HAL open science

Multidimensional ESPRIT: Algorithm, Computations and Perturbation Analysis

Souleymen Sahnoun, Konstantin Usevich, Pierre Comon

To cite this version:

Souleymen Sahnoun, Konstantin Usevich, Pierre Comon. Multidimensional ESPRIT: Algorithm, Computations and Perturbation Analysis. 2016. hal-01360438v1

HAL Id: hal-01360438

https://hal.science/hal-01360438v1

Preprint submitted on 5 Sep 2016 (v1), last revised 7 Jun 2017 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multidimensional ESPRIT: Algorithm, Computations and Perturbation Analysis

Souleymen Sahnoun, Konstantin Usevich*, Pierre Comon

***Abstract*—In this paper we present and analyse the performance of multidimensional ESPRIT (N-D ESPRIT) method for estimating parameters of N-D superimposed damped 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], Shaped ESPRIT [6], improved multidimensional folding (IMDF) [7], [8], Tensor-ESPRIT [9], principal-singular-vector utilization for modal analysis (PUMA) [10], [11] and the methods proposed in [12], [13].

It is generally admitted that these methods yield accurate estimates at high SNR and/or when the frequencies are well separated. Statistical performances of some of these methods have been studied in the case of undamped sinusoids [7], [8]. Recently, analytical performances of tensor-based ESPRIT-type algorithms have been assessed for undamped signals [14]. These analyses are based on the results published in [15]. Statistical performances of subspace 1-D estimation methods have been extensively studied in the case of undamped sinusoids [16], [17] and damped ones [18], [15]. More recently, a new study was presented for the case of 1-D damped single-tone [19], resulting in new closed-from expressions. An extension of the results of [19] to the 2-D ESPRIT case was initiated in [20].

In this paper, we focus our attention on the 2-D ESPRIT algorithm of [5]. Although in [5] it is mentioned that 2-D ESPRIT can be extended to handle N-D signals, to our knowledge, it has never been done explicitly in the literature. Moreover, no perturbation analysis of this extension has yet been conducted. A related approach called IMDF was proposed in [7], [8], but it is not a direct extension of the 2-D ESPRIT of [5]. In sensor array processing, this extension corresponds to the case of a single snapshot via spatial smoothing [9].

b) Contributions: Firstly, we extend 2-D ESPRIT algorithm to N-D signals (we call this extension N-D ESPRIT).

We use an approach simpler than in [5] for describing the algorithm; our description is close in spirit to the approach used in [7], [8]. We also discuss the difference between the N-D ESPRIT and the IMDF method of [8]. Next, we propose a fast version of the N-D ESPRIT algorithm which uses the truncated SVD, which we call Fast N-D ESPRIT. The Fast N-D ESPRIT algorithm has low computational complexity and allows handling large signals and large matrices.

Finally, the main contribution 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. We propose a simple formula for first-order perturbation that does not involve the factors of the SVD. We derive closed-form expressions for the variances of the perturbations in the N-D damped single-tone case.

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 partial SVD of multilevel Hankel matrices and the gain in computational complexity is showed. The difference between N-D ESPRIT and IMDF is also pointed out. In Section V, a first-order perturbation analysis for ND-ESPRIT is performed and simplified expressions are

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

S. Sahnoun, K. Usevich, and P. Comon are with CNRS, GIPSA-Lab, Univ. Grenoble Alpes, F-38000 Grenoble, France. Email: firstname.name@gipsa-lab.grenoble-inp.fr . Fax: +33476574790. Tel.: +33476826304 (S. Sahnoun), +33476574575 (K. Usevich), +33476826271 (P. Comon).

*Corresponding author.

derived in the multiple tones case. In Section IV, the single tone case is analyzed and closed form expressions are derived for damped and undamped signals. 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.

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-Penroce 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 $\mathrm{vec_r}\{\mathcal{A}\}$ its "row-major" vectorization, i.e.

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

The row-major vectorization is used because it is compatible with the Kronecker product [21], i.e.

$$\mathrm{vec_r}\{\mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_N\} = \mathbf{a}_1 \boxtimes \cdots \boxtimes \mathbf{a}_N.$$

Unlike in [21], we use a special notation for row-major vectorisation in order to distinguish it from the conventional column-major vectorisation.

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

$$\mathbf{x}^{(M)} \stackrel{\mathrm{def}}{=} \begin{bmatrix}1 & x & x^2 & \cdots & x^{(M-1)}\end{bmatrix}^\mathsf{T} \tag{1}$$

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 $m_n = 0,\ldots,M_n-1$:

$$\tilde{y}(m_1,\ldots,m_N) = y(m_1,\ldots,m_N) + \varepsilon(m_1,\ldots,m_N), \tag{2}$$

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:

$$y(m_1,\ldots,m_N) = \sum_{r=1}^{R} c_r \prod_{n=1}^{N} (a_{r,n})^{m_n}, \tag{3}$$

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 and positive) damping factors,
- $\{\omega_{r,n} = 2\pi\nu_{r,n}\}_{r=1,n=1}^{R,N}$ are angular frequencies.

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

C. Tensor formulation

It is often convenient to rewrite the signal in tensor notation. Let the tensor $\mathcal{Y} \in \mathbb{C}^{M_1\times\cdots\times M_N}$ be given as[1]

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

We also define similarly the tensors $\tilde{\mathcal{Y}}, \mathcal{E} \in \mathbb{C}^{M_1\times\cdots\times M_N}$. Then (2) can be compactly written as $\tilde{\mathcal{Y}} = \mathcal{Y} + \mathcal{E}$, and (3) is the canonical polyadic (CP) tensor decomposition

$$\mathcal{Y} = \sum_{r=1}^{R} c_r \mathbf{a}_{r,1}^{(M_1)} \otimes \cdots \otimes \mathbf{a}_{r,N}^{(M_N)} \tag{4}$$

where $\mathbf{a}_{r,n}^{(M_n)}$ are Vandermonde-structured vectors for $a_{r,n}$ defined in (1).

