Multiple Invariance ESPRIT: A Coupled Matrix-Tensor Factorization Approach

Mikael Sørensen and Lieven De Lathauwer, *Fellow, IEEE*

***Abstract*—The Canonical Polyadic Decomposition (CPD) of higher-order tensors has proven to be an important tool for array processing. CPD approaches have so far assumed regular array geometries such as uniform linear arrays. However, in the case of sparse arrays such as Nonuniform Linear Arrays (NLAs), the CPD approach is not suitable anymore. Using the Coupled Matrix-Tensor Factorization (CMTF) we propose in this paper a multiple invariance ESPRIT method for both one- and multi-dimensional NLA processing. We obtain a multiresolution ESPRIT method for sparse arrays with multiple baselines. The CMTF framework also yields a new uniqueness condition that is relaxed compared to the CPD approach. It also leads to an eigenvalue decomposition based algorithm that is guaranteed to reduce the multi-source NLA problem into decoupled single-source NLA problems in the noiseless case. Finally, we present a new polynomial rooting procedure for the latter problem, which again is guaranteed to find the solution in the noiseless case.**

***Index Terms*—array processing, direction-of-arrival estimation, tensor, coupled decomposition, sparse array, multiple baseline, multiresolution, nonuniform sampling.**

I. Introduction

The connection between Uniform Linear Arrays (ULAs) and the Canonical Polyadic Decomposition (CPD) is well-known, see [17] for a discussion. In order to increase the aperture or to deal with malfunctioning sensors in ULAs, Nonuniform Linear Arrays (NLAs) have been used, see [10], [27], [5], [12], [25], [13] and references therein. Because of the nonuniform sampling, the CPD framework is not suitable for NLA processing. Many of the developed methods for NLA make the assumption that the impinging signals are uncorrelated (orthogonal). This assumption often makes it possible to transform the original NLA problem into a uniformly sampled (multidimensional) harmonic retrieval problem.

As our contribution, we present a link between the Coupled Matrix-Tensor Factorization (CMTF) model [20], [21] and NLA processing. A key feature of the CMTF approach is that it works directly on the data (e.g. no data imputation/interpolation or signal statistics are used and orthogonality is not assumed). Interestingly, the CMTF approach can be interpreted as a Multiple Invariance ESPRIT (MI-ESPRIT) method [23] for one- and multi-dimensional NLA. We obtain a new uniqueness condition and derive an algorithm for NLA processing. In particular, we adapt some of the ULA results in [18] to NLAs.

A related problem concerns Direction-Of-Arrival (DOA) estimation based on sparse arrays embedded with multiple baselines (e.g. [6], [29], [8], [1]). The multiple baseline property can make sparse arrays ambiguity-free when several of the involved shift-invariance structures are exploited together. However, the existing results have so far been limited to dual shift-invariance structures and only CPD based methods have been considered. We explain that the CMTF model provides an algebraic framework for sparse array processing in the case of multiple baselines. This CMTF approach can be interpreted as a multiresolution ESPRIT method [8].

Even though we focus on sensor array processing, the results presented in this paper may also be used in, e.g., system identification and parametric spectral analysis.

The paper is organized as follows. The rest of the introduction will present the notation. Section II reviews the necessary algebraic prerequisites. Section III presents an algebraic CMTF framework for 1D NLA processing including a relaxed uniqueness condition. An algebraic algorithm is developed in Section IV. A nice feature of the CMTF approach is that it can easily be generalized to the multidimensional case. This is illustrated in Section V for two-dimensional (2D) arrays composed of superimposed NLAs (e.g. L-shaped arrays with missing sensors). Numerical experiments are reported in section VI. Section VII concludes the paper.

Notation: Vectors, matrices and tensors are denoted by lower case boldface, upper case boldface and upper case calligraphic letters, respectively. The rth column vector and transpose of a matrix $\mathbf{A}$ are denoted by $\mathbf{a}_r$ and $\mathbf{A}^T$, respectively. The symbols $\otimes$ and $\odot$ denote the Kronecker and Khatri-Rao product, defined as

$$\mathbf{A}\otimes\mathbf{B} := \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \mathbf{A}\odot\mathbf{B} := [\mathbf{a}_1\otimes\mathbf{b}_1 \;\; \mathbf{a}_2\otimes\mathbf{b}_2 \;\ldots],$$

in which $(\mathbf{A})_{mn} = a_{mn}$. The outer product of three vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is denoted by $\mathbf{a}\circ\mathbf{b}\circ\mathbf{c}$, such that $(\mathbf{a}\circ\mathbf{b}\circ\mathbf{c})_{ijk} = a_i b_j c_k$. The range, kernel, Frobenius norm, real part and imaginary part of matrix $\mathbf{A}$ are denoted by $\text{range}(\mathbf{A})$, $\ker(\mathbf{A})$, $\|\mathbf{A}\|_F$, $\text{Re}\{\mathbf{A}\}$ and $\text{Im}\{\mathbf{A}\}$. From the context it should be clear when i denotes the imaginary

M. Sørensen and L. De Lathauwer are with KU Leuven - E.E. Dept. (ESAT) - STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium, the Group Science, Engineering and Technology, KU Leuven Kulak, E. Sabbelaan 53, 8500 Kortrijk, Belgium, and iMinds Medical IT, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium, {Mikael.Sorensen, Lieven.DeLathauwer}@kuleuven-kulak.be.

unit number, i.e., $i = \sqrt{-1}$.

Matlab index notation will be used for submatrices of a given matrix. For example, $\mathbf{A}(1{:}k,:)$ represents the submatrix of $\mathbf{A}$ consisting of the rows from 1 to k of $\mathbf{A}$. Let $\mathbf{A} \in \mathbb{C}^{I\times R}$, then $\overline{\mathbf{A}} = \mathbf{A}(2:I,:) \in \mathbb{C}^{(I-1)\times R}$ and $\underline{\mathbf{A}} = \mathbf{A}(1:I-1,:) \in \mathbb{C}^{(I-1)\times R}$, i.e., $\overline{\mathbf{A}}$ and $\underline{\mathbf{A}}$ are obtained by deleting the top and bottom row of $\mathbf{A}$, respectively.

Let $\mathbf{C}_{m,n} \in \mathbb{C}^{mn\times mn}$ denote the commutation matrix property with $\mathbf{C}_{m,n}(\mathbf{a}\otimes\mathbf{b}) = \mathbf{b}\otimes\mathbf{a}$, where $\mathbf{a}\in\mathbb{C}^m$ and $\mathbf{b}\in\mathbb{C}^n$. We will also make use of the linear operator $\mathbf{S}_n(\mathbf{X}) = \mathbf{C}_{2,n}\left[\begin{smallmatrix}\mathbf{I}_n\mathbf{X}\\ \bar{\mathbf{I}}_n\mathbf{X}\end{smallmatrix}\right] \in \mathbb{C}^{2(n-1)\times p}$, where $\mathbf{X}\in\mathbb{C}^{n\times p}$ and where $\mathbf{I}_n \in \mathbb{C}^{n\times n}$ denotes the identity matrix.

The cardinality of a set S is denoted by $\mathrm{card}(S)$. The binomial coefficient is denoted by $C_m^k = \frac{m!}{k!(m-k)!}$. The k-th compound matrix of $\mathbf{A}\in\mathbb{C}^{I\times R}$ is denoted by $C_k(\mathbf{A}) \in \mathbb{C}^{C_I^k\times C_R^k}$. It is the matrix containing the determinants of all $k\times k$ submatrices of $\mathbf{A}$, arranged with the submatrix index sets in lexicographic order, see [3] and references therein for a discussion. The modulus and argument of $a\in\mathbb{C}$ are denoted by $|a|$ and $\arg(a)$, respectively. The greatest common divisor of $M, N \in \mathbb{N}$ is denoted by $\mathrm{GCD}(M,N)$.

II. Algebraic Prerequisites

A. Canonical Polyadic Decomposition (CPD)

Consider the third-order tensor $\mathcal{X}\in\mathbb{C}^{I\times J\times K}$. We say that $\mathcal{X}$ is a rank-1 tensor if it is equal to the outer product of non-zero vectors $\mathbf{a}\in\mathbb{C}^I$, $\mathbf{b}\in\mathbb{C}^J$ and $\mathbf{c}\in\mathbb{C}^K$ such that $x_{ijk} = a_i b_j c_k$. A Polyadic Decomposition (PD) is a decomposition of $\mathcal{X}$ into a sum of rank-1 terms:

$$\mathbb{C}^{I\times J\times K} \ni \mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r\,. \tag{1}$$

The rank of a tensor $\mathcal{X}$ is equal to the minimal number of rank-1 tensors that yield $\mathcal{X}$ in a linear combination. Assume that the rank of $\mathcal{X}$ is R, then (1) is called the CPD of $\mathcal{X}$. Let us stack the vectors $\{\mathbf{a}_r\}$, $\{\mathbf{b}_r\}$ and $\{\mathbf{c}_r\}$ into the matrices $\mathbf{A} = [\mathbf{a}_1,\ldots,\mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1,\ldots,\mathbf{b}_R]$ and $\mathbf{C} = [\mathbf{c}_1,\ldots,\mathbf{c}_R]$. The matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ will be referred to as the factor matrices of the (C)PD of $\mathcal{X}$ in (1).

1) Matrix Representation: Let $\mathbf{X}^{(i\cdot\cdot)} \in \mathbb{C}^{J\times K}$ denote the matrix such that $\left(\mathbf{X}^{(i\cdot\cdot)}\right)_{jk} = x_{ijk}$, then $\mathbf{X}^{(i\cdot\cdot)} = \mathbf{B}D_i(\mathbf{A})\mathbf{C}^T$ and we obtain the following matrix representation of (1):

$$\mathbb{C}^{IJ\times K} \ni \mathbf{X} := \left[\mathbf{X}^{(1\cdot\cdot)T} \;\ldots\; \mathbf{X}^{(I\cdot\cdot)T}\right]^T = (\mathbf{A}\odot\mathbf{B})\mathbf{C}^T. \tag{2}$$

2) Uniqueness: The uniqueness properties of the CPD have been intensely studied in recent years, see [3], [4], [16] and references therein. In this paper we consider the case where one factor matrix has full column rank. For this case the following sufficient uniqueness condition has been obtained.

Theorem II.1. *Consider the PD of $\mathcal{X}\in\mathbb{C}^{I\times J\times K}$ in* (1)*. If*

$$\begin{cases}\boldsymbol{C} \text{ has full column rank,}\\ C_2(\boldsymbol{A})\odot C_2(\boldsymbol{B}) \text{ has full column rank,}\end{cases} \tag{3}$$

then the rank of $\mathcal{X}$ is R and the CPD of $\mathcal{X}$ is unique [7], [2], [3]. Generically[1]*, condition* (3) *is satisfied if $2R(R-1) \le I(I-1)J(J-1)$ and $R\le K$ [2], [22].*

The uniqueness condition (3) is fairly easy to check and yet very powerful. Moreover, under condition (3) the CPD can be computed by algebraic means [2]. More precisely, the CPD problem was reduced to a Simultaneous matrix Diagonalization (SD) problem. The SD problem can in the exact case be solved via a standard Generalized EigenValue Decomposition (GEVD).

B. Coupled CPD and Coupled Matrix-Tensor Factorization

We say that a collection of tensors $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n\times J_n\times K}$, $n\in\{1,\ldots,N\}$, admits an R-term coupled PD if each tensor $\mathcal{X}^{(n)}$ can be written as [20]:

$$\mathcal{X}^{(n)} = \sum_{r=1}^{R} \mathbf{a}_r^{(n)} \circ \mathbf{b}_r^{(n)} \circ \mathbf{c}_r\,, \quad n\in\{1,\ldots,N\}, \tag{4}$$

with factor matrices $\mathbf{A}^{(n)} = \left[\mathbf{a}_1^{(n)},\ldots,\mathbf{a}_R^{(n)}\right]$, $\mathbf{B}^{(n)} = \left[\mathbf{b}_1^{(n)},\ldots,\mathbf{b}_R^{(n)}\right]$ and $\mathbf{C} = [\mathbf{c}_1,\ldots,\mathbf{c}_R]$. We define the coupled rank of $\{\mathcal{X}^{(n)}\}$ as the minimal number of coupled rank-1 tensors $\mathbf{a}_r^{(n)}\circ\mathbf{b}_r^{(n)}\circ\mathbf{c}_r$ that yield $\{\mathcal{X}^{(n)}\}$ in a linear combination. Assume that the coupled rank of $\{\mathcal{X}^{(n)}\}$ is R, then (4) will be called the coupled CPD of $\{\mathcal{X}^{(n)}\}$.

The coupled (C)PD of $\{\mathcal{X}^{(n)}\}$ given by (4) has the following matrix representation

$$\mathbf{X} = \left[\mathbf{X}^{(1)T},\ldots,\mathbf{X}^{(N)T}\right]^T = \mathbf{F}\mathbf{C}^T \in \mathbb{C}^{(\sum_{n=1}^N I_n J_n)\times K}, \tag{5}$$

where $\mathbf{F} = [(\mathbf{A}^{(1)}\odot\mathbf{B}^{(1)})^T,\ldots,(\mathbf{A}^{(N)}\odot\mathbf{B}^{(N)})^T]^T \in \mathbb{C}^{(\sum_{n=1}^N I_n J_n)\times R}$.

