Tensor methods for multisensor signal processing

Sebastian Miron*, *Université de Lorraine, CNRS, CRAN,* F-54000 Nancy, France
Yassine Zniyed, *Université de Lorraine, CNRS, CRAN,* F-54000 Nancy, France
Rémy Boyer, *Université de Lille, CNRS, CRIStAL,* 59655 Lille, France
André L.F. de Almeida, *Federal University of Ceará,* Fortaleza, CE 60440-900, Brazil
Gérard Favier, *Université de Côte d'Azur, CNRS, I3S Laboratory,* Sophia Antipolis, France
David Brie, *Université de Lorraine, CNRS, CRAN,* F-54000 Nancy, France
Pierre Comon, *Université Grenoble Alpes, CNRS, GIPSA-Lab,* 38000 Grenoble, France

***Abstract*—Over the last two decades, tensor-based methods have received growing attention in the signal processing community. In this work, we propose a comprehensive overview of tensor-based models and methods for multisensor signal processing. We present for instance the Tucker decomposition, the Canonical Polyadic Decomposition (CPD), the Tensor-Train Decomposition (TTD), the Structured TTD, including Nested Tucker Train (NTT), as well as the associated optimization strategies. More precisely, we give synthetic descriptions of state-of-art estimators as the Alternating Least Square (ALS) algorithm, the High-Order SVD (HOSVD), and of more advanced algorithms as the Rectified ALS, the TT-SVD/TT-HSVD and the Joint dImensionally Reduction And Factor retrieval Estimator (JIRAFE) scheme. We illustrate the efficiency of the introduced methodological and algorithmic concepts in the context of three important and timely signal processing-based applications: the Direction-Of-Arrival (DOA) estimation based on sensor arrays, multidimensional harmonic retrieval and MIMO wireless communication systems.**

I. Introduction

Tensors are, roughly speaking, generalizations of matrices to more than 2 dimensions (orders larger than 2). Tensors present several advantages over matrices, such as uniqueness of rank decomposition under mild conditions, or preservation of the local structure in N-way data processing. This is what motivated the advent of tensors in data analysis (phonetics [1], psychometrics [2]) in the early '70s and later on, in the signal processing community. Over the last fifteen years approximately, the computer science community actively participated in the popularization of tensors as data analysis tools. Due to their ability to model highly structured data, tensors quickly became a staple of data mining and machine learning fields (see, *e.g.* [3], [4]).

In multisensor signal processing, tensors are being used for almost two decades in various applications such as Direction Of Arrival (DOA) estimation [5], harmonic retrieval [6], communications systems [7] or RADAR [8]. However, despite a large number of publications on this topic, we were unable to find in the literature a comprehensive "point of entry" article on the use of tensors in multisensor signal processing. The aim of the present work is to fill this need. This paper does not intend to draw an exhaustive picture of the existing methods and models but rather provides an overview of the principles and main ideas used in tensor-based multisensor signal processing. We illustrate our presentation by three popular multisensor applications: DOA estimation, multidimensional harmonic retrieval and MIMO systems design.

The paper is organized as follows: in Section II we define the main tensor concepts and notations and in Section III we introduce some of the most commonly used tensor decompositions and algorithms. These tensor decompositions are then applied to DOA estimation problem (Section IV), multidimensional harmonic retrieval (Section V) and MIMO wireless communication system (Section VI). Conclusions are drawn in Section VII.

II. Tensor preliminaries: notations and operations

Tensors are *multilinear maps* between two sets of linear spaces [9]. Once these linear spaces are fixed as well as their bases, a tensor is characterized by its array of coordinates. In the present contribution, we shall assume that this is the case and shall mainly work on array coordinates. Hence, with usual abuse of vocabulary, we shall assimilate *tensors* with their array representations. The number of dimensions (indices) of such an array is called the *order* of the tensor.

Tensors are typeset with a bold calligraphic font (*e.g.*, $\boldsymbol{\mathcal{T}}$), matrices in bold caps (*e.g.*, $\boldsymbol{Q}$); the elements of tensors or matrices are typeset as $\mathcal{T}_{i,j,k}$ and $Q_{i,j}$ respectively; vectors are denoted with bold lower cases (*e.g.* $\boldsymbol{v}$), with entries v_i. The Frobenius norm of a matrix/tensor is denoted $\|\cdot\|_F$; $(.)^{\mathsf{T}}$, $(.)^{\mathsf{H}}$ and $(.)^{\dagger}$ symbolyze the matrix transpose, conjugate transpose and the Moore-Penrose inverse, respectively, and "$\star$" denotes the convolution operator.

The tensor (outer) product between tensors, matrices, or vectors, is denoted by $\otimes$, as in *e.g.* $\boldsymbol{\mathcal{T}} = \boldsymbol{v} \otimes \boldsymbol{M}$. The contraction on the pth index of a tensor is denoted as $\underset{p}{\bullet}$; by convention, when contracted with a matrix, the summation always applies on the first index of the matrix if it appears after, or to the second index if it appears first. In other words, the product $\boldsymbol{A}\,\boldsymbol{B}$ between two matrices can be written $\boldsymbol{A} \bullet \boldsymbol{B}$ as well, without further index specification.