By the properties of CP decomposition, eqn. (4) after vectorization can be rewritten with the help of Khatri-Rao products:

$$\mathrm{vec_r}\{\mathcal{Y}\} = \left(\mathbf{A}_1^{(M_1)} \odot \mathbf{A}_2^{(M_2)} \odot \cdots \odot \mathbf{A}_N^{(M_N)}\right)\mathbf{c}, \tag{5}$$

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

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

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 $K_n \stackrel{\mathrm{def}}{=} M_n - L_n + 1$. Define by $\mathbf{y}^{(i_1,\ldots,i_N)} \in \mathbb{C}^{L_1\times\cdots\times L_N}$ the vectorized subarray

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

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

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

$$\mathbf{H} \stackrel{\text{def}}{=} \Big[\mathbf{y}^{(1,\ldots,1)} \;\; \mathbf{y}^{(1,\ldots,1,2)} \;\; \ldots \;\; \mathbf{y}^{(1,\ldots,1,K_N)} \;\; \mathbf{y}^{(1,\ldots,1,2,1)} \ldots \;\; \cdots \mathbf{y}^{(K_1,K_2,\ldots,K_N-1)} \;\; \mathbf{y}^{(K_1,K_2,\ldots,K_N)}\Big]. \tag{6}$$

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

Remark 1: The matrix (6) 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 [22].

It can be verified that in the absence of noise, MH matrix (6) admits a factorization of the form

$$\mathbf{H} = \mathbf{P}\,\mathrm{diag}(\mathbf{c})\mathbf{Q}^{\mathsf{T}}, \tag{7}$$

where

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

The factorization (7) directly follows from (5). The proof can be also found in [7].

B. Shift properties of subspaces

Let us define the selection matrices

$$\overline{{}^{n}\mathbf{I}} \stackrel{\text{def}}{=} \mathbf{I}_{L_1} \boxtimes \mathbf{I}_{L_2} \boxtimes \cdots \boxtimes \overline{\mathbf{I}}_{L_n} \boxtimes \cdots \boxtimes \mathbf{I}_{L_N} \tag{8}$$
$$= \mathbf{I}_{\prod_{i=1}^{n-1} L_i} \boxtimes \overline{\mathbf{I}}_{L_n} \boxtimes \mathbf{I}_{\prod_{i=n+1}^{N} L_i} \tag{9}$$
$${}_{n}\underline{\mathbf{I}} \stackrel{\text{def}}{=} \mathbf{I}_{L_1} \boxtimes \mathbf{I}_{L_2} \boxtimes \cdots \boxtimes \underline{\mathbf{I}}_{L_n} \boxtimes \cdots \boxtimes \mathbf{I}_{L_N} \tag{10}$$
$$= \mathbf{I}_{\prod_{i=1}^{n-1} L_i} \boxtimes \underline{\mathbf{I}}_{L_n} \boxtimes \mathbf{I}_{\prod_{i=n+1}^{N} L_i} \tag{11}$$

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 $\overline{{}^{n}\mathbf{X}} = \overline{{}^{n}\mathbf{I}}\,\mathbf{X}$ and ${}_{n}\underline{\mathbf{X}} = {}_{n}\underline{\mathbf{I}}\,\mathbf{X}$. Then the shifted versions of $\mathbf{P}$ satisfy the following equation:

$${}_{n}\underline{\mathbf{P}}\,\mathbf{\Psi}_n = \overline{{}^{n}\mathbf{P}}, \tag{12}$$

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

$${}_{n}\underline{\mathbf{U}}_s\mathbf{F}_n = \overline{{}^{n}\mathbf{U}}_s, \tag{13}$$

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

$$\mathbf{F}_n = \left({}_{n}\underline{\mathbf{I}}\,\mathbf{U}_s\right)^{\dagger} \left(\overline{{}^{n}\mathbf{I}}\,\mathbf{U}_s\right) := \left({}_{n}\underline{\mathbf{U}}_s\right)^{\dagger} \left(\overline{{}^{n}\mathbf{U}}_s\right) \tag{14}$$

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.

IV. N-D ESPRIT FOR MULTILEVEL HANKEL MATRICES

A. N-D ESPRIT algorithm

The N-D ESPRIT algorithm can be summarized 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 $\widetilde{\mathbf{H}}$ from the noisy signal, in the same format as (6).
3) Perform the SVD of $\widetilde{\mathbf{H}}$, 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 $\widetilde{\mathbf{F}}_n$
$$\widetilde{\mathbf{F}}_n := \left({}_{n}\underline{\widetilde{\mathbf{U}}}_s\right)^{\dagger} \left(\overline{{}^{n}\widetilde{\mathbf{U}}}_s\right). \tag{15}$$
5) Compute a linear combination of matrices, where $\beta_1, \ldots, \beta_n$ are given parameters.
$$\widetilde{\mathbf{K}} = \sum_{n=1}^{N} \beta_n \widetilde{\mathbf{F}}_n \tag{16}$$
6) Compute a diagonalizing matrix $\mathbf{T}$ of $\widetilde{\mathbf{K}}$ (from its eigenvalue decomposition):
$$\widetilde{\mathbf{K}} = \mathbf{T}\,\mathrm{Diag}(\boldsymbol{\eta})\mathbf{T}^{-1}. \tag{17}$$
7) Apply the transformation $\mathbf{T}$ to $\mathbf{F}_n$:
$$\widetilde{\mathbf{D}}_n = \mathbf{T}^{-1}\widetilde{\mathbf{F}}_n\mathbf{T}, \quad \text{for} \quad n = 1, \ldots, N \tag{18}$$
8) Extract $\{[\widehat{a}_{1,n}, \ldots, \widehat{a}_{R,n}]\}_{n=1}^{N}$ from $\mathrm{diag}(\widetilde{\mathbf{D}}_n)$, $n = 1, \ldots, N$

B. Recovery conditions

There are some essential assumptions which guarantee that in the noiseless case the algorithm recovers the modes correctly.