The coupled rank-1 tensors in (4) can be arbitrarily permuted and the vectors within the same coupled rank-1 tensor can be arbitrarily scaled provided that the overall coupled rank-1 term remains the same. We say that the coupled CPD is unique when it is only subject to these trivial indeterminacies. Uniqueness conditions for the coupled CPD were derived in [20]. Theorem II.2 extends Theorem II.1 to coupled CPD.

Theorem II.2. *Consider the coupled PD of $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n\times J_n\times K}$, $n\in\{1,\ldots,N\}$ in* (4)*. Define*

$$\boldsymbol{G} = \begin{bmatrix} C_2\left(\boldsymbol{A}^{(1)}\right)\odot C_2\left(\boldsymbol{B}^{(1)}\right)\\ \vdots \\ C_2\left(\boldsymbol{A}^{(N)}\right)\odot C_2\left(\boldsymbol{B}^{(N)}\right)\end{bmatrix} \in \mathbb{C}^{\left(\sum_{n=1}^N C_{I_n}^2 C_{J_n}^2\right)\times C_R^2}.$$

If

$$\begin{cases}\boldsymbol{C} \text{ has full column rank,}\\ \boldsymbol{G} \text{ has full column rank,}\end{cases} \tag{6}$$

then the coupled rank of $\{\mathcal{X}^{(n)}\}$ is R and the coupled CPD of $\{\mathcal{X}^{(n)}\}$ is unique [20].

[1] A tensor decomposition property is called generic if it holds with probability one when the entries of the factor matrices are drawn from absolutely continuous probability density measures.

As in ordinary CPD, the uniqueness condition (6) is fairly easy to check and yet very powerful. An extension of the SD method in [2] to the coupled CPD has been developed in [21]. Again, in the noise-free case the SD problem for coupled CPD can be solved via a GEVD.

A special case of (4) is the CMTF of a matrix $\mathbf{X} \in \mathbb{C}^{I\times K}$ and a set of tensors $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n\times J_n\times K}$, $n \in \{1,\dots,N\}$, given by

$$\begin{cases} \mathbf{X} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{c}_r\,, \quad (\mathbf{b}_r = 1), \\ \mathcal{X}^{(n)} = \sum_{r=1}^{R} \mathbf{a}_r^{(n)} \circ \mathbf{b}_r^{(n)} \circ \mathbf{c}_r\,, \quad n \in \{1,\dots,N\}. \end{cases} \tag{7}$$

III. New uniqueness condition for NLA via CMTF

Assume that R sources are impinging on an I element NLA. Let us additionally assume that the sources are located in the far-field, that the narrowband assumption holds and that K snapshots are available such that $\mathbf{s}_r \in \mathbb{C}^K$ denotes the signal vector associated with the rth source. In this section we also assume that the sources are located in the "xy"-plane such that only the azimuth angles are of interest. Let us choose the first sensor in the array as the reference sensor. Under these assumptions the observed data matrix admits the factorization

$$\mathbf{X} = \mathbf{A}\mathbf{S}^T \in \mathbb{C}^{I\times K}, \quad \mathbf{A} = [\mathbf{a}_1\ \dots\ \mathbf{a}_R]\,,\ \mathbf{S} = [\mathbf{s}_1\ \dots\ \mathbf{s}_R]\,, \tag{8}$$

in which

$$\mathbf{a}_r = \left[1\ e^{-i\omega_c\tau_{r2}}\ \dots\ e^{-i\omega_c\tau_{rI}}\right]^T = \left[1\ x_r^{n_2}\ \dots\ x_r^{n_I}\right]^T \in \mathbb{C}^I, \tag{9}$$

where $\tau_{ri} = n_i d_x \cos(\theta_r)/c$ corresponds to the propagation delay associated with the ith sensor and the rth source, ω_c is the carrier frequency, c is the propagation speed, d_x is the nominal unit measure along the sensor axis ("x-axis"), $n_i d_x$ is the nominal distance between the reference sensor ($n_1 = 0$) and the ith sensor in the NLA in which $n_i \in \mathbb{N}$, and $x_r = e^{-i\omega_c d_x \cos(\theta_r)/c}$ in which $0 \le \theta_r < \pi$ denotes the azimuth angle associated with the rth source. Clearly $\mathbf{a}_r$ is an incomplete Vandermonde vector where not all powers appear. This is in contrast to the ULA case where all powers are present, e.g., $\mathbf{a}_r = [1\ x_r\ x_r^2\ \dots\ x_r^{I-1}]^T$.

We define the NLA constrained rank of $\mathbf{X}$ in (8) as the minimal number of NLA constrained rank-1 matrices $\mathbf{a}_r\mathbf{s}_r^T$, where $\mathbf{a}_r$ is of the form (9), that yield $\mathbf{X}$ in a linear combination. We say that the NLA decomposition of $\mathbf{X}$ is unique if for every pair $(\widehat{\mathbf{A}}, \widehat{\mathbf{S}})$ satisfying the relation (8) we have $\widehat{\mathbf{A}} = \mathbf{A}\Pi$, where Π is a permutation matrix.

In DOA applications the goal is to estimate $\{\theta_r\}$ via $\mathbf{A}$, observing $\mathbf{X}$. On the other hand, in source separation the goal is to estimate $\mathbf{S}$, observing $\mathbf{X}$. For simplicity, we will limit the exposition to the case where $\mathbf{S}$ has full column rank. However, the results presented in this paper can be extended to cases where $\mathbf{S}$ does not have full column rank.

In this section we establish a link between NLA processing and the CMTF. The approach can be interpreted as a MI-ESPRIT method [23] for NLA. More precisely, we explain that by simultaneously exploiting several shift-invariance structures inherent in the given NLA it is under certain conditions possible to transform the NLA problem into a CMTF problem. The idea of exploiting several shift-invariance structures was first considered in [23] and later also in [13].

In order to clarify the idea we first provide an informative example in subsection III-A. Next, in subsection III-B, we formalize the reasoning and establish the link between 1D NLA and CMTF. Subsection III-C briefly explains that the proposed CMTF approach also provides a multiresolution interpretation of sparse arrays with multiple baselines.

A. Illustration

Consider the NLA observation matrix

$$\mathbf{X} = \mathbf{A}\mathbf{S}^T \in \mathbb{C}^{I\times K}, \quad \mathbf{a}_r = \left[1\ x_r\ x_r^2\ x_r^4\ x_r^5\ x_r^6\ x_r^8\right]^T,\ I = 7. \tag{10}$$

The shift-invariance structures within $\mathbf{A}$ can be visualized as follows:

$$\begin{array}{c} \\ x \\ x \\ x^2 \\ x^3 \\ x^3 \\ x^4 \\ x^4 \\ x^4 \\ x^5 \\ x^5 \\ x^6 \\ x^6 \\ x^7 \\ x^8 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1 & x & x^2 & - & x^4 & x^5 & x^6 & - & x^8 \end{array} \\ \begin{pmatrix} 1 & x & x^2 & & & & & & \\ & & & & x^4 & x^5 & x^6 & & \\ 1 & & x^2 & & x^4 & & x^6 & & x^8 \\ & x & & & x^4 & & & & \\ & & x^2 & & & x^5 & & & x^8 \\ 1 & & & & x^4 & & & & x^8 \\ & x & & & & x^5 & & & \\ & & x^2 & & & & x^6 & & \\ 1 & & & & & x^5 & & & \\ & x & & & & & x^6 & & \\ 1 & & & & & x^6 & & & \\ & & x^2 & & & & & & x^8 \\ & x & & & & & & & x^8 \\ 1 & & & & & & & & x^8 \end{pmatrix} \end{array}, \tag{11}$$

where the outer column contains the generators from x up to x^8, the outer row denotes the incomplete Vandermonde vector $\mathbf{a}_r$ (in which '-' denotes a missing power of x) and each row inside the matrix denotes the sub-ULA associated with a particular partial shift-invariance. By inspection of (11) it becomes clear that the NLA can be interpreted as a set of superimposed ULAs. From $\mathbf{X}$ in (10) we extract the submatrices $\mathbf{X}^{(1,1)} \in \mathbb{C}^{3\times K}$ and $\mathbf{X}^{(2,1)} \in \mathbb{C}^{3\times K}$ associated with x, and the submatrix $\mathbf{X}^{(1,2)} \in \mathbb{C}^{5\times K}$ associated with x^2 as follows:

$$\mathbf{X}^{(1,1)} = \mathbf{X}([1,2,3],:) = \mathbf{A}^{(1,1)}\mathbf{S}^T, \tag{12}$$

$$\mathbf{X}^{(2,1)} = \mathbf{X}([4,5,6],:) = \mathbf{A}^{(2,1)}\mathbf{S}^T, \tag{13}$$

$$\mathbf{X}^{(1,2)} = \mathbf{X}([1,3,4,6,7],:) = \mathbf{A}^{(1,2)}\mathbf{S}^T, \tag{14}$$

where the submatrices of $\mathbf{A}$ associated with the generators x and x^2, denoted by $\mathbf{A}^{(1,1)} \in \mathbb{C}^{3\times R}$, $\mathbf{A}^{(2,1)} \in \mathbb{C}^{3\times R}$ and $\mathbf{A}^{(1,2)} \in \mathbb{C}^{5\times R}$, have columns given by

$$\mathbf{a}_r^{(1,1)} = \left[1\ x_r\ x_r^2\right]^T,\ \mathbf{a}_r^{(2,1)} = \left[x_r^4\ x_r^5\ x_r^6\right]^T, \tag{15}$$

$$\mathbf{a}_r^{(1,1)} = \left[1\ x_r^2\ x_r^4\ x_r^6\ x_r^8\ \right]^T. \tag{16}$$

Note that the vectors (15)–(16) are (column-wise scaled) Vandermonde vectors. Exploiting the shift-invariance, we obtain from (12)–(14) tensors $\mathcal{Y}^{(1)} \in \mathbb{C}^{4\times 2\times K}$ and

$\mathcal{Y}^{(2)} \in \mathbb{C}^{4\times 2\times K}$ of which the matrix representation admits the coupled decomposition

$$\begin{aligned}\mathbf{Y}^{(1)} &= [S_2(\mathbf{X}^{(1,1)})^T, S_2(\mathbf{X}^{(2,1)})^T]^T = (\mathbf{A}^{(1)} \odot \mathbf{B}^{(1)})\mathbf{S}^T,\\ \mathbf{Y}^{(2)} &= S_2(\mathbf{X}^{(1,2)}) = (\mathbf{A}^{(2)} \odot \mathbf{B}^{(2)})\mathbf{S}^T,\end{aligned}$$

where

$$\begin{aligned}\mathbf{a}_r^{(1)} &= [\underline{\mathbf{a}}_r^{(1,1)T} \mid \underline{\mathbf{a}}_r^{(2,1)T}] = [1\ x_r \mid x_r^4\ x_r^5]^T,\ \mathbf{b}_r^{(1)} = [1\ x_r]^T,\\ \mathbf{a}_r^{(2)} &= \underline{\mathbf{a}}_r^{(1,2)} = [1\ x_r^2\ x_r^4\ x_r^6]^T,\ \mathbf{b}_r^{(2)} = [1\ x_r^2]^T.\end{aligned}$$

We now obtain the coupled CPD of $\{\mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}\}$ with matrix representation

$$\mathbf{Y} = \begin{bmatrix}\mathbf{Y}^{(1)}\\ \mathbf{Y}^{(2)}\end{bmatrix} = \begin{bmatrix}\mathbf{A}^{(1)} \odot \mathbf{B}^{(1)}\\ \mathbf{A}^{(2)} \odot \mathbf{B}^{(2)}\end{bmatrix}\mathbf{S}^T. \quad (17)$$

Thanks to Theorem II.2 the coupled CPD of $\{\mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}\}$ is generically unique if $R \leq 5$. Note that if we only use the CPD of $\mathcal{Y}^{(1)}$, then DOA identifiability can only be guaranteed for $R \leq 4$ (e.g. Theorem II.1).

B. 1D NLA via CMTF

Let us now formalize the reasoning behind the previous example. The first step is to identify ULA structures within $\mathbf{A}$ in (8). In the second step we turn the matrix decomposition into a tensor decomposition via spatial smoothing. In the third and final step we combine the generated tensors, leading to a CMTF problem.

