Multidimensional ESPRIT for Damped and Undamped Signals: Algorithm, Computations, and Perturbation Analysis

Souleymen Sahnoun, Konstantin Usevich, Pierre Comon

Multidimensional ESPRIT for Damped and Undamped Signals: Algorithm, Computations and Perturbation Analysis

Souleymen Sahnoun, Konstantin Usevich*, Pierre Comon

Abstract—**In this paper we present and analyze the performance of multidimensional ESPRIT (N-D ESPRIT) method for estimating parameters of N-D superimposed damped and/or undamped exponentials. N-D ESPRIT algorithm is based on low-rank decomposition of multilevel Hankel matrices formed by the N-D data. In order to reduce the computational complexity for large signals, we propose a fast N-D ESPRIT using truncated singular value decomposition (SVD). Then, through a first-order perturbation analysis, we derive simple expressions of the variance of the estimates in N-D multiple-tones case. These expressions do not involve the factors of the SVD. We also derive closed-form expressions of the variances of the complex modes, frequencies and damping factors estimates in the N-D single-tone case. Computer results are presented to show effectiveness of the fast version of N-D ESPRIT and verify theoretical expressions.**

Index Terms—**Frequency estimation, harmonic retrieval, multilevel Hankel matrix, 2-D ESPRIT, perturbation analysis, truncated SVD.**

EDICS: SSP-PARE, SSP-PERF, SSP-SPEC, SAM-DOAE

I. INTRODUCTION

Parameter estimation from bidimensional (2-D) and multidimensional (N-D) signals finds many applications in signal processing and communications such as magnetic resonance (NMR) spectroscopy [1], wireless communication channel estimation, antenna array processing and radar [2]. In these applications, signals are modeled by a superposition of damped or undamped N-D complex exponentials.

a) State of art: To deal with this problem, several parametric methods have been proposed. They include linear prediction-based methods such as 2-D TLS-Prony [3], and subspace approaches such as matrix enhancement and matrix pencil (MEMP) [4], 2-D ESPRIT [5], R-D ESPRIT [6], Shaped ESPRIT [7], improved multidimensional folding (IMDF) [8], [9], Tensor-ESPRIT [10], principal-singular-vector utilization for modal analysis (PUMA) [11], [12] and the methods proposed in [13], [14].

Other approaches were presented recently to address the N-D harmonic retrieval problem. The coupled Canonical Polyadic Decomposition (CPD) formulation was investigated in [15], [16] and an algorithm based on simultaneous matrix diagonalization was applied. Also, the authors in [17] proposed two methods based on multilevel Toeplitz matrices. The first one (called MaPP) is similar to the MEMP algorithm [4], and requires an extra step for pairing N-D modes. The second algorithm of [17] (called RWTM) belongs to the class of sparse recovery methods based on convex relaxations [18], [19], and has a prohibitive computational complexity.

In this paper, we consider multidimensional ESPRIT methods that generalize the well-known ESPRIT [20] algorithm. In [6] the multidimensional ESPRIT algorithm was proposed for undamped signals in the context of antenna array processing. In [5] the 2-D ESPRIT algorithm was proposed; it can handle damped and/or undamped bi-dimensional signals and works in presence of identical modes in all dimensions. The methods of [6] and [5] employ different joint diagonalization schemes for shift-invariance matrices: approximate simultaneous Schur decomposition in [6] versus diagonalization of a linear combination of matrices in [5]. The difference also is that [6] treats the case of several temporal samples (so-called snapshots) of the signal (which is common in array processing), whereas [5] treats a single temporal sample. Therefore, in [5] an extended Hankel-block-Hankel matrix is first constructed from data, which corresponds to so-called spatial smoothing in antenna array processing literature.

It is generally admitted that ESPRIT-type (and, in general, subspace-based methods) methods yield accurate estimates at high SNR and/or when the frequencies are well separated. Statistical performances of subspace 1-D estimation methods have been extensively studied in the case of undamped sinusoids [21], [22], [23] and damped ones [24]. Analytical performances of tensor-based ESPRIT-type algorithms have been assessed for undamped signals [25] , and more recently, for the case of spatial smoothing [26]. Statistical performance of some related methods have been also studied, but only in the case of undamped sinusoids [8], [9]. For damped signals, a new study was presented for the case of 1-D damped single-tone [27], resulting in new closed-from expressions. An extension of the results of [27] to the case of 2-D ESPRIT was initiated in [28] independently of [26].

Despite many advantages, multidimensional ESPRIT-type algorithms, especially in the case of spatial smoothing, are often considered as slow. This happens due to the fact that a naive implementation often uses the full SVD, whose complexity grows very fast with the size of the involved matrices.

This work is funded by the European Research Council under the Seventh Framework Programme FP7/2007–2013 Grant Agreement no. 320594.

S. Sahnoun is with Situ8ed SA, 655 Avenue de l'Europe, 38330 Montbonnot-Saint-Martin, France. K. Usevich, and P. Comon are with CNRS, GIPSA-Lab, Univ. Grenoble Alpes, F-38000 Grenoble, France. Email: souleymen.sahnoun@situ8ed.com, konstantin.usevich@gipsa-lab.fr, pierre.comon@gipsa-lab.fr . Fax: +33476574790. Tel.: +33476826304 (K. Usevich), +33476826271 (P. Comon).

*Corresponding author.

b) Contributions: In this paper, we focus on the 2-D ESPRIT algorithm of [5] and its generalization to N dimensions, since we are interested in possibly damped signals and the case of single snapshot. First, we give an explicit description of the extension [5] of the 2-D ESPRIT algorithm to N-D signals (we call it N-D ESPRIT[1]. We use an approach simpler than in [5] for describing the algorithm, using tensor formalism and multilevel Hankel (MH) matrices, that is also useful for deriving other results of the paper. We discuss the difference between the N-D ESPRIT and the multidimensional ESPRIT of [6], and other methods, such as IMDF [9] and MEMP [4]. We also give explicitly recovery (identifiability) conditions for the N-D ESPRIT algorithm, that are, up to our knowledge, not discussed in [6], [26].

Next, we propose a fast version of the N-D ESPRIT algorithm (which we call Fast N-D ESPRIT) that utilizes the multilevel Hankel structure of the involved matrices and uses the truncated SVD. It enjoys a low computational complexity and allows handling large signals and large matrices.

One of the main contributions of our paper is the perturbation analysis of the N-D ESPRIT algorithm. Through a first-order perturbation analysis, we derive expressions of the variance of the complex modes, frequencies and damping factors estimates in the N-D damped multiple tones case. Our derivations of the first-order perturbations are self-contained and are based on rigorous proofs. In particular, we base our results on the recent full expressions for the first-order perturbations of the SVD [29, Theorem 1], [30, Proposition 9], unlike the state-of-the-art papers [25], [26] (and earlier papers [21], [31]) that neglect the term containing the change of basis of the signal subspace. In our paper, we fill this gap and provide a rigorous proof that the aforementioned term does not influence the first-order perturbation of the modes (similarly to 1-D ESPRIT as shown in [29], [30]). Moreover, we propose a simplified formula for first-order perturbation that does not involve the factors of the SVD, which allows for easier analysis and interpretability. Finally, we derive closed-form expressions for the variances of the perturbations in the N-D damped and undamped single-tone case. For the single tone case for undamped signals, we obtain the results as in [26, Theorem 3]. However, our final formula is simpler than the one of [26].

c) Organisation of the paper: In Section II, we introduce notation and present the N-D modal retrieval problem. In Section III, we describe construction of multilevel Hankel matrices and their subspace properties are recalled. In Section VI, the N-D ESPRIT algorithm is presented and recovery conditions are discussed. Then a fast implementation of N-D ESPRIT is proposed using truncated SVD of MH matrices and the gain in computational complexity is shown. The difference with related methods is also pointed out. In Section V, a first-order perturbation analysis for N-D ESPRIT is performed and simplified expressions are derived in the multiple tones case. In Section IV, the single tone case is analyzed and closed form expressions are derived. In Section VII, computer results are presented to verify the theoretical expressions and to compare N-D ESPRIT, fast N-D ESPRIT and IMDF algorithms.

[1] In our terminology, N-D ESPRIT stands for the algorithm of [5] and its generalization (the case of single snapshot) and "multidimensional ESPRIT" for the algorithm of [6].

II. Background and problem statement

A. Notation

In this paper we use the following fonts: lowercase (a) for scalars, boldface lowercase ($\mathbf{a}$) for vectors, uppercase boldface ($\mathbf{A}$) for matrices, and calligraphic ($\mathcal{A}$) for N-D arrays (tensors). Vectors are, by convention, one-column matrices. The elements of vectors/matrices/tensors are accessed as $(\mathbf{a})_i$, $(\mathbf{A})_{i,j}$ and $(\mathcal{A})_{i_1,\ldots,i_N}$ respectively. We use MATLAB-like notation for taking subarrays, i.e. $(\mathcal{A})_{i_1:j_1,\ldots,i_N:j_N}$.

We denote by $\mathbf{a}^*$, $\mathbf{A}^*$ and $\mathcal{A}^*$ elementwise conjugation of vectors matrices and tensors. For a matrix $\mathbf{A}$, we denote its transpose, Hermitian transpose and Moore-Penrose pseudoinverse as $\mathbf{A}^{\mathsf{T}}$, $\mathbf{A}^{\mathsf{H}}$ and $\mathbf{A}^{\dagger}$ respectively. The notation $\mathbf{I}_M$ is used for the $M \times M$ identity matrix.

Given a collection of vectors $\mathbf{a}_1 \in \mathbb{C}^{I_1}, \ldots, \mathbf{a}_N \in \mathbb{C}^{I_N}$, the outer product $\mathcal{A} = \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_N$ is the tensor

$$(\mathcal{A})_{i_1,\ldots,i_N} = (\mathbf{a}_1)_{i_1}(\mathbf{a}_2)_{i_2}\cdots(\mathbf{a}_N)_{i_N}.$$

We use the symbol $\boxtimes$ for the Kronecker product of matrices in order to distinguish it from the outer product, and $\odot$ for the Khatri-Rao (column-wise Kronecker) product.

For a tensor (or matrix) $\mathcal{A} \in \mathbb{C}^{I_1\times\cdots\times I_N}$ we denote by $\mathrm{vec}_{\mathrm{r}}\{\mathcal{A}\}$ its "row-major" vectorization, i.e.

$$\mathrm{vec}_{\mathrm{r}}\{\mathcal{A}\} \overset{\text{def}}{=} [(\mathcal{A})_{1,\ldots,1}, (\mathcal{A})_{1,\ldots,2}, \cdots, (\mathcal{A})_{1,\ldots,1,I_N} \\ (\mathcal{A})_{1,\ldots,2,1}, \cdots, (\mathcal{A})_{I_1,\ldots,I_N}]^{\mathsf{T}}.$$

The row-major vectorization is used because it is compatible with the Kronecker product [32], i.e.

$$\mathrm{vec}_{\mathrm{r}}\{\mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_N\} = \mathbf{a}_1 \boxtimes \cdots \boxtimes \mathbf{a}_N. \tag{1}$$

Unlike in [32], we use a special notation for row-major vectorisation in order to distinguish it from the conventional column-major vectorisation.

Given a scalar a and a natural number M we will use the notation $\mathbf{a}^{(M)}$ for the Vandermonde-structured vector

$$\mathbf{a}^{(M)} \overset{\text{def}}{=} \begin{bmatrix} 1 & a & a^2 & \cdots & a^{(M-1)} \end{bmatrix}^{\mathsf{T}}. \tag{2}$$

For a vector $\mathbf{v} \in \mathbb{C}^M$ we denote by $\mathrm{Diag}(\mathbf{v})$ the $M \times M$ diagonal matrix with the elements of $\mathbf{v}$ on the diagonal; for a matrix $\mathbf{A} \in \mathbb{C}^{M\times M}$, $\mathrm{diag}(\mathbf{A})$ stands for the vector of the elements on its main diagonal.

B. Signal model

Denote N the number of dimensions and M_n, $n = 1, \ldots, N$, the size of the sampling grid in each dimension. We consider the model below, for $m_n = 0, \ldots, M_n - 1$:

$$\tilde{y}(m_1, \ldots, m_N) = y(m_1, \ldots, m_N) + \varepsilon(m_1, \ldots, m_N), \tag{3}$$

where $\varepsilon(\cdot)$ is random noise (we leave the assumptions on the noise for later), and the signal $y(m_1, \ldots, m_N)$ is a superposition of R N-D damped complex sinusoids:

$$y(m_1, \ldots, m_N) = \sum_{r=1}^{R} c_r \prod_{n=1}^{N} (a_{r,n})^{m_n}, \tag{4}$$

where

- c_r are complex amplitudes,
- $a_{r,n} = e^{-\alpha_{r,n}+\jmath\omega_{r,n}}$ are modes in the n-th dimension,
- $\{\alpha_{r,n}\}_{r=1,n=1}^{R,N}$ are real damping factors[2] (not necessary positive),
- $\{\omega_{r,n} = 2\pi\nu_{r,n}\}_{r=1,n=1}^{R,N}$ are angular frequencies.

