymbol{w}_n^H\boldsymbol{u}_i\right\}\boldsymbol{u}_w^H\boldsymbol{y}_n\boldsymbol{y}_k^H\boldsymbol{u}_v\\ &\quad+\sum_{k,\,n} E\left\{\boldsymbol{u}_j^H\boldsymbol{w}_k\boldsymbol{w}_n^H\boldsymbol{u}_i\right\}E\left\{\boldsymbol{u}_w^H\boldsymbol{w}_n\boldsymbol{w}_k^H\boldsymbol{u}_v\right\}. \qquad (75)\end{aligned}$$

Note that the first term is equal to (74), which is identical to zero. Thus, we are left with the last three terms only. Consider the second term (henceforth denoted by $t2$). Let $\boldsymbol{A}_m \in \mathbb{C}^{mM_L,\,d}$ be the extended array steering matrix, the d-vector $\boldsymbol{\rho}$ be the complex amplitudes of the d signals, and $\boldsymbol{B} := \text{diag}\{\boldsymbol{\rho}\}$. The spatio-temporally smoothed noise free-data matrix $\boldsymbol{Y}_{m,\,L} \in \mathbb{C}^{mM_L,\,L(N-m+1)}$ can then be expressed as

$$\boldsymbol{Y}_{m,\,L} = \boldsymbol{A}_m\boldsymbol{B}\left[\boldsymbol{F}_s \quad \boldsymbol{\Theta}\boldsymbol{F}_s \quad \cdots \quad \boldsymbol{\Theta}^{L-1}\boldsymbol{F}_s\right] =: \boldsymbol{A}_m\boldsymbol{S}.$$

For $p,\ q \in [0,\ L]$ and $g,\ h \in [0,\ N-m]$, let k and n be written as

$$k = p(N-m+1) + k'$$

$$n = q(N-m+1) + n'.$$

In addition, let $\boldsymbol{\Phi}^n := \text{diag}\,\{\phi_i^n\}_{i=1}^d$, and $\boldsymbol{\Theta}^n := \text{diag}\,\{\theta_i^n\}_{i=1}^d$. Then, the kth and nth columns of $\boldsymbol{Y}_{m,\,L}$, which are denoted by $\boldsymbol{y}_k$ and $\boldsymbol{y}_n$, are given by

$$\boldsymbol{s}_k = \boldsymbol{A}_m\boldsymbol{\Phi}^{k'}\boldsymbol{\Theta}^p\boldsymbol{\rho}$$

$$\boldsymbol{s}_n = \boldsymbol{A}_m\boldsymbol{\Phi}^{n'}\boldsymbol{\Theta}^q\boldsymbol{\rho}.$$

Let $\boldsymbol{D} = \boldsymbol{\rho}\boldsymbol{\rho}^H$; then, we have

$$\begin{aligned}t2 = \sum_{k',\,p,\,n',\,q} & \boldsymbol{u}_j^H\boldsymbol{A}_m\boldsymbol{\Phi}^{k'}\boldsymbol{\Theta}^p\boldsymbol{D}\boldsymbol{\Phi}^{-n'}\boldsymbol{\Theta}^{-q}\boldsymbol{A}_m^H\\ &\cdot E\left\{\boldsymbol{u}_w^H\boldsymbol{w}_n\boldsymbol{w}_k^H\boldsymbol{u}_v\right\}. \qquad (76)\end{aligned}$$