1) Multiple generators lead to multiple sub-ULAs: Consider the NLA observation matrix $\mathbf{X} \in \mathbb{C}^{I\times K}$ given by (8) where $\mathbf{A} \in \mathbb{C}^{I\times R}$ is a NLA matrix (e.g. of the form (10)). We call x_r^n a generator if there exist $I_{m,n} \geq 2$ entries of $\mathbf{a}_r$ denoted by $a_{\sigma_{m,n}(1),r}, a_{\sigma_{m,n}(2),r}, \ldots, a_{\sigma_{m,n}(I_{m,n}),r}$ such that $a_{\sigma_{m,n}(i),r} = a_{\sigma_{m,n}(1),r} x_r^{(i-1)n}$, $\forall i \in \{1,\ldots,I_{m,n}\}$. The first index of $I_{m,n}$ denotes the mth subvector extracted from $\mathbf{a}_r$ while the second index of $I_{m,n}$ denotes the power of x_r in the generator (i.e., x_r^n). As an example, in the cases where $n \in \{1,2\}$, the matrix $\mathbf{A}$ in (10) has ULA structured submatrices with properties:

$$\begin{array}{llllll} I_{1,1}=3 & \sigma_{1,1}(1)=1 & \sigma_{1,1}(2)=2 & \sigma_{1,1}(3)=3 & & \\ I_{2,1}=3 & \sigma_{2,1}(1)=4 & \sigma_{2,1}(2)=5 & \sigma_{2,1}(3)=6 & & \\ I_{1,2}=5 & \sigma_{1,2}(1)=1 & \sigma_{1,2}(2)=3 & \sigma_{1,2}(3)=4 & \sigma_{1,2}(4)=6 & \\ & \sigma_{1,2}(5)=7 & & & & \end{array}$$

Let $\Omega \subseteq \{1,\ldots,I\}$ denote the set of powers of x in the generators. Assume that there are M_n submatrices of $\mathbf{A}$ associated with the generator x^n, denoted by $\mathbf{A}^{(1,n)} \in \mathbb{C}^{I_{1,n}\times R}, \ldots, \mathbf{A}^{(M_n,n)} \in \mathbb{C}^{I_{M_n,n}\times R}$, with property $a_{p+1,r}^{(m,n)} = a_{p,r}^{(m,n)} x_r^n$. In other words, the submatrices $\{\mathbf{A}^{(M_n,n)}\}$ are (column-wise scaled) Vandermonde matrices with generators $\{x_r^n\}$. By way of illustration, consider $\mathbf{A}$ in (10) and the subset $\Omega = \{1,2\}$. Then we get $M_1 = 2$ with $\mathbf{A}^{(1,1)}$ and $\mathbf{A}^{(2,1)}$ given by (15) and $M_2 = 1$ with $\mathbf{A}^{(1,2)}$ given by (16).

For each pair (m,n) we extract the corresponding submatrix $\mathbf{X}^{(m,n)} \in \mathbb{C}^{I_{m,n}\times K}$ from $\mathbf{X}$ in (8) as follows

$$\mathbf{X}^{(m,n)} = \mathbf{X}([\sigma_{m,n}(1),\ldots,\sigma_{m,n}(I_{m,n})],:) = \mathbf{A}^{(m,n)}\mathbf{S}^T, \quad (18)$$

where $\mathbf{A}^{(m,n)} = \mathbf{A}([\sigma_{m,n}(1),\ldots,\sigma_{m,n}(I_{m,n})],:)$, $m \in \{1,\ldots,M_n\}$ and $n \in \Omega$.

2) From matrices to tensors: The mapping $S_2(\mathbf{X}^{(m,n)})$ produces $\mathcal{Y}^{(m,n)} \in \mathbb{C}^{(I_{m,n}-1)\times 2\times K}$ with matrix representation

$$\mathbf{Y}^{(m,n)} = S_2(\mathbf{X}^{(m,n)}) = \left(\underline{\mathbf{A}}^{(m,n)} \odot \mathbf{B}^{(n)}\right)\mathbf{S}^T, \quad (19)$$

where the rth column of $\underline{\mathbf{A}}^{(m,n)} \in \mathbb{C}^{(I_{m,n}-1)\times R}$ is given by

$$\underline{\mathbf{a}}_r^{(m,n)} = \mathbf{a}_r^{(m,n)}(1 : I_{m,n}-1, :), \quad (20)$$

and

$$\mathbf{B}^{(n)} = \begin{bmatrix}1 & \cdots & 1\\ x_1^n & \cdots & x_R^n\end{bmatrix} \in \mathbb{C}^{2\times R}. \quad (21)$$

Define $I_n = \sum_{m=1}^{M_n} I_{m,n} - M_n$. Stacking $\mathbf{Y}^{(1,n)}, \ldots, \mathbf{Y}^{(M_n,n)}$, yields $\mathcal{Y}^{(n)} \in \mathbb{C}^{I_n\times 2\times K}$ for which the matrix representation $\mathbf{Y}^{(n)} \in \mathbb{C}^{I_n 2\times K}$ admits the decomposition

$$\mathbf{Y}^{(n)} = \left[\mathbf{Y}^{(1,n)T}\ \ldots\ \mathbf{Y}^{(M_n,n)T}\right]^T = \left(\mathbf{A}^{(n)} \odot \mathbf{B}^{(n)}\right)\mathbf{S}^T, \quad (22)$$

where

$$\mathbf{A}^{(n)} = \left[\underline{\mathbf{A}}^{(1,n)T}\ \ldots\ \underline{\mathbf{A}}^{(M_n,n)T}\right]^T \in \mathbb{C}^{I_n\times R}. \quad (23)$$

3) From NLA to CMTF: Let us index the elements in Ω by $\omega(1),\ldots,\omega(\text{card}\,(\Omega))$. For every element $n \in \Omega$ it is assumed that $I_{\omega(n)} \geq 2$. This condition is necessary for the definition of $\mathbf{G}^{(\Omega)}$ given by (25). We now obtain the coupled CPD of $\{\mathcal{Y}^{(n)}\}$ with matrix representation $\mathbf{Y} \in \mathbb{C}^{(\sum_{n\in\Omega} I_n)2\times K}$ given by

$$\mathbf{Y}_{\text{tot}} = \begin{bmatrix}\mathbf{Y}^{(\omega(1))}\\ \vdots\\ \mathbf{Y}^{(\omega(\text{card}(\Omega)))}\end{bmatrix} = \begin{bmatrix}\mathbf{A}^{(\omega(1))} \odot \mathbf{B}^{(\omega(1))}\\ \vdots\\ \mathbf{A}^{(\omega(\text{card}(\Omega)))} \odot \mathbf{B}^{(\omega(\text{card}(\Omega)))}\end{bmatrix}\mathbf{S}^T. \quad (24)$$

Proposition III.1 makes use of the following matrix

$$\mathbf{G}^{(\Omega)} = \begin{bmatrix}C_2\left(\mathbf{A}^{(\omega(1))}\right) \odot C_2\left(\mathbf{B}^{(\omega(1))}\right)\\ \vdots\\ C_2\left(\mathbf{A}^{(\omega(\text{card}(\Omega)))}\right) \odot C_2\left(\mathbf{B}^{(\omega(\text{card}(\Omega)))}\right)\end{bmatrix}. \quad (25)$$

Proposition III.1. *Consider the NLA factorization of $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}^T \in \mathbb{C}^{I\times K}$ in (8) where $\boldsymbol{A}$ is a NLA structured matrix with columns $\boldsymbol{a}_r = [1\ x_r^{n_2}\ \ldots\ x_r^{n_I}]^T$ and $d_x \leq \lambda/2 = \frac{\pi c}{\omega_c}$. If*

$$\begin{cases}\boldsymbol{S} \text{ in (8) has full column rank,} & (26a)\\ \boldsymbol{G}^{(\Omega)} \text{ in (25) has full column rank,} & (26b)\\ GCD(n_i, n_j) = 1 \text{ for some } i \neq j, & (26c)\end{cases}$$

then the NLA constrained rank of $\boldsymbol{X}$ is R and the NLA decomposition of $\boldsymbol{X}$ is unique.

Proof: Since $\mathbf{G}^{(\Omega)}$ and $\mathbf{S}$ have full column rank, we know from Theorem II.2 that the coupled rank of $\{\mathcal{Y}^{(n)}\}$ is R and the CMTF of $\{\mathbf{X}, \mathcal{Y}^{(n)}\}$ is unique. This in turn implies that $\mathbf{A}$ is unique up to a column permutation and scaling. In other words, the CMTF of $\{\mathbf{X}, \mathcal{Y}^{(n)}\}$ has split the problem into a set of single source DOA estimation problems. It is now easy to show that we can recover θ_r from $\mathbf{z}_r = \alpha_r \mathbf{a}_r$, where $\alpha_r \in \mathbb{C}$ is an arbitrary scaling factor. We add the details for completeness.

For simplicity, consider first the special case where $n_j - n_i = 1$ for some $i < j$. This implies that $\begin{bmatrix}z_{ir}\\ z_{jr}\end{bmatrix} = \alpha_r \begin{bmatrix}x_r^{n_i}\\ x_r^{n_j}\end{bmatrix} \Rightarrow$

$z_{jr}/z_{ir} = x_r^{n_j-n_i} = x_r$ for some $i < j$. Since $d_x \leq \frac{\pi c}{\omega_c}$, $n_j - n_i = 1$ and $0 \leq \theta_r < \pi$, we can uniquely recover θ_r from $x_r = e^{-i\omega_c d_x(n_j-n_i)\cos(\theta_r)/c} = e^{-i\omega_c d_x \cos(\theta_r)/c}$.

Consider now the general case in which $n_j - n_i \geq 1$, $\forall i < j$ and $\text{GCD}(n_i, n_j) = 1$ for some $i \neq j$. We have

$$\begin{bmatrix} z_{1r} \\ z_{ir} \end{bmatrix} = \alpha_r \begin{bmatrix} 1 \\ x_r^{n_i} \end{bmatrix} \Rightarrow b_r = z_{ir}/z_{1r} = x_r^{n_i}, \quad (27)$$

$$\begin{bmatrix} z_{1r} \\ z_{jr} \end{bmatrix} = \alpha_r \begin{bmatrix} 1 \\ x_r^{n_j} \end{bmatrix} \Rightarrow c_r = z_{jr}/z_{1r} = x_r^{n_j}. \quad (28)$$

The solutions to (27) are $x_k^{(b_r)} = |b_r|^{1/n_i} e^{\frac{i(2\pi k + \arg(b_r))}{n_i}}$, $k \in \{0, 1, \ldots, n_i - 1\}$. Similarly, the solutions to (28) are $x_m^{(c_r)} = |c_r|^{1/n_j} e^{\frac{i(2\pi m + \arg(c_r))}{n_j}}$, $m \in \{0, 1, \ldots, n_j - 1\}$. Hence, any solution that satisfies both (27) and (28) must also satisfy the relation

$$(2\pi k + \arg(b_r))/n_i = (2\pi m + \arg(c_r))/n_j. \quad (29)$$

Note that $b_r^{n_j} = x_r^{n_i n_j} = |b_r|^{n_j} e^{i\arg(b_r)n_j}$ and $c_r^{n_i} = x_r^{n_i n_j} = |c_r|^{n_i} e^{i\arg(c_r)n_i}$, implying that $\arg(b_r)n_j = \arg(c_r)n_i$. Together with relation (29) this implies that we have $k\frac{n_j}{n_i} = m$. Since $\text{GCD}(n_i, n_j) = 1$ and $k < n_i$ we conclude that the only common solution to (27) and (28) is $k = m = 0$, corresponding to $\arg(b_r)/n_i = \arg(c_r)/n_j = -\omega_c d_x \cos(\theta_r)/c$. Since $d_x \leq \frac{\pi c}{\omega_c}$ and $0 \leq \theta_r < \pi$ we can uniquely recover θ_r from $\arg(b_r)$ and $\arg(c_r)$.

We conclude that if conditions (26a)–(26c) are satisfied, then the NLA constrained rank of $\mathbf{X}$ is R and the NLA decomposition of $\mathbf{X}$ is unique. ∎

Note that condition (26c) is necessary while conditions (26a) and (26b) are only sufficient. Even though we have only discussed the case of unit norm generators, Proposition III.1 is also valid for cases where the generators are not necessarily located on the unit circle.