The problem is to estimate $\{a_{r,n}\}_{r=1}^{R}$ and $\{c_r\}_{r=1}^{R}$ from noisy observations $\tilde{y}(m_1, \ldots, m_N)$.

C. Tensor formulation

It is often convenient to rewrite the signal model in tensor notation. The tensor representation is particularly useful in the proofs contained in Appendices B and C. Let the tensor $\mathcal{Y} \in \mathbb{C}^{M_1\times\cdots\times M_N}$ be given as[3]

$$(\mathcal{Y})_{i_1,\ldots,i_N} = y(i_1-1, \ldots, i_N-1).$$

We also define similarly the tensors $\widetilde{\mathcal{Y}}, \mathcal{E} \in \mathbb{C}^{M_1\times\cdots\times M_N}$. Then (3) can be compactly written as $\widetilde{\mathcal{Y}} = \mathcal{Y} + \mathcal{E}$, and (4) is the canonical polyadic (CP) tensor decomposition

$$\mathcal{Y} = \sum_{r=1}^{R} c_r\, \mathbf{a}_{r,1}^{(M_1)} \otimes \cdots \otimes \mathbf{a}_{r,N}^{(M_N)} \tag{5}$$

where $\mathbf{a}_{r,n}^{(M_n)}$ are Vandermonde-structured vectors for $a_{r,n}$ defined in (2).

By the properties of CP decomposition, eqn. (5) after vectorization can be rewritten with the help of Khatri-Rao products:

$$\mathrm{vec_r}\{\mathcal{Y}\} = \left(\mathbf{A}_1^{(M_1)} \odot \mathbf{A}_2^{(M_2)} \odot \cdots \odot \mathbf{A}_N^{(M_N)}\right)\mathbf{c}, \tag{6}$$

where $\mathbf{c} = \begin{bmatrix} c_1 & \cdots & c_R \end{bmatrix}^{\mathsf{T}}$ is the vector of amplitudes, and $\mathbf{A}_n^{(M_n)}$ defines the Vandermonde matrix of the modes in the n-th dimension

$$\mathbf{A}_n^{(M_n)} \stackrel{\text{def}}{=} \left[\mathbf{a}_{1,n}^{(M_n)} \cdots \mathbf{a}_{R,n}^{(M_n)}\right] \in \mathbb{C}^{M_n\times R}.$$

III. Multilevel Hankel matrices and their subspaces

A. Definition and factorization

In this section, we describe the construction of the multilevel Hankel matrix, which is used in many subspace-based methods. Assume that $(L_n)_{n=1}^{N}$ is chosen such that $1 \le L_n \le M_n$ and define $K_n \stackrel{\text{def}}{=} M_n - L_n + 1$. Define by $\mathbf{y}^{(i_1,\ldots,i_N)} \in \mathbb{C}^{L_1\times\cdots\times L_N}$ the vectorized subarray

$$\mathbf{y}^{(i_1,\ldots,i_N)} \stackrel{\text{def}}{=} \mathrm{vec_r}\{(\mathcal{Y})_{i_1:i_1+L_1-1,\ldots,i_N:i_N+L_N-1}\}.$$

Then the multilevel Hankel (MH) matrix[4] $\mathbf{H} \in \mathbb{C}^{(L_1\cdots L_N)\times(K_1\cdots K_N)}$ is defined by stacking the vectorized subarrays in the vectorization order

$$\begin{aligned}\mathbf{H} \stackrel{\text{def}}{=} \Big[&\mathbf{y}^{(1,\ldots,1)}\ \ \mathbf{y}^{(1,\ldots,1,2)}\ \ \cdots\ \ \mathbf{y}^{(1,\ldots,1,K_N)} \\ &\mathbf{y}^{(1,\ldots,1,2,1)} \cdots \\ &\cdots \mathbf{y}^{(K_1,K_2,\ldots,K_N-1)}\ \ \mathbf{y}^{(K_1,K_2,\ldots,K_N)}\Big].\end{aligned} \tag{7}$$

By $\widetilde{\mathbf{H}}$ we denote the noisy version of the signal constructed upon noisy observations $\widetilde{y}$.

Remark 1: The matrix (7) has nested structure of Hankel blocks inside each other, as shown in Appendix A. Such matrices are conventionally called "multilevel Hankel matrices" in the linear algebra literature [33].

It can be verified that in the absence of noise, MH matrix (7) admits a factorization of the form

$$\mathbf{H} = \mathbf{P}\,\mathrm{diag}(\mathbf{c})\mathbf{Q}^{\mathsf{T}}, \tag{8}$$

where

$$\begin{aligned}\mathbf{P} &= \mathbf{A}_1^{(L_1)} \odot \mathbf{A}_2^{(L_2)} \odot \cdots \odot \mathbf{A}_N^{(L_N)},\\ \mathbf{Q} &= \mathbf{A}_1^{(K_1)} \odot \mathbf{A}_2^{(K_2)} \odot \cdots \odot \mathbf{A}_N^{(K_N)}.\end{aligned}$$

The factorization (8) directly follows from (6). The proof can be also found in [8].

B. Shift properties of subspaces

Let us define the selection matrices

$$\overset{n}{\overline{\mathbf{I}}} \stackrel{\text{def}}{=} \mathbf{I}_{L_1} \boxtimes \mathbf{I}_{L_2} \boxtimes \cdots \boxtimes \overline{\mathbf{I}}_{L_n} \boxtimes \cdots \boxtimes \mathbf{I}_{L_N} \tag{9}$$

$$= \mathbf{I}_{\prod_{i=1}^{n-1} L_i} \boxtimes \overline{\mathbf{I}}_{L_n} \boxtimes \mathbf{I}_{\prod_{i=n+1}^{N} L_i}, \tag{10}$$

$$\underset{n}{\underline{\mathbf{I}}} \stackrel{\text{def}}{=} \mathbf{I}_{L_1} \boxtimes \mathbf{I}_{L_2} \boxtimes \cdots \boxtimes \underline{\mathbf{I}}_{L_n} \boxtimes \cdots \boxtimes \mathbf{I}_{L_N} \tag{11}$$

$$= \mathbf{I}_{\prod_{i=1}^{n-1} L_i} \boxtimes \underline{\mathbf{I}}_{L_n} \boxtimes \mathbf{I}_{\prod_{i=n+1}^{N} L_i}, \tag{12}$$

where $\underline{\mathbf{X}}$ (resp. $\overline{\mathbf{X}}$) represents $\mathbf{X}$ without the last (resp. first) row.

Next, for a matrix $\mathbf{X}$ we define $\overset{n}{\overline{\mathbf{X}}} = \overset{n}{\overline{\mathbf{I}}}\,\mathbf{X}$ and $\underset{n}{\underline{\mathbf{X}}} = \underset{n}{\underline{\mathbf{I}}}\,\mathbf{X}$. Then the shifted versions of $\mathbf{P}$ satisfy the following equation:

$$\underset{n}{\underline{\mathbf{P}}}\,\mathbf{\Psi}_n = \overset{n}{\overline{\mathbf{P}}}, \tag{13}$$

where $\mathbf{\Psi}_n = \mathrm{diag}(\mathbf{a}_{(n)})$, $\mathbf{a}_{(n)} = [a_{1,n}, \ldots, a_{R,n}]^{\mathsf{T}}$.

Now consider $\mathbf{U}_s$ the matrix of the leading R left singular vectors of the noiseless matrix $\mathbf{H}$. Since the ranges of $\mathbf{U}_s$ and $\mathbf{P}$ coincide, they are linked by a nonsingular transformation:

$$\mathbf{P} = \mathbf{U}_s\mathbf{T}.$$

Hence, the matrix $\mathbf{F}_n \stackrel{\text{def}}{=} \mathbf{T}\mathbf{\Psi}_n\mathbf{T}^{-1}$ satisfies the equation

$$\underset{n}{\underline{\mathbf{U}}}_s\mathbf{F}_n = \overset{n}{\overline{\mathbf{U}}}_s, \tag{14}$$

If the matrix $\underset{n}{\underline{\mathbf{U}}}_s$ is full-column rank, then the matrix $\mathbf{F}_n$ satisfies the following equation:

$$\mathbf{F}_n = \left(\underset{n}{\underline{\mathbf{I}}}\,\mathbf{U}_s\right)^{\dagger}\left(\overset{n}{\overline{\mathbf{I}}}\,\mathbf{U}_s\right) := \left(\underset{n}{\underline{\mathbf{U}}}_s\right)^{\dagger}\left(\overset{n}{\overline{\mathbf{U}}}_s\right) \tag{15}$$

Hence, the matrices $\mathbf{F}_n$ can be computed from the signal subspace $\mathbf{U}_s$, and the modes of each dimension n can be estimated by the eigenvalues of $\mathbf{F}_n$.

Remark 2: Instead of $\mathbf{U}_s$, any basis of the signal subspace can be used.

[2]The damping factors are important in a number of applications, including NMR spectroscopy [1].

[3]The subtraction is needed because the indices in the tensor start from 1.

[4]The definition of the MH matrix corresponds in some other works to the spatial smoothing operation [10], to "smoothed data matrix" [9] or to "enhanced matrix" [4].

IV. N-D ESPRIT FOR MULTILEVEL HANKEL MATRICES

A. N-D ESPRIT algorithm

We formulate the N-D ESPRIT algorithm as an extension of the 2-D ESPRIT algorithm of [5]. The N-D ESPRIT algorithm consists of the following steps:

1) Choose $L_1, \ldots, L_N$ and set $K_n = M_n - L_n + 1$.
2) Construct the MH matrix $\tilde{\mathbf{H}}$ from the noisy signal, in the same format as (7).
3) Perform the SVD of $\tilde{\mathbf{H}}$, and form the matrix $\widetilde{\mathbf{U}}_s \in \mathbb{C}^{(L_1 \cdots L_N)\times R}$ of the R dominant singular vectors.
4) Compute the matrices $\widetilde{\mathbf{F}}_n$

$$\widetilde{\mathbf{F}}_n := \left(\underset{n-}{\widetilde{\mathbf{U}}_s}\right)^{\dagger} \left(\overset{n-}{\widetilde{\mathbf{U}}_s}\right). \tag{16}$$

5) Compute a linear combination of matrices, where $\beta_1, \ldots, \beta_n$ are given parameters.

$$\widetilde{\mathbf{K}} = \sum_{n=1}^{N} \beta_n \widetilde{\mathbf{F}}_n \tag{17}$$

6) Compute a diagonalizing matrix $\mathbf{T}$ of $\widetilde{\mathbf{K}}$ (from its eigenvalue decomposition):

$$\widetilde{\mathbf{K}} = \mathbf{T}\,\mathrm{Diag}(\boldsymbol{\eta})\mathbf{T}^{-1}. \tag{18}$$

7) Apply the transformation $\mathbf{T}$ to $\mathbf{F}_n$:

$$\widetilde{\mathbf{D}}_n = \mathbf{T}^{-1}\widetilde{\mathbf{F}}_n\mathbf{T}, \quad \text{for} \quad n = 1, \ldots, N \tag{19}$$

8) Extract $\{[\widehat{a}_{1,n}, \ldots, \widehat{a}_{R,n}]\}_{n=1}^{N}$ from $\mathrm{diag}(\widetilde{\mathbf{D}}_n)$, $n = 1, \ldots, N$

Note that in N-D ESPRIT there is no separate step of pairing of the modes. The modes are paired automatically because the same diagonalizing transformation $\mathbf{T}$ is used. Note that, by proper choice of β_n, the N-D ESPRIT algorithm can handle the case of identical modes in one or several dimensions. The conditions for the correct recovery of modes depend on M_n, L_n and β_n, and are described in Section IV-C.

B. Variants and related algorithms

First, there are several variants of N-D ESPRIT.

- There is a well-known multidimensional ESPRIT algorithm[5] proposed in [6] and [26]. In fact, the algorithm described in Section IV-A corresponds to the version of the algorithm of [6], [26] with spatial smoothing (because only a single snapshot is available). The main difference is in steps 5–8. The matrices in [6], [26] $\widetilde{\mathbf{F}}_n$ are jointly triangularized using simultaneous Schur decomposition [6] and the modes $a_{r,n}$ are extracted from the diagonals of the triangular matrices.
- In [5], in addition to 2D-ESPRIT, an algorithm under the name "2D-MEMP with improved pairing step" was proposed. The difference is only in steps 7–8: the modes $a_{r,n}$ are extracted from individual eigenvalue decompositions of matrices $\widetilde{\mathbf{F}}_n$ and the matrix $\mathbf{T}$ is used just to perform the pairing of the modes.

[5]Note that in the original paper of [6] only the unitary version of N-D ESPRIT was considered that is not applicable to damped signals.

As we will see, our first-order perturbation analysis also applies to these two variants.

Second, the algorithm IMDF of [9] is related to N-D ESPRIT, but it is not an extension of 2-D ESPRIT. The first difference is that the selection matrices (analogues of $\overset{n-}{\mathbf{I}}$ and $\underset{n-}{\mathbf{I}}$) are defined in a slightly different way. The second difference is that in [9] modes are estimated from $\underset{n-}{\mathbf{P}}$. However, in the N-D ESPRIT algorithm defined in Section IV-A, the modes are estimated from (19).