From the spatio-temporal smoothing process, we see that the factor

$$E\left\{\boldsymbol{u}_w^H\boldsymbol{w}_n\boldsymbol{w}_k^H\boldsymbol{u}_v\right\}$$

is a function of $k-n$ only. More precisely, let $\boldsymbol{Z}^h$ be a Toeplitz matrix with all the elements equal to zero, except for those unity-valued entries on the hth parallel to the diagonal, and let σ_n^2 be the variance of the noise. Let $L_o = \min(L,\ M_L)$ and $m_o = \min(m,\ N-m+1)$; then

$$\begin{aligned}E & \left\{\boldsymbol{u}_w^H\boldsymbol{w}_n\boldsymbol{w}_k^H\boldsymbol{u}_v\right\}\\ &= \begin{cases}\sigma_n^2\boldsymbol{Z}^{(k'-n')M_L+(p-q)}, & |k'-n'| < m_o\\ & \quad \&|p-q| < L_o\\ 0, & \text{otherwise.}\end{cases} \qquad (77)\end{aligned}$$

Making change of variables and, for $r \in [-L_o,\ L_o]$ and $h = \in [-m_o,\ m_o]$, setting $n' = k' - h$ and $q = p - r$, (76) can be expressed as

$$\begin{aligned}t2 = \sigma_n^2 \sum_{k',\,p,\,r,\,h} & \boldsymbol{u}_j^H\boldsymbol{A}_m\boldsymbol{\Phi}^{k'}\boldsymbol{\Theta}^p\boldsymbol{D}\boldsymbol{\Phi}^{h-k'}\boldsymbol{\Theta}^{r-p}\\ &\cdot\boldsymbol{A}_m^H\boldsymbol{u}_i\boldsymbol{u}_w^H\boldsymbol{Z}^{hM_L+r}\boldsymbol{u}_v. \qquad (78)\end{aligned}$$

Rearranging the terms, we obtain

$$\begin{aligned}t2 = & \left(\sigma_n^2\sum_{r,\,h}\boldsymbol{u}_w^H\boldsymbol{Z}^{rM_L+h}\boldsymbol{u}_v\boldsymbol{u}_j^H\boldsymbol{A}_m\right)\\ &\times\sum_{k',\,p}\left(\boldsymbol{\Phi}^{k'}\boldsymbol{\Theta}^p\boldsymbol{D}\boldsymbol{\Phi}^{-k'}\boldsymbol{\Theta}^{-p}\right)\boldsymbol{\Phi}^h\boldsymbol{\Theta}^r\boldsymbol{A}_m^H\boldsymbol{u}_i.\end{aligned}$$

Note that

$$\sum_{k',p} \boldsymbol{\Phi}^{k'}\boldsymbol{\Theta}^{p}\boldsymbol{D}\boldsymbol{\Phi}^{-k'}\boldsymbol{\Theta}^{-p} = L(N-m+1)\boldsymbol{R}_{ss}$$

where $\boldsymbol{R}_{ss} \in \mathbb{C}^{d,d}$ is the signal covariance matrix. Thus, we have

$$t2 = \left(L(N-m+1)\sigma_n^2\right) \times \sum_{r,h} \boldsymbol{u}_w^H \boldsymbol{Z}^{hM_L+r}\boldsymbol{u}_v\boldsymbol{u}_j^H\boldsymbol{A}_m\boldsymbol{R}_{ss}\boldsymbol{\Phi}^h\boldsymbol{\Theta}^r\boldsymbol{A}_m^H\boldsymbol{u}_i.$$

Since $\boldsymbol{U}_s$ and $\boldsymbol{A}_m$ span the same column space, there exists an invertible matrix $\boldsymbol{T} \in \mathbb{C}^{d,d}$ such that $\boldsymbol{U} = \boldsymbol{A}_m\boldsymbol{T}^{-1}$. Let $\boldsymbol{Q} = \boldsymbol{T}^{-1}$; then, noting that $\boldsymbol{R}_{ss} = \boldsymbol{Q}\boldsymbol{\Sigma}^2\boldsymbol{Q}^H$, for $x \le d$, it follows that

$$\boldsymbol{u}_x\boldsymbol{A}_m\boldsymbol{R}_{ss} = \overline{\sigma}_x^2\boldsymbol{q}_x^H$$
$$\boldsymbol{A}_m^H\boldsymbol{u}_x = \boldsymbol{t}_x$$