Since each tensor $\mathcal{Y}^{(n)} \in \mathbb{C}^{I_n \times 2 \times K}$, $n \in \Omega$, only consists of two matrix slices, the coupled CPD of $\{\mathcal{Y}^{(n)}\}$ can be interpreted as a coupled Matrix-Pencil (MP) problem. More precisely, let $\mathbf{P}^{(n)} \in \mathbb{C}^{2I_n \times 2I_n}$ denote a row-permutation matrix such that $\mathbf{Z}^{(n)} := \mathbf{P}^{(n)}\mathbf{Y}^{(n)} = (\mathbf{B}^{(n)} \odot \mathbf{A}^{(n)})\mathbf{S}^T$. Consider the partitioning $\mathbf{Z}^{(n)} = \begin{bmatrix} \mathbf{Z}^{(1,n)} \\ \mathbf{Z}^{(2,n)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{(n)}\mathbf{S}^T \\ \mathbf{A}^{(n)}D_2(\mathbf{B}^{(n)})\mathbf{S}^T \end{bmatrix}$ with $\mathbf{Z}^{(1,n)}, \mathbf{Z}^{(2,n)} \in \mathbb{C}^{I_n \times K}$. Define $\mathbf{\Gamma} = [\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_R] := (\mathbf{S}^T)^\dagger$ and $\mathbf{\Upsilon} = D_1([\lambda_1, \ldots, \lambda_R]) := (D_1([x_1, \ldots, x_R]))^{-1}$. Then we obtain a coupled MP problem

$$\mathbf{Z}^{(1,n)}\mathbf{\Gamma} = \mathbf{Z}^{(2,n)}\mathbf{\Gamma}\mathbf{\Upsilon}^n \Leftrightarrow (\mathbf{Z}^{(1,n)} - \lambda_r^n \mathbf{Z}^{(2,n)})\boldsymbol{\gamma}_r = \mathbf{0}_{I_n}, \quad (30)$$

where $r \in \{1, \ldots, R\}$ and $n \in \Omega$. In other words, the factorization problem (8) amounts to finding the common generalized eigenvectors of a set of MPs $\{(\mathbf{Z}^{(1,n)}, \mathbf{Z}^{(2,n)})\}_{n\in\Omega}$ subject to a power constraint on the associated generalized eigenvalues ($\{\lambda_r^n\}_{n\in\Omega}$). The power constraint can make the generalized eigenvalues (λ_r) unique despite $\min(\Omega) > 1$. Likewise, the coupling between the MPs can make the common generalized eigenvectors ($\boldsymbol{\gamma}_r$) in (30) unique, even in cases where for certain n, $\mathbf{Z}^{(1,n)}$ and $\mathbf{Z}^{(2,n)}$ have overlapping kernels ($\ker\left(\mathbf{Z}^{(1,n)}\right) \cap \ker\left(\mathbf{Z}^{(2,n)}\right) \neq \emptyset$) so that the eigenvector in (30) is not unique. Note also that the ordinary linear MP case ($\Omega = \{1\}$) corresponds to classical ESPRIT [15].

Let us illustrate that the aforementioned power constraint (λ_r^n) leads to improved DOA identifiability. Consider the NLA for which $\mathbf{A}$ has columns of the form:

$$\mathbf{a}_r = [1 \; x_r^3 \; x_r^6 \; x_r^8]^T. \quad (31)$$

Consider also the sub-ULAs $[1 \; x_r^3 \; x_r^6]^T$ and $[x_r^6 \; x_r^8]^T$ of (31) with inter-element spacings $n = 3$ and $n = 2$, respectively. Since $\Omega = \{2, 3\}$ and $\text{GCD}(2, 3) = 1$, Proposition III.1 guarantees the uniqueness of a generic NLA decomposition in which $\mathbf{A}$ has columns of the form (31) if $R \leq 2$. On the other hand, if only the sub-ULA $[1 \; x_r^3 \; x_r^6]^T$ is used and the element x_r^8 in (31) is ignored, then unique DOA recovery is not possible since now $\Omega = \{3\}$ and $\min(\Omega) > 1$. Indeed, the associated MP involves only x_r^3 and not x_r itself. Similarly, if only the sub-ULA $[x_r^6 \; x_r^8]^T$ is used, then unique DOA recovery is not possible since $\Omega = \{2\}$ and $\min(\Omega) > 1$.

To illustrate that coupling can reduce the dimension of the eigenspace, consider the NLA for which $\mathbf{A}$ has columns of the form:

$$\mathbf{a}_r = [1 \; x_r \; x_r^2 \; x_r^{11} \; x_r^{15} \; x_r^{18} \; x_r^{21} \; x_r^{23}]^T. \quad (32)$$

By simultaneously exploiting the multiple shift-invariance structures in (32) that are associated with the generators x_r, x_r^2, x_r^3 and x_r^{10}, Proposition III.1 guarantees the uniqueness of a generic NLA decomposition with columns of the form (32) if $R \leq 3$. Note that if we only exploit one of the sub-ULAs in (32), say $[1 \; x_r \; x_r^2]^T$, then the more restrictive NLA uniqueness condition $R \leq 2$ is obtained.

This does not not mean that identifiability properties for all possible NLAs can be derived via the CMTF framework. For instance, nonredundant arrays [26] such as $\mathbf{a}_r = [1 \; x_r \; x_r^4 \; x_r^9 \; x_r^{11}]^T$ do not contain any shift-invariance structure and hence the CMTF approach does not apply if $R > 1$.

C. Multiresolution interpretation of CMTF for NLA

Consider the NLA response matrix $\mathbf{A}$ with columns of the form (32). We note that the partial sub-ULAs within $\mathbf{A}$ have generators with small and large powers, i.e., $\mathbf{a}_r(1)x_r = \mathbf{a}_r(2)$, $\mathbf{a}_r([1,7])x_r^2 = \mathbf{a}_r([3,8])$, $\mathbf{a}_r([1,7])x_r^2 = \mathbf{a}_r([3,8])$, $\mathbf{a}_r([5,6])x_r^3 = \mathbf{a}_r([6,7])$ and $\mathbf{a}_r([2,4])x_r^{10} = \mathbf{a}_r([4,7])$. We say that the array is equipped with multiple baselines. This property has been used to develop arrays with multiresolution properties (e.g. [6], [29], [8], [1]). An example taken from [1] is the case of three separated ULAs where the response matrix $\mathbf{A}$ has columns of the form $\mathbf{a}_r = [1 \; x_r \; x_r^2 \; x_r^{14} \; x_r^{15} \; x_r^{21} \; x_r^{22} \; x_r^{23}]^T$. We note that $\mathbf{a}_r$ has generators x_r, x_r^7, x_r^8, x_r^{13}, x_r^{14} and x_r^{20}. Roughly speaking, a small baseline can lead to coarse but unique DOA estimates while a large baseline can lead to accurate but ambiguous DOA estimates. The idea behind array processing involving multiple baselines is to combine the advantages of small and large baselines.

A related example is nested arrays or co-prime arrays composed of superimposed or concatenated ULAs with

different inter-element spacings [12], [25], [11], [24], [14]. An example could be two superimposed ULAs where the response matrix $\mathbf{A}$ has columns of the form $\mathbf{a}_r = [1\ x_r^5\ x_r^7\ x_r^{10}\ x_r^{14}\ x_r^{15}\ x_r^{20}\ x_r^{21}\ x^{25}\ x_r^{28}\ x_r^{30}]^T$. From $\mathbf{A}$ we can extract $\mathbf{A}^{(1)}$, $\mathbf{A}^{(5)}$ and $\mathbf{A}^{(7)}$ with columns of the form $\mathbf{a}_r^{(1)} = [x_r^{14}\ x_r^{15}\ x_r^{20}\ x_r^{21}]^T$, $\mathbf{a}_r^{(5)} = [1\ x_r^5\ x_r^{10}\ x_r^{15}\ x_r^{20}\ x_r^{25}\ x_r^{30}]^T$ and $\mathbf{a}_r^{(7)} = [1\ x_r^7\ x_r^{14}\ x_r^{21}\ x_r^{28}]^T$, respectively. It is now clear that the CMTF approach for NLA can also be used for nested arrays or co-prime arrays without making orthogonality assumptions on $\mathbf{S}$, or by making use of any kind of data imputation/interpolation.

The CMTF approach also supports the use of hybrid arrays composed of sparse arrays with periodic subarrays [28].

Actually, in the case of sparse array processing, the CMTF approach for NLA can be interpreted as a multiresolution variant of ESPRIT. A distinct advantage of the CMTF approach compared to the existing methods (e.g. [6], [29], [8], [1], [24]) is that it simultaneously exploits the structure of the involved subarrays. As such, none of the involved subarrays is required to yield unique DOAs and also a more relaxed bound on R may be obtained.

As an example, consider the NLA response matrix $\mathbf{A} = [\mathbf{A}^{(2)T}\ \mathbf{A}^{(11)T}\ \mathbf{A}^{(14)T}]^T \in \mathbb{C}^{15\times R}$ with columns $\mathbf{a}_r^{(2)} = [1\ x_r^2\ x_r^4\ x_r^6\ x_r^8]^T$, $\mathbf{a}_r^{(11)} = [x_r^{11}\ x_r^{21}\ x_r^{31}\ x_r^{41}\ x_r^{51}]^T$ and $\mathbf{a}_r^{(14)} = [x_r^{14}\ x_r^{28}\ x_r^{46}\ x_r^{62}\ x_r^{76}]^T$. Methods that process the sub-ULAs individually can identify at most $R \leq 4$ DOAs. However, Proposition III.1 implies that this NLA factorization is generically unique if $R \leq 6$.

IV. New Simultaneous matrix Diagonalization (SD) method for MI-ESPRIT

In this section we present an SD method for MI-ESPRIT that will be used to compute the NLA factorization of $\mathbf{X}$.

a) Step 0: Dimensionality reduction: Assume that the full column rank conditions (26a) and (26b) in Proposition III.1 are satisfied, then $\mathbf{A}$ in (8) has full column rank. Indeed, if $\mathbf{A}$ does not have full column rank, then there exists a vector $\mathbf{y} \in \ker(\mathbf{A})$ and a vector $\mathbf{x} \in \mathbb{C}^K$ such that $\mathbf{X} = \mathbf{A}(\mathbf{S} + \mathbf{x}\mathbf{y}^T)^T$ has rank $R-1$. This in turn contradicts Proposition III.1. Let $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ denote the compact SVD of $\mathbf{X}$, where $\mathbf{U} \in \mathbb{C}^{I\times R}$ and $\mathbf{V} \in \mathbb{C}^{K\times R}$ are columnwise orthonormal and $\mathbf{\Sigma} \in \mathbb{C}^{R\times R}$ is diagonal. We know that $\text{range}(\mathbf{X}) = \text{range}(\mathbf{U})$ and that there exists a nonsingular matrix $\mathbf{F} \in \mathbb{C}^{R\times R}$ such that

$$\mathbf{U}\mathbf{\Sigma} = \mathbf{A}\mathbf{F}^T \text{ and } \mathbf{V}^*\mathbf{F}^{-1} = \mathbf{S}. \tag{33}$$

b) Step 1: From NLA factorization to coupled CPD: For every $m \in \{1,\ldots,M_n\}$ and for every $n \in \Omega$ we extract from $\mathbf{U}\mathbf{\Sigma}$ in (33) the compressed variant of (18):

$$\mathbf{U}^{(m,n)} = \mathbf{U}([\sigma_{m,n}(1),\ldots,\sigma_{m,n}(I_{m,n})],:)\mathbf{\Sigma} = \mathbf{A}^{(m,n)}\mathbf{F}^T. \tag{34}$$

From the matrices $\{\mathbf{U}^{(m,n)}\}$ in (34) we obtain the following compressed variant of (22):

$$\mathbf{Y}_{\text{red}}^{(n)} = \begin{bmatrix} S_2(\mathbf{U}^{(1,n)}) \\ \vdots \\ S_2(\mathbf{U}^{(M_n,n)}) \end{bmatrix} = \left(\mathbf{A}^{(n)} \odot \mathbf{B}^{(n)}\right)\mathbf{F}^T, \quad n \in \Omega, \tag{35}$$

where 'red' stands for reduced. Stacking for all $n \in \Omega$ yields the following compressed variant of (24):

$$\mathbf{Y}_{\text{tot,red}} = \begin{bmatrix} \mathbf{Y}_{\text{red}}^{(\omega(1))} \\ \vdots \\ \mathbf{Y}_{\text{red}}^{(\omega(\text{card}(\Omega)))} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{(\omega(1))} \odot \mathbf{B}^{(\omega(1))} \\ \vdots \\ \mathbf{A}^{(\omega(\text{card}(\Omega)))} \odot \mathbf{B}^{(\omega(\text{card}(\Omega)))} \end{bmatrix} \mathbf{F}^T. \tag{36}$$

Clearly the low rank structured matrix decomposition (36) corresponds to a matrix representation of the coupled CPD of the tensors $\mathcal{Y}_{\text{red}}^{(n)} \in \mathbb{C}^{I_n\times 2\times K}$, $n \in \Omega$.

c) Step 2: From coupled CPD to SD: Assume that the full column rank conditions (26a) and (26b) in Proposition III.1 are satisfied, then there exist R symmetric matrices $\{\mathbf{M}^{(r)}\}$ built from the matrix (36), admitting the factorizations (see [21], [18] for the precise construction):

$$\mathbf{M}^{(r)} = \mathbf{G}\mathbf{\Lambda}^{(r)}\mathbf{G}^T, \quad r \in \{1,\ldots,R\}, \tag{37}$$

where $\mathbf{G} = \mathbf{F}^{-1}$ and where $\mathbf{\Lambda}^{(r)} \in \mathbb{C}^{R\times R}$ are diagonal matrices. It is explained in [21], [18] that the complexity of the construction of $\{\mathbf{M}^{(r)}\}$ is only of order $4(\sum_{n\in\Omega} I_n)R^2$. In the exact case, the SD problem (37) can be solved by means of an eigenvalue decomposition (e.g. [9]), with the columns of $\mathbf{G}^{-T}$ as generalized eigenvectors.

d) Step 3: From SD to single-source DOA retrieval: After $\mathbf{G} = \mathbf{F}^{-1}$ has been found from (37), we obtain $\mathbf{A}$ and $\mathbf{S}$ via (33). The generator x_r can now be obtained from $\mathbf{a}_r$. More precisely, define

$$\mathbf{e}_r^{(n)} = \begin{bmatrix} \underline{\mathbf{a}}_r^{(1,n)} \\ \vdots \\ \underline{\mathbf{a}}_r^{(M_n,n)} \end{bmatrix} \quad \text{and} \quad \mathbf{f}_r^{(n)} = \begin{bmatrix} \overline{\mathbf{a}}_r^{(1,n)} \\ \vdots \\ \overline{\mathbf{a}}_r^{(M_n,n)} \end{bmatrix}, \quad n \in \Omega. \tag{38}$$

If $x_r \in \mathbb{C}$, then due to conditions (26a)–(26c) in Proposition III.1 we know that the generator can be estimated as the minimizer of the univariate polynomial $f(x_r) = \sum_{n\in\Omega} \left\|\mathbf{e}_r^{(n)}x_r^n - \mathbf{f}_r^{(n)}\right\|_F^2$. However, in DOA estimation x_r is located on the unit circle. In that case x_r must satisfy the conditions

$$\mathbf{e}_r^{(n)}x_r^n = \mathbf{f}_r^{(n)}, \ n \in \Omega \text{ subject to } x_r = e^{i\cdot\phi_r}, \ 0 \leq \phi_r < \pi.$$

Proposition IV.1 below states that ϕ_r can be found via the rooting of a real polynomial.

