Multidimensional ESPRIT for Damped and Undamped Signals: Algorithm, Computations, and Perturbation Analysis
Souleymen Sahnoun, Konstantin Usevich, Pierre Comon

To cite this version:

HAL Id: hal-01360438
https://hal.archives-ouvertes.fr/hal-01360438v3
Submitted on 7 Jun 2017

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 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 single-tone 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, wireless communication channel estimation, antenna array processing and radar. 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, and subspace approaches such as matrix enhancement and matrix pencil (MEMP), 2-D ESPRIT, R-D ESPRIT, Shaped ESPRIT, improved multidimensional folding (IMDF), Tensor-ESPRIT, principal-singular-vector utilization for modal analysis (PUMA) and the methods proposed in [13] and [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 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 [25].

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.
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 formalism and multilevel Hankel (MH) matrices, that is also to description of the extension [5] of the 2-D ESPRIT algorithm to N-D signals, since we are interested in possibly damped signals\(^1\).

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 [23], [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.

II. BACKGROUND AND PROBLEM STATEMENT

A. Notation

In this paper we use the following fonts: lowercase (a) for scalars, boldface (A) for matrices, and calligraphic (A) for N-D arrays (tensors). Vectors are, by convention, one-column matrices. The elements of vectors/matrices/tensors are accessed as (a)\(^i,j\).\(^{mxn}\), where (tensors). Vectors are, by convention, one-column matrices. The row-major vectorization is used because it is compatible with the Kronecker product. Unlike in [32], we use a special notation for row-major vectorisation, i.e.

\[
\text{vec}(A) = (A)_{1,1},..., (A)_1,..., (A)_{1,1}\]

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

\[
\text{vec}(a_1 \otimes \cdots \otimes a_N) = a_1 \otimes \cdots \otimes a_N.
\]  \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 \(a(M)\) for the Vandermonde-structured vector

\[
a^{(M)} = \begin{bmatrix} \begin{array}{c} a \ a^2 \ \cdots \ a^{(M-1)} \end{array} \end{bmatrix}\]

For a vector \(v \in \mathbb{C}^M\) we denote by \(\text{Diag}(v)\) the \(M \times M\) diagonal matrix with the elements of \(v\) on the diagonal; for a matrix \(A \in \mathbb{C}^{M \times M}\), \(\text{Diag}(A)\) 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 \(m_n = 0, \ldots, M_n - 1\):

\[y(m_1, \ldots, m_N) = y(m_1, \ldots, m_N) + \varepsilon(m_1, \ldots, m_N),\]

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:

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

\(^1\)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].
where
- $c_r$ are complex amplitudes,
- $a_{r,n} = e^{-\pi \nu r + j \omega r n}$ are modes in the $n$-th dimension,
- $\{a_{r,n}\}_{r=1}^{R_N} \in \mathbb{C}^{M_N}$ are real damping factors (not necessary positive),
- $\{\nu_{r,n} = 2\pi \nu r_n\}_{r=1}^{R_N} \in \mathbb{R}^{M_N}$ are angular frequencies.

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

### 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

$$\mathcal{Y} = \sum_{r=1}^{R} c_r \mathcal{A}_{r}^{(M_1)} \otimes \cdots \otimes \mathcal{A}_{r}^{(M_N)}$$

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

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

$$\text{vec}(\mathcal{Y}) = \left( \mathcal{A}_{1}^{(M_1)} \circ \mathcal{A}_{2}^{(M_2)} \circ \cdots \circ \mathcal{A}_{N}^{(M_N)} \right) \mathbf{c},$$

where $\mathbf{c} = [c_1, \ldots, c_R]^T$ is the vector of amplitudes, and $\mathcal{A}_{r}^{(M_n)}$ defines the Vandermonde matrix of the modes in the $n$-th dimension

$$\mathcal{A}_{r}^{(M_n)} \triangleq \left[ \mathcal{a}_{1,n}^{(M_n)} \cdots \mathcal{a}_{R_N,n}^{(M_n)} \right] \in \mathbb{C}^{M_n \times R}.$$

### 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 $1 \leq L_{n} \leq M_{n}$ and define $K_{n} \triangleq M_{n} - L_{n} + 1$. Define by $y^{(i_1, \ldots, i_N)} \in \mathbb{C}^{L_1 \times \cdots \times L_N}$ the vectorized subarray

$$y^{(i_1, \ldots, i_N)} \triangleq \text{vec}\left\{ \mathcal{Y}^{i_1, \ldots, i_N} \right\} = \mathcal{Y}^{(i_1, \ldots, i_N)}.$$ 