C. Recovery conditions

There are some essential assumptions which guarantee that in the noiseless case the N-D ESPRIT algorithm recovers the modes correctly. These are not the recovery conditions for the multidimensional harmonic retrieval problem [34], but they give the limits of applicability of N-D ESPRIT.

Assumption 1: For every n, the matrices $\underset{n-}{\mathbf{P}}$ and $\mathbf{Q}$ are full column rank (their rank is equal to R).

Assumption 2: The coefficients $\beta_n, n = 1, \ldots, N$ should satisfy the condition that all the numbers η_r defined as

$$\eta_r = \sum_{n=1}^{N} \beta_n a_{r,n}$$

are distinct.

Remark 3: The conditions can be explained as follows:

1) The first assumption is to guarantee that (16) gives the unique solution to (14). Thus the matrices $\mathbf{F}_n = \widetilde{\mathbf{F}}_n$, i.e. the matrices $\mathbf{F}_n$ are recovered correctly.
2) The η_r are exactly the eigenvalues of $\mathbf{K} = \sum_{n=1}^{N} \beta_n \mathbf{F}_n$. Thus the eigenvalue decomposition of $\mathbf{K}$ is unique (up to permutation of columns), and therefore the step of the algorithm retrieves the correct $\mathbf{T}$.

Now we establish some results on when these assumptions are satisfied. We start from Assumption 2.

Lemma 1: For any set of modes, a generic (random) choice of β_k satisfies the Assumption 2 almost surely.

Proof Since a projection of R points in $\mathbb{C}^N$ on a random line separates the points, the lemma holds true. ■

The following lemma establishes conditions for generic identifiability.

Lemma 2: Let the number of modes satisfy

$$R \leq \min_{n} \Big\{ (L_n - 1) \prod_{\substack{i=1 \\ i \neq n}}^{N} L_i, \ \prod_{i=1}^{N} K_i \Big\}.$$

Then for a generic choice of modes, $\mathrm{rank}\,\underset{n-}{\mathbf{P}} = \mathrm{rank}\,\mathbf{Q} = R$

Proof The proof follows from [34, Proposition 4]. ■

D. Fast N-D ESPRIT

Let us define $L = L_1 \cdots L_N$ and $K = K_1 \cdots K_N$ (the sizes of the MH matrix) and $M = M_1 \cdots M_N$. The main bottleneck of N-D ESPRIT (and, in general, all subspace-based methods for multidimensional harmonic retrieval) is the SVD of the multilevel Hankel matrix, which can be very large. The classic

Golub-Reinsch algorithm [35, Ch. 8] for the full SVD requires $O(L^2K)$ flops [35] (in the case $L \leq K$), and also the matrix itself needs to be stored in memory.

In this paper, we propose to compute the truncated SVD (TSVD), i.e., to find only the R leading singular values/vectors. Let $T_{\mathbf{A}}$ be the number of flops needed to compute the matrix-vector products $\mathbf{A}\mathbf{v}$ and $\mathbf{A}^{\mathsf{H}}\mathbf{u}$ for given vectors $\mathbf{u}$ and $\mathbf{v}$. Then the leading R singular values/vectors of a matrix $\mathbf{A}$ can be found using, for example, Lanczos bidiagonalization [35, Ch. 9] with partial reorthogonalization [36] is $O(RT_{\mathbf{A}} + R^2(L+K))$ (see [37, §3] for an overview of Lanczos-based methods).

The Lanczos-based methods were first used for 1-D ESPRIT in [38], [39]. However, in [38], [39], just the original Lanczos iterations [35, Ch. 9] are used, which may have poor performance due to loss of orthogonality and/or slow convergence of the iterations. This can be remedied by using partial reorthogonalization and/or restarting schemes [36], [40], [41], which yield accurate computations for singular values and vectors and have a stable and efficient implementation [42].

In the case of N-D ESPRIT, for MH matrices, the matrix-vector product can be computed using the N-D Fast Fourier Transform (FFT) in $O(M\log(M))$ flops using the N-D FFT, as we show in Appendix B. This fact was used in [38], [37] for truncated SVD of Hankel matrices, and independently in [43] and [44] for special cases of MH matrices. Although the matrix-vector multiplication in the general MH case is a straightforward extension of algorithms [43, eqn. (22)] and [44, Lemma 2], we provide in Appendix B a description of the algorithm for several reasons: in [43] Toeplitz matrices are treated and in [44] only real Hankel-block-Hankel matrices are treated (also, the proof of [44, Lemma 2] contains misprints). As a result, when R is small compared with $\log(M)$, the cost of the TSVD is $O(RM\log(M))$ flops.

Therefore, we have the following flop counts for the N-D ESPRIT algorithm:

1–3) $O(RMlog(M))$ flops;

4) $O(RLN)$ flops (can be further reduced to $O(rL)$ if the selection matrices (10) and (12) are used);

5–6) $O(R^3)$ flops (computing eigenvalues up to required precision);

The total computational complexity is $O(RM\log(M))$ (compared with $O(L^2K)$ when using the full SVD).

V. Perturbation analysis

A. Basic expressions

The SVD of the noiseless MH matrix $\mathbf{H}$ is given by:

$$\mathbf{H} = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^{\mathsf{H}} + \mathbf{U}_n\mathbf{\Sigma}_n\mathbf{V}_n^{\mathsf{H}} \tag{20}$$

where $\mathbf{\Sigma}_n = \mathbf{0}$. The perturbed $\widetilde{\mathbf{H}}$ is expressed as

$$\widetilde{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}.$$

In this section, we derive first-order perturbations with respect to $\Delta\mathbf{H}$ for the quantities in the N-D ESPRIT algorithm. The first-order perturbations are equal to complex matrix differentials (see [45] for a definition and a summary of properties).

First, we recall the expression for the perturbation of $\mathbf{U}_s$.

Lemma 3 ([29, Theorem 1] and [30, Proposition 9]): Let

$$\widetilde{\mathbf{H}} = \widetilde{\mathbf{U}}_s\widetilde{\mathbf{\Sigma}}_s\widetilde{\mathbf{V}}_s^{\mathsf{H}} + \widetilde{\mathbf{U}}_n\widetilde{\mathbf{\Sigma}}_n\widetilde{\mathbf{V}}_n^{\mathsf{H}} \tag{21}$$

be the subspace decomposition of $\widetilde{\mathbf{H}}$. Then the first-order approximation of the $\widetilde{\mathbf{U}}_s - \mathbf{U}_s$ is given by

$$\Delta\mathbf{U}_s = \mathbf{U}_n\mathbf{U}_n^{\mathsf{H}}\Delta\mathbf{H}\,\mathbf{V}_s\mathbf{\Sigma}_s^{-1} + \mathbf{U}_s\mathbf{R}, \tag{22}$$

where $\mathbf{R}$ is an antihermitian matrix (*i.e.* $\mathbf{R}^H = -\mathbf{R}$) that depends on $\Delta\mathbf{H}$ (the precise expression of the matrix $\mathbf{R}$ can be found in [29, Theorem 1] or [30, Proposition 9]).

Remark 4: In earlier papers on perturbation analysis of the SVD [21], [31], as well as in the state-of-the art literature on perturbation analysis for multidimensional ESPRIT-type algorithms [25] the term $\mathbf{U}_s\mathbf{R}$ was often neglected. In this paper, we derive perturbations based on the full formula (22). First, we give expressions first-order perturbations of the matrices $\mathbf{F}_n$.

Lemma 4: The first-order perturbation of $\mathbf{F}_n$ is given by

$$\Delta\mathbf{F}_n = (\underset{\mathrm{n}-}{\mathbf{U}}_s)^{\dagger}(\Delta\overset{\mathrm{n}-}{\mathbf{U}}_s - \Delta\underset{\mathrm{n}-}{\mathbf{U}}_s\mathbf{F}_n). \tag{23}$$

Proof The proof can be found in Appendix C.

Next, let $\mathbf{t}_r$ denote the eigenvectors of $\mathbf{K}$ (the columns of $\mathbf{T}$) and $\boldsymbol{\tau}_r^{\mathsf{T}}$ denote the rows of $\mathbf{T}^{-1}$:

$$\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_R], \quad \mathbf{T}^{-1} = [\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_R]^{\mathsf{T}}.$$

Then the following result holds true.

Lemma 5: The first-order perturbations of the modes given by steps 7-8 of the N-D ESPRIT algorithm are given by

$$\Delta a_{r,n} = \boldsymbol{\tau}_r^{\mathsf{T}}\Delta\mathbf{F}_n\mathbf{t}_r. \tag{24}$$

Proof The proof can be found in Appendix C.

Remark 5: An immediate consequence of Lemma 5 is that the first-order perturbation does not depend on the way the matrix $\mathbf{F}_n$ is diagonalized (in particular it does not depend on the coefficients β_r). In fact, it depends only on the perturbations of the matrices $\mathbf{F}_n$. Hence, in particular, the first-order perturbations for 2D-ESPRIT and 2D-MEMP with improved pairing step coincide.

A substitution of (23) into (24) leads to the following formula for the perturbation of the modes.

Corollary 1: The first order perturbation of the modes can be given as

$$\Delta a_{r,n} = \boldsymbol{\tau}_r^{\mathsf{T}}(\underset{\mathrm{n}-}{\mathbf{U}}_s)^{\dagger}(\overset{\mathrm{n}-}{\mathbf{I}} - a_{r,n}\,\underset{\mathrm{n}-}{\mathbf{I}})\Delta\mathbf{H}\,\mathbf{V}_s\mathbf{\Sigma}_s^{-1}\mathbf{t}_r \tag{25}$$

Proof The proof can be found in Appendix C.

Note that the term $\mathbf{U}_s\mathbf{R}$ from (22) does not affect the expression (25).

B. A simplified formula for the perturbations

The expression of first order perturbation (25) is widely used in the literature. It corresponds to the expressions given in [25], which is the state-of-the-art perturbation analysis. The main problem is that in (25) knowledge of the singular value decomposition of the MH matrix $\mathbf{H}$ is needed. This complicates a further analysis, since for $R \geq 2$ it becomes difficult to obtain the components of the SVD analytically. In what follows we give a simplified expression that does not require knowledge of the SVD.

Proposition 1: Let the matrix $\mathbf{H}$ satisfy (8), where the matrix $\mathbf{P}$ satisfies (13) for $n = 1, \ldots, N$. Further, by $\mathbf{b}_r \in \mathbb{C}^R$ we denote the r-th unit vector. Then the first order perturbation of the modes obtained by N-D ESPRIT admits an expansion

$$\Delta a_{r,n} = \frac{1}{c_r} \mathbf{b}_r^\top \underset{n}{\underline{\mathbf{P}}}^\dagger (\overset{n}{\overline{\mathbf{I}}} - a_{r,n} \underset{n}{\underline{\mathbf{I}}}) \Delta \mathbf{H} \, (\mathbf{Q}^\top)^\dagger \mathbf{b}_r. \qquad (26)$$

Proof The proof can be found in Appendix C.

The main advantage of the formula (26) is that it allows for an *a priori* perturbation analysis, *i.e.*, we do not need the SVD of the MH matrix to compute the perturbation. Yet another advantage of (26) is that it clearly shows the perturbation of the r-th tone ($\Delta a_{r,n}$) does not depend on the amplitudes of other tones (the coefficients c_k, $k \neq r$), and depends only on angles between the columns of matrices $\underset{n}{\underline{\mathbf{P}}}$ and $\mathbf{Q}$. This is a remarkable feature of N-D ESPRIT (a similar fact for 1D ESPRIT can be found in [30, Proposition 12]).

Remark 6: The formula (26) can be extended to the case of multiple snapshots and other subspace-based methods.

C. Computation of moments of the perturbation

First, we rewrite the perturbation (26) in the form

$$\Delta a_{r,n} = \mathbf{v}_{r,n}^{\mathsf{H}} \Delta \mathbf{H} \mathbf{x}_r^*, \qquad (27)$$

where

$$\mathbf{v}_{r,n}^{\mathsf{H}} = \frac{1}{c_r} \mathbf{b}_r^{\mathsf{T}} \underset{n}{\underline{\mathbf{P}}}^\dagger (\overset{n}{\overline{\mathbf{I}}} - a_{r,n} \underset{n}{\underline{\mathbf{I}}}), \quad \mathbf{x}_r^* = (\mathbf{Q}^{\mathsf{T}})^\dagger \mathbf{b}_r. \qquad (28)$$

Since the equation (27) is linear in $\Delta\mathbf{H}$, there is the following alternative way to compute the perturbation.

Lemma 6: Let $\mathbf{e} = \mathrm{vec_r}\{\mathcal{E}\}$ be the vectorization of the tensor of the noise term in (3). Then the product (27) is equal to

$$\Delta a_{r,n} = \mathbf{z}_{r,n}^{\mathsf{H}} \mathbf{e}, \qquad (29)$$

where the vector $\mathbf{z}_{r,n}$ is defined as

$$\mathbf{z}_{r,n} = \mathrm{vec_r}\{\mathcal{X} \star \mathcal{V}\},$$

where $\mathcal{V} \in \mathbb{C}^{L_1 \times \cdots \times L_N}$ and $\mathcal{X} \in \mathbb{C}^{K_1 \times \cdots \times K_N}$ are the tensorizations of $\mathbf{v}_{r,n}$ and $\mathbf{x}_r$, and $\mathcal{X} \star \mathcal{V} \in \mathbb{C}^{M_1 \times \cdots \times M_N}$ is the multidimensional convolution[6] of tensors.