For instance, the products $\mathcal{X}_{i,j,k} = \sum_q \mathcal{T}_{q,j,k} M_{i,q}$, $\mathcal{Y}_{i,j,k} = \sum_q \mathcal{T}_{i,q,k} M_{j,q}$ and $\mathcal{Z}_{i,j,k} = \sum_q \mathcal{T}_{i,j,q} M_{k,q}$ can be written in compact form as $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{T}} \underset{1}{\bullet} \boldsymbol{M}$, $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{T}} \underset{2}{\bullet} \boldsymbol{M}$ and $\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{T}} \underset{3}{\bullet} \boldsymbol{M}$, respectively. Note that we also have $\boldsymbol{\mathcal{X}} = \boldsymbol{M}^{\mathsf{T}} \underset{1}{\bullet} \boldsymbol{\mathcal{T}}$.

*Corresponding author.
E-mail: sebastian.miron@univ-lorraine.fr

The Kronecker product between two matrices is denoted $\boldsymbol{A} \boxtimes \boldsymbol{B}$, and is defined by the block matrix:

$$\boldsymbol{A} \boxtimes \boldsymbol{B} = \begin{bmatrix} A_{1,1}\boldsymbol{B} & A_{1,2}\boldsymbol{B} & \dots & A_{1,J}\boldsymbol{B} \\ A_{2,1}\boldsymbol{B} & A_{2,2}\boldsymbol{B} & \dots & A_{1,J}\boldsymbol{B} \\ \vdots & \vdots & & \\ A_{I,1}\boldsymbol{B} & A_{I,2}\boldsymbol{B} & \dots & A_{I,J}\boldsymbol{B} \end{bmatrix},$$

if matrix $\boldsymbol{A}$ is of size $I \times J$. Once bases of linear spaces are well defined, the tensor product between operators can be represented by Kronecker products [9]. The Khatri-Rao (column-wise Kronecker) product between two matrices with the same number J of columns is written as:

$$\boldsymbol{A} \odot \boldsymbol{B} = [\boldsymbol{a}_1 \boxtimes \boldsymbol{b}_1, \dots \boldsymbol{a}_2 \boxtimes \boldsymbol{b}_2, \dots \boldsymbol{a}_J \boxtimes \boldsymbol{b}_J],$$

if $\boldsymbol{a}_j$ and $\boldsymbol{b}_j$ represent the columns of $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively.

Analogously to the Frobenius norm of a matrix, the *Frobenius norm* of a tensor $\boldsymbol{\mathcal{T}}$ is defined as the square root of the sum of all the squares of its elements, *i.e.*:

$$\|\boldsymbol{\mathcal{T}}\|_F = \sqrt{\sum_i \sum_j \sum_k \mathcal{T}_{i,j,k}^2}.$$

Contraction between two tensors

The product $\overset{p}{\underset{q}{\bullet}}$ between two tensors $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ of size $I_1 \times \cdots \times I_Q$ and $J_1 \times \cdots \times J_P$, respectively, where $I_q = J_p$, is a tensor of order $(Q + P - 2)$ denoted by:

$$[\boldsymbol{\mathcal{A}} \overset{p}{\underset{q}{\bullet}} \boldsymbol{\mathcal{B}}]_{i_1,\dots,i_{q-1},i_{q+1},\dots,i_Q,j_1,\dots,j_{p-1},j_{p+1},\dots,j_P}$$
$$= \sum_{k=1}^{I_q} [\boldsymbol{\mathcal{A}}]_{i_1,\dots,i_{q-1},k,i_{q+1},\dots,i_Q} [\boldsymbol{\mathcal{B}}]_{j_1,\dots,j_{p-1},k,j_{p+1},\dots,j_P}.$$

Tensor reshaping

Tensor reshaping transforms a Q-order tensor $\boldsymbol{\mathcal{T}}$ of dimensions $I_1 \times \cdots \times I_Q$ into a matrix $\boldsymbol{T}_{(q)}$ having the product of the first q dimensions of $\boldsymbol{\mathcal{T}}$, say $I_1 \cdots I_q$, as the number of rows, and the product of the remaining dimensions, $I_{q+1} \cdots I_Q$, as the number of columns. In MATLAB, this reshaping can be obtained using the native `reshape` function, such that

$$\boldsymbol{T}_{(q)} = \text{reshape}\left(\boldsymbol{\mathcal{T}}, \prod_{s=1}^{q} I_s, \prod_{s=q+1}^{Q} I_s\right). \tag{1}$$

For example, let $\boldsymbol{\mathcal{T}}$ be a 4-order tensor of dimensions $2 \times 2 \times 2 \times 2$, defined by:

$$\boldsymbol{\mathcal{T}}(:,:,1,1) = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \boldsymbol{\mathcal{T}}(:,:,2,1) = \begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix},$$
$$\boldsymbol{\mathcal{T}}(:,:,1,2) = \begin{bmatrix} 9 & 11 \\ 10 & 12 \end{bmatrix}, \boldsymbol{\mathcal{T}}(:,:,2,2) = \begin{bmatrix} 13 & 15 \\ 14 & 16 \end{bmatrix}.$$