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

$$\mathbf{H} \triangleq \begin{bmatrix} y^{(1, \ldots, 1)} & y^{(1, \ldots, 2)} & \cdots & y^{(1, \ldots, K_N)} \\ y^{(1, 2, \ldots, 1)} & \cdots & \cdots & y^{(K_1, K_2, \ldots, K_N - 1)} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

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

**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

$$\mathbf{H} = \mathbf{P} \text{diag}(\mathbf{c}) \mathbf{Q}^T,$$

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

#### B. Shift properties of subspaces

Let us define the selection matrices

$$\mathbf{I}_1 \triangleq \begin{bmatrix} \mathbf{I}_{L_1} \otimes \mathbf{I}_{L_2} \otimes \cdots \otimes \mathbf{I}_{L_N} \end{bmatrix} \quad (9)$$

$$\mathbf{I}_n \triangleq \begin{bmatrix} \mathbf{I}_{1, \ldots, 1} \otimes \mathbf{I}_{L_n} \otimes \cdots \otimes \mathbf{I}_{L_N} \end{bmatrix} \quad (10)$$

$$\mathbf{I}_n \triangleq \begin{bmatrix} \mathbf{I}_{1, \ldots, 1} \otimes \mathbf{I}_{L_n} \otimes \cdots \otimes \mathbf{I}_{L_N} \end{bmatrix} \quad (11)$$

$$\mathbf{I}_n \triangleq \begin{bmatrix} \mathbf{I}_{1, \ldots, 1} \otimes \mathbf{I}_{L_n} \otimes \mathbf{I}_{L_N} \end{bmatrix} \quad (12)$$

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

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

$$\mathbf{P} \Psi_n \triangleq \mathbf{P},$$

where $\Psi_n = \text{diag}(\mathbf{a}_{1}, \ldots, \mathbf{a}_{n}) = [a_{1,1}, \ldots, a_{R,n}]^T$.

Now consider $\mathbf{U}_s$ the matrix of the leading $R_s$ 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 $\mathbf{F}_n \triangleq \mathbf{T} \Psi_n \mathbf{T}^{-1}$ satisfies the equation

$$\mathbf{U}_s \mathbf{F}_n \triangleq \mathbf{U}_s.$$

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

$$\mathbf{F}_n = \left( \mathbf{I} \mathbf{U}_s \right) \mathbf{T} \Psi_n \mathbf{T}^{-1} \left( \mathbf{I} \mathbf{U}_s \right)^{-1} = \left( \mathbf{U}_s \right)^{T} \left( \mathbf{U}_s \mathbf{U}_s \right) \left( \mathbf{U}_s \right)^{-1} \mathbf{T} \left( \mathbf{I} \mathbf{U}_s \right)^{-1}.$$

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.

---

2The damping factors are important in a number of applications, including NMR spectroscopy [11].

3The subtractions are needed because the indices in the tensor start from 1.

4The definition of the MH matrix corresponds in some other works to the spatial smoothing operation [10], to “smoothed data matrix” [2] or to “enhanced matrix” [1].
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 \(\hat{H}\) from the noisy signal, in the same format as (7).
3) Perform the SVD of \(\hat{H}\), and form the matrix \(\hat{U}_s \in \mathbb{C}^{(L_1\cdots L_N) \times R}\) of the \(R\) dominant singular vectors.
4) Compute the matrices \(\tilde{F}_n\)
   \[\tilde{F}_n := \left(\hat{U}_s^{\dagger}R\hat{U}_s\right)^{\frac{a_n}{2}}.\] (16)
5) Compute a linear combination of matrices, where \(\beta_1, \ldots, \beta_n\) are given parameters.
   \[\tilde{K} = \sum_{n=1}^{N} \beta_n \tilde{F}_n\] (17)
6) Compute a diagonalizing matrix \(T\) of \(\tilde{K}\) (from its eigenvalue decomposition):
   \[\tilde{K} = T \text{Diag}(\eta)T^{-1}.\] (18)
7) Apply the transformation \(T\) to \(F_n\):
   \[\tilde{D}_n = T^{-1}\tilde{F}_nT\], for \(n = 1, \ldots, N\) (19)
8) Extract \(\{\tilde{a}_{1,n}, \ldots, \tilde{a}_{R,n}\}\) from \(\text{diag}(\tilde{D}_n)\), \(n = 1, \ldots, N\)

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 \(T\) is used. Note that, by proper choice of \(\beta_n\), 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 \(\beta_n\), 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] \(F_n\) are jointly triangularized using simultaneous Schur decomposition \(\hat{F}\) and the modes \(a_{r,n}\) are extracted from the diagonals of the triangular matrices.