Assumption 1: For every n, the matrices ${}_{n}\underline{\mathbf{P}}$ 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 (15) gives the unique solution to (13). 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 $\mathbf{T}$.
   Note that there is no problem of pairing of the modes [23].

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

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

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

Proof The proof follows from [24, Proposition 4]. ■

C. Variants and related algorithms

There are several related algorithms to N-D ESPRIT. First, instead of steps 5–8 of the algorithm there are other ways to extract the modes from matrices $\widetilde{\mathbf{F}}_n$.

- In [5], in addition to 2D-ESPRIT, another algorithm was proposed under the name "2D-MEMP with improved pairing step". The difference is that the modes $a_{r,n}$ are extracted from individual eigenvalue decompositions of matrices $\widetilde{\mathbf{F}}_n$. The matrix $\mathbf{T}$ is used to perform the pairing of the modes.
- The matrices $\widetilde{\mathbf{F}}_n$ can be jointly diagonalized using simultaneous Schur diagonalization [25].

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

Second, the algorithm IMDF of [8] is related to N-D ESPRIT, but it is not an extension of 2-D ESPRIT. The first difference lies in the matrices ${}^{\mathrm{n}\overline{\ }}\mathbf{I}$ and ${}_{\mathrm{n}\underline{\ }}\mathbf{I}$. In fact, in [8] they are defined by

$$ {}^{\mathrm{n}\overline{\ }}\mathbf{I} = \underline{\mathbf{I}}_{L_1} \boxtimes \underline{\mathbf{I}}_{L_2} \boxtimes \cdots \boxtimes \overline{\mathbf{I}}_{L_n} \boxtimes \cdots \boxtimes \underline{\mathbf{I}}_{L_N} \tag{19}$$

$$ {}_{\mathrm{n}\underline{\ }}\mathbf{I} = \underline{\mathbf{I}}_{L_1} \boxtimes \underline{\mathbf{I}}_{L_2} \boxtimes \cdots \boxtimes \underline{\mathbf{I}}_{L_n} \boxtimes \cdots \boxtimes \underline{\mathbf{I}}_{L_N} \tag{20}$$

The second difference is that in [8] modes are estimated from ${}_{\mathrm{n}\underline{\ }}\mathbf{P}$), however in the N-D ESPRIT algorithm defined in Section IV-A the modes are estimated from (18).

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 [26, Ch. 8] for the full SVD requires $O(L^2K)$ flops [26] (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 $T_{\mathbf{A}}$ is 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 $\mathbf{A}$ using Lanczos bidiagonalization [26, Ch. 9] with partial reorthogonalization [27] is $O(RT_{\mathbf{A}} + R^2(L+K))$ (see [28, §3] for an overview of Lanczos-based methods).

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 the Appendix B. This fact was used in [28] for truncated SVD of Hankel matrices, and independently in [29] and [30] for special cases of MH matrices. Although the matrix-vector multiplication in the general MH case is a straightforward extension of algorithms [29, eqn. (22)] and [30, Lemma 2], we provide in Appendix B a description of the algorithm for several reasons: in [29] Toeplitz matrices are treated and in [30] only the real Hankel-block-Hankel matrices are treated (also, the proof of [30, 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 (19)–(20) 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 the complexity $O(L^2K)$ when using the full SVD).

V. Perturbation analysis

A. Basic expressions

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

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

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

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

whose subspace decomposition is given by

$$\tilde{\mathbf{H}} = \tilde{\mathbf{U}}_s \tilde{\mathbf{\Sigma}}_s \tilde{\mathbf{V}}_s^{\mathsf{H}} + \tilde{\mathbf{U}}_n \tilde{\mathbf{\Sigma}}_n \tilde{\mathbf{V}}_n^{\mathsf{H}} \tag{22}$$

We use the following lemma on the first-order approximation.

Lemma 3 ([15] and [31]): The perturbed signal subspace is $\tilde{\mathbf{U}}_s = \mathbf{U}_s + \Delta\mathbf{U}_s$, $\tilde{\mathbf{V}}_s = \mathbf{V}_s + \Delta\mathbf{V}_s$ and $\tilde{\mathbf{\Sigma}}_s = \mathbf{\Sigma}_s + \Delta\mathbf{\Sigma}_s$. A first order approximation of the perturbation is given by

$$\Delta\mathbf{U}_s = \mathbf{U}_n \mathbf{U}_n^{\mathsf{H}} \Delta\mathbf{H}\, \mathbf{V}_s \mathbf{\Sigma}_s^{-1} \tag{23}$$

$$\Delta\mathbf{V}_s^{\mathsf{H}} = \mathbf{\Sigma}_s^{-1} \mathbf{U}_s^{\mathsf{H}} \Delta\mathbf{H}\, \mathbf{V}_n \mathbf{V}_n^{\mathsf{H}} \tag{24}$$

$$\Delta\mathbf{\Sigma}_s = \mathbf{U}_s^{\mathsf{H}} \Delta\mathbf{H}\, \mathbf{V}_s \tag{25}$$

Next, we express the first-order perturbations of the matrices $\mathbf{F}_n$. Similar expressions can be found in the literature, but we give a short proof in the appendix for completeness.

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

$$\Delta\mathbf{F}_n = ({}_{\mathrm{n}\underline{\ }}\mathbf{U}_s)^{\dagger} (\Delta {}^{\mathrm{n}\overline{\ }}\mathbf{U}_s - \Delta {}_{\mathrm{n}\underline{\ }}\mathbf{U}_s \mathbf{F}_n). \tag{26}$$

Proof The proof can be found in Appendix C.

Next, let $\mathbf{t}_r$ denote the eigenvectors of $\mathbf{K}$ (the columns of $\mathbf{T}$):

$$\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_R],$$

and $\boldsymbol{\tau}_r^{\mathsf{T}}$ denote the rows of $\mathbf{T}^{-1}$

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

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

$$\Delta a_{r,n} = \tau_r^{\mathsf{T}} \Delta \mathbf{F}_n t_r. \tag{27}$$

Proof The proof can be found in Appendix C.