where $\overline{\sigma}_x$ is the xth noise-free eigenvalue, and $\boldsymbol{q}_x^H$ and $\boldsymbol{t}_x$ are the xth row and column of $\boldsymbol{Q}^H$ and $\boldsymbol{T}^H$, respectively. Putting these in to above, we get

$$t2 = L(N-m+1)\overline{\sigma}_j^2\sigma_n^2\sum_{r,h}\boldsymbol{u}_w^H\boldsymbol{Z}^{hM_L+r}\boldsymbol{u}_v\boldsymbol{q}_j^H\boldsymbol{\Phi}^h\boldsymbol{\Theta}^r\boldsymbol{t}_i.$$

In a similar way, for the third term in (69) (denoted by $t3$), we get

$$t3 = L(N-m+1)\overline{\sigma}_w^2\sigma_n^2\sum_{r,h}\boldsymbol{u}_j^H\boldsymbol{Z}^{-hM_L-r}\boldsymbol{u}_i\boldsymbol{q}_w^H\boldsymbol{\Phi}^{-h}\boldsymbol{\Theta}^{-r}\boldsymbol{t}_v$$

and the fourth term becomes

$$t4 = \sigma_n^4\sum_k\sum_{r,h}\boldsymbol{u}_w^H\boldsymbol{Z}^{hM_L+r}\boldsymbol{u}_v\boldsymbol{u}_j^H\boldsymbol{Z}^{-hM_L-r}\boldsymbol{u}_i.$$

Since $\boldsymbol{Z}^{hM_L+r}$ is independent of k, this reduces to

$$t4 = L(N-m+1)\sigma_n^4\sum_{r,h}\boldsymbol{u}_w^H\boldsymbol{Z}^{hM_L+r}\boldsymbol{u}_v\boldsymbol{u}_j^H\boldsymbol{Z}^{-hM_L-r}\boldsymbol{u}_i.$$

Combining all the three terms, we obtain

$$\begin{aligned} &E\left\{\boldsymbol{u}_j^H\boldsymbol{R}_{xx}\boldsymbol{u}_v\boldsymbol{u}_w^H\boldsymbol{R}_{xx}\boldsymbol{u}_i\right\} \\ &= \sum_{r=-L_o}^{L_o}\sum_{h=-m_o}^{m_o}\left(\frac{\overline{\sigma}_j^2\sigma_n^2}{L(N-m+1)}\right. \\ &\quad\cdot\left(\boldsymbol{u}_w^H\boldsymbol{Z}^{rM_L+h}\boldsymbol{u}_v\boldsymbol{q}_j^H\boldsymbol{\Phi}^h\boldsymbol{\Theta}^r\boldsymbol{t}_i\right) \\ &\quad+\frac{\overline{\sigma}_w^2\sigma_n^2}{L(N-m+1)}\left(\boldsymbol{u}_j^H\boldsymbol{Z}^{-rM_L-h}\boldsymbol{u}_i\boldsymbol{q}_w^H\boldsymbol{\Phi}^{-h}\boldsymbol{\Theta}^{-r}\boldsymbol{t}_v\right) \\ &\quad\left.+\frac{\sigma_n^4}{L(N-m+1)}\left(\boldsymbol{u}_w^H\boldsymbol{Z}^{rM_L+h}\boldsymbol{u}_v\boldsymbol{u}_j^H\boldsymbol{Z}^{-rM_L-h}\boldsymbol{u}_i\right)\right). \end{aligned} \tag{79}$$

Putting this into (69) and noting that $t2$ is nonzero only for $i, j \le d$, (24) follows. This concludes the proof.