Proposition IV.1. *Consider* $f(\phi_r) = \sum_{n\in\Omega} \left\|\boldsymbol{e}_r^{(n)}e^{i\cdot\phi_r\cdot n} - \boldsymbol{f}_r^{(n)}\right\|_F^2$, *where* $\boldsymbol{e}_r^{(n)}$ *and* $\boldsymbol{f}_r^{(n)}$ *are given by* (38) *and* $\Omega \subset \{1,\ldots,I\}$. *The largest element in* Ω *is denoted by* $N_{\max}$. *Consider also the*

real polynomial of degree $2N_{\max}$*:*

$$g(t) = \sum_{n\in\Omega}\Bigg(Re\left\{\beta_r^{(n)}\right\}\sum_{p=0}^{2n}\sum_{k=0}^{N_{max}-n}\binom{2n}{p}\binom{N_{max}-n}{k}\Big(-pt^{(2n-p+1+2k)} + (2n-p)t^{(2n-p-1+2k)}\Big)\cdot\cos(\frac{1}{2}(2n-p)\pi) + Im\{\beta_r^{(n)}\}\sum_{q=0}^{2n}\sum_{m=0}^{N_{max}-n}\binom{2n}{q}\binom{N_{max}-n}{m}\Big(-qt^{(2n-q+1+2m)} + (2n-q)t^{(2n-q-1+2m)}\Big)\cdot\sin(\frac{1}{2}(2n-q)\pi)\Bigg), \quad (39)$$

where $\beta_r^{(n)} = \mathbf{e}_r^{(n)H}\mathbf{f}_r^{(n)}$*. Let* $\hat{t}$ *denote the real root of* $g(t)$ *that minimizes* $f(2\tan^{-1}(\hat{t}))$*, then*

$$2\tan^{-1}(\hat{t}) = \arg\min_{0\le\phi_r<\pi} f(\phi_r).$$

Proof: The proof is technical and is given in the supplementary material of [19]. ∎

We now explain that Proposition IV.1 is guaranteed to find the true unit norm generator $x_r = e^{i\cdot\phi_r}$ in the exact case. Recall that conditions (26a) and (26b) in Proposition III.1 guarantee that the columns $\{\mathbf{a}_r\}$ of $\mathbf{A}$ are unique up to individual scaling factors. This means that there exists a unit norm generator $x_r = e^{i\cdot\phi_r}$ such that $\mathbf{e}_r^{(n)}x_r = \mathbf{f}_r^{(n)}$, $\forall n \in \Omega$. Finally, condition (26c) in Proposition III.1 guarantees the uniqueness of this unit norm generator.

e) Summary of algorithm: The overall procedure is summarized as Algorithm 1. Note that if conditions (26a)–(26c) in Proposition III.1 are satisfied, then Algorithm 1 is guaranteed to find the exact decomposition in the noiseless case.

Algorithm 1 SD method for MI-ESPRIT (Prop. III.1).

Input: $\mathbf{X} = \mathbf{A}\mathbf{S}^T$ of the form (8).
1. Determine subset $\Omega \subseteq \{1,\dots,I\}$.
2. Compute SVD $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ (and find R from $\mathbf{\Sigma}$).
3. Extract $\mathbf{U}^{(m,n)}$ in (34) from $\mathbf{U}\mathbf{\Sigma}$, $\forall m \in \{1,\dots,M_n\}$, $\forall n \in \Omega$.
4. Build $\mathbf{Y}_{\text{red}}^{(n)}$ in (35) for every $n \in \Omega$.
5. Build matrices $\{\mathbf{M}^{(r)}\}$ from $\{\mathbf{Y}_{\text{red}}^{(n)}\}_{n\in\Omega}$ [21], [18].
6. Solve SD problem $\mathbf{M}^{(r)} = \mathbf{G}\mathbf{\Lambda}^{(r)}\mathbf{G}^T$, $r \in \{1,\dots,R\}$.
7. Compute $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{G}^T$ and $\mathbf{S} = \mathbf{V}^*\mathbf{G}$.
8. Obtain $\{x_r\}$ from $\mathbf{A}$ (e.g. via Proposition IV.1).

Output: $\{x_r\}$, $\mathbf{A}$ and $\mathbf{S}$.

V. Extension to the multidimensional case

A simple extension of the 1D NLA is an array built from superimposed NLAs. Here we adapt some of the results in [18] to the NLA case. Subsection V-A provides an illustrative example. Next, in Subsection V-B we formalize the extension to the multidimensional case.

A. Illustration

Consider an L-shaped array composed of $I^{(1)}$ sensors along the "x_1-axis" and $I^{(2)}$ sensors along the "x_2-axis" such that (8) extends to

$$\mathbf{X}_{(1)} = \mathbf{A}_{(1)}\mathbf{S}^T \in \mathbb{C}^{I^{(1)}\times K}, \quad \mathbf{X}_{(2)} = \mathbf{A}_{(2)}\mathbf{S}^T \in \mathbb{C}^{I^{(2)}\times K}, \quad (40)$$

where $\mathbf{X}_{(1)}$ denotes the observation matrix associated with the sensors along the "x_1-axis" in which $(\mathbf{A}_{(1)})_r = \begin{bmatrix}1 & x_{r,1}^{n_{2,1}} & \dots & x_{r,1}^{n_{I^{(1)},1}}\end{bmatrix}^T \in \mathbb{C}^{I^{(1)}}$, where $x_{r,1} = e^{-i\omega_c d_1 \sin(\phi_r)\cos(\theta_r)/c}$, d_1 is the unit spacing along the "x_1-axis" and the new variable ϕ_r denotes the elevation angle associated with the rth source. Similarly, $\mathbf{X}_{(2)}$ denotes the observation matrix associated with the sensors along the "x_2-axis" in which $(\mathbf{A}_{(2)})_r = \begin{bmatrix}1 & x_{r,2}^{n_{2,2}} & \dots & x_{r,2}^{n_{I^{(2)},2}}\end{bmatrix}^T \in \mathbb{C}^{I^{(2)}}$, where $x_{r,2} = e^{-i\omega_c d_2 \sin(\phi_r)\sin(\theta_r)/c}$ and d_2 is the unit spacing along the "x_2-axis".

As in the 1D case, we define the NLA constrained and coupled rank of $\{\mathbf{X}_{(1)}, \mathbf{X}_{(2)}\}$ in (40) as the minimal number of NLA constrained and coupled rank-1 matrices $\{(\mathbf{A}_{(1)})_r\mathbf{s}_r^T, (\mathbf{A}_{(2)})_r\mathbf{s}_r^T\}$ that yield $\{\mathbf{X}_{(1)}, \mathbf{X}_{(2)}\}$ in a linear combination. We also say that the coupled NLA decomposition of $\{\mathbf{X}_{(1)}, \mathbf{X}_{(2)}\}$ is unique if for every triplet $(\widehat{\mathbf{A}}_{(1)}, \widehat{\mathbf{A}}_{(2)}, \widehat{\mathbf{S}})$ satisfying the relation (40) we have $\widehat{\mathbf{A}}_{(1)} = \mathbf{A}_{(1)}\Pi$ and $\widehat{\mathbf{A}}_{(2)} = \mathbf{A}_{(2)}\Pi$, where Π is a permutation matrix.

By way of example, let $\mathbf{X}_{(1)}$ in (40) be of the form (10) while $\mathbf{X}_{(2)}$ in (40) admits the factorization

$$\mathbf{X}_{(2)} = \mathbf{A}_{(2)}\mathbf{S}^T \in \mathbb{C}^{I^{(2)}\times K}, \quad (\mathbf{A}_{(2)})_r = \begin{bmatrix}1 & x_{r,2} & x_{r,2}^4 & x_{r,2}^5 & x_{r,2}^8 & x_{r,2}^{12}\end{bmatrix}^T. \quad (41)$$

From $\mathbf{X}_{(1)}$ we obtain the coupled CPD of $\{\mathcal{Y}_{(1)}^{(1)}, \mathcal{Y}_{(1)}^{(2)}\}$ with matrix representation (17). We will now capitalize on the NLA structure of $\mathbf{A}_{(2)}$. Note that the vector $(\mathbf{A}_{(2)})_r$ in $\mathbf{A}_{(2)}$ contains the generators $x_{r,2}$, $x_{r,2}^3$ and $x_{r,2}^4$:

$$\begin{array}{c} \\ x \\ x \\ x^3 \\ x^3 \\ x^4 \\ x^4 \end{array}\!\!\begin{array}{c} \begin{array}{cccccccccccc} 1 & x & - & - & x^4 & x^5 & - & - & x^8 & - & - & - & x^{12} \end{array} \\ \begin{pmatrix} 1 & x & & & & & & & & & & & \\ & & & & x^4 & y^5 & & & & & & & \\ & x & & & x^4 & & & & & & & & \\ & & & & & x^5 & & & x^8 & & & & \\ 1 & & & & x^4 & & & & x^8 & & & & x^{12} \\ & x & & & & x^5 & & & & & & & \end{pmatrix} \end{array}. \quad (42)$$

We consider the following submatrices of $\mathbf{A}_{(2)}$:

$$\mathbf{X}_{(2)}^{(1,1)} = \mathbf{X}_{(2)}([1,2],:) = \mathbf{A}_{(2)}^{(1,1)}\mathbf{S}^T \in \mathbb{C}^{2\times K}, \qquad I_{1,1}^{(2)} = 2, \quad (43)$$
$$\mathbf{X}_{(2)}^{(2,1)} = \mathbf{X}_{(2)}([3,4],:) = \mathbf{A}_{(2)}^{(2,1)}\mathbf{S}^T \in \mathbb{C}^{2\times K}, \qquad I_{2,1}^{(2)} = 2, \quad (44)$$
$$\mathbf{X}_{(2)}^{(1,3)} = \mathbf{X}_{(2)}([2,3],:) = \mathbf{A}_{(2)}^{(1,3)}\mathbf{S}^T \in \mathbb{C}^{2\times K}, \qquad I_{1,3}^{(2)} = 2, \quad (45)$$
$$\mathbf{X}_{(2)}^{(2,3)} = \mathbf{X}_{(2)}([4,5],:) = \mathbf{A}_{(2)}^{(2,3)}\mathbf{S}^T \in \mathbb{C}^{2\times K}, \qquad I_{2,3}^{(2)} = 2, \quad (46)$$
$$\mathbf{X}_{(2)}^{(1,4)} = \mathbf{X}_{(2)}([1,3,5,6],:) = \mathbf{A}_{(2)}^{(1,4)}\mathbf{S}^T \in \mathbb{C}^{4\times K}, \qquad I_{1,4}^{(2)} = 4, \quad (47)$$
$$\mathbf{X}_{(2)}^{(2,4)} = \mathbf{X}_{(2)}([2,4],:) = \mathbf{A}_{(2)}^{(2,4)}\mathbf{S}^T \in \mathbb{C}^{2\times K}, \qquad I_{2,4}^{(2)} = 2, \quad (48)$$

where

$$\mathbf{A}_{(2)}^{(1,1)} = \begin{bmatrix}1 & \cdots & 1\\ x_{1,2} & \cdots & x_{R,2}\end{bmatrix}, \qquad \mathbf{A}_{(2)}^{(2,1)} = \begin{bmatrix}x_{1,2}^4 & \cdots & x_{R,2}^4\\ x_{1,2}^5 & \cdots & x_{R,2}^5\end{bmatrix},$$
$$\mathbf{A}_{(2)}^{(1,3)} = \begin{bmatrix}x_{1,2} & \cdots & x_{R,2}\\ x_{1,2}^4 & \cdots & x_{R,2}^4\end{bmatrix}, \qquad \mathbf{A}_{(2)}^{(2,3)} = \begin{bmatrix}x_{1,2}^5 & \cdots & x_{R,2}^5\\ x_{1,2}^8 & \cdots & x_{R,2}^8\end{bmatrix},$$
$$\mathbf{A}_{(2)}^{(1,4)} = \begin{bmatrix}1 & \cdots & 1\\ x_{1,2}^4 & \cdots & x_{R,2}^4\\ x_{1,2}^8 & \cdots & x_{R,2}^8\\ x_{1,2}^{12} & \cdots & x_{R,2}^{12}\end{bmatrix}, \qquad \mathbf{A}_{(2)}^{(2,4)} = \begin{bmatrix}x_{1,2} & \cdots & x_{R,2}\\ x_{1,2}^5 & \cdots & x_{R,2}^5\end{bmatrix}.$$