- In [3], in addition to 2D-ESPRIT, an algorithm under the name “2D-MEMPR 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 \(F_n\), and the matrix \(T\) is used just to perform the pairing of the modes.

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 \(\hat{I}\) and \(I\)) are defined in a slightly different way. The second difference is that in [9] modes are estimated from \(P\). 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 \(P\) and \(Q\) 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 \(\eta_r\) 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 \(\tilde{F}_n = F_n\), i.e. the matrices \(F_n\) are recovered correctly.
2) The \(\eta_r\) are exactly the eigenvalues of \(\tilde{K} = \sum_{n=1}^{N} \beta_n F_n\), Thus the eigenvalue decomposition of \(\tilde{K}\) is unique (up to permutation of columns), and therefore the step of the algorithm retrieves the correct \(T\).

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 \(\beta_k\) 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
\[R \leq \min_n \left\{(L_n - 1) \prod_{i=1}^{N} L_i \prod_{i=1}^{N} K_i\right\}\]
Then for a generic choice of modes, \(\text{rank } P = \text{rank } Q = R\)

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

\(^5\)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.
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 \(T_A\) be the number of flops needed to compute the matrix-vector products \(Av\) and \(A^H u\) for given vectors \(u\) and \(v\). Then the leading \(R\) singular values/vectors of a matrix \(A\) can be found using, for example, Lanczos bidiagonalization [35, Ch. 9] with partial reorthogonalization [36] or \(O(RT_A + R^2(L + K))\) (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 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) \(O(RM \log(M))\) flops;

4) \(O(RLN)\) (can be further reduced to \(O(rL)\) if the selection matrices \(T\) and \(L\) are used);

5-6) \(O(R^3)\) (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 \(H\) is given by:

\[
H = U_s \Sigma_s V_s^H + U_n \Sigma_n V_n^H
\]

where \(\Sigma_n = 0\). The perturbed \(\tilde{H}\) is expressed as

\[
\tilde{H} = H + \Delta H.
\]

First, we recall the expression for the perturbation of \(U_s\): Let

\[
\tilde{H} = U_s \Sigma_s V_s^H + \tilde{U}_n \Sigma_n V_n^H
\]

be the subspace decomposition of \(\tilde{H}\). Then the first-order approximation of the \(U_s - \tilde{U}_s\) is given by

\[
\Delta U_s = U_n U_n^H \Delta H V_s \Sigma_n^{-1} + U_s R,
\]

where \(R\) is an antihermitian matrix (i.e., \(R^H = -R\)) that depends on \(\Delta H\) (the precise expression of the matrix \(R\) 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 \(U_s R\) 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 \(F_n\).

Lemma 4: The first-order perturbation of \(F_n\) is given by

\[
\Delta F_n = (U_n^H)\Delta U_n - \Delta U_n^H F_n.
\]

Proof The proof can be found in Appendix C.

Next, let \(t_r\) denote the eigenvectors of \(K\) (the columns of \(T\)) and \(\tau_r^T\) denotes the rows of \(T^{-1}\):

\[
T = [t_1, \ldots, t_n], \quad T^{-1} = [\tau_1, \ldots, \tau_R]^T.
\]

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

\[
\Delta a_{r,n} = \tau_r^T \Delta F_n \tau_r.
\]

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 \(F_n\) is diagonalized (in particular it does not depend on the coefficients \(\beta_n\)). In fact, it depends only on the perturbations of the matrices \(F_n\). 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

\[
\Delta a_{r,n} = \tau_r^T (U_n^H)\Delta H V_s \Sigma_n^{-1} t_r.
\]

Proof The proof can be found in Appendix C.

Note that the term \(U_s R\) 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 \(H\) 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 $H$ satisfy (3), where the matrix $P$ satisfies (13) for $n = 1, \ldots, N$. Further, by $b_r \in \mathbb{C}^R$ we denote the $r$-th unit vector. Then the first order perturbation of the modes obtained by $N$-DESPRIT admits an expansion

$$
\Delta a_{r,n} = \frac{1}{c_r} b_r^T P_n \left( I - a_{r,n} \mathbf{1} I \right) \Delta H (Q^T)^* b_r. 
$$

(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 $P$ and $Q$. This is a remarkable feature of $N$-DESPRIT (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

$$
\Delta a_{r,n} = v_{r,n}^H \Delta H x_r^*,
$$

(27)

where

$$
v_{r,n}^H = \frac{1}{c_r} b_r^T P_n \left( I - a_{r,n} \mathbf{1} I \right) \Delta H (Q^T)^* b_r.
$$