C. Proof of Lemma IV.1

For white noise, $\boldsymbol{Z}^{hM_L+r}$ in (79) is nonzero only for $r = 0$ and $h = 0$. In this case, $\boldsymbol{Z} = \boldsymbol{I}$, and (79) reduces to

$$\begin{aligned} &E\left\{\boldsymbol{u}_j^H\boldsymbol{R}_{xx}\boldsymbol{u}_v\boldsymbol{u}_w^H\boldsymbol{R}_{xx}\boldsymbol{u}_i\right\} \\ &= \frac{\overline{\sigma}_j^2\sigma_n^2}{L(N-m+1)}\left(\boldsymbol{u}_w^H\boldsymbol{u}_v\boldsymbol{q}_j^H\boldsymbol{t}_i\right) \\ &\quad+\frac{\overline{\sigma}_w^2\sigma_n^2}{L(N-m+1)}\left(\boldsymbol{u}_j^H\boldsymbol{u}_i\boldsymbol{q}_w^H\boldsymbol{t}_v\right) \\ &\quad+\frac{\sigma_n^4}{L(N-m+1)}\left(\boldsymbol{u}_w^H\boldsymbol{u}_v\boldsymbol{u}_j^H\boldsymbol{u}_i\right). \end{aligned} \tag{80}$$

Note that for $i, j > d$, the first term is zero; for $i \ne j$, $\boldsymbol{u}_j^H\boldsymbol{u}_i = \boldsymbol{q}_j^H\boldsymbol{t}_i = 0$, and for $v \ne w$, $\boldsymbol{u}_w^H\boldsymbol{u}_v = \boldsymbol{q}_w^H\boldsymbol{t}_v = 0$. Putting (80) into (69), we thus obtain

$$E\left\{\boldsymbol{u}_j^H\boldsymbol{R}_{xx}\boldsymbol{u}_v\boldsymbol{u}_w^H\boldsymbol{R}_{xx}\boldsymbol{u}_i\right\} = 0$$

for $w \ne v$ or $i \ne j$, and

$$E\left\{\boldsymbol{u}_j^H\boldsymbol{R}_{xx}\boldsymbol{u}_v\boldsymbol{u}_w^H\boldsymbol{R}_{xx}\boldsymbol{u}_i\right\} = \frac{1}{L(N-m+1)} \times\left(\sum_{\substack{j=1\\ j\ne v}}^{d}\frac{\overline{\sigma}_j^2\sigma_n^2+\overline{\sigma}_w^2\sigma_n^2+\sigma_n^4}{(\sigma_w^2-\sigma_j^2)^2}+\sum_{j=d+1}^{mM_L}\frac{\overline{\sigma}_w^2\sigma_n^2+\sigma_n^4}{(\sigma_w^2-\sigma_j^2)^2}\right) \tag{81}$$

for $w = v$ and $i = j$. Since $\sigma_x^2 = \overline{\sigma}_x^2 + \sigma_n^2$, we may write

$$\overline{\sigma}_j^2\sigma_n^2+\overline{\sigma}_w^2\sigma_n^2+\sigma_n^4 = \sigma_w^2\sigma_j^2-\overline{\sigma}_w^2\overline{\sigma}_j^2$$
$$\overline{\sigma}_w^2\sigma_n^2+\sigma_n^4 = \sigma_w^2\sigma_n^2.$$

Placing these into (81), we obtain (25). This concludes the proof.

REFERENCES

[1] R. Roy, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1987.

[2] A.-J. van der Veen, P. Ober, and E. Deprettere, "Azimuth and elevation computation in high resolution DOA estimation," *IEEE Trans. Signal Processing*, vol. 40, pp. 1828–1832, July 1992.

[3] M. D. Zoltowski and C. P. Mathews, "Real-time frequency and 2-D angle estimation with sub-nyquist spatio-temporal sampling," *IEEE Trans. Signal Processing*, vol. 42, pp. 2781–2794, Oct. 1994.

[4] M. Haardt and J. A. Nossek, "3-D unitary ESPRIT for joint angle and carrier estimation," in *Proc. ICASSP*, Munich, Germany, Apr. 1997, pp. 255–258.