Remark 4: 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 (26) into (27) leads to the following formula for the perturbation of the modes after some simplifications.

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

$$\Delta a_{r,n} = \boldsymbol{\tau}_r^{\mathsf{T}} (\underline{\mathbf{U}}_{s_n})^{\dagger} (\overline{\mathbf{I}}_n - a_{r,n} \underline{\mathbf{I}}_n) \Delta \mathbf{H}\, \mathbf{V}_s \boldsymbol{\Sigma}_s^{-1} \mathbf{t}_r \tag{28}$$

Proof The proof can be found in Appendix C.

B. A simplified formula for the perturbations

The expression of first order perturbation (28) is widely used in the literature. It corresponds to the expressions given in [14], which is the state-of-the-art perturbation analysis. The main problem is that the expression (28) requires the knowledge of the singular value decomposition of the multilevel Hankel matrix $\mathbf{H}$. This complicates the 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 (7), where the matrix $\mathbf{P}$ satisfies (12) 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 the N-D ESPRIT algorithm admits an expansion

$$\Delta a_{r,n} = \frac{1}{c_r} \mathbf{b}_r^{\top} \underline{\mathbf{P}}_n^{\dagger} (\overline{\mathbf{I}}_n - a_{r,n} \underline{\mathbf{I}}_n) \Delta \mathbf{H}\, (\mathbf{Q}^{\mathsf{T}})^{\dagger} \mathbf{b}_r. \tag{29}$$

Proof The proof can be found in Appendix C.

In the formula (29), no coefficients of the SVD are needed. We should note that a similar formula for 1D ESPRIT can be found in [32, Proposition 12]. Moreover the following remarks can be made.

Remark 5: The first-order perturbation of $\Delta a_{r,n}$ depends only on the contribution (c_r) of the r-th tone and the angle between the subspace of r-th tone and the others. It does not depend on the coefficients c_k, $k \neq r$ of other tones.

Remark 6: The formula (29) can be extended to the case of multiple snapshots.

C. Computing the first-order perturbation and its moments

We see that the perturbation (29) is of the form

$$\Delta a_{r,n} = \mathbf{v}_{r,n}^{\top} \Delta \mathbf{H} \mathbf{x}_r,$$

where

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

Remark 7 (Computational details): Vectors $\mathbf{v}_{r,n}^{\top}$ 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 $\mathbf{v}_{r,n}^{\top}$.

Since the MH matrix $\Delta \mathbf{H}$ depends linearly on the elements of $\mathbf{e} = \mathrm{vec}_{\mathrm{r}}\{\mathcal{E}\}$, the perturbation can be expressed as

$$\Delta a_{r,n} = \mathbf{z}_{r,n}^{\mathsf{H}} \mathbf{e}. \tag{30}$$

In Appendix B, we show how $\mathbf{z}_{r,n}$ can be computed from $\mathbf{v}_{r,n}^{\top}$ and $\mathbf{x}_r$ efficiently using the N-D convolution.

Therefore, we have the following:

1) $\mathbb{E}\{\Delta a_{r,n}\} = 0$ if $\mathbf{e}$ is zero-mean.
2) $\mathbb{E}\{\Delta a_{r,n}^2\} = 0$ if $\mathbf{e}$ is circular.
3) $\mathbb{E}\{|\Delta a|_{r,n}^2\} = \mathbf{z}^{\mathsf{H}} \Gamma \mathbf{z}$ if $\mathbf{e}$ has covariance matrix $\Gamma = \mathbb{E}\{\mathbf{e}\mathbf{e}^{\mathsf{H}}\}$.
4) $\mathbb{E}\{|\Delta a|_{r,n}^2\} = \sigma_e^2 \|\mathbf{z}\|_2^2$ if $\mathbf{e}$ is white with variance σ_e^2.
5) Finally,

$$\mathrm{var}(\Delta \omega_n) = \mathrm{var}(\Delta \alpha_n) = \frac{\mathbb{E}\{|\Delta a_n|^2\}}{2|a_n|^2}$$

if $\mathbf{e}$ is complex circular Gaussian.

VI. Single-tone case

In this section, we calculate the perturbations of the parameter estimates for the signal $y(m_1, \ldots, m_N) = c \prod_{n=1}^{N} a_n^{m_n}$.

A. Specialising the general formula

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

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

Since $\mathbf{a}^{\dagger} = \frac{1}{\|\mathbf{a}\|_2^2} \mathbf{a}^*$ for any vector $\mathbf{a}$, (29) can be specialised to

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

By using the link between matrix multiplication and convolution, we get

Lemma 6: The first-order perturbation is expressed as

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

where

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

where $\star$ denotes convolution of vectors.

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 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 $f(L, M, x)$ and $g(L, M, x)$ are defined as

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

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

C. Closed form expressions

Here we provide expressions that can be also found in [20]. We also give the full proofs that are absent in [20].

First, we note that the formula for $f(L, M, x)$ coincides with the formula for the variance of the first-order perturbation of the 1D-ESPRIT. Hence, the results from [33], [32], [19] can be used.

Proposition 2 ([19, eqn. (26)], [33, eqn. (4.16)], [32, Corollary 7]): In the undamped case ($|x| = 1$), the function $f(L, M, x)$ has the form

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

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

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

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).

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

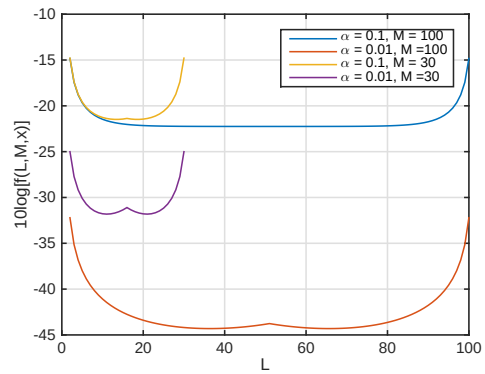

Fig. 1. Behavior of function $f(L, M, x)$ as a function of L for different values of M and damping factors.