(28)

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

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

$$
\Delta a_{r,n} = z_{r,n}^H e,
$$

(29)

where the vector $z_{r,n}$ is defined as

$$
z_{r,n} = \text{vec}_r \{ \mathcal{X} \} \ast \mathcal{V},
$$

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 $v_{r,n}$ and $x_r$, and $\mathcal{X} \ast \mathcal{V} \in \mathbb{C}^{M_1 \times \cdots \times M_N}$ is the multidimensional convolution 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. $\mathbb{E} \{ \Delta a_{r,n} \} = 0$ if $e$ is zero-mean.
2. $\mathbb{E} \{ \Delta a_{r,n}^2 \} = 0$ if $e$ is circular.
3. If $e$ has covariance matrix $\Gamma = \mathbb{E} \{ ee^H \}$, then

$$
\mathbb{E} \{ |\Delta a_{r,n}|^2 \} = z_{r,n}^H \Gamma z_{r,n}.
$$

4) If $e$ is complex circular Gaussian,

$$
\mathbb{V}ar(\Delta a_{r,n}) = \mathbb{V}ar(\Delta x_r) = \mathbb{E} \{ |\Delta a_{r,n}|^2 \} /
$$

(30)

In particular, if $e$ is white with variance $\sigma_e^2$, then

$$
\mathbb{E} \{ |\Delta a_{r,n}|^2 \} = \sigma_e^2 ||z_{r,n}||_2^2.
$$

**Remark 7:** As in the previous subsection, the new formula for the variance $\mathbb{E} \{ |\Delta a_{r,n}|^2 \}$ allows for an *a priori* perturbation analysis. It also shows a remarkable feature of $N$-DESPRIT: 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 perturbations),** the vectors $v_{r,n}$ and $x_r$ do not require the computation of pseudo-inverses. Indeed, $x_r$ can be obtained by the QR decomposition of $Q$, followed by solving a triangular system. It is similar for $v_{r,n}$. Finally $z_{r,n}$ 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

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

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 $a^T = 1 / \|a\|_2 a^H$ for any vector $a$, and the matrices $\tilde{F}_n$ defined in (16) are just scalars, the steps 4–8 of $N$-DESPRIT are equivalent to defining the estimates $\tilde{a}_n$ as

$$
\tilde{a}_n = \tilde{F}_n = \frac{1}{\|a\|_2} (\tilde{u}^H)^n \tilde{u},
$$

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

$$
p = \left( a_1^{(L_1)} \otimes \cdots \otimes a_n^{(L_n)} \otimes \cdots \otimes a_N^{(L_N)} \right)
$$

and

$$
q = \left( a_1^{(K_1)} \otimes \cdots \otimes a_n^{(K_n)} \otimes \cdots \otimes a_N^{(K_N)} \right).
$$

(30)

Hence (26) becomes to

$$
\Delta a_n = \frac{1}{c_p \|p\|_2 \|q\|_2^2} p^H \left( I - a_{n} \mathbf{1} I \right) \Delta H q^*.
$$

From Lemma 6 we get the following expression.

**Lemma 7:** The first-order perturbation is expressed as

$$
\Delta a_n = z_{n}^H e,
$$

(31)
Proposition 3: In the undamped case the expression is given in (35).

\[ g(L, M, a) = \begin{cases} \frac{1}{K} \left( 1, \frac{L-1}{2}, \frac{L-2}{2} \right), & \text{if } L \leq \frac{M}{2} + 1 \text{ and } |a| = 1, \\ \frac{1}{2} \left( 1, \frac{L-2}{2}, \frac{L-1}{2} \right), & \text{if } L \geq \frac{M}{2} + 1 \text{ and } |a| = 1. \end{cases} \] (35)

In the damped case (|a| ≠ 1), the function \( g(L, M, a) \) can be found as in 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 \( \mathbb{E} \{ |\Delta \omega_a|^2 \} \), 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 \( a_n = e^{-\alpha + j \omega} \) and a white Gaussian noise it is given by

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

- 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)).