From (43)–(48) we obtain the tensors $\mathcal{Y}_{(2)}^{(1)} \in \mathbb{C}^{2\times 2\times K}$, $\mathcal{Y}_{(2)}^{(3)} \in \mathbb{C}^{2\times 2\times K}$ and $\mathcal{Y}_{(2)}^{(4)} \in \mathbb{C}^{4\times 2\times K}$ for which the associated

matrix representations $\mathbf{Y}^{(1)}_{(2)} \in \mathbb{C}^{4\times K}$, $\mathbf{Y}^{(3)}_{(2)} \in \mathbb{C}^{4\times K}$ and $\mathbf{Y}^{(4)}_{(2)} \in \mathbb{C}^{8\times K}$ admit the following CPDs:

$$\begin{aligned}
\mathbf{Y}^{(1)}_{(2)} &= \left[S_2(\mathbf{X}^{(1,1)}_{(2)})^T, S_2(\mathbf{X}^{(2,1)}_{(2)})^T\right]^T = (\mathbf{A}^{(1)}_{(2)} \odot \mathbf{B}^{(1)}_{(2)})\mathbf{S}^T,\\
\mathbf{Y}^{(3)}_{(2)} &= \left[S_2(\mathbf{X}^{(1,3)}_{(2)})^T, S_2(\mathbf{X}^{(2,3)}_{(2)})^T\right]^T = \left(\mathbf{A}^{(3)}_{(2)} \odot \mathbf{B}^{(3)}_{(2)}\right)\mathbf{S}^T,\\
\mathbf{Y}^{(4)}_{(2)} &= \left[S_2(\mathbf{X}^{(1,4)}_{(2)})^T, S_2(\mathbf{X}^{(2,4)}_{(2)})^T\right]^T = \left(\mathbf{A}^{(4)}_{(2)} \odot \mathbf{B}^{(4)}_{(2)}\right)\mathbf{S}^T,
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{A}^{(1)}_{(2)} &= \begin{bmatrix}\underline{\mathbf{A}}^{(1,1)}_{(2)}\\ \underline{\mathbf{A}}^{(2,1)}_{(2)}\end{bmatrix} = \begin{bmatrix}1 & \cdots & 1\\ x^4_{1,2} & \cdots & x^4_{R,2}\end{bmatrix}, & \mathbf{B}^{(1)}_{(2)} &= \begin{bmatrix}1 & \cdots & 1\\ x_{1,2} & \cdots & x_{R,2}\end{bmatrix},\\
\mathbf{A}^{(3)}_{(2)} &= \begin{bmatrix}\underline{\mathbf{A}}^{(1,3)}_{(2)}\\ \underline{\mathbf{A}}^{(2,3)}_{(2)}\end{bmatrix} = \begin{bmatrix}1 & \cdots & 1\\ x^5_{1,2} & \cdots & x^5_{R,2}\end{bmatrix}, & \mathbf{B}^{(3)}_{(2)} &= \begin{bmatrix}1 & \cdots & 1\\ x^3_{1,2} & \cdots & x^3_{R,2}\end{bmatrix},\\
\mathbf{A}^{(4)}_{(2)} &= \begin{bmatrix}\underline{\mathbf{A}}^{(1,4)}_{(2)}\\ \underline{\mathbf{A}}^{(2,4)}_{(2)}\end{bmatrix} = \begin{bmatrix}1 & \cdots & 1\\ x^4_{1,2} & \cdots & x^4_{R,2}\\ x^8_{1,2} & \cdots & x^8_{R,2}\\ x_{1,2} & \cdots & x_{R,2}\end{bmatrix}, & \mathbf{B}^{(4)}_{(2)} &= \begin{bmatrix}1 & \cdots & 1\\ x^4_{1,2} & \cdots & x^4_{R,2}\end{bmatrix}.
\end{aligned}$$

We now obtain the coupled CPD of $\{\mathcal{Y}^{(1)}_{(2)}, \mathcal{Y}^{(3)}_{(2)}, \mathcal{Y}^{(4)}_{(2)}\}$ with matrix representation

$$\mathbf{Y}_{(2)} = \begin{bmatrix}\mathbf{Y}^{(1)}_{(2)}\\ \mathbf{Y}^{(2)}_{(2)}\\ \mathbf{Y}^{(3)}_{(2)}\end{bmatrix} = \begin{bmatrix}\mathbf{A}^{(1)}_{(2)} \odot \mathbf{B}^{(1)}_{(2)}\\ \mathbf{A}^{(3)}_{(2)} \odot \mathbf{B}^{(3)}_{(2)}\\ \mathbf{A}^{(4)}_{(2)} \odot \mathbf{B}^{(4)}_{(2)}\end{bmatrix}\mathbf{S}^T. \tag{49}$$

Overall, from (17) and (49) we obtain the coupled CPD of $\{\mathcal{Y}^{(1)}_{(1)}, \mathcal{Y}^{(2)}_{(1)}, \mathcal{Y}^{(1)}_{(2)}, \mathcal{Y}^{(3)}_{(2)}, \mathcal{Y}^{(4)}_{(2)}\}$. By exploiting the coupling between the two NLAs imposed by $\mathbf{S}$, improved uniqueness conditions are obtained. Indeed, from subsection III-A we know that the coupled CPD structure of only $\{\mathcal{Y}^{(1)}_{(1)}, \mathcal{Y}^{(2)}_{(1)}\}$ yields the upper bound $R \leq 5$. Likewise, from Theorem II.2 the coupled CPD structure of $\{\mathcal{Y}^{(1)}_{(2)}, \mathcal{Y}^{(3)}_{(2)}, \mathcal{Y}^{(4)}_{(2)}\}$ yields the upper bound $R \leq 3$. On the other hand, by exploiting the overall structure of $\{\mathcal{Y}^{(1)}_{(1)}, \mathcal{Y}^{(2)}_{(1)}, \mathcal{Y}^{(1)}_{(2)}, \mathcal{Y}^{(3)}_{(2)}, \mathcal{Y}^{(4)}_{(2)}\}$, Theorem II.2 relaxes the bound to $R \leq 6$.

B. CMTF framework for multidimensional NLA

Let us now formalize the reasoning behind the previous example. Consider the decomposition

$$\mathbf{X} = \mathbf{A}\mathbf{S}^T \in \mathbb{C}^{I\times K}, \tag{50}$$

where the matrix $\mathbf{A} \in \mathbb{C}^{I\times R}$ is shift-invariance structured while $\mathbf{S} \in \mathbb{C}^{K\times R}$ is unstructured. More precisely, let $\mathbf{J}_{(p)} \in \mathbb{C}^{I^{(p)}\times R}$ denote a row-selection matrix such that $\mathbf{A}_{(p)} = \mathbf{J}_{(p)}\mathbf{A} \in \mathbb{C}^{I^{(p)}\times R}$ consists of $I^{(p)}$ rows of $\mathbf{A}$. It is assumed that each of the following P decompositions

$$\mathbf{X}_{(p)} = \mathbf{J}_{(p)}\mathbf{X} = \mathbf{A}_{(p)}\mathbf{S}^T \in \mathbb{C}^{I^{(p)}\times K}, \quad p \in \{1,\dots,P\}, \tag{51}$$

corresponds to an NLA factorization of the form (8). This in turn means that $\mathbf{A}_{(p)} \in \mathbb{C}^{I^{(p)}\times R}$ is an NLA matrix with columns $(\mathbf{A}_{(p)})_r = [x_{r,p}^{n_{1,p}}\ x_{r,p}^{n_{2,p}}\ \dots\ x_{r,p}^{n_{I^{(p)},p}}]^T$ in which $x_{r,p} \in \mathbb{C}$ and $n_{i,p} \in \mathbb{Z}$. We can now proceed as in Section III. Briefly, let $\Omega_p \subseteq \{1,\dots,I^{(p)}\}$ denote the set of powers of $x_{r,p}$ in the generators associated with pth NLA matrix $\mathbf{A}_{(p)}$. Relation (18) now becomes

$$\mathbf{X}^{(m,n)}_{(p)} = \mathbf{X}_{(p)}([\sigma_{m,n,p}(1),\dots,\sigma_{m,n,p}(I^{(p)}_{m,n})],:) = \mathbf{A}^{(m,n)}_{(p)}\mathbf{S}^T \in \mathbb{C}^{I^{(p)}_{m,n}\times K},$$

where $\mathbf{A}^{(m,n)}_{(p)} = \mathbf{A}_{(p)}([\sigma_{m,n,p}(1),\dots,\sigma_{m,n,p}(I^{(p)}_{m,n})],:)$, $m \in \{1,\dots,M_{n,p}\}$, $n \in \Omega_p$ and the new index 'p' refers to the associated NLA factorization $\mathbf{X}_{(p)}$. Likewise, relation (19) extends to $\mathbf{Y}^{(m,n)}_{(p)} = S_2(\mathbf{X}^{(m,n)}_{(p)}) = \left(\underline{\mathbf{A}}^{(m,n)}_{(p)} \odot \mathbf{B}^{(n)}_{(p)}\right)\mathbf{S}^T$, where $\mathbf{B}^{(n)}_{(p)} = \begin{bmatrix}1 & \cdots & 1\\ x^n_{1,p} & \cdots & x^n_{R,p}\end{bmatrix}$. Let us index the elements in Ω_p by $\omega_p(1),\dots,\omega_p(\text{card}\left(\Omega_p\right))$. We now obtain the augmented version of (24):

$$\mathbf{Y}_{\text{tot}} = \begin{bmatrix}\mathbf{Y}_{\text{tot},1}\\ \vdots\\ \mathbf{Y}_{\text{tot},P}\end{bmatrix},\ \mathbf{Y}_{\text{tot},p} = \begin{bmatrix}\mathbf{Y}^{(\omega_p(1))}_{(p)}\\ \vdots\\ \mathbf{Y}^{(\omega_p(\text{card}(\Omega_p)))}_{(p)}\end{bmatrix},\ \mathbf{Y}^{(n)}_{(p)} = \begin{bmatrix}\mathbf{Y}^{(1,n)}_{(p)}\\ \vdots\\ \mathbf{Y}^{(M_{n,p},n)}_{(p)}\end{bmatrix},$$

in which $\mathbf{Y}^{(\omega_p(n))}_{(p)} = (\mathbf{A}^{(\omega_p(n))}_{(p)} \odot \mathbf{B}^{(\omega_p(n))}_{(p)})\mathbf{S}^T$ and $\mathbf{A}^{(n)}_{(p)} = \left[\underline{\mathbf{A}}^{(1,n)T}_{(p)}\ \dots\ \underline{\mathbf{A}}^{(M_{n,p},n)T}_{(p)}\right]^T$. In the same way, the augmented version of (25) is given by

$$\mathbf{G}_{\text{tot}} = \left[\mathbf{G}^{(\Omega_1,1)T}\ \dots\ \mathbf{G}^{(\Omega_P,P)T}\right]^T, \tag{52}$$

where the index 'p' indicates that the matrix $\mathbf{G}^{(\Omega_p,p)}$, which is of the form (25), is built from the factor matrices $\{\mathbf{A}^{(\omega_p(n),p)}_{(p)}, \mathbf{B}^{(\omega_p(n),p)}_{(p)}\}_{n=1}^{\text{card}(\Omega_p)}$ associated with the pth NLA.

Proposition V.1 below summarizes the extension of Proposition III.1 to the multidimensional case.

Proposition V.1. *Consider the multidimensional NLA factorization of $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}^T \in \mathbb{C}^{I\times K}$ in (50). If*

$$\begin{cases}\boldsymbol{S} \text{ in (50) has full column rank},\\ \boldsymbol{G}_{\text{tot}} \text{ in (52) has full column rank},\\ \forall p \in \{1,\dots,P\},\ \exists\ i,j \in \Omega_p\ :\ GCD(n_{i,p}, n_{j,p}) = 1\,,\end{cases} \tag{53}$$

then the multidimensional NLA constrained rank of $\boldsymbol{X}$ is R and the multidimensional NLA decomposition of $\boldsymbol{X}$ is unique.

Furthermore, if the conditions (53) in Proposition V.1 are satisfied, then Algorithm 1 can easily be extended to the multidimensional case. We summarize the extension as Algorithm 2.