Proof The proof can be found in Appendix B.

From the representation (29) of the perturbation, it follows that we can compute the moments of the perturbation as follows.

Corollary 2:

1) $\mathbb{E}\left\{\Delta a_{r,n}\right\} = 0$ if $\mathbf{e}$ is zero-mean.
2) $\mathbb{E}\left\{\Delta a_{r,n}^2\right\} = 0$ if $\mathbf{e}$ is circular.
3) If $\mathbf{e}$ has covariance matrix $\Gamma = \mathbb{E}\left\{\mathbf{e}\mathbf{e}^{\mathsf{H}}\right\}$,
$$\mathbb{E}\left\{|\Delta a|_{r,n}^2\right\} = \mathbf{z}_{r,n}^{\mathsf{H}} \Gamma \mathbf{z}_{r,n}.$$
In particular, if $\mathbf{e}$ is white with variance σ_e^2, then
$$\mathbb{E}\left\{|\Delta a|_{r,n}^2\right\} = \sigma_e^2 \|\mathbf{z}_{r,n}\|_2^2.$$
4) If $\mathbf{e}$ is complex circular Gaussian,
$$\mathrm{var}(\Delta\omega_n) = \mathrm{var}(\Delta\alpha_n) = \frac{\mathbb{E}\left\{|\Delta a_{r,n}|^2\right\}}{2|a_{r,n}|^2}.$$

[6] See Appendix B for a definition of multidimensional convolution.

Remark 7: As in the previous subsection, the new formula for the variance $\mathbb{E}\left\{|\Delta a|_{r,n}^2\right\}$ allows for an *a priori* perturbation analysis. It also shows a remarkable feature of N-D ESPRIT: the variance of the perturbation of the r-th tone does not depend on the amplitudes of other tones. In particular, it depends on the partial SNR with respect to each tone.

Remark 8 (On computation of the $\mathbf{v}_{r,n}$, $\mathbf{x}_r$): The vectors $\mathbf{v}_{r,n}$ and $\mathbf{x}_r$ do not require the computation of pseudo-inverses. Indeed, $\mathbf{x}_r$ can be obtained by the QR decomposition of $\mathbf{Q}$, followed by solving a triangular system. It is similar for $\mathbf{v}_{r,n}$. Finally $\mathbf{z}_{r,n}$ can be computed efficiently using FFT, as shown in Appendix B.

VI. Single-tone case

In this section, we calculate the perturbations of the parameter estimates for the single-tone signal

$$y(m_1, \ldots, m_N) = c \prod_{n=1}^{N} a_n^{m_n}.$$

As in [28], [26], we analyze the single-tone case in order to gain more insight in the optimal choice of the parameters L_n.

A. Specialising the general formulas

Since $\mathbf{a}^\dagger = \frac{1}{\|\mathbf{a}\|_2^2}\mathbf{a}^{\mathsf{H}}$ for any vector $\mathbf{a}$, and the matrices $\widetilde{\mathbf{F}}_n$ defined in (16) are just scalars, the steps 4–8 of N-D ESPRIT are equivalent to defining the estimates $\widehat{a}_n$ as

$$\widehat{a}_n = \widetilde{\mathbf{F}}_n = \frac{1}{\|\underset{n}{\underline{\widetilde{\mathbf{u}}}}\|_2^2} (\underset{n}{\underline{\widetilde{\mathbf{u}}}})^{\mathsf{H}} \overset{n}{\overline{\widetilde{\mathbf{u}}}},$$

where $\widetilde{\mathbf{u}}$ is the leading left singular vector of $\widetilde{\mathbf{H}}$. For the perturbations, the expression (26) can be also simplified. In this case, the matrices $\mathbf{P}$ and $\mathbf{Q}$ consist of a single column, which we denote by $\mathbf{p}$ and $\mathbf{q}$, respectively:

$$\mathbf{p} = \left(\mathbf{a}_1^{(L_1)} \boxtimes \cdots \boxtimes \mathbf{a}_n^{(L_n)} \boxtimes \cdots \boxtimes \mathbf{a}_N^{(L_N)}\right)$$

and

$$\mathbf{q} = \left(\mathbf{a}_1^{(K_1)} \boxtimes \cdots \boxtimes \mathbf{a}_n^{(K_n)} \boxtimes \cdots \boxtimes \mathbf{a}_N^{(K_N)}\right). \qquad (30)$$

Hence (26), becomes to

$$\Delta a_n = \frac{1}{c\|\underset{n}{\underline{\mathbf{p}}}\|_2^2 \|\mathbf{q}\|_2^2} \underset{n}{\underline{\mathbf{p}}}^{\mathsf{H}} (\overset{n}{\overline{\mathbf{I}}} - a_n \underset{n}{\underline{\mathbf{I}}}) \Delta \mathbf{H} \, \mathbf{q}^*.$$

From Lemma 6, we get the following expression.

Lemma 7: The first-order perturbation is expressed as

$$\Delta a_n = \mathbf{z}_n^{\mathsf{H}} \mathbf{e},$$

where

$$\mathbf{z}_n = \frac{1}{c}\Bigg\{ \frac{\left(\mathbf{a}_1^{(L_1)} \star \mathbf{a}_1^{(K_1)}\right)}{\|\mathbf{a}_1^{(L_1)}\|_2^2 \|\mathbf{a}_1^{(K_1)}\|_2^2} \boxtimes \cdots \boxtimes \frac{\left(\mathbf{a}_1^{(L_{n-1})} \star \mathbf{a}_1^{(K_{n-1})}\right)}{\|\mathbf{a}_{n-1}^{(L_{n-1})}\|_2^2 \|\mathbf{a}_{n-1}^{(K_{n-1})}\|_2^2}$$
$$\boxtimes \frac{\left(\left((\bar{\mathbf{I}}_{L_n} - a_n^* \underline{\mathbf{I}}_{L_n})^\mathsf{T} \mathbf{a}_n^{(L_n-1)} \right) \star \mathbf{a}_1^{(K_n)} \right)}{\|\mathbf{a}_1^{(L_n-1)}\|_2^2 \|\mathbf{a}_1^{(K_n)}\|_2^2}$$
$$\boxtimes \frac{\left(\mathbf{a}_1^{(L_{n+1})} \star \mathbf{a}_1^{(K_{n+1})}\right)}{\|\mathbf{a}_1^{(L_{n+1})}\|_2^2 \|\mathbf{a}_1^{(K_{n+1})}\|_2^2} \boxtimes \cdots \boxtimes \frac{\left(\mathbf{a}_1^{(L_N)} \star \mathbf{a}_1^{(K_N)}\right)}{\|\mathbf{a}_1^{(L_N)}\|_2^2 \|\mathbf{a}_1^{(K_N)}\|_2^2} \Bigg\},$$

where $\star$ denotes convolution of vectors.

Proof The proof can be found in Appendix C.

B. Expressions for the moments of the perturbations

Here we assume that $\mathbf{e}$ is zero-mean, and $\mathbb{E}\left\{\mathbf{e}\mathbf{e}^\mathsf{H}\right\} = \sigma_e^2 \mathbf{I}_{\prod_{n=1}^N M_n}$. Then, by Lemma 7, the variance of Δa_n can be expressed as

$$\mathbb{E}\left\{|\Delta a_n|^2\right\} = \sigma_e^2 \|\mathbf{z}_n\|_2^2 \tag{31}$$
$$= \frac{\sigma_e^2}{|c|^2} f(L_n, M_n, a_n) \prod_{\substack{i=1 \\ i \neq n}}^{N} g(L_i, M_i, a_i), \tag{32}$$

where the functions $f(L,M,a)$ and $g(L,M,a)$ are defined as

$$f(L,M,a) = \frac{\left\| \left((\bar{\mathbf{I}}_L - a^* \underline{\mathbf{I}}_L)^\mathsf{T} \mathbf{a}^{(L-1)} \right) \star \mathbf{a}^{(K)} \right\|^2}{\|\mathbf{a}^{(L-1)}\|_2^2 \|\mathbf{a}^{(K)}\|_2^4},$$
$$g(L,M,a) = \frac{\|\mathbf{a}^{(L)} \star \mathbf{a}^{(K)}\|^2}{\|\mathbf{a}^{(L)}\|^4 \|\mathbf{a}^{(K)}\|^4},$$

with $K = M - L + 1$ and $\mathbf{a}^{(\cdot)}$ defined as in (2).

C. Closed form expressions

Here we provide expressions that can be also found in [28]. We also give the full proofs that are absent in [28].

First, we note that the formula for $f(L,M,a)$ coincides with the formula for the variance of the first-order perturbation of the 1-D ESPRIT. Hence, the results from [46], [30], [27] can be used.

Proposition 2 ([27, eqn. (26)], [46, eqn. (4.16)], [30, Corollary 7]): In the undamped case ($|a| = 1$), the function $f(L,M,a)$ has the form

$$f(L,M,a) = \begin{cases} \frac{2}{K^2(L-1)}, & \text{if } L-1 \le \frac{M}{2} \text{ and } |a| = 1, \\ \frac{2}{K(L-1)^2}, & \text{if } L-1 \ge \frac{M}{2} \text{ and } |a| = 1. \end{cases} \tag{33}$$

In the damped case ($|a| \neq 1$), we have that

$$f(L,M,a) = (1-|a|^2)^3 \times$$
$$\begin{cases} \frac{1+|a|^{2K}}{(1-|a|^{2K})^2 \left(1-|a|^{2(L-1)}\right)}, & \text{if } L-1 \le \frac{M}{2} \text{ and } |a| \neq 1, \\ \frac{1+|a|^{2(L-1)}}{(1-|a|^{2K})\left(1-|a|^{2(L-1)}\right)^2}, & \text{if } L-1 \ge \frac{M}{2} \text{ and } |a| \neq 1. \end{cases} \tag{34}$$

We note that the function f is symmetric with respect to $L = \frac{M}{2} + 1$.

Proposition 3: In the undamped case the expression is given in (35).

$$g(L,M,a) = \begin{cases} \frac{1}{K}\left(1\text{-}\frac{L^2\text{-}1}{3LK}\right), & \text{if } L \le \frac{M+1}{2} \text{ and } |a| = 1, \\ \frac{1}{L}\left(1\text{-}\frac{K^2\text{-}1}{3LK}\right), & \text{if } L \ge \frac{M+1}{2} \text{ and } |a| = 1. \end{cases} \tag{35}$$

In the damped case ($|a| \neq 1$), the function $g(L,M,a)$ can be found as Eq. (36)

Proof The proof can be found in Appendix C.

The behavior of $f(L,M,a)$ and $g(L,M,a)$ for typical examples are shown in Figures 1 and 2. In Figure 3 and Figure 4 the analytic variances $\text{var}(\Delta\omega_a)$ and $\text{var}(\Delta\omega_b)$ are plotted. In Figure 5, total mean square error is plotted.

Remark 9: Based on Propositions 2–3, the optimal values for L_i can be obtained, in the same manner as it was shown in [28] for the 2-D case.

- If one wishes to minimize an individual variance $\mathbb{E}\left\{|\Delta a_n|^2\right\}$, then the optimal window sizes are chosen as follows: take L_j, $j \neq n$ as small as possible, and take the optimal L_n as in 1-D ESPRIT [27]. For $a_n = e^{-\alpha + \jmath\omega}$ and a white Gaussian noise it is given by

$$L_n^{(opt)} = \begin{cases} \frac{M_n}{3} + 1, & \alpha = 0, \\ \frac{M_n}{2} + \frac{1}{\alpha}\ln(\tan\frac{\pi - \arctan e^{\alpha M_n}}{3}) + 1, & \alpha \neq 0. \end{cases}$$

- If one wishes to minimize the total MSE for all modes, the optimal window sizes seem to be difficult to describe analytically.

Note that the optimal L_i depends on the type of noise, as in the 1D case [47].

Remark 10: For the undamped case ($|a_n| = 1$), the expression of the variance similar to (32) was independently and almost simultaneously obtained in [26, Theorem 3]. The question of finding optimal L_r for the undamped case is also discussed in [26]. Nevertheless, the expression for $g(L,M,a)$ given in (35) is much simpler than in [26, eqn. (40)].

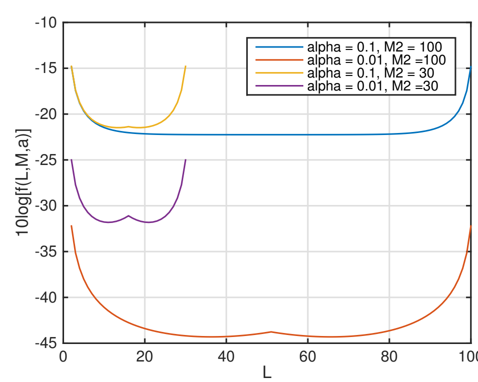

Fig. 1. Behavior of function $f(L,M,a)$ as a function of L for different values of M and damping factors.