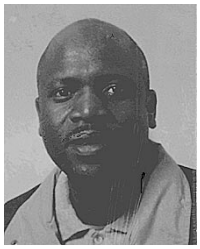
[5] A. van der Veen, M. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Processing*, vol. 46, pp. 405–418, Feb. 1998.

[6] A. N. Lemma, A.-J. van der Veen, and E. F. Deprettere, "Joint angle-frequency estimation using multi-resolution ESPRIT," in *Proc. ICASSP*, Seattle, WA, May 1998.

[7] A.-J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1155, May 1996.

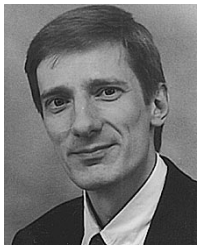
[8] L. de Lathauwer, "Signal processing based on multilinear algebra," Ph.D. dissertation, Katholieke Univ. Leuven, Leuven, Belgium, 1997.

[9] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 806–811, Aug. 1985.
[10] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 8–15, Jan. 1989.
[11] B. Rao and K. V. S. Hari, "Weighted subspace methods and spatial smoothing: Analysis and comparison," *IEEE Trans. Signal Processing*, vol. 41, pp. 788–803, Feb. 1993.
[12] R. Bachl, "The forward-backward averaging technique applied to TLS-ESPRIT processing," *IEEE Trans. Signal Processing*, vol. 43, no. Nov., pp. 2691–2699, 1995.
[13] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 8–15, Jan. 1989.
[14] K. C. Huarng and C. C. Yeh, "A unitary transformation method for angle of arrival estimation," *IEEE Trans. Signal Processing*, vol. 39, pp. 975–977, Apr. 1991.
[15] D. A. Linebarger, R. D. deGroat, and E. M. Dowling, "Efficient direction finding methods employing forward/backward averaging," *IEEE Trans. Signal Processing*, vol. 42, pp. 2136–2145, Aug. 1994.
[16] A. Lee, "Centro-Hermitian and skew-centroHermitian matrices," *Linear Algebra Applicat.*, vol. 29, pp. 205–210, 1980.
[17] G. Xu, R. H. Roy, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Trans. Signal Processing*, vol. 42, pp. 102–112, Jan. 1994.
[18] M. Haardt, "Efficient one-, two-, and multidimensional high-resolution array signal processing," Ph.D. dissertation, Munich Univ. Technol., Munich, Germany, 1997.
[19] J. E. Hudson, *Adaptive Array Principles*. London, U.K.: Inst. Electr. Eng., 1981.
[20] A. N. Lemma, A.-J. van der Veen, and E. F. Deprettere, "On the multi-resolution ESPRIT algorithm," in *Proc. 9th SP Workshop SSAP*, Portland, OR, Sept. 1998.
[21] ——, "Multi-resolution ESPRIT algorithm," *IEEE Trans. Signal Processing*, vol. 47, pp. 1722–1726, June 1999.
[22] T. W. Anderson, "Asymptotic theory for principal component analysis," *Annals Math. Stat.*, vol. 34, pp. 122–148, 1963.
[23] D. R. Brillinger, *Time Series: Data Analysis and Theory*. New York: Holt, Rhinehart, and Winston, 1975.
[24] D. J. Jeffries and D. R. Farrier, "Asymptotic results for eigenvector methods," *Proc. Inst. Elect. Eng.*, vol. 132, pp. 589–594, June 1985.
[25] M. Kaveh and A. J. Barabell, "The statistical performance of MUSIC and minimum-norm algorithms in resolving plane waves," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 331–341, Apr. 1986.
[26] H. Lee and F. Li, "An eigenvector technique for detecting the number of emitters in a cluster," *IEEE Trans. Signal Processing*, vol. 42, pp. 2380–2388, Sept. 1994.
[27] B. D. Rao and K. V. S. Hari, "Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp. 1990–1995, Dec. 1989.
[28] P. Stoica and A. Nehorai, "Performance comparison of subspace rotation and MUSIC methods for direction estimation," *IEEE Tran. Signal Processing*, vol. 39, pp. 446–453, Feb. 1991.
[29] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Signal Processing*, vol. 38, pp. 814–824, May 1990.
[30] M. Viberg, B. Ottersten, and A. Nehorai, "Performance analysis of direction finding with large arrays and finite data," *IEEE Trans. Signal Processing*, vol. 43, pp. 469–477, Feb. 1995.
[31] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 1110–1121, May 1991.
[32] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Trans. Signal Processing*, vol. 39, pp. 1122–1135, May 1991.
[33] D. B. Rao and K. Hari, "Performance analysis of root-MUSIC," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1939–1949, Dec. 1989.
[34] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[35] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. London, U.K.: Oxford Univ. Press, 1965.