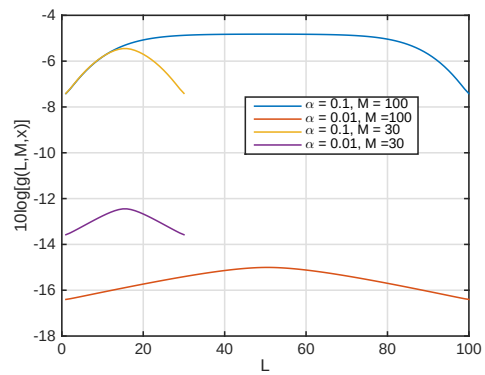

Fig. 2. Behavior of function $g(L, M, x)$ as a function of L for different values of M and damping factors.

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

Proof The proof can be found in Appendix C.

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

Remark 8: Based on Propositions 2–3, the optimal values for L_i can be obtained, in the same manner as it was shown in [20] for the 2-D case. Note that the optimal L_i should depend on the type of the noise, as in the 1D case [34].

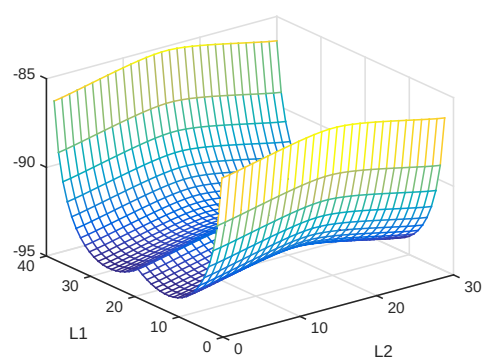

Fig. 3. Variance of $\Delta\omega_a$ as a function of L_1 and L_2

$$g(L,M,x) = (1-|x|^2) \times \begin{cases} \frac{-2L(1-|x|^2)\left(|x|^{2K}+|x|^{2L}\right)}{(1-|x|^{2L})^2(1-|x|^{2K})^2} + \frac{(1+|x|^{2K})(1+|x|^2)}{(1-|x|^{2L})(1-|x|^{2K})^2}, & \text{if } L \leq \frac{M+1}{2} \text{ and } |x| \neq 1 \\ \frac{-2K_2(1-|x|^2)\left(|x|^{2L}+|x|^{2K}\right)}{(1-|x|^{2L})^2(1-|x|^{2K})^2} + \frac{(1+|x|^{2L})(1+|x|^2)}{(1-|x|^{2L})^2(1-|x|^{2K})}, & \text{if } L \geq \frac{M+1}{2} \text{ and } |x| \neq 1 \end{cases} \tag{36}$$

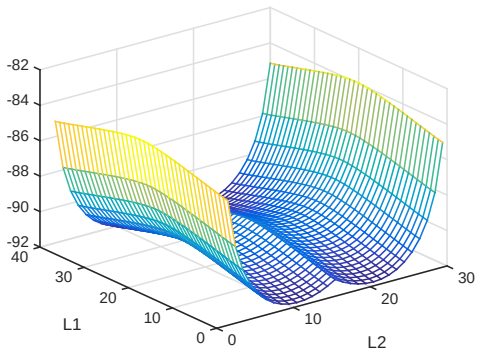

Fig. 4. Variance of $\Delta\omega_b$ as a function of L_1 and L_2

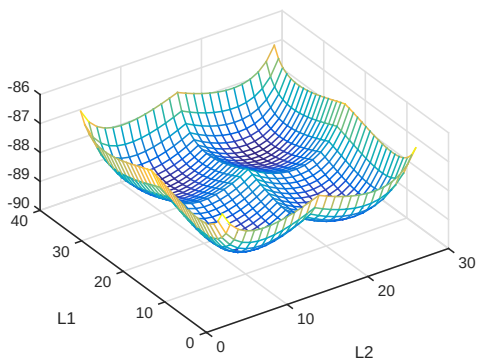

Fig. 5. tMSE as a function of L_1 and L_2

VII. Simulations

Numerical simulations have been carried out to verify theoretical expressions and compare the performances of N-D ESPRIT, Fast N-D ESPRIT and IMDF algorithms 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

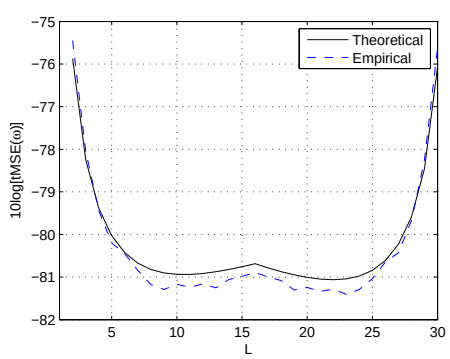

Fig. 6. Theoretical and empirical MSEs for 2-D ESPRIT versus L, ($L = L_1 = L_2$). $(\alpha_a, \omega_a) = (0.1, 0.2\pi)$, $(\alpha_b, \omega_b) = (0.1, 0.4\pi)$, $(M_1, M_2) = (30, 30)$, SNR = 40 dB.

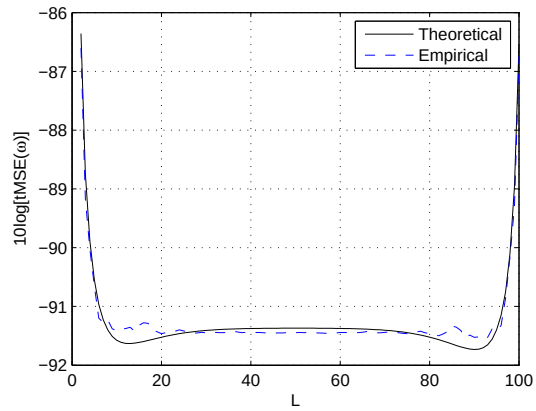

Fig. 7. Theoretical and empirical MSEs for 2-D ESPRIT (fast SVD) versus L, ($L = L_1 = L_2$). $(\alpha_a, \omega_a) = (0.1, 0.2\pi)$, $(\alpha_b, \omega_b) = (0.1, 0.4\pi)$, $(M_1, M_2) = (100, 100)$, SNR = 40 dB.