VII. Simulations

Numerical simulations have been carried out to verify theoretical expressions and compare the performances of N-D ESPRIT and Fast N-D ESPRIT with the state-of the art

$$g(L,M,a) = (1-|a|^2) \times \begin{cases} \frac{-2L(1-|a|^2)\left(|a|^{2K}+|a|^{2L}\right)}{(1-|a|^{2L})^2(1-|a|^{2K})^2} + \frac{(1+|a|^{2K})(1+|a|^2)}{(1-|a|^{2L})(1-|a|^{2K})^2}, & \text{if } L \leq \frac{M+1}{2} \text{ and } |a| \neq 1 \\ \frac{-2K(1-|a|^2)\left(|a|^{2L}+|a|^{2K}\right)}{(1-|a|^{2L})^2(1-|a|^{2K})^2} + \frac{(1+|a|^{2L})(1+|a|^2)}{(1-|a|^{2L})^2(1-|a|^{2K})}, & \text{if } L \geq \frac{M+1}{2} \text{ and } |a| \neq 1 \end{cases} \tag{36}$$

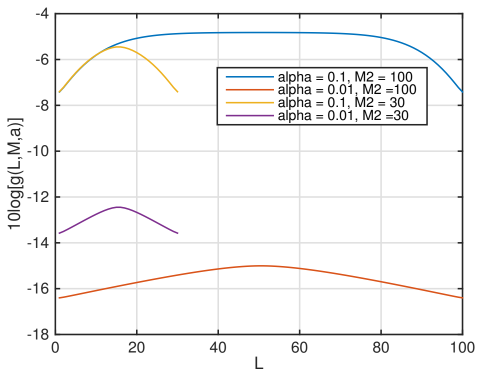

Fig. 2. Behavior of function $g(L,M,a)$ as a function of L for different values of M and damping factors.

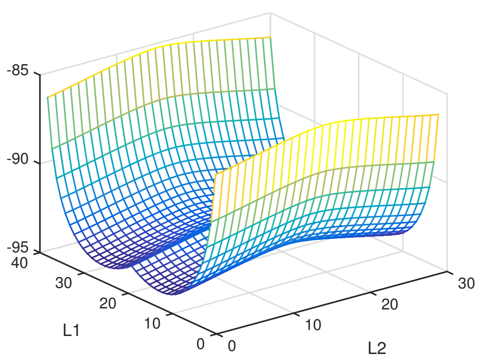

Fig. 3. Variance of $\Delta\omega_a$ as a function of L_1 and L_2

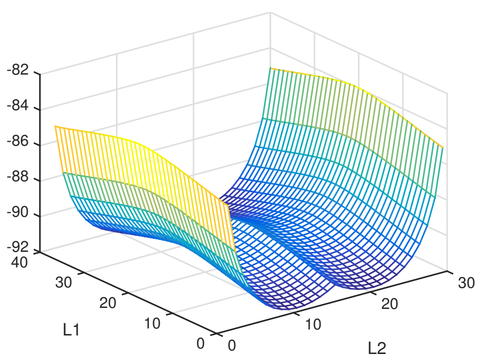

Fig. 4. Variance of $\Delta\omega_b$ as a function of L_1 and L_2

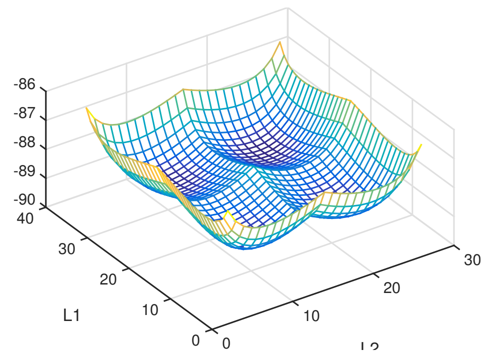

Fig. 5. tMSE as a function of L_1 and L_2

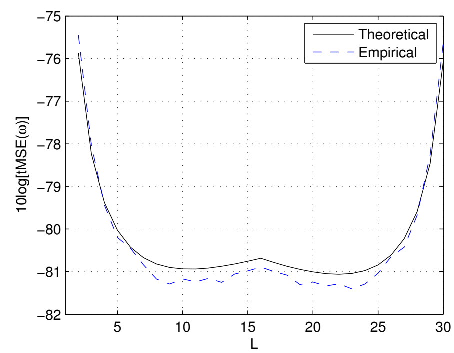

Fig. 6. Theoretical and empirical MSEs for 2-D ESPRIT versus L, ($L = L_1 = L_2$). $(\alpha_a, \omega_a) = (0.1, 0.2\pi)$, $(\alpha_b, \omega_b) = (0.1, 0.4\pi)$, $(M_1, M_2) = (30, 30)$, SNR = 40 dB.

methods, such as IMDF[7] [8], [9] and Tensor-ESPRIT [10], in the presence of white Gaussian noise. The performances are measured by the total mean square error (tMSE) on estimated parameters and the computational time. The total MSE is defined as $\text{tMSE} = \frac{1}{RF}\mathbb{E}_p\left\{\sum_{r=1}^{R}\sum_{f=1}^{F}(\xi_{f,r} - \hat{\xi}_{f,r})^2\right\}$ where $\hat{\xi}_{f,r}$ is an estimate of $\xi_{f,r}$, and $\mathbb{E}_p$ is the average on p Monte-Carlo trials. In our simulations, $\xi_{f,r}$ can be either a frequency or a damping factor.

A. N-D single-tone

In the first three experiments, we tend to verify the obtained closed-form expressions in the case of N-D single tone. We consider a 2-D damped single-tone signal with parameters $(\alpha_a, \omega_a) = (0.1, 0.2\pi)$ and $(\alpha_b, \omega_b) = (0.1, 0.4\pi)$. The SNR is fixed to 40 dB. Figure 6 shows the total MSE and its theoretical value obtained from 200 Monte Carlo trials with $(M_1, M_2) = (30, 30)$. Since it is difficult to see the difference between the two curves in a 3-D plot, we show only one diagonal slice of the 3-D plot corresponding to $L_1 = L_2$. We can observe that the theoretical tMSEs are close to the estimated ones. In the second example, we repeat the same experience with $(M_1, M_2) = (100, 100)$ using the fast N-D ESPRIT method. The obtained results are reported in Figure 7, where it can bee seen that theoretical tMSEs are again close to the estimated ones.

In the third example, the same parameters of the modes are used but the SNR is varying. The parameters (L_1, L_2) are

[7] IMDF is an improved version of the multidimensional folding approach initially proposed to maximize identifiability in [48].

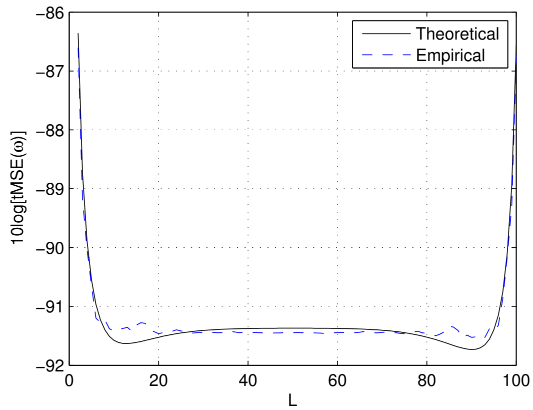

Fig. 7. Theoretical and empirical MSEs for 2-D ESPRIT (fast SVD) versus L, $(L = L_1 = L_2)$. $(\alpha_a, \omega_a) = (0.1, 0.2\pi)$, $(\alpha_b, \omega_b) = (0.1, 0.4\pi)$, $(M_1, M_2) = (100, 100)$, SNR = 40 dB.

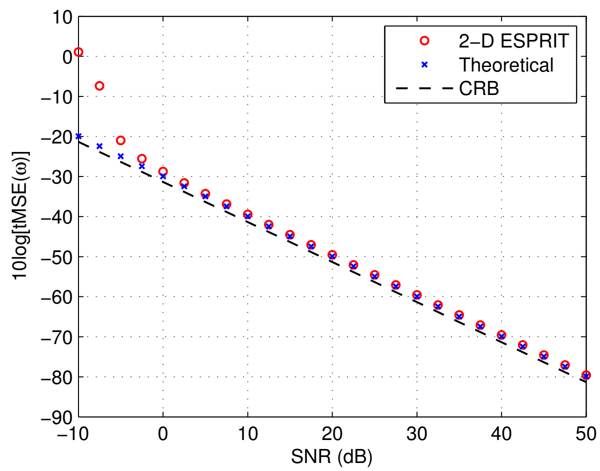

Fig. 8. Theoretical and empirical tMSEs for 2-D ESPRIT versus SNR. $(L_1, L_2) = (4, 4)$. $(\alpha_a, \omega_a) = (0.1, 0.2\pi)$, $(\alpha_b, \omega_b) = (0.1, 0.4\pi)$, $(M_1, M_2) = (10, 10)$.

set to $(4, 4)$. The obtained results are depicted in Figure 8. We observe that the theoretical results are almost equal to empirical ones beyond a threshold, which is here -5 dB.

For a fast implementation of N-D ESPRIT (denoted as "Fast N-D ESPRIT"), we use the implementation of the TSVD in the PROPACK package [42] developed within the PhD thesis [49]. We use the updated version of the PROPACK package available as a part of the SVT software [50].

B. Multiple tones N-D modal signals

Experiments of this section verify theoretical expressions of the variances in the multiple tones case, and compare them with empirical results of N-D ESPRIT, Fast N-D ESPRIT, IMDF and Tensor ESPRIT. CRB are also reported.

d) Experiment 4: In this experiment, we simulate a 2-D signal of size 10×10 containing two modes whose parameters are given by $(\alpha_{1,1}, \omega_{1,1}) = (0.01, 0.2\pi)$, $(\alpha_{2,1}, \omega_{2,1}) = (0.01, 0.6\pi)$, $(\alpha_{1,2}, \omega_{1,2}) = (0.01, 0.3\pi)$, $(\alpha_{2,2}, \omega_{2,2}) = (0.01, 0.8\pi)$, $(M_1, M_2) = (10, 10)$. (L_1, L_2) are set to $(4, 4)$. Figure 9 shows the obtained results. We can see that N-D

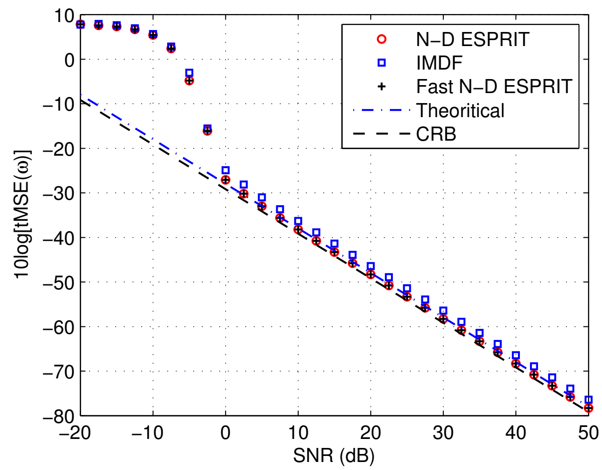

Fig. 9. Theoretical and empirical tMSEs versus SNR. 2-D damped signal containing two tones. $(L_1, L_2) = (4, 4)$. $(\alpha_{1,1}, \omega_{1,1}) = (0.01, 0.2\pi)$, $(\alpha_{2,1}, \omega_{2,1}) = (0.01, 0.6\pi)$, $(\alpha_{1,2}, \omega_{1,2}) = (0.01, 0.3\pi)$, $(\alpha_{2,2}, \omega_{2,2}) = (0.01, 0.8\pi)$, $(M_1, M_2) = (10, 10)$.

ESPRIT and Fast N-D ESPRIT have the the same results, which are almost equal to theoretical ones beyond 0 dB. We can also remark that N-D ESPRIT outperforms slightly IMDF.

e) Experiment 5: a 3-D signal of size $10 \times 10 \times 10$ containing two modes is simulated with the parameters given in table I. The results are shown on Figure 10. In this experiment,

TABLE I
3-D SIGNAL WITH TWO MODES

r	$\omega_{r,1}$	$\alpha_{r,1}$	$\omega_{r,2}$	$\alpha_{r,2}$	$\omega_{r,3}$	$\alpha_{r,3}$	c_r
1	0.2π	0.01	0.3π	0.01	0.26π	0.01	1
2	0.6π	0.01	0.8π	0.015	0.2π	0.01	1

N-D ESPRIT outperforms IMDF and and have almost similar results as those obtained by theoretical expressions.

f) Experiment 6: Results on a 3-D signal of size $10 \times 10 \times 10$ containing three modes are given in Figure 11. Parameters of the simulated modes are given in table II. Here we observe

TABLE II
3-D SIGNAL WITH THREE MODES

r	$\omega_{r,1}$	$\alpha_{r,1}$	$\omega_{r,2}$	$\alpha_{r,2}$	$\omega_{r,3}$	$\alpha_{r,3}$	c_r
1	0.2π	0.01	0.3π	0.01	0.26π	0.01	1
2	0.6π	0.01	0.8π	0.015	0.2π	0.01	1
3	0.4π	0.01	π	0.01	0.6π	0.01	1
4	0.5π	0.01	0.7π	0.01	0.8π	0.01	1
5	0.7π	0.01	0.9π	0.01	0.5π	0.01	1

that N-D ESPRIT outperforms IMDF and the gap between them become bigger compared to the previous experiment (experiment with two tones). As shown by the results in Figure 11, the Tensor ESPRIT algorithm does not yield an improvement in our case.