**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, α) = (1 − |α|^2) \times \begin{cases} 
-2L(|1−|α|^2)(|α|^2K+|α|^{2L}) + \frac{1+|α|^2}{(1−|α|^2)^2(1−|α|^2K)^2} & \text{if } L \leq \frac{M+1}{2} \text{ and } |α| \neq 1 \\
2K(|1−|α|^2)^2(|α|^2K+|α|^{2L}) + \frac{1+|α|^2}{(1−|α|^2)^2(1−|α|^2K)^2} & \text{if } L \geq \frac{M+1}{2} \text{ and } |α| \neq 1 
\end{cases} 
(36)

Fig. 7. Theoretical and empirical MSEs for 2-D ESPRIT (fast SVD) versus $L$, $(L = L_1 = L_2)$. $(\alpha_0, \omega_0) = (0.1, 0.2\pi)$, $(\alpha_0, \omega_0) = (0.1, 0.4\pi)$, $(M_1, M_2) = (100, 100)$, SNR = 40 dB.

Fig. 8. Theoretical and empirical tMSEs for 2-D ESPRIT versus SNR. $(L_1, L_2) = (4, 4)$. $(\omega_0, \omega_0) = (0.1, 0.2\pi)$, $(\alpha_0, \omega_0) = (0.1, 0.4\pi)$, $(M_1, M_2) = (10, 10)$.

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 [60].

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 $10 \times 10$ 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 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
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 [V-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.

### 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

\[
J_{k_n}^{L_n} = \begin{bmatrix} 0_{L_n \times (k_n-1)} & I_{L_n} & 0_{L_n \times (K_n-k_n)} \end{bmatrix}
\]

\[
J_{k_1,k_2,\ldots,k_N} = J_{k_1}^{L_1} \boxtimes J_{k_2}^{L_2} \boxtimes \cdots \boxtimes J_{k_N}^{L_N}
\]

where \( J_{k_n}^{L_n} \) and \( J_{k_1,k_2,\ldots,k_N} \) 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

\[
J_{k_1,k_2,\ldots,k_N} \mathbf{y} = \mathbf{y}^{(k_1,\ldots,k_N)},
\]

where \( \mathbf{y} = \text{vec}(\mathbf{Y}) \) 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:

\[
H = \begin{bmatrix} H_0 & H_1 & \cdots & H_{K_1-1} \\
H_1 & H_2 & \cdots & H_{K_1} \\ \vdots & \vdots & \ddots & \vdots \\
H_{K_1-1} & H_{K_1} & \cdots & H_{M_1-1} \end{bmatrix},
\]

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

\[
\mathbf{H}_{m_1,\ldots,m_r} = \begin{bmatrix} \mathbf{H}_{m_1,\ldots,m_r,0} & \mathbf{H}_{m_1,\ldots,m_r,1} & \cdots & \mathbf{H}_{m_1,\ldots,m_r,K_{r+1}-1} \\
\mathbf{H}_{m_1,\ldots,m_r,1} & \mathbf{H}_{m_1,\ldots,m_r,2} & \cdots & \mathbf{H}_{m_1,\ldots,m_r,K_{r+1}} \\ \vdots & \vdots & \ddots & \vdots \\
\mathbf{H}_{m_1,\ldots,m_r,K_{r+1}-1} & \mathbf{H}_{m_1,\ldots,m_r,K_{r+1}} & \cdots & \mathbf{H}_{m_1,\ldots,m_r,M_{r+1}-1} \end{bmatrix}
\]

and the blocks of the last level are just scalars

\[
\mathbf{H}_{m_1,\ldots,m_N} = \mathbf{y}(m_1,\ldots,m_N).
\]
Algorithm 1: MH matrix-vector multiplication

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

1) Construct the tensorization \( \mathcal{X} \in \mathbb{C}^{K_1 \times \cdots \times K_N} \) of \( x \).
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 
\[
\mathcal{X}_{1,1,1,\ldots,1} = \mathcal{X}_{1,1,1,\ldots,1}^* = \mathcal{X}_{1,1,1,\ldots,1}^*,
\]
and other elements are zeros.
3) Compute the N-D FFT of \( \mathcal{Y} \) and \( \mathcal{X}^* \)
\[
\mathcal{X}^{	ext{FFT}} := \mathcal{X}^* \bullet F_{M_1} \bullet F_{M_2} \cdots \bullet F_{M_N},
\]
\[
\mathcal{Y}^{	ext{FFT}} := \mathcal{Y} \bullet F_{M_1} \bullet F_{M_2} \cdots \bullet F_{M_N},
\]
4) Compute the following tensor
\[
\mathcal{W}^{	ext{FFT}} := (\mathcal{Y}^* \odot \mathcal{X}^*),
\]
where \( \odot \) is the elementwise (Hadamard) product.
5) Compute the inverse FFT of \( \mathcal{W}^{	ext{FFT}} \)
\[
\mathcal{W} := \mathcal{W}^{	ext{FFT}} \bullet F_{M_1} \bullet F_{M_2} \cdots \bullet F_{M_N}.
\]
6) Extract \( w \) by truncating and vectorizing \( \mathcal{W} \)
\[
w = \text{vec}_i(\{\mathcal{W}\}_{1,1,1,\ldots,1}). \]

Algorithm 2: MH bilinear transform

input : Vector \( x \in \mathbb{C}^{K_1 \times \cdots \times K_N}, v \in \mathbb{C}^{L_1 \times \cdots \times L_N} \).
output: The vector \( z \) in the bilinear operation (22).