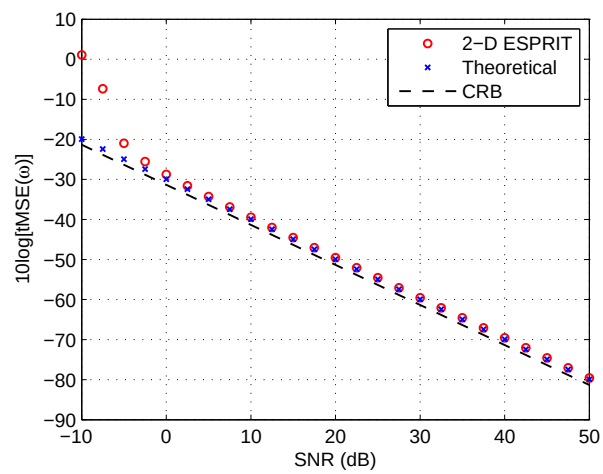

Fig. 8. Theoretical and empirical tMSEs for 2-D ESPRIT versus SNR. $(L_1, L_2) = (4, 4)$. $(\alpha_a, \omega_a) = (0.1, 0.2\pi)$, $(\alpha_b, \omega_b) = (0.1, 0.4\pi)$, $(M_1, M_2) = (10, 10)$.

$\text{tMSE} = \frac{1}{RF}\mathbb{E}_p\left\{\sum_{r=1}^{R}\sum_{f=1}^{F}(\xi_{f,r} - \hat{\xi}_{f,r})^2\right\}$ 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

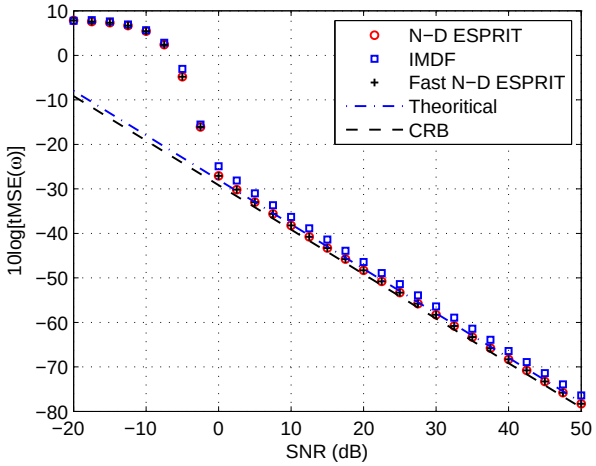

Fig. 9. Theoretical and empirical tMSEs versus SNR. 2-D damped signal containing two tones. $(L_1, L_2) = (4, 4)$. $(\alpha_{1,1}, \omega_{1,1}) = (0.01, 0.2\pi)$, $(\alpha_{2,1}, \omega_{2,1}) = (0.01, 0.6\pi)$, $(\alpha_{1,2}, \omega_{1,2}) = (0.01, 0.3\pi)$, $(\alpha_{2,2}, \omega_{2,2}) = (0.01,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

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 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 [35] developed within the PhD thesis [36]. We use the updated version of the PROPACK package available as a part of the SVT software [37].

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

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

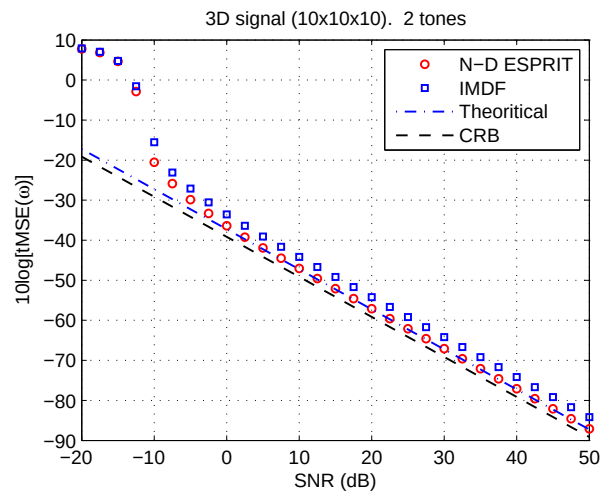

Fig. 10. Theoretical and empirical tMSEs versus SNR. 3-D damped signal containing two tones. $(L_1, L_2, L_3) = (4, 4, 4)$, $(M_1, M_2, M_3) = (10, 10, 10)$.

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

that N-D ESPRIT outperforms IMDF and the gap between them become bigger compared to the previous experiment (experiment with two tones).

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 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 components (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:

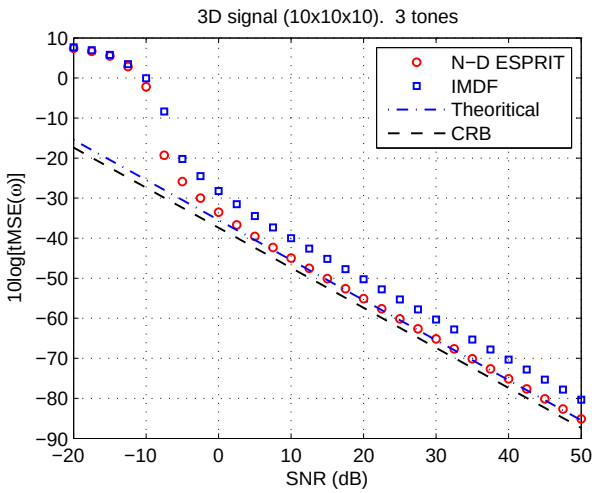

Fig. 11. Theoretical and empirical tMSEs versus SNR. 3-D damped signal containing two tones. $(L_1, L_2, L_3) = (4, 4, 4)$, $(M_1, M_2) = (10, 10, 10)$.