C. Computational time

Figure 12 shows the CPU time results of N-D ESPRIT, FAST N-D ESPRIT and IMDF algorithms versus M_1 for a

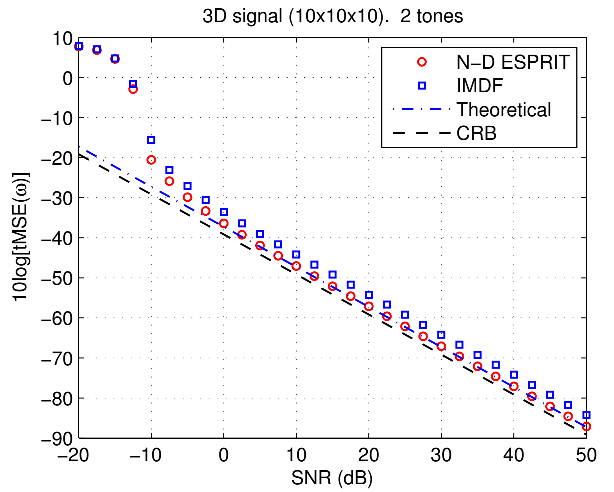

Fig. 10. Theoretical and empirical tMSEs versus SNR. 3-D damped signal containing two tones. $(L_1, L_2, L_3) = (4, 4, 4)$, $(M_1, M_2, M_3) = (10, 10, 10)$.

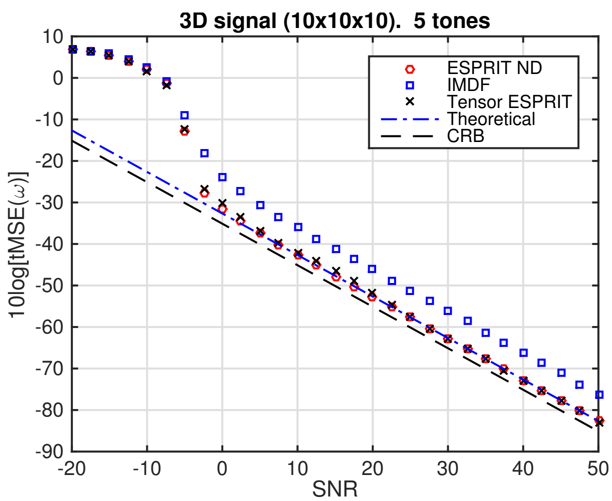

Fig. 11. Theoretical and empirical tMSEs versus SNR. 3-D damped signal containing five tones. $(L_1, L_2, L_3) = (4, 4, 4)$, $(M_1, M_2) = (10, 10, 10)$.

2-D damped signal containing two modes with $M_2 = 10$. We observe that the FAST N-D ESPRIT involves a low computational complexity compared to TPUMA and Tensor-ESPRIT when M_1 is large. This is due to the fast computation of the truncated SVD (see Section IV-D).

VIII. Conclusion

The N-D ESPRIT algorithm is implemented by storing the multidimensional data into a multilevel Hankel matrix. A fast version of N-D ESPRIT based on partial SVD has been proposed to handle efficiently large N-D signals. A first-order perturbation analysis has been carried out, which led to: i) simpler expressions that do not involve the SVD factors ii) closed form expression of the variances of parameters (damping factors and frequencies) in the single tone case. It has then been shown that variables $L_n, n = 1, \ldots, N$ separate in each of these variances.

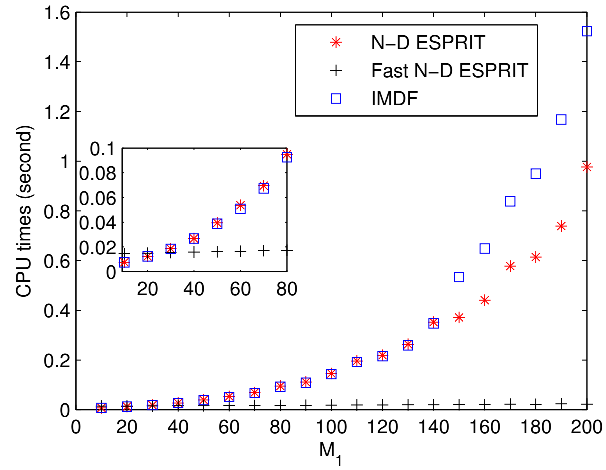

Fig. 12. Average CPU time for a single run under $M_2 = 10$ and $R = 2$.

IX. Acknowledgement

We would like to thank the associate editor and anonymous reviewers for very important remarks that helped to improve considerably the quality of the manuscript.

Appendix

A. Properties of multilevel Hankel matrices

It is often convenient to use the selection matrices to construct the HbH matrix.

Given $M_n, n = 1, \ldots, N$, let us define a set of selection matrices

$$\mathbf{J}_{k_n}^{L_n} = \begin{bmatrix} \mathbf{0}_{L_n \times (k_n - 1)} & \mathbf{I}_{L_n} & \mathbf{0}_{L_n \times (K_n - k_n)} \end{bmatrix} \quad (37)$$

$$\mathbf{J}_{k_1,k_2,\ldots,k_N} = \mathbf{J}_{k_1}^{L_1} \boxtimes \mathbf{J}_{k_2}^{L_2} \boxtimes \cdots \boxtimes \mathbf{J}_{k_N}^{L_N} \quad (38)$$

where $\mathbf{J}_{k_n}^{L_n}$ and $\mathbf{J}_{k_1,k_2,\ldots,k_N}$ are of sizes $L_n \times M_n$ and $\prod_{n=1}^{N} L_n \times \prod_{n=1}^{N} M_n$, respectively; and K_n are defined as previously. It is easy to verify that

$$\mathbf{J}_{k_1,k_2,\ldots,k_N} \mathbf{y} = \mathbf{y}^{(k_1,\ldots,k_N)},$$

where $\mathbf{y} = \mathrm{vec_r}\{\mathcal{Y}\}$ and $\mathbf{y}^{(k_1,\ldots,k_N)}$ is defined as in the previous subsection.

The multilevel Hankel matrix has also the following multi-level structure:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_{K_1-1} \\ \mathbf{H}_1 & \mathbf{H}_2 & \cdots & \mathbf{H}_{K_1} \\ \vdots & \vdots & & \vdots \\ \mathbf{H}_{L_1-1} & \mathbf{H}_{L_1} & \cdots & \mathbf{H}_{M_1-1} \end{bmatrix}, \quad (39)$$

where for $r = 1, \ldots, N-1$ the block matrices $\mathbf{H}_{m_1,\ldots,m_r}$ are defined recursively

$$\mathbf{H}_{m_1,\ldots,m_r} = \begin{bmatrix} \mathbf{H}_{m_1,\ldots,m_r,0} & \mathbf{H}_{m_1,\ldots,m_r,1} & \cdots & \mathbf{H}_{m_1,\ldots,m_r,K_{r+1}-1} \\ \mathbf{H}_{m_1,\ldots,m_r,1} & \mathbf{H}_{m_1,\ldots,m_r,2} & \cdots & \mathbf{H}_{m_1,\ldots,m_r,K_{r+1}} \\ \vdots & \vdots & & \vdots \\ \mathbf{H}_{m_1,\ldots,m_r,L_{r+1}-1} & \mathbf{H}_{m_1,\ldots,m_r,L_{r+1}} & \cdots & \mathbf{H}_{m_1,\ldots,m_r,M_{r+1}-1} \end{bmatrix} \quad (40)$$

and the blocks of the last level are just scalars

$$\mathbf{H}_{m_1,\ldots,m_N} = y(m_1, \ldots, m_N).$$

Algorithm 1: MH matrix-vector multiplication

input : An N-D signal $\mathcal{Y}$, vector $\mathbf{x} \in \mathbb{C}^{K_1 \cdots K_N}$.
output: Matrix-vector product $\mathbf{w} = \mathbf{H}\mathbf{x}$.

1) Construct the tensorization $\mathcal{X} \in \mathbb{C}^{K_1 \times \cdots \times K_N}$ of $\mathbf{x}$.
2) Construct the element wise conjugate to $\mathcal{X}$ and padded by zeros tensor $\mathcal{X}'^* \in \mathbb{C}^{M_1 \times \cdots \times M_N}$, so that
$$\mathcal{X}'^*_{1:K_1,1:K_2,\ldots,1:K_N} = \mathcal{X}^*_{1:K_1,1:K_2,\ldots,1:K_N}$$
and other elements are zeros.
3) Compute the N-D FFT of $\mathcal{Y}$ and $\mathcal{X}'$
$$\widehat{\mathcal{X}'^*} := \mathcal{X}'^* \bullet_1 \mathcal{F}_{M_1} \bullet_2 \mathcal{F}_{M_1} \cdots \bullet_N \mathcal{F}_{M_N},$$
$$\widehat{\mathcal{Y}} := \mathcal{Y} \bullet_1 \mathcal{F}_{M_1} \bullet_2 \mathcal{F}_{M_1} \cdots \bullet_N \mathcal{F}_{M_N},$$
4) Compute the following tensor
$$\widehat{\mathcal{W}'} := (\widehat{\mathcal{Y}} \boxdot \widehat{\mathcal{X}'^*})^*,$$
where $\boxdot$ is the elementwise (Hadamard) product.
5) Compute the inverse FFT of $\widehat{\mathcal{W}'}$:
$$\mathcal{W}' := \widehat{\mathcal{W}'} \bullet_1 \mathcal{F}^{\mathsf{H}}_{M_1} \bullet_2 \mathcal{F}^{\mathsf{H}}_{M_1} \cdots \bullet_N \mathcal{F}^{\mathsf{H}}_{M_N}.$$
6) Extract $\mathbf{w}$ by truncating and vectorizing $\mathcal{W}'$
$$\mathbf{w} = \mathrm{vec}_{\mathrm{r}}\{(\mathcal{W}')_{1:L_1,1:L_2,\ldots,1:L_N}\}.$$

B. Matrix-vector products for MH matrices

Let the MH matrix $\mathbf{H}$ be given in (7). Assume that $\mathbf{x} \in \mathbb{C}^{K_1 \cdots K_N}$. Consider the matrix-vector product
$$\mathbf{w} = \mathbf{H}\mathbf{x} \in \mathbb{C}^{L_1 \cdots L_N}.$$
If we tensorize the vectors $\mathbf{w}$ and $\mathbf{x}$ to $\mathcal{W} \in \mathbb{C}^{L_1 \times \cdots \times L_N}$ and $\mathcal{X} \in \mathbb{C}^{K_1 \times \cdots \times K_N}$ (such that $\mathbf{x} = \mathrm{vec}_{\mathrm{r}}\{\mathcal{X}\}$ and $\mathbf{w} = \mathrm{vec}_{\mathrm{r}}\{\mathcal{W}\}$), then we see that the tensor $\mathcal{W}$ is obtained via a "filtering operation", i.e.
$$(\mathcal{W})_{i_1,\ldots,i_N} = \sum_{j_1,\ldots,j_N=1}^{K_1,\ldots,K_N} (\mathcal{Y})_{i_1+j_1,\ldots,i_N+j_N} (\mathcal{X})_{j_1,\ldots,j_N}. \quad (41)$$
Hence, the matrix-vector product can be computed using the FFT. In what follows, we define by $\mathcal{F}_M \in \mathbb{C}^{M \times M}$ the matrix of the Discrete Fourier Transfrom.

Algorithm 1 has the following advantages:

- It has computational complexity $O(M \log M)$ due to the use of FFT.
- It does not require the storage of the matrix $\mathbf{H}$, so the storage complexity is minimal.
- The N-D FFT of $\mathcal{Y}$ can be precomputed, and can be used for matrix-vector multiplication with different choice of $L_1, \ldots, L_N$.
- The matrices can be padded by zeros, so that the tensor size can be adjusted for better performance of the FFT.

In a similar way we can compute the bilinear transform of type (29), as shown by Lemma 6.

Proof of Lemma 6 Let $\mathbf{x} \in \mathbb{C}^{K_1 \cdots K_N}$ and $\mathbf{v} \in \mathbb{C}^{L_1 \cdots L_N}$. We need to show that for any vector $\mathbf{y} \in \mathbb{C}^{M_1 \cdots M_N}$ and the corresponding matrix $\mathbf{H}$ given in (7), it holds that
$$\mathbf{v}^{\mathsf{H}} \mathbf{H} \mathbf{x}^* = \mathbf{z}^{\mathsf{H}} \mathbf{y}, \quad (42)$$

Algorithm 2: MH bilinear transform

input : Vector $\mathbf{x} \in \mathbb{C}^{K_1 \cdots K_N}$, $\mathbf{v} \in \mathbb{C}^{L_1 \cdots L_N}$
output: The vector $\mathbf{z}$ in the bilinear operation (42).