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 \( \mathcal{X}', \mathcal{V}' \in \mathbb{C}^{M_1 \times \cdots \times M_N} \)
\[
\mathcal{X}'_{1,1,1,\ldots,1} = \mathcal{X}_{1,1,1,\ldots,1} = \mathcal{X}_{1,1,1,\ldots,1}^*,
\]
\[
\mathcal{V}'_{1,1,1,\ldots,1} = \mathcal{V}_{1,1,1,\ldots,1} = \mathcal{V}_{1,1,1,\ldots,1}^*,
\]
and other elements are zeros.
3) Compute the N-D FFT of \( \mathcal{X}' \) and \( \mathcal{V}' \)
\[
\mathcal{X}'^{	ext{FFT}} := \mathcal{X}' \bullet F_{M_1} \bullet F_{M_2} \cdots \bullet F_{M_N},
\]
\[
\mathcal{V}'^{	ext{FFT}} := \mathcal{V}' \bullet F_{M_1} \bullet F_{M_2} \cdots \bullet F_{M_N},
\]
4) Compute the inverse FFT of the Hadamard product
\[
\mathcal{Z} := (\mathcal{V}^* \circ \mathcal{X}^*) \bullet F_{M_1} \bullet F_{M_2} \cdots \bullet F_{M_N}.
\]
5) Extract \( w \) by vectorizing \( \mathcal{Z} \)
\[
z = \text{vec}_i(\{\mathcal{Z}\}).
\]

\( \mathcal{X}, \mathcal{V} \) are tensorizations of \( x, v \), and \( z = \text{vec}_i(\{\mathcal{Z}\}) \), where \( \mathcal{Z} := \mathcal{X}^* \mathcal{V} \), i.e.
\[
(\mathcal{Z})_{i_1,\ldots,i_N} := \sum_{i_1+j_1=1}^{K_1,\ldots,K_N} (\mathcal{X})_{i_1,\ldots,i_N} (\mathcal{V})_{j_1,\ldots,j_N}.
\]
Indeed, from (41), we have that
\[
v^\top H x^* = \sum_{i_1+j_1=1}^{K_1,\ldots,K_N} \sum_{i_1,\ldots,i_N=1}^{i_1+j_1,\ldots,i_N} (\mathcal{X})_{i_1,\ldots,i_N} (\mathcal{V})_{j_1,\ldots,j_N}.
\]
which completes the proof. \( \square \)

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 \( A = U_s \) and \( B = \tilde{U} \). By the rule of differentiation of the product, \( \Delta (A B) = A^\dagger \Delta B + \Delta (A^\dagger ) B \). Next, since \( A \) is full column rank, \( \Delta (A^\dagger ) \) can be expressed using (43) Proposition 1:
\[
\Delta (A^\dagger B) = A^\dagger \Delta B + (-A^\dagger A A^\dagger )^{-1} \Delta A (I - A A^\dagger ) B.
\]
Since \( A \) and \( B \) span the same column space, \( (I - A A^\dagger ) B = 0 \), and (43) becomes
\[
\Delta (A^\dagger B) = A^\dagger (\Delta B - A A^\dagger ) B.
\]
Finally, since \( F_n = A^\dagger B \), (44) can be simplified to (23). \( \square \)

Proof of Lemma 5 First, from (19) and the rule of differentiation of inverses \( \Delta (T^{-1}) = -T^{-1} \Delta TT^{-1} \), we have that:
\[
\Delta D_n = T^{-1} \Delta F_n T + T^{-1} F_n \Delta T - T^{-1} \Delta TT^{-1} F_n T
\]
\[
= T^{-1} \Delta F_n T + D_n T^{-1} \Delta T - T^{-1} \Delta T D_n.
\]
Next, we denote by \( b_r \) the \( r \)-th unit vector, and write
\[
\Delta a_{r,n} = b_r^T \Delta D_n b_r \\
= \tau_r^T \Delta F_n t_r + b_r^T D_n T^{-1} \Delta T b_r - \tau_r^T T^{-1} \Delta T D_n b_r. 
\] (45)
Since \( b_r^T D_n = a_{r,n} b_r^T \) and \( D_n b_r = a_{r,n} b_r \), the last two terms in (45) cancel, and eqn. (24) takes place. □