Aweke N. Lemma was born in Arba Minch, Ethiopia, on September 7, 1965. He received the B.Sc. degree in 1988 (with great distinction) from the Department of Electrical Engineering, Addis Ababa University, Addis Ababa, Ethiopia, the M.Sc. degree in 1994 (with great distinction) from the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, and the Chartered Designer degree in 1996 and the Ph.D. degree in 2000 from the Department of Electrical Engineering, Delft University of Technology, Delft, The Netherlands.

Between 1988 and 1992, he was a Lecturer at Addis Ababa University. From 1994 to 2000, he worked as a Researcher and Assistant Lecturer with the Signal Processing Group, Delft University of Technology, in the fields of multirate signal processing, speech coding, and statistical and array signal processing. Currently, he is with the Philips Digital Systems Laboratory, Eindhoven. His current research focuses on watermarking techniques for multimedia signals.



Alle-Jan van der Veen (S'87–M'94–SM'02) was born in The Netherlands in 1966. He graduated (cum laude) from the Department of Electrical Engineering, Delft University of Technology, Delft, The Netherlands, in 1988, and received the Ph.D. degree (cum laude) from the same institute in 1993.

Throughout 1994, he was a postdoctoral scholar with the Scientific Computing/ Computational Mathematics Group and in the Information Systems Lab, Stanford University, Stanford, CA. At present, he is a Full Professor with the Signal Processing Group of the Department of Information Technology, Delft University of Technology. His research interests are in the general area of system theory applied to signal processing and, in particular, algebraic methods for array signal processing.

Dr. van der Veen was the recipient of a 1994 and a 1997 IEEE SPS Young Author Paper Award and was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1998 to 2001. He is currently chairman of the IEEE SPS SPCOM Technical Committee, and Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS.

Ed F. Deprettere (M'80–SM'88–F'96) was born in Roeselare, Belgium, on August 10, 1944. He received the M.Sc. degree from the University of Ghent, Ghent, Belgium, in 1968 and the Ph.D. Degree from the Delft University of Technology (DUT), Delft, The Netherlands, in 1981.

In 1970, he became a Research Assistant and Lecturer at DUT. From 1980 to 1999, he was Professor with the Department of Electrical Engineering, Circuits and Systems Section, Signal Processing Group, DUT. Since January 1, 2000, he has been a Professor at the Leiden Institute of Advances Computer Sciences, Leiden University, Leiden, The Netherlands, where he is Head of the Leiden Embedded Research Center. His current research interests are in system-level design of embedded systems, in particular, for signal, image, and video processing applications, including wireless communications and multimedia. He is editor of the books *SVD and Signal Processing: Algorithms, Architectures and Applications and Algorithms and Parallel VLSI Architectures*. He was on the Editorial Board of the *Journal of VLSI Signal Processing and Integration, the VLSI Journal*. He was a Guest Professor with Philips Research, Eindhoven, The Netherlands, from September 1, 1993 to September 30, 1994.

Dr. Deprettere has been on the editorial boards of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS. He coauthored papers that received IEEE SP Awards in 1989 and 1995.