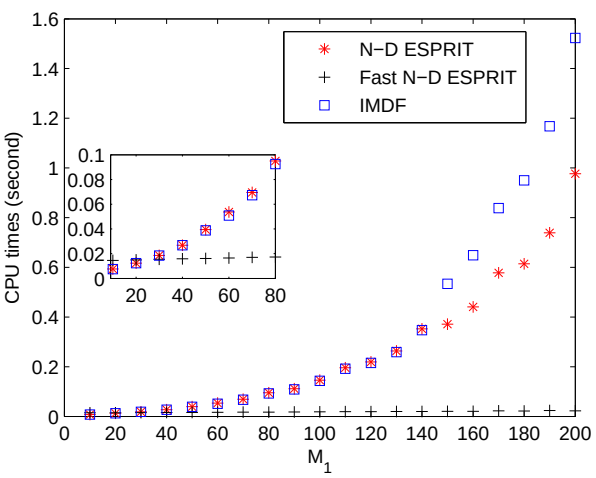

Fig. 12. Average CPU time for a single run under $M_2 = 10$ and $R = 2$.

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.

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

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

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

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 multilevel structure:

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

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

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

and the blocks of the last level are just scalars

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

B. Matrix-vector products for MH matrices

Let the MH matrix $\mathbf{H}$ be given in (6). 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 $\mathbf{x} = \mathrm{vec_r}\{\mathcal{X}\}$ and $\mathbf{w} = \mathrm{vec_r}\{\mathcal{W}\}$), then we see that the tensor $\mathcal{W}$ is obtained via a "filtering operation", i.e.

$$(\mathcal{W})_{i_1,\ldots,i_N} = \sum_{j_1,\ldots,j_N=1}^{K_1,\ldots,K_N} (\mathcal{Y})_{i_1+j_1,\ldots,i_N+j_N} (\mathcal{X})_{j_1,\ldots,j_N}. \tag{41}$$

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 by $\mathbf{H}$. Let $\mathbf{x} \in \mathbb{C}^{K_1 \cdots K_N}$ and $\mathbf{v} \in \mathbb{C}^{L_1 \cdots L_N}$. Assume that we wish to evaluate the product $\mathbf{v}^\top \mathbf{H}\mathbf{x}$ for an arbitrary matrix $\mathbf{H}$. By linearity of the operation, there exists a vector $\mathbf{z} \in \mathbb{C}^{M_1 \cdots M_N}$, such that the product is given as follows:

$$\mathbf{v}^\top \mathbf{H}\mathbf{x} = \mathbf{z}^\mathsf{T} \mathbf{y}, \tag{42}$$

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 $\mathcal{W}'$
$$\mathbf{w} = \mathrm{vec_r}\{(\mathcal{W}')_{1:L_1,1:L_2,\ldots,1:L_N}\}.$$

where $\mathbf{y}$ is the vectorized tensor $\mathcal{Y}$. The following lemma shows the structure of $\mathbf{z}$.

Lemma 7: Let $\mathcal{V} \in \mathbb{C}^{L_1 \times \cdots \times L_N}$ and $\mathcal{X} \in \mathbb{C}^{K_1 \times \cdots \times K_N}$ be the tensorizations of $\mathbf{w}$ and $\mathbf{v}$ respectively. Then the vector $\mathbf{z} = \mathrm{vec_r}\{\mathcal{Z}\}$, where the tensor $\mathcal{Z} \in \mathbb{C}^{M_1 \times \cdots \times M_N}$ is the convolution of $\mathcal{V}$ and $\mathcal{X}$:

$\mathcal{Z} := \mathcal{X} \star \mathcal{V}$, i.e.

$$(\mathcal{Z})_{m_1,\ldots,m_r} := \sum_{i_1+j_1=m_1,\ldots,i_N+j_N=m_N}^{K_1,\ldots,K_N} (\mathcal{X})_{i_1,\ldots,i_N} (\mathcal{V})_{j_1,\ldots,j_N}.$$

Proof From (41), we have that

$$\mathbf{v}^\top \mathbf{H}\mathbf{x} = \sum_{i_1,\ldots,i_N=1}^{L_1,\ldots,L_N} \sum_{j_1,\ldots,j_N=1}^{K_1,\ldots,K_N} (\mathcal{Y})_{i_1+j_1,\ldots,i_N+j_N} (\mathcal{X})_{j_1,\ldots,j_N} (\mathcal{V})_{i_1,\ldots,i_N},$$

which completes the proof. □

The convolution can be also implemented in a standard way using the FFT, see Algorithm 2.

C. Proofs of lemmas and propositions

Proof of Lemma 4 For simplicity, we denote $\mathbf{A} = \underset{\mathrm{n}-}{\mathbf{U}}_s$ and $\mathbf{B} = \overset{\mathrm{n}-}{\mathbf{U}}$. Next, since $\mathbf{A}$ is full column rank, we can use [38, Theorem 2.4.6] (its complexified version) and the rule of differentiation of the product, and obtain

$$\begin{aligned}\Delta(\mathbf{A}^\dagger \mathbf{B}) &= \mathbf{A}^\dagger \Delta\mathbf{B} + \Delta(\mathbf{A}^\dagger)\mathbf{B} \\ &= \mathbf{A}^\dagger \Delta\mathbf{B} - \mathbf{A}^\dagger \Delta\mathbf{A}\mathbf{A}^\dagger \mathbf{B} \\ &\quad + (\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}\Delta\mathbf{A}^{\mathsf{H}}(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{B}.\end{aligned} \tag{43}$$

Since $\mathbf{A}$ and $\mathbf{B}$ span the same column space, the last term in (43) vanishes. Finally, using the fact that $\mathbf{F}_n = \mathbf{A}^\dagger \mathbf{B}$, (43) can be simplified to (26). □

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 $\mathcal{Z}$
$$\mathbf{z} = \mathrm{vec_r}\{\mathcal{Z}\}.$$