**Proof of Corollary** [I] Next, we combine (22) and (23). For simplicity, denote \( C = \Delta H V_s \Sigma_s^{-1} \). Then
\[
\Delta F_n = (U_n)^{\dagger} \left( \frac{n}{n} I \Delta U_s - I \Delta U_s \Sigma_s \right) (U_n)^{\dagger} F_n, \\
= (U_n)^{\dagger} \left( \frac{n}{n} I (I - U_s U^H_s) C - I (I - U_s U^H_s) C F_n \right) \\
+ (U_n)^{\dagger} \left( U_s R - U_s \Sigma_s \right) F_n. 
\] By expanding the parentheses and using the identities
\[
(U_n)^{\dagger} U_n = F_n, \quad (U_n)^{\dagger} U_n = I,
\] we get
\[
\Delta F_n = (U_n)^{\dagger} \left( \frac{n}{n} I C - I \Delta U_s \Sigma_s \right) (U_n)^{\dagger} F_n + (U_n)^{\dagger} \left( \frac{n}{n} \Delta F_n - F_n \Sigma_s \right) F_n. 
\] (46)
where \( G = (U_s^H C - R) \). Next, we combine (46) and (24).
\[
\Delta a_{r,n} = \tau_r^T \left( U_n \right)^{\dagger} \left( \frac{n}{n} I C - I \Delta U_s \Sigma_s \right) (U_n)^{\dagger} F_n \\
+ \tau_r^T \left( G \Delta F_n - F_n \Sigma_s \right) F_n. 
\] (47)
Since \( F_n t_r = a_{r,n} t_r \) and \( \tau_r^T F_n = a_{r,n} \tau_r \), the last term in (47) vanishes, and eqn. (47) is simplified to (35). □

**Proof of Proposition** [I] From equation between (12) and (14), we have that
\[
U_s = PT^{-1}, \quad \Sigma_s V_s^H = T \text{Diag}(c) Q^T. 
\] where the matrices \( P \) and \( Q \) are defined in (6). Next, since
\[
(\Sigma_s V_s^H)^{\dagger} = V_s \Sigma_s^{-1}, 
\] and from properties of the pseudoinverse, we have that
\[
\Delta a_{r,n} = \tau_r^T \left( U_n \right)^{\dagger} \left( \frac{n}{n} I - a_{r,n} a_n \right) \Delta H \left( \Sigma_s V_s^H \right)^{\dagger} t_r \\
= \tau_r^T \left( P T^{-1} \right)^{\dagger} \left( \frac{n}{n} I - a_{r,n} a_n \right) \Delta H \left( T \text{Diag}(c) Q^T \right)^{\dagger} t_r \\
= \tau_r^T \left( T P \right)^{\dagger} \left( \frac{n}{n} I - a_{r,n} a_n \right) \Delta H \left( Q^T \right)^{\dagger} t_r \\
= \frac{1}{\epsilon_r} a_{r,n} \left( P \right)^{\dagger} \left( \frac{n}{n} I - a_{r,n} a_n \right) \Delta H \left( Q^T \right)^{\dagger} t_r. 
\] □

**Proof of Lemma** [I] First, we remark that
\[
\left( I_n - a_n a_n \right)^{\dagger} = \left( a_n^{(L_1)} \otimes \cdots \otimes b \otimes \cdots \otimes a_n^{(L_N)} \right)^{\dagger} \Delta H, 
\] (48)
where \( b = (I_{L_n} - a_n a_n) \). Next, the convolution of two rank-one tensors is a rank-one tensor:
\[
(x_1 \otimes \cdots \otimes x_N) \ast (v_1 \otimes \cdots \otimes v_N) = (x_1 \ast v_1) \otimes \cdots \otimes (x_N \ast v_N). 
\] Hence, by applying (1) and Lemma 5 to (48) and (50), we get the desired result. □

**Proof of Proposition** [I] We define \( L_{**} = \min(L, K) \) and \( K_{**} = \max(L, K) \). Then the vector in the denominator can be explicitly written as
\[
a^{(L)} \ast a^{(K)} = \left[ 1, 2a, \ldots, L_{**} a^{(L_{**} - 1)}, L_{**} a^{(L_{**})}, \ldots, L_{**} a^{(K_{**} - 2)}, L_{**} a^{(K_{**})}, \ldots, 2a^{(M - 1)}, a^{(M - 1)} \right]. 
\] Next, we consider each case separately.