1) Construct the tensorizations $\mathcal{X} \in \mathbb{C}^{K_1 \times \cdots \times K_N}$ and $\mathcal{V} \in \mathbb{C}^{L_1 \times \cdots \times L_N}$.
2) Construct the padded by zeros tensors $\mathcal{X}', \mathcal{V}' \in \mathbb{C}^{M_1 \times \cdots \times M_N}$
$$\mathcal{X}'_{1:K_1,1:K_2,\ldots,1:K_N} = \mathcal{X}_{1:K_1,1:K_2,\ldots,1:K_N},$$
$$\mathcal{V}'_{1:L_1,1:L_2,\ldots,1:L_N} = \mathcal{V}_{1:L_1,1:L_2,\ldots,1:L_N},$$
and other elements are zeros.
3) Compute the N-D FFT of $\mathcal{V}'$ and $\mathcal{X}'$
$$\widehat{\mathcal{X}'} := \mathcal{X}' \bullet_1 \mathcal{F}_{M_1} \bullet_2 \mathcal{F}_{M_1} \cdots \bullet_N \mathcal{F}_{M_N},$$
$$\widehat{\mathcal{V}'} := \mathcal{V}' \bullet_1 \mathcal{F}_{M_1} \bullet_2 \mathcal{F}_{M_1} \cdots \bullet_N \mathcal{F}_{M_N},$$
4) Compute the inverse FFT of the Hadamard product
$$\mathcal{Z} := (\mathcal{V}' \circ \mathcal{X}') \bullet_1 \mathcal{F}^{\mathsf{H}}_{M_1} \bullet_2 \mathcal{F}^{\mathsf{H}}_{M_1} \cdots \bullet_N \mathcal{F}^{\mathsf{H}}_{M_N}.$$
5) Extract $\mathbf{w}$ by vectorizing $\mathcal{Z}$
$$\mathbf{z} = \mathrm{vec}_{\mathrm{r}}\{\mathcal{Z}\}.$$

$\mathcal{X}$, $\mathcal{V}$ are tensorizations of $\mathbf{x}$, $\mathbf{v}$, and $\mathbf{z} = \mathrm{vec}_{\mathrm{r}}\{\mathcal{Z}\}$, where $\mathcal{Z} := \mathcal{X} \star \mathcal{V}$, i.e.
$$(\mathcal{Z})_{m_1,\ldots,m_r} := \sum_{i_1+j_1=m_1,\ldots,i_N+j_N=m_N}^{K_1,\ldots,K_N} (\mathcal{X})_{i_1,\ldots,i_N} (\mathcal{V})_{j_1,\ldots,j_N}.$$
Indeed, from (41), we have that
$$\mathbf{v}^{\mathsf{H}} \mathbf{H} \mathbf{x}^* = \sum_{i_1,\ldots,i_N=1}^{L_1,\ldots,L_N} \sum_{j_1,\ldots,j_N=1}^{K_1,\ldots,K_N} (\mathcal{Y})_{i_1+j_1,\ldots,i_N+j_N} (\mathcal{X})_{j_1,\ldots,j_N} (\mathcal{V})_{i_1,\ldots,i_N},$$
which completes the proof. □

Finally, using the relation with the convolution, the vector-matrix-vector product (42) can be computed efficiently using the FFT, as shown in Algorithm 2.

C. Proofs of lemmas and propositions

Proof of Lemma 4 For simplicity, we denote $\mathbf{A} = \underset{\text{n—}}{\mathbf{U}_s}$ and $\mathbf{B} = \overset{\text{n—}}{\mathbf{U}}$. By the rule of differentiation of the product, $\Delta(\mathbf{A}^\dagger \mathbf{B}) = \mathbf{A}^\dagger \Delta \mathbf{B} + \Delta(\mathbf{A}^\dagger)\mathbf{B}$. Next, since $\mathbf{A}$ is full column rank, $\Delta(\mathbf{A}^\dagger)$ can be expressed using [45, Proposition 1]:
$$\begin{aligned}\Delta(\mathbf{A}^\dagger \mathbf{B}) &= \mathbf{A}^\dagger \Delta \mathbf{B} \\ &+ \left(-\mathbf{A}^\dagger \Delta \mathbf{A} \mathbf{A}^\dagger + (\mathbf{A}^{\mathsf{H}} \mathbf{A})^{-1} \Delta \mathbf{A}^{\mathsf{H}} (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\right) \mathbf{B}.\end{aligned} \quad (43)$$
Since $\mathbf{A}$ and $\mathbf{B}$ span the same column space, $(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{B} = 0$, and (43) becomes
$$\Delta(\mathbf{A}^\dagger \mathbf{B}) = \mathbf{A}^\dagger (\Delta \mathbf{B} - \Delta \mathbf{A} \mathbf{A}^\dagger \mathbf{B}). \quad (44)$$
Finally, since $\mathbf{F}_n = \mathbf{A}^\dagger \mathbf{B}$, (44) can be simplified to (23). □

Proof of Lemma 5 First, from (19) and the rule of differentiation of inverses $\Delta(\mathbf{T}^{-1}) = -\mathbf{T}^{-1} \Delta \mathbf{T} \mathbf{T}^{-1}$, we have that:
$$\begin{aligned}\Delta \mathbf{D}_n &= \mathbf{T}^{-1} \Delta \mathbf{F}_n \mathbf{T} + \mathbf{T}^{-1} \mathbf{F}_n \Delta \mathbf{T} - \mathbf{T}^{-1} \Delta \mathbf{T} \mathbf{T}^{-1} \mathbf{F}_n \mathbf{T} \\ &= \mathbf{T}^{-1} \Delta \mathbf{F}_n \mathbf{T} + \mathbf{D}_n \mathbf{T}^{-1} \Delta \mathbf{T} - \mathbf{T}^{-1} \Delta \mathbf{T} \mathbf{D}_n.\end{aligned}$$

Next, we denote by $\mathbf{b}_r$ the r-th unit vector, and write

$$\begin{aligned}\Delta a_{r,n} &= \mathbf{b}_r^{\mathsf{T}} \Delta \mathbf{D}_n \mathbf{b}_r \\ &= \boldsymbol{\tau}_r^{\mathsf{T}} \Delta \mathbf{F}_n \mathbf{t}_r + \mathbf{b}_r^{\mathsf{T}} \mathbf{D}_n \mathbf{T}^{-1} \Delta \mathbf{T} \mathbf{b}_r - \mathbf{b}_r^{\mathsf{T}} \mathbf{T}^{-1} \Delta \mathbf{T} \mathbf{D}_n \mathbf{b}_r\end{aligned} \tag{45}$$

Since $\mathbf{b}_r^{\mathsf{T}} \mathbf{D}_n = a_{r,n} \mathbf{b}_r^{\mathsf{T}}$ and $\mathbf{D}_n \mathbf{b}_r = a_{r,n} \mathbf{b}_r$, the last two terms in (45) cancel, and eqn. (24) takes place. □

Proof of Corollary 1 Next, we combine (22) and (23). For simplicity, denote $\mathbf{C} = \Delta \mathbf{H}\, \mathbf{V}_s \boldsymbol{\Sigma}_s^{-1}$. Then

$$\begin{aligned}\Delta \mathbf{F}_n &= (\underset{n}{\underline{\mathbf{U}}}_s)^{\dagger} (\overset{n}{\overline{\mathbf{I}}}\, \Delta \mathbf{U}_s - \underset{n}{\underline{\mathbf{I}}}\, \Delta \mathbf{U}_s \mathbf{F}_n) \\ &= (\underset{n}{\underline{\mathbf{U}}}_s)^{\dagger} (\overset{n}{\overline{\mathbf{I}}}\, (\mathbf{I} - \mathbf{U}_s \mathbf{U}_s^{\mathsf{H}}) \mathbf{C} - \underset{n}{\underline{\mathbf{I}}}\, (\mathbf{I} - \mathbf{U}_s \mathbf{U}_s^{\mathsf{H}}) \mathbf{C} \mathbf{F}_n) \\ &\quad + (\underset{n}{\underline{\mathbf{U}}}_s)^{\dagger} (\overset{n}{\overline{\mathbf{U}}}_s \mathbf{R} - \underset{n}{\underline{\mathbf{U}}}_s \mathbf{R} \mathbf{F}_n).\end{aligned}$$

By expanding the parentheses and using the identities

$$(\underset{n}{\underline{\mathbf{U}}}_s)^{\dagger} \overset{n}{\overline{\mathbf{U}}}_s = \mathbf{F}_n, \quad (\underset{n}{\underline{\mathbf{U}}}_s)^{\dagger} \underset{n}{\underline{\mathbf{U}}}_s = \mathbf{I},$$

we get

$$\Delta \mathbf{F}_n = (\underset{n}{\underline{\mathbf{U}}}_s)^{\dagger} (\overset{n}{\overline{\mathbf{I}}}\, \mathbf{C} - \underset{n}{\underline{\mathbf{I}}}\, \mathbf{C} \mathbf{F}_n) + \mathbf{G} \mathbf{F}_n - \mathbf{F}_n \mathbf{G}, \tag{46}$$

where $\mathbf{G} = (\mathbf{U}_s^{\mathsf{H}} \mathbf{C} - \mathbf{R})$. Next, we combine (46) and (24).

$$\begin{aligned}\Delta a_{r,n} &= \boldsymbol{\tau}_r^{\mathsf{T}} (\underset{n}{\underline{\mathbf{U}}}_s)^{\dagger} (\overset{n}{\overline{\mathbf{I}}}\, \mathbf{C} - \underset{n}{\underline{\mathbf{I}}}\, \mathbf{C} \mathbf{F}_n) \mathbf{t}_r \\ &\quad + \boldsymbol{\tau}_r^{\mathsf{T}} (\mathbf{G} \mathbf{F}_n - \mathbf{F}_n \mathbf{G}) \mathbf{t}_r.\end{aligned} \tag{47}$$

Since $\mathbf{F}_n \mathbf{t}_r = a_{r,n} \mathbf{t}_r$ and $\boldsymbol{\tau}_r^{\mathsf{T}} \mathbf{F}_n = a_{r,n} \boldsymbol{\tau}_r^{\mathsf{T}}$, the last term in (47) vanishes, and eqn. (47) is simplified to (25). □

Proof of Proposition 1 From equation between (13) and (14), we have that

$$\mathbf{U}_s = \mathbf{P} \mathbf{T}^{-1}, \quad \boldsymbol{\Sigma}_s \mathbf{V}_s^{\mathsf{H}} = \mathbf{T}\, \mathrm{Diag}(\mathbf{c}) \mathbf{Q}^{\mathsf{T}}.$$

where the matrices $\mathbf{P}$ and $\mathbf{Q}$ are defined in (8). Next, since

$$(\boldsymbol{\Sigma}_s \mathbf{V}_s^{\mathsf{H}})^{\dagger} = \mathbf{V}_s \boldsymbol{\Sigma}_s^{-1},$$

and from properties of the pseudoinverse, we have that

$$\begin{aligned}\Delta a_{r,n} &= \boldsymbol{\tau}_r^{\mathsf{T}} (\underset{n}{\underline{\mathbf{U}}}_s)^{\dagger} (\overset{n}{\overline{\mathbf{I}}} - a_{r,n} \underset{n}{\underline{\mathbf{I}}}) \Delta \mathbf{H}\, (\boldsymbol{\Sigma}_s \mathbf{V}_s^{\mathsf{H}})^{\dagger} \mathbf{t}_r \\ &= \boldsymbol{\tau}_r^{\mathsf{T}} (\underset{n}{\underline{\mathbf{P}}}\, \mathbf{T}^{-1})^{\dagger} (\overset{n}{\overline{\mathbf{I}}} - a_{r,n} \underset{n}{\underline{\mathbf{I}}}) \Delta \mathbf{H}\, (\mathbf{T}\, \mathrm{Diag}(\mathbf{c}) \mathbf{Q}^{\mathsf{T}})^{\dagger} \mathbf{t}_r \\ &= \boldsymbol{\tau}_r^{\mathsf{T}} \mathbf{T} \underset{n}{\underline{\mathbf{P}}}^{\dagger} (\overset{n}{\overline{\mathbf{I}}} - a_{r,n} \underset{n}{\underline{\mathbf{I}}}) \Delta \mathbf{H}\, (\mathbf{Q}^{\mathsf{T}})^{\dagger}\, \mathrm{Diag}(\mathbf{c})^{-1} \mathbf{T}^{-1} \mathbf{t}_r \\ &= \frac{1}{c_r} \mathbf{b}_r^{\top} \underset{n}{\underline{\mathbf{P}}}^{\dagger} (\overset{n}{\overline{\mathbf{I}}} - a_{r,n} \underset{n}{\underline{\mathbf{I}}}) \Delta \mathbf{H}\, (\mathbf{Q}^{\mathsf{T}})^{\dagger} \mathbf{b}_r.\end{aligned}$$

□

Proof of Lemma 7 First, we remark that

$$\underset{n}{\underline{\mathbf{P}}}^{\mathsf{H}} (\overset{n}{\overline{\mathbf{I}}} - a_n \underset{n}{\underline{\mathbf{I}}}) = \left(\mathbf{a}_1^{(L_1)} \boxtimes \cdots \boxtimes \mathbf{b} \boxtimes \cdots \boxtimes \mathbf{a}_N^{(L_N)}\right)^{\mathsf{H}}, \tag{48}$$

where $\mathbf{b} = (\overline{\mathbf{I}}_{L_n} - a_n^* \underline{\mathbf{I}}_{L_n})^{\mathsf{T}} \mathbf{a}_n^{(L_n-1)}$. Next, the convolution of two rank-one tensors is a rank-one tensor:

$$(\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_N) \star (\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_N) = (\mathbf{x}_1 \star \mathbf{v}_1) \otimes \cdots \otimes (\mathbf{x}_N \star \mathbf{v}_N).$$