Algorithm 2 SD method for multidimensional MI-ESPRIT (Proposition V.1).

Input: $\mathbf{X} = \mathbf{A}\mathbf{S}^T$ of the form (50).

1. Compute SVD $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ (and find R from $\mathbf{\Sigma}$).
2. Choose row-selection matrices $J_{(1)},\dots,J_{(P)}$.
3. Extract $\mathbf{U}_{(p)} = J_{(p)}\mathbf{U}\mathbf{\Sigma}$ for every $p \in \{1,\dots,P\}$
4. Determine the subsets $\Omega_p \subseteq \{1,\dots,I^{(p)}\}$, $\forall p \in \{1,\dots,P\}$.
5. Extract $\{\mathbf{U}^{(m,n)}_{(p)}\}_{n\in\Omega_p}^{m\in\{1,\dots,M_{n,p}\}}$ of the form (34), $\forall p \in \{1,\dots,P\}$.
6. Build $\mathbf{Y}^{(n)}_{\text{red},p}$ of the form (35), $\forall n \in \Omega_p$, $\forall p \in \{1,\dots,P\}$.
7. Build $\{\mathbf{M}^{(r)}\}$ in (37) from $\{\mathbf{Y}^{(n)}_{\text{red},p}\}_{p\in\{1,\dots,P\}}^{n\in\Omega_p}$ [21], [18].
8. Solve SD problem $\mathbf{M}^{(r)} = \mathbf{G}\mathbf{\Lambda}^{(r)}\mathbf{G}^T$, $r \in \{1,\dots,R\}$.
9. Compute $\mathbf{A}_{(p)} = \mathbf{J}_{(p)}\mathbf{U}\mathbf{\Sigma}\mathbf{G}^T$, $\forall p \in \{1,\dots,P\}$ and $\mathbf{S} = \mathbf{V}^*\mathbf{G}$.
10. Obtain $\{x_{r,p}\}$ via $\mathbf{A}_{(p)}$ (e.g. via Proposition IV.1).

Output: $\{x_{r,p}\}$, $\{\mathbf{A}_{(p)}\}$ and $\mathbf{S}$.

VI. Numerical experiments

Consider the factorization $\mathbf{X} = \mathbf{A}\mathbf{S}^T$ in (8) or (50). The goal is to estimate the angles $\{\theta_r\}$ or $\{\theta_r, \phi_r\}$ from $\mathbf{T} = \mathbf{X} + \beta\mathbf{N}$, where $\mathbf{N}$ is an unstructured perturbation matrix and $\beta \in \mathbb{R}$ controls the noise level. The azimuth angles $\{\theta_r\}$ and elevation angles $\{\phi_r\}$ are randomly chosen on the intervals $[0, \pi]$ and $[-\pi/2, \pi/2]$, respectively. The real and imaginary entries of the perturbation matrices are randomly drawn from a Gaussian distribution with zero mean and unit variance. The following Signal-to-Noise Ratio (SNR) measure will be used: SNR [dB] $= 10\log(\|\mathbf{X}\|_F^2/\|\beta\mathbf{N}\|_F^2)$. The performance evaluation will be based on the distance between the true angles $\{\theta_r, \phi_r\}$ and their estimates $\{\hat{\theta}_r, \hat{\phi}_r\}$. The distance is measured according to: $\mathrm{E_{RMS}} = \min_{\sigma\in\Lambda_R} \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\theta_r - \hat{\theta}_{\sigma(r)}\right)^2 + \left(\phi_r - \hat{\phi}_{\sigma(r)}\right)^2}$, where Λ_R denotes the set of $R!$ possible permutations of the elements $\{1, \ldots, R\}$ and $\sigma(r)$ denotes the rth element of the corresponding permuted set.

A. One-dimensional (1D) NLA

Consider the NLA problem with $I = 8$, $K = 50$, $R = 3$ and NLA response matrix $\mathbf{A}$ with columns of the form (32). We compute the NLA factorization of (32) via Algorithm 1 in which the generators x_r, x_r^2, x_r^3 and x_r^{10} are used and the associated subvectors of (32) are $[1\ x_r\ x_r^2]^T$, $[1\ x_r^2 \mid x_r^{21}\ x_r^{23}]^T$, $[x_r^{15}\ x_r^{18}\ x_r^{21}]^T$ and $[1\ x_r^{11}\ x_r^{21}]^T$, respectively. Note that methods that only exploit one shift-invariance structure within the NLA require that $R \leq 2$ and hence cannot be used. In order to demonstrate that the multiresolution property of $\mathbf{A}$ leads to improved performance, we compare the $I = 8$ element NLA with an $I = 8$ (number of elements in the NLA) element ULA and an $I = 23$ (aperture of the NLA) element ULA. The DOA estimation in the ULA cases will be based on the ESPRIT method [15]. In this experiment we only estimate the azimuth angles and consequently we set $\phi_r = \hat{\phi}_{\sigma(r)}$ in $\mathrm{E_{RMS}}$. The mean $\mathrm{E_{RMS}}$ value with $\phi_r = \hat{\phi}_{\sigma(r)}$ over 500 trials as a function of SNR can be seen in figure 1. We observe that the NLA performs significantly better than the $I = 8$ element ULA and almost as well as the $I = 23$ element ULA.

B. L-shaped two-dimensional (2D) NLA

Consider the L-shaped 2D NLA in subsection V-A with $I^{(1)} = 7$, $I^{(2)} = 6$, $K = 50$ and $R = 4$. Consider also an $(I^{(1)}, I^{(2)}) = (7, 6)$ (number of elements in the L-shaped 2D NLA) element 2D ULA and an $(I^{(1)}, I^{(2)}) = (9, 13)$ (aperture of the L-shaped 2D NLA) element 2D ULA. The DOA estimation in the 2D NLA case and in the 2D ULA cases will be based on Algorithm 2 and the SD algorithm developed in [18], respectively. The mean $\mathrm{E_{RMS}}$ value over 500 trials as a function of SNR can be seen in figure 2. As in the 1D case, 2D NLA performs better than the $(I^{(1)}, I^{(2)}) = (7, 6)$ element 2D ULA and almost as well as the $(I^{(1)}, I^{(2)}) = (9, 13)$ element 2D ULA.

VII. Conclusion

Sparse array processing presents many important algebraic challenges. We first made a link between (multidimensional) NLA and the CMTF. This led to a new uniqueness condition for sparse array processing problems involving shift-invariance structures. To the authors' knowledge, this is the first formal proof that improved identifiability results for a sparse array can be obtained by simultaneously exploiting several shift-invariance structures within the sparse array.

Second, the CMTF approach also provided us with an algebraic SD method that is guaranteed to find the decomposition in the exact case if the presented uniqueness condition is satisfied. In the inexact case, it can be used as a computationally inexpensive initialization procedure for a more costly optimization method. The low cost property of the SD method also makes it interesting for shift-invariant structured large scale and sparse arrays.

Finally, numerical experiments indicated that by exploiting the multiresolution property of sparse arrays, a good compromise between performance, identifiability, complexity and the number of sampling points (e.g. antennas) can be obtained.

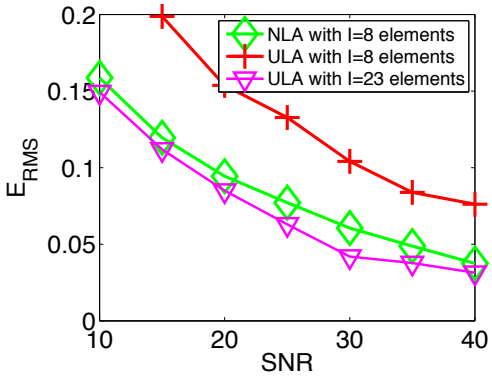

Fig. 1. Mean $\mathrm{E_{RMS}}$ over 500 trials, 1D NLA and ULAs.

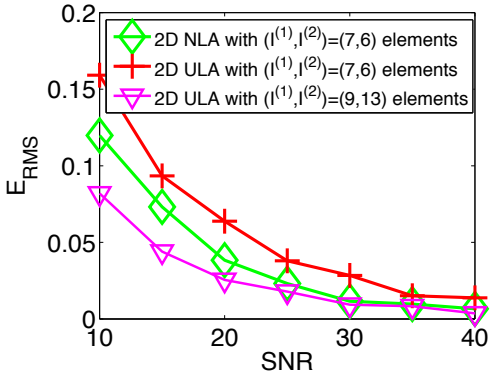

Fig. 2. Mean $\mathrm{E_{RMS}}$ over 500 trials, L-shaped 2D NLA and ULAs.

References

[1] F. Athley, C. Engdahl, and P. Sunnergren, "On radar detection and direction finding using sparse arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 4, pp. 1319–1333, Oct. 2007.

[2] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 3, pp. 642–666, 2006.

[3] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part I: Basic results and uniqueness of one factor matrix," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 855–875, 2013.

[4] ——, "On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part II: Overall uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 876–903, 2013.

[5] C. El Kassis, J. Picheral, and C. Mokbel, "Advantages of nonuniform arrays using root-MUSIC," *Signal Processing*, vol. 90, pp. 689–695, 2010.

[6] E. Jacobs and E. W. Ralston, "Ambiguity resolution in interferometry," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 17, no. 6, pp. 766–780, Nov. 1981.

[7] T. Jiang and N. D. Sidiropoulos, "Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear model with constant modulus constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2625–2636, Sep. 2004.

[8] A. N. Lemma, A.-J. van der Veen, and E. F. Deprettere, "Multiresolution ESPRIT algorithm," *IEEE Trans. Signal Process.*, vol. 47, no. 6, pp. 1722–1726, Jun. 1999.

[9] S. Leurgans, R. T. Ross, and R. B. Abel, "A decomposition for three-way arrays," *SIAM J. Matrix Anal. Appl.*, vol. 14, no. 4, pp. 1064–1083, 1993.

[10] A. T. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-16, no. 2, pp. 172–175, Mar. 1968.

[11] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the MUSIC algorithm," in *Proc. of the 14th/6th IEEE DSP/SPE Workshop*, Jan 4-7, Sedona AZ, USA, 2011.

[12] ——, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[13] P. Parvazi, M. Pesavento, and A. B. Gershman, "Rooting-based harmonic retrieval using multiple shift-invariances: The complete and the incomplete cases," *IEEE Trans. Signal Processing*, vol. 60, no. 4, pp. 1556–1570, Apr. 2012.

[14] S. Qin, Y. D. Zhang, and M. G. Amin, "Generalized coprime array configurations for direction-of-arrival estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1377–1390, Jun. 2015.

[15] R. Roy and T. Kailath, "Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 7, pp. 984–995, Jul. 1989.

[16] M. Sørensen and L. De Lathauwer, "New uniqueness conditions for the canonical polyadic decomposition of third-order tensors," *Accepted for publication in SIAM J. Matrix Anal. Appl.*

[17] ——, "Multidimensional harmonic retrieval via coupled canonical polyadic decompositions," ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep. 13-240, 2013.

[18] ——, "Coupled tensor decompositions in array processing," ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep. 13-241, 2014.

[19] ——, "Multiple invariance ESPRIT: A coupled matrix-tensor factorization approach," ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep. 13-242, 2014.

[20] ——, "Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-$(L_{r,n}, L_{r,n}, 1)$ terms — Part I: Uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 2, pp. 496–522, 2015.

[21] M. Sørensen, I. Domanov, and L. De Lathauwer, "Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-$(L_{r,n}, L_{r,n}, 1)$ terms — Part II: Algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 3, pp. 1015–1045, 2015.

[22] A. Stegeman, J.M.F. ten Berge, and L. De Lathauwer, "Sufficient conditions for uniqueness in Candecomp/Parafac and Indscal with random component matrices," *Psychometrika*, vol. 71, pp. 219–229, 2006.

[23] A. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple invariance ESPRIT," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 868–881, Apr. 1992.

[24] P. P. Vaidyanathan and P. Pal, "Why does direct-MUSIC on sparse-arrays work ?" in *Proc. Asilomar 2013*, Nov 3-6, Pacific Grove, CA, USA, 2013.

[25] ——, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb. 2011.

[26] E. Vertatschitsch and S. Haykin, "Nonredundant arrays," *Proc. of the IEEE*, vol. 74, no. 1, p. 217, Jan. 1986.

[27] A. J. Weiss, A. S. Wilsky, and B. C. Levy, "Nonuniform array processing via the polynomial approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, no. 1, pp. 48–55, Jan. 1989.

[28] M. J. Wilson and R. McHugh, "Sparse-periodic hybrid array beamformer," *IET Radar Sonar Navig.*, vol. 1, no. 2, pp. 116–123, 2007.

[29] K. T. Wong and M. D. Zoltowski, "Direction-finding with sparse rectangular dual-size spatial invariance array," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 17, no. 4, pp. 1320–1335, Oct. 1998.

Acknowledgment

Research supported by: (1) Research Council KU Leuven: CoE EF/05/006 Optimization in Engineering (OPTEC), C1 project c16/15/059-nD, (2) F.W.O.: project G.0830.14N, G.0881.14N, (3) the Belgian Federal Science Policy Office: IUAP P7 (DYSCO II, Dynamical systems, control and optimization, 2012–2017), (4) EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no. 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.