**a) Undamped case:** In this case,
\[
\|a^{(L)}\|^2 = L_{**}^2, \quad \|a^{(K)}\|^2 = K_{**}^2, \quad \text{and}
\]
\[
\|a^{(L)} \ast a^{(K)}\|^2 = L_{**}^2 (M - 2L_{**}) + 2 \sum_{k=1}^{L_{**}} k^2 \\
= L_{**}^2 (K_{**} - L_{**} - 1) + \frac{2L_{**}^3}{3} + L_{**} + \frac{L_{**}}{3} \\
= L_{**}^2 K_{**} - \frac{L_{**}^2 (L_{**} - 1)}{3}. 
\]
By combining all these expressions together, we get eqn. (35).

**b) Damped case:** Then the squared norm of the convolution is equal to
\[
\|a^{(L)} \ast a^{(K)}\|^2 = \sum_{k=1}^{L_{**}} k^2 |a|^{2(k-1)} \\
+ L_{**}^{K_{**}-1} \sum_{k=L_{**}+1}^{M} |a|^{2(k-1)} + \sum_{k=K_{**}}^{M} (M - k + 1)^2 |a|^{2(k-1)} \\
= \sum_{k=1}^{L_{**}} k^2 |a|^{2(k-1)} + L_{**} \sum_{k=L_{**}+1}^{M} |a|^{2(k-1)} \\
+ \sum_{j=1}^{M-L_{**}} L_{**}^2 \sum_{k=L_{**}+1}^{M} |a|^{2(k-1)} \\
+ |a|^{2(M-1)} \sum_{j=1}^{M-L_{**}} \frac{1}{|a|^{2(j-1)}}. 
\]
In order to simplify the expression, we use the fact that for \( \rho \neq 1 \)
\[
\sum_{k=1}^{L} k^2 \rho^{k-1} = L^2 \rho - \frac{\rho^L}{\rho - 1} - 2L \frac{\rho^L}{(\rho - 1)^2} + \frac{(\rho^L - 1)(\rho + 1)}{(\rho - 1)^3}. 
\]
By substituting \( \rho = |a|^{2} \) and get
\[
\sum_{k=1}^{L_{**}} k^2 |a|^{2(k-1)} = \frac{L_{**}^2 |a|^{2L_{**}}}{|a|^{2} - 1} - 2L_{**} |a|^{2L_{**}} \frac{1}{(|a|^{2} - 1)^2} + \frac{|a|^{2L_{**} - 1}(|a|^{2} + 1)}{(|a|^{2} - 1)^3}. 
\]
Next we take \( \rho = |\alpha|^{-2} \), and get
\[
|\alpha|^{2(M-1)} \sum_{j=1}^{L_*} \left( \frac{1}{|\alpha|} \right)^{2(2j-1)} = |\alpha|^{2(M-1)} L_*^2 \left( \frac{|\alpha|^{-2L_*}}{|\alpha|^{-2} - 1} - 2L_* \frac{|\alpha|^{-2L_*}}{(|\alpha|^{-2} - 1)^2} \right) + \frac{1}{(|\alpha|^{-2} - 1)^3} (|\alpha|^{-2} - 1 + |\alpha|^{2K_*} - 1) (|\alpha|^2 + 1).
\]
Next, due to the fact that
\[
\sum_{k=L_*+1}^{K_*} |\alpha|^{2(k-1)} = L_*^2 \frac{|\alpha|^{2(K_*-1)} - |\alpha|^{2L_*}}{|\alpha|^2 - 1},
\]
and after cancellations of some terms, we have
\[
\frac{1}{|\alpha|^2} - 2L_* \frac{|\alpha|^{2K_*}}{(|\alpha|^2 - 1)^2} + \frac{1}{(|\alpha|^2 - 1)^3} (|\alpha|^{2K_*} - 1) (|\alpha|^{2L_*} - 1) (|\alpha|^2 + 1).
\]
Finally, combining (49) with
\[
\|\alpha^{(L_*)-1}\|^2 = \frac{|\alpha|^{2L_*} - 1}{|\alpha|^2 - 1}, \quad \|\alpha^{(K_*)}\|^2 = \frac{|\alpha|^{2K_*} - 1}{|\alpha|^2 - 1},
\]
yields eqn. (50).

REFERENCES