Proof of Lemma 5 First, from (18) and from the rules of differentiation, we have that the first-order perturbation has the form:

$$\begin{aligned}\Delta\mathbf{D}_n &= \mathbf{T}^{-1}\Delta\mathbf{F}_n\mathbf{T} + \mathbf{T}^{-1}\mathbf{F}_n\Delta\mathbf{T} - \mathbf{T}^{-1}\Delta\mathbf{T}\mathbf{T}^{-1}\mathbf{F}_n\mathbf{T} \\ &= \mathbf{T}^{-1}\Delta\mathbf{F}_n\mathbf{T} + \mathbf{D}_n\mathbf{T}^{-1}\Delta\mathbf{T} - \mathbf{T}^{-1}\Delta\mathbf{T}\mathbf{D}_n.\end{aligned}$$

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

$$\begin{aligned}&\Delta a_{r,n} = \mathbf{b}_r^\mathsf{T}\Delta\mathbf{D}_n\mathbf{b}_r \\ &= \tau_r^\mathsf{T}\Delta\mathbf{F}_n t_r + \mathbf{b}_r^\mathsf{T}\mathbf{D}_n\mathbf{T}^{-1}\Delta\mathbf{T}\mathbf{b}_r - \mathbf{b}_r^\mathsf{T}\mathbf{T}^{-1}\Delta\mathbf{T}\mathbf{D}_n\mathbf{b}_r\end{aligned} \tag{44}$$

Noting that $\mathbf{b}_r^\mathsf{T}\mathbf{D}_n = a_{r,n}\mathbf{b}_r^\mathsf{T}$ and $\mathbf{D}_n\mathbf{b}_r = a_{r,n}\mathbf{b}_r$, we see that the last two terms in (44) cancel, and we get that the equation (27) takes place. □

Proof of Corollary 1 Next, we combine (23) and (26). For simplicity, denote $\mathbf{C} = \Delta\mathbf{H}\,\mathbf{V}_s\mathbf{\Sigma}_s^{-1}$. Then

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

By expanding the parentheses and using the identities

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

we get

$$\begin{aligned}\Delta\mathbf{F}_n &= (\underset{\mathrm{n}-}{\mathbf{U}}_s)^\dagger(\overset{\mathrm{n}-}{\mathbf{I}}\,\mathbf{C} - \underset{\mathrm{n}-}{\mathbf{I}}\,\mathbf{C}\mathbf{F}_n) \\ &\quad + \mathbf{U}_s^{\mathsf{H}}\mathbf{C}\mathbf{F}_n - \mathbf{F}_n\mathbf{U}_s^{\mathsf{H}}\mathbf{C}.\end{aligned} \tag{45}$$

Next, we combine (45) and (27).

$$\begin{aligned}\Delta a_{r,n} &= \boldsymbol{\tau}_r^\mathsf{T}(\underset{\mathrm{n}-}{\mathbf{U}}_s)^\dagger(\overset{\mathrm{n}-}{\mathbf{I}}\,\mathbf{C} - \underset{\mathrm{n}-}{\mathbf{I}}\,\mathbf{C}\mathbf{F}_n)\mathbf{t}_r \\ &\quad + \boldsymbol{\tau}_r^\mathsf{T}(\mathbf{U}_s^{\mathsf{H}}\mathbf{C}\mathbf{F}_n - \mathbf{F}_n\mathbf{U}_s^{\mathsf{H}}\mathbf{C})\mathbf{t}_r.\end{aligned} \tag{46}$$

Using the fact that $\mathbf{F}_n\mathbf{t}_r = a_{r,n}\mathbf{t}_r$ and $\boldsymbol{\tau}_r^\mathsf{T}\mathbf{F}_n = a_{r,n}\boldsymbol{\tau}_r^\mathsf{T}$, we see that the last term in (46) vanishes, and eqn. (46) is simplified to (28). □

Proof of Proposition 1 From equation between (12) and (13), we have that

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

where the matrices $\mathbf{P}$ and $\mathbf{Q}$ are defined in (7). Next, since

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

and from properties of the pseudoinverse, we have that

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

□

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

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

Next, we consider each case separately.

a) Undamped case: In this case,

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

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

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

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

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

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 $\rho = |x|^2$ and get

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

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

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

Next, due to the fact that

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

after cancellations of some terms, we have

$$\begin{aligned}\|\mathbf{x}^{(L)} \star \mathbf{x}^{(K)}\|_2^2 &= -2L_{**}\frac{|x|^{2K_{**}} + |x|^{2L_{**}}}{(|x|^2-1)^2} \\ &+ \frac{(|x|^{2K_{**}} + 1)(|x|^{2L_{**}} - 1)(|x|^2 + 1)}{(|x|^2 - 1)^3}.\end{aligned} \tag{47}$$

Finally, combining (47) with

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

yields eqn. (36).

References

[1] Y. Li, J. Razavilar, and K. Liu, "A high-resolution technique for multidimensional NMR spectroscopy," *Biomedical Engineering, IEEE Transactions on*, 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] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] S. Sahnoun, K. Usevich, and P. Comon, "Optimal choice of Hankel-block-Hankel matrix shape in 2-D parameter estimation," in *24th European Signal Processing Conference (EUSIPCO)*, 2016, available from https://hal.archives-ouvertes.fr/hal-01288994.

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

[22] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0024379599002517

[23] Y. Hua and K. Abed-Meraim, "Techniques of eigenvalues estimation and association," *Digital signal processing*, vol. 7, no. 4, pp. 253–259, 1997.

[24] 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.

[25] 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.

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

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

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

[29] 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.

[30] 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.

[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] 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.

[33] 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.

[34] 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.

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

[36] ——, "Efficient algorithms for helioseismic inversion," Ph.D. dissertation, University of Aarhus, Denmark, 1998.

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

[38] R. J. Hanson and C. L. Lawson, "Extensions and applications of the Householder algorithm for solving linear least squares problems," *Mathematics of Computation*, vol. 23, no. 108, pp. 787–812, 1969.