Mikael Sørensen received the Master's degree from Aalborg University, Denmark, and the Ph.D. degree from University of Nice, France, in 2006 and 2010, respectively, both in electrical engineering. Since 2010 he has been a Postdoctoral Fellow with the KU Leuven, Belgium. His research interests include applied linear algebra, tensor decompositions, and tensor-based (array) signal processing.



Lieven De Lathauwer received the Master's degree in electromechanical engineering and the Ph.D. degree in applied sciences from KU Leuven, Belgium, in 1992 and 1997, respectively. He is currently Professor with KU Leuven, Belgium. Dr. De Lathauwer is an Associate Editor of the SIAM Journal on Matrix Analysis and Applications and has served as an Associate Editor for the IEEE Transactions on Signal Processing. His research concerns the development of tensor tools for engineering applications.

SUPPLEMENTARY MATERIAL — PROOF OF PROPOSITION IV.1

MIKAEL SØRENSEN *† AND LIEVEN DE LATHAUWER *†

Proposition IV.1. Consider $f(\phi_r) = \sum_{n\in\Omega} \left\| \mathbf{e}_r^{(n)} e^{i\cdot\phi_r\cdot n} - \mathbf{f}_r^{(n)} \right\|_F^2$, where $\mathbf{e}_r^{(n)}$ and $\mathbf{f}_r^{(n)}$ are given by [1, eq. (38)] and $\Omega \subset \{1,\ldots,I\}$. The largest element in Ω is denoted by $N_{\max}$. Consider also the real polynomial of degree $2N_{\max}$:

$$\begin{aligned} g(t) = \sum_{n\in\Omega} \Bigg(& \mathrm{Re}\left\{\beta_r^{(n)}\right\} \sum_{p=0}^{2n} \sum_{k=0}^{N_{max}-n} \binom{2n}{p} \binom{N_{max}-n}{k} \Big(-pt^{(2n-p+1+2k)} \\ & + (2n-p)t^{(2n-p-1+2k)} \Big) \cdot \cos(\frac{1}{2}(2n-p)\pi) \\ & + \mathrm{Im}\{\beta_r^{(n)}\} \sum_{q=0}^{2n} \sum_{m=0}^{N_{max}-n} \binom{2n}{q} \binom{N_{max}-n}{m} \Big(-qt^{(2n-q+1+2m)} \\ & + (2n-q)t^{(2n-q-1+2m)} \Big) \cdot \sin(\frac{1}{2}(2n-q)\pi) \Bigg), \end{aligned} \tag{S.0.1}$$

where $\beta_r^{(n)} = \mathbf{e}_r^{(n)H} \mathbf{f}_r^{(n)}$. Let $\hat{t}$ denote the real root of $g(t)$ that minimizes $f(2\tan^{-1}(\hat{t}))$, then

$$2\tan^{-1}(\hat{t}) = \arg \min_{0\le\phi_r<\pi} f(\phi_r).$$

Proof. Let $\mathbf{e}_r^{(n)}$ and $\mathbf{f}_r^{(n)}$ be given by [1, eq. (38)] and $x_r^n = e^{i\cdot\phi_r\cdot n}$ where $0 \le \phi_r < \pi$. Consider

$$\begin{aligned} f(\phi_r) &= \sum_{n\in\Omega} \left\| \mathbf{e}_r^{(n)} x_r^n - \mathbf{f}_r^{(n)} \right\|_F^2 = \sum_{n\in\Omega} \left(\alpha_r^{(n)} - \beta_r^{(n)} e^{-i\cdot\phi_r\cdot n} - \beta_r^{(n)*} e^{i\cdot\phi_r\cdot n} \right) \\ &= \sum_{n\in\Omega} \left(\alpha_r^{(n)} - 2\mathrm{Re}\left\{\beta_r^{(n)}\right\} \cos(\frac{\phi_r}{2}\cdot 2n) - 2\mathrm{Im}\{\beta_r^{(n)}\} \sin(\frac{\phi_r}{2}\cdot 2n) \right) \\ &= \sum_{n\in\Omega} \left(\alpha_r^{(n)} - 2\hat{\beta}_r^{(n)} \sum_{p=0}^{2n} g(\frac{\phi_r}{2}, 2n, p) - 2\check{\beta}_r^{(n)} \sum_{q=0}^{2n} h(\frac{\phi_r}{2}, 2n, q) \right), \end{aligned} \tag{S.0.2}$$

*Group Science, Engineering and Technology, KU Leuven - Kulak, E. Sabbelaan 53, 8500 Kortrijk, Belgium, KU Leuven - E.E. Dept. (ESAT) - STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, and iMinds Medical IT Department, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium. {Mikael.Sorensen, Lieven.DeLathauwer}@kuleuven-kulak.be.

†Research supported by: (1) Research Council KU Leuven: CoE EF/05/006 Optimization in Engineering (OPTEC), C1 project c16/15/059-nD, (2) F.W.O.: project G.0830.14N, G.0881.14N, (3) the Belgian Federal Science Policy Office: IUAP P7 (DYSCO II, Dynamical systems, control and optimization, 2012–2017), (4) EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no. 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

where $\alpha_r^{(n)} = \mathbf{e}_r^{(n)H}\mathbf{e}_r^{(n)} + \mathbf{f}_r^{(n)H}\mathbf{f}_r^{(n)}$, $\beta_r^{(n)} = \mathbf{e}_r^{(n)H}\mathbf{f}_r^{(n)}$, $\hat{\beta}_r^{(n)} = \mathrm{Re}\left\{\beta_r^{(n)}\right\}$, $\check{\beta}_r^{(n)} = \mathrm{Im}\{\beta_r^{(n)}\}$ and the sine and cosine multiple angle formulas have been used, i.e.,

$$g(\frac{\phi_r}{2}, 2n, p) = m(\frac{\phi_r}{2}, 2n, p) \cdot \hat{\gamma}(n, p) \ ,$$
$$h(\frac{\phi_r}{2}, 2n, q) = m(\frac{\phi_r}{2}, 2n, q) \cdot \check{\gamma}(n, q) \ ,$$

where $m(\frac{\phi_r}{2}, 2n, p) = \binom{2n}{p}\cos^p(\frac{\phi_r}{2})\sin^{(2n-p)}(\frac{\phi_r}{2})$, $\hat{\gamma}(n, p) = \cos(\frac{1}{2}(2n-p)\pi)$ and $\check{\gamma}(n, q) = \sin(\frac{1}{2}(2n-q)\pi)$. The derivative of $f(\phi_r)$ is given by

$$\frac{\mathrm{d}f(\phi_r)}{\mathrm{d}\phi_r} = -2\sum_{n\in\Omega}\big(\hat{\beta}_r^{(n)}\sum_{p=0}^{2n} g'(\frac{\phi_r}{2}, 2n, p) + \check{\beta}_r^{(n)}\sum_{q=0}^{2n} h'(\frac{\phi_r}{2}, 2n, q)\big), \tag{S.0.3}$$

where

$$g'(\frac{\phi_r}{2}, 2n, p) = \frac{1}{2}\cdot\binom{2n}{p}\cdot s(\frac{\phi_r}{2}, 2n, p)\cdot\hat{\gamma}(n, p) \ ,$$
$$h'(\frac{\phi_r}{2}, 2n, q) = \frac{1}{2}\cdot\binom{2n}{q}\cdot s(\frac{\phi_r}{2}, 2n, q)\cdot\check{\gamma}(n, q) \ ,$$

in which

$$s(\frac{\phi_r}{2}, 2n, p) = -p\cos^{(p-1)}(\frac{\phi_r}{2})\sin^{(2n-p+1)}(\frac{\phi_r}{2}) + (2n-p)\cos^{(p+1)}(\frac{\phi_r}{2})\sin^{(2n-p-1)}(\frac{\phi_r}{2}).$$

Since $\big(\cos^2(\phi_r/2) + \sin^2(\phi_r/2)\big)^{(N_{max}-n)} = 1$, equation (S.0.3) can be made homogeneous as follows

$$\begin{aligned}\frac{\mathrm{d}f(\phi_r)}{\mathrm{d}\phi_r} = \\ -2\sum_{n\in\Omega}\Bigg(\hat{\beta}_r^{(n)}\sum_{p=0}^{2n} g'(\frac{\phi_r}{2}, 2n, p)\big(\cos^2(\frac{\phi_r}{2}) + \sin^2(\frac{\phi_r}{2})\big)^{(N_{max}-n)} \\ + \check{\beta}_r^{(n)}\sum_{q=0}^{2n} h'(\frac{\phi_r}{2}, 2n, q)\big(\cos^2(\frac{\phi_r}{2}) + \sin^2(\frac{\phi_r}{2})\big)^{(N_{max}-n)}\Bigg).\end{aligned} \tag{S.0.4}$$

Using

$$\left(\cos^2(\frac{\phi_r}{2}) + \sin^2(\frac{\phi_r}{2})\right)^{(N_{max}-n)} = \sum_{k=0}^{N_{max}-n} n(\frac{\phi_r}{2}, n, N_{max}, k),$$

where

$$n(\frac{\phi_r}{2}, n, N_{max}, k) = \binom{N_{max}-n}{k}\cos^{2(N_{max}-n-k)}(\frac{\phi_r}{2})\sin^{(2k)}(\frac{\phi_r}{2}),$$

we obtain

$$g'(\frac{\phi_r}{2},2n,p)\cdot\left(\cos^2(\frac{\phi_r}{2})+\sin^2(\frac{\phi_r}{2})\right)^{(N_{max}-n)}$$
$$=\frac{1}{2}\sum_{k=0}^{N_{max}-n}\binom{2n}{p}\cdot\binom{N_{max}-n}{k}\cdot t(\frac{\phi_r}{2},p,N_{max},k)\cdot\hat{\gamma}(n,p),$$
$$h'(\frac{\phi_r}{2},2n,q)\cdot\left(\cos^2(\frac{\phi_r}{2})+\sin^2(\frac{\phi_r}{2})\right)^{(N_{max}-n)}$$
$$=\frac{1}{2}\sum_{k=0}^{N_{max}-n}\binom{2n}{q}\cdot\binom{N_{max}-n}{k}\cdot t(\frac{\phi_r}{2},q,N_{max},k)\cdot\check{\gamma}(n,q),$$

where

$$t(\frac{\phi_r}{2},p,N_{max},k)=-p\cos^{(2N_{max}+p-1-2n-2k)}(\frac{\phi_r}{2})\sin^{(2n-p+1+2k)}(\frac{\phi_r}{2})$$
$$+(2n-p)\cos^{(2N_{max}+p+1-2n-2k)}(\frac{\phi_r}{2})\sin^{(2n-p-1+2k)}(\frac{\phi_r}{2}).$$

Since $0\leq\phi_r<\pi$, we note that

$$g'(\frac{\phi_r}{2},2n,p)\left(\cos^2(\frac{\phi_r}{2})+\sin^2(\frac{\phi_r}{2})\right)^{(N_{max}-n)}/\cos^{(2N_{max})}(\frac{\phi_r}{2})$$
$$=\frac{1}{2}\sum_{k=0}^{N_{max}-n}\binom{2n}{p}\cdot\binom{N_{max}-n}{k}\cdot\Big(-p\tan^{(2n-p+1+2k)}(\frac{\phi_r}{2})$$
$$+(2n-p)\tan^{(2n-p-1+2k)}(\frac{\phi_r}{2})\Big)\cdot\hat{\gamma}(n,p)\ , \tag{S.0.5}$$
$$h'(\frac{\phi_r}{2},2n,q)\left(\cos^2(\frac{\phi_r}{2})+\sin^2(\frac{\phi_r}{2})\right)^{(N_{max}-n)}/\cos^{(2N_{max})}(\frac{\phi_r}{2})$$
$$=\frac{1}{2}\sum_{k=0}^{N_{max}-n}\binom{2n}{q}\cdot\binom{N_{max}-n}{k}\cdot\Big(-q\tan^{(2n-q+1+2k)}(\frac{\phi_r}{2})$$
$$+(2n-q)\tan^{(2n-q-1+2k)}(\frac{\phi_r}{2})\Big)\cdot\check{\gamma}(n,q)\ . \tag{S.0.6}$$

Combination of (S.0.4), (S.0.5) and (S.0.6) yields that we can find the critical values as the roots of the real polynomial of degree $2N_{\max}$ given in (S.0.1):

$$\frac{\mathrm{d}f(\phi_r)}{\mathrm{d}\phi_r}/\cos^{(2N_{\max})}(\phi_r/2)=0\Leftrightarrow g(t)=0\,,$$

where we have substituted $t=\tan(\phi_r/2)\in\mathbb{R}$. After computing the real roots of $g(t)$, the critical value that minimizes (S.0.2) is retained. □

REFERENCES

[1] M. Sørensen and L. De Lathauwer, *Multiple Invariance ESPRIT: A Coupled Matrix-Tensor Factorization Approach*, ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep. 13-242, 2014.