Hence, by applying (1) and Lemma 6 to (48) and (30), we get the desired result. □

Proof of Proposition 3 We define $L_{**} = \min(L, K)$ and $K_{**} = \max(L, K)$. Then the vector in the denominator can be explicitly written as

$$\begin{aligned}\mathbf{a}^{(L)} \star \mathbf{a}^{(K)} = \Big[&1, 2a, \ldots, L_{**} a^{(L_{**}-1)}, \\ &L_{**} a^{(L_{**})}, \ldots, L_{**} a^{(K_{**}-2)}, \\ &L_{**} a^{(K_{**}-1)}, \ldots, 2a^{(M-1)}, a^{(M-1)}\Big].\end{aligned}$$

Next, we consider each case separately.

a) Undamped case: In this case,

$$\|\mathbf{a}^{(L)}\|^4 = L^2, \quad \|\mathbf{a}^{(K)}\|^4 = K^2, \quad \text{and}$$

$$\begin{aligned}\|\mathbf{a}^{(L)} \star \mathbf{a}^{(K)}\|^2 &= L_{**}^2 (M - 2L_{**}) + 2 \sum_{k=1}^{L_{**}} k^2 \\ &= L_{**}^2 (K_{**} - L_{**} - 1) + \frac{2L_{**}^3}{3} + L_{**}^2 + \frac{L_{**}}{3} \\ &= L_{**}^2 K_{**} - \frac{L_{**}(L_{**}^2 - 1)}{3}.\end{aligned}$$

By combining all these expressions together, we get eqn. (35).

b) Damped case: Then the squared norm of the convolution is equal to

$$\begin{aligned}&\|\mathbf{a}^{(L)} \star \mathbf{a}^{(K)}\|^2 = \sum_{k=1}^{L_{**}} k^2 |a|^{2(k-1)} \\ &+ L_{**}^2 \sum_{k=L_{**}+1}^{K_{**}-1} |a|^{2(k-1)} + \sum_{k=K_{**}}^{M} (M - k + 1)^2 |a|^{2(k-1)} \\ &= \sum_{k=1}^{L_{**}} k^2 |a|^{2(k-1)} + L_{**}^2 \sum_{k=L_{**}+1}^{M-L_{**}} |a|^{2(k-1)} \\ &\quad + |a|^{2(M-1)} \sum_{j=1}^{L_{**}} j^2 \frac{1}{|a|^{2(j-1)}}.\end{aligned}$$

In order to simplify the expression, we use the fact that for $\rho \neq 1$

$$\sum_{k=1}^{L} k^2 \rho^{k-1} = L^2 \frac{\rho^L}{\rho - 1} - 2L \frac{\rho^L}{(\rho - 1)^2} + \frac{(\rho^L - 1)(\rho + 1)}{(\rho - 1)^3}.$$

By subsituting $\rho = |a|^2$ and get

$$\begin{aligned}&\sum_{k=1}^{L_{**}} k^2 |a|^{2(k-1)} = \\ &L_{**}^2 \frac{|a|^{2L_{**}}}{|a|^2 - 1} - 2L_{**} \frac{|a|^{2L_{**}}}{(|a|^2 - 1)^2} + \frac{(|a|^{2L_{**}} - 1)(|a|^2 + 1)}{(|a|^2 - 1)^3}.\end{aligned}$$

Next we take $\rho = |a|^{-2}$, and get

$$
\begin{aligned}
&|a|^{2(M-1)} \sum_{j=1}^{L_{**}} j^2 \left(\frac{1}{|a|}\right)^{2(j-1)} = \\
&= |a|^{2(M-1)} \Big(L_{**}^2 \frac{|a|^{-2L_{**}}}{|a|^{-2}-1} - 2L_{**} \frac{|a|^{-2L_{**}}}{(|a|^{-2}-1)^2} \\
&\qquad + \frac{(|a|^{-2L_{**}}-1)(|a|^{-2}+1)}{(|a|^{-2}-1)^3} \Big) \\
&= -L_{**}^2 \frac{|a|^{2(K_{**}-1)}}{|a|^2-1} - 2L_{**} \frac{|a|^{2K_{**}}}{(|a|^2-1)^2} \\
&+ \frac{|a|^{2K_{**}}(|a|^{2L_{**}}-1)(|a|^2+1)}{(|a|^2-1)^3}.
\end{aligned}
$$

Next, due to the fact that

$$
L_{**}^2 \sum_{k=L_{**}+1}^{K_{**}-1} |a|^{2(k-1)} = L_{**}^2 \frac{|a|^{2(K_{**}-1)} - |a|^{2L_{**}}}{|a|^2-1},
$$

after cancellations of some terms, we have

$$
\begin{aligned}
\|\mathbf{a}^{(L)} \star \mathbf{a}^{(K)}\|_2^2 &= -2L_{**} \frac{|a|^{2K_{**}} + |a|^{2L_{**}}}{(|a|^2-1)^2} \\
&+ \frac{(|a|^{2K_{**}}+1)(|a|^{2L_{**}}-1)(|a|^2+1)}{(|a|^2-1)^3}.
\end{aligned} \tag{49}
$$

Finally, combining (49) with

$$
\|\mathbf{a}^{(L_{**})}\|_2^2 = \frac{|a|^{2L_{**}}-1}{|a|^2-1}, \quad \|\mathbf{a}^{(K_{**})}\|_2^2 = \frac{|a|^{2K_{**}}-1}{|a|^2-1},
$$

yields eqn. (36).

References

[1] Y. Li, J. Razavilar, and K. Liu, "A high-resolution technique for multidimensional NMR spectroscopy," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 1, pp. 78–86, 1998.

[2] A. B. Gershman and N. D. Sidiropoulos, *Space-time processing for MIMO communications*. Wiley Online Library, 2005.

[3] J. Sacchini, W. Steedly, and R. Moses, "Two-dimensional Prony modeling and parameter estimation," *IEEE Trans. Signal Process.*, vol. 41, no. 11, pp. 3127–3137, 1993.

[4] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2267–2280, 1992.

[5] S. Rouquette and M. Najim, "Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 237–245, 2001.

[6] M. Haardt and J. A. Nossek, "Simultaneous Schur decomposition of several nonsymmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 161–169, 1998.

[7] A. Shlemov and N. Golyandina, "Shaped extension of singular spectrum analysis," in *Proceedings of the 21st International Symposium on Mathematical Theory of Networks and Systems (MTNS 2014)*, July 7-11, 2014, Groningen, The Netherlands, 2014.

[8] J. Liu and X. Liu, "An eigenvector-based approach for multidimensional frequency estimation with improved identifiability," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4543–4556, 2006.

[9] J. Liu, X. Liu, and X. Ma, "Multidimensional frequency estimation with finite snapshots in the presence of identical frequencies," *IEEE Transactions on Signal Processing*, vol. 55, pp. 5179–5194, 2007.

[10] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3198–3213, 2008.

[11] W. Sun and H.-C. So, "Accurate and computationally efficient tensor-based subspace approach for multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5077–5088, 2012.

[12] H.-C. So, F. Chan, W. Lau, and C.-F. Chan, "An efficient approach for two-dimensional parameter estimation of a single-tone," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 1999–2009, 2010.

[13] L. Huang, Y. Wu, H. So, Y. Zhang, and L. Huang, "Multidimensional sinusoidal frequency estimation using subspace and projection separation approaches," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5536–5543, 2012.

[14] C. Lin and W. Fang, "Efficient multidimensional harmonic retrieval: A hierarchical signal separation framework," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 427–430, May 2013.

[15] M. Sørensen and L. De Lathauwer, "Multidimensional harmonic retrieval via coupled canonical polyadic decomposition—part i: Model and identifiability," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 517–527, 2017.

[16] ——, "Multidimensional harmonic retrieval via coupled canonical polyadic decomposition—part ii: Algorithm and multirate sampling," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 528–539, 2017.

[17] Z. Yang, L. Xie, and P. Stoica, "Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3685–3701, 2016.

[18] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[19] Y. Chi and Y. Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1030–1042, 2015.

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

[21] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DOA estimation algorithms: unification, simplification, and observations," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 29, no. 4, pp. 1170–1184, 1993.

[22] B. D. Rao and K. Hari, "Performance analysis of esprit and tam in determining the direction of arrival of plane waves in noise," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 12, pp. 1990–1995, 1989.

[23] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, 1989.

[24] A. Okhovat and J. Cruz, "Statistical analysis of the Tufts-Kumaresan and principal Hankel components methods for estimating damping factors of single complex exponentials," in *International Conference on Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989*. IEEE, 1989, pp. 2286–2289.

[25] F. Roemer, M. Haardt, and G. Del Galdo, "Analytical performance assessment of multi-dimensional matrix-and tensor-based ESPRIT-type algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2611–2625, 2014.

[26] J. Steinwandt, F. Roemer, M. Haardt, and G. D. Galdo, "Performance analysis of multi-dimensional ESPRIT-type algorithms for arbitrary and strictly non-circular sources with spatial smoothing," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2262–2276, May 2017.

[27] E.-H. Djermoune and M. Tomczak, "Perturbation analysis of subspace-based methods in estimating a damped complex exponential," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4558–4563, 2009.

[28] S. Sahnoun, K. Usevich, and P. Comon, "Optimal choice of Hankel-block-Hankel matrix shape in 2-D parameter estimation: The rank-one case," in *2016 24th European Signal Processing Conference (EUSIPCO)*, Aug 2016, pp. 321–325.

[29] J. Liu, X. Liu, and X. Ma, "First-order perturbation analysis of singular vectors in singular value decomposition," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3044–3049, 2008.

[30] R. Badeau, G. Richard, and B. David, "Performance of ESPRIT for estimating mixtures of complex exponentials modulated by polynomials," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 492–504, 2008.

[31] Z. Xu, "Perturbation analysis for subspace decomposition with applications in subspace-based algorithms," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2820–2830, 2002.

[32] J. E. Cohen, "About notations in multiway array processing," *arXiv preprint arXiv:1511.01306*, 2015.

[33] D. Fasino and P. Tilli, "Spectral clustering properties of block multilevel Hankel matrices," *Linear Algebra and its Applications*, vol. 306, no. 13, pp. 155–163, 2000.

[34] T. Jiang, N. D. Sidiropoulos, and J. M. Ten Berge, "Almost-sure identifiability of multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1849–1859, 2001.

[35] G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed. Johns Hopkins University Press, 1996.

[36] H. D. Simon, "The Lanczos algorithm with partial reorthogonalization," *Mathematics of Computation*, vol. 42, no. 165, pp. 115–142, 1984.

[37] A. Korobeynikov, "Computation- and space-efficient implementation of SSA," *Statistics and Its Interface*, vol. 3, no. 3, pp. 357–368, 2010.

[38] G. Xu, R. H. Roy, and T. Kailath, "Fast ESPRIT — application of fast signal-subspace decomposition in array signal processing," in *1990 Conference Record Twenty-Fourth Asilomar Conference on Signals, Systems and Computers, 1990.*, vol. 2, Nov 1990, pp. 961–965.

[39] G. Xu and T. Kailath, "Fast subspace decomposition," *IEEE Transactions on Signal Processing*, vol. 42, no. 3, pp. 539–551, Mar 1994.

[40] R. Larsen, "Lanczos bidiagonalization with partial reorthogonalization," *DAIMI Report Series*, vol. 27, no. 537, 1998.

[41] K. Wu and H. Simon, "Thick-restart lanczos method for large symmetric eigenvalue problems," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 2, pp. 602–616, May 2000.

[42] R. M. Larsen, *PROPACK — software for large and sparse SVD calculations*, available from http://sun.stanford.edu/r̃munk/PROPACK/.

[43] J. Gao, M. D. Sacchi, and X. Chen, "A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions," *Geophysics*, vol. 78, no. 1, pp. V21–V30, 2013.

[44] N. Golyandina, A. Korobeynikov, A. Shlemov, and K. Usevich, "Multivariate and 2D extensions of Singular Spectrum Analysis with the Rssa package," *Journal of Statistical Software*, vol. 67, no. 1, 2015.

[45] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2740–2746, June 2007.

[46] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, pp. 814–824, May 1990.

[47] N. Golyandina, "On the choice of parameters in singular spectrum analysis and related subspace-based methods," *Statistics and Its Interface*, vol. 3, no. 3, pp. 259–279, 2010.

[48] X. Liu and N. D. Sidiropoulos, "On constant modulus multidimensional harmonic retrieval," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, May 2002, pp. III–2977–III–2980.

[49] R. M. Larsen, "Efficient algorithms for helioseismic inversion," Ph.D. dissertation, University of Aarhus, Denmark, 1998.

[50] E. Candès and S. Becker, *SVT: Singular Value Thresholding software*, available from http://svt.stanford.edu/code.html.