Matrix $\boldsymbol{T}_{(2)}$ of dimensions 4×4 is then given by:

$$\boldsymbol{T}_{(2)} = \text{reshape}\,(\boldsymbol{\mathcal{T}}, 4, 4) = \begin{bmatrix} 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \\ 4 & 8 & 12 & 16 \end{bmatrix}.$$

The so-called *flattening* or *unfolding* operation is a special case of the reshaping operation [9]. The most used flattenings are those keeping one of the original dimensions, say N_q, as the number of rows. Hence, the so-called qth matrix unfolding of $\boldsymbol{\mathcal{T}}$ is of dimension $I_q \times I_1 \dots I_{q-1} I_{q+1} \dots I_Q$. Compared to the flattening operation, the `reshape` function can be considered as a more general flattening operation, in the sense that

$$\text{unfold}_q \boldsymbol{\mathcal{T}} = \text{reshape}\left(\boldsymbol{\mathcal{T}}^{'}, I_q, \frac{\prod_{s=1}^{Q} I_s}{I_q}\right), \tag{2}$$

with

$$\boldsymbol{\mathcal{T}}^{'} = \text{permute}\,(\boldsymbol{\mathcal{T}}, [q, 1, 2, \cdots, q-1, q+1, q+2, \cdots, Q]),$$

where `permute` is a native `MATLAB` function that rearranges the dimensions of the Q-order tensor $\boldsymbol{\mathcal{T}}$.

III. Tensor decompositions and algorithms

We introduce hereafter the tensor decompositions (Section III-A) used in the applications presented in Sections IV, V and VI, as well as some basic algorithms to compute these decompositions (Section III-B).

The Tucker decomposition along with two of its variants (High-Order SVD and partial Tucker), are introduced in Subsection III-A1. Tucker decomposition is a generic tensor tool that decomposes a tensor into a set of non-structured factor matrices and a core tensor. A simple algorithm for computing the High-Order SVD (Tucker decomposition with semi-unitary factor matrices) for third-order tensors is given in Subsection III-B2. The Canonical Polyadic Decomposition (CPD) (Subsection III-A2) - probably the most widely used tensor decomposition - can be seen as a Tucker decomposition with a diagonal core. Its existence and uniqueness issues are discussed in Subsection III-A3. The pseudo-code for the popular Alternating Least Squares (ALS) algorithm for estimating the CPD is given in Subsection III-B1. Subsection III-A4 introduces the Tensor-Train Decomposition (TTD), a tool designed to efficiently handle high-order tensors (order higher than 3), by splitting them into an interconnected set ("train") of lower-order tensors. A particular class of TTD, the Structured Tensor-Train (STT) models, is presented in Subsection III-A6. The Tensor-Train SVD (TT-SVD) algorithm for computing TTD is illustrated in Subsection III-B3. A link between the TTD and the CPD of high-order tensors (TT-CPD) is developed in Subsection III-B4; the JIRAFE (Joint dImensionality Reduction And Factors rEtrieval) method for estimating the parameters of the TT-CPD model is also detailed. Two additional tools used in MIMO application (Section VI), the least squares Kronecker (Subsection III-B5) and least squares Khatri-Rao (Subsection III-B6) factorizations, conclude this section.

A. Tensor decompositions

Any matrix can always be diagonalized by congruent transformation. In addition, the two linear transforms involved can

be imposed to be unitary: this is the Singular Value decomposition (SVD). When we move from matrices to tensors, it is generally impossible to use unitary transforms and obtain a diagonal tensor. Consequently, depending on which property is desired, we end up with two different decompositions. This is what is explained in the next two subsections, where we limit ourselves to third order tensors to ease the presentation.

1) Tucker decomposition, HOSVD, and multilinear rank: Given a tensor $\boldsymbol{\mathcal{T}}$ of dimensions $I \times J \times K$, it is always possible to find three matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ of dimensions $I \times R_1$, $J \times R_2$ and $K \times R_3$, respectively, such that

$$\mathcal{T}_{i,j,k} = \sum_{m=1}^{R_1} \sum_{n=1}^{R_2} \sum_{p=1}^{R_3} A_{i,m} B_{j,n} C_{k,p}\, \mathcal{G}_{m,n,p}, \tag{3}$$

where R_1, R_2, R_3 are minimal and $R_1 \leq I$, $R_2 \leq J$, $R_3 \leq K$. Tensor $\boldsymbol{\mathcal{G}}$, often called *core tensor*, is of dimensions $R_1 \times R_2 \times R_3$. The triplet of minimal values of $\{R_1, R_2, R_3\}$ forms the *multilinear rank* of $\boldsymbol{\mathcal{T}}$. This is referred to as the *Tucker decomposition* of $\boldsymbol{\mathcal{T}}$, and can be written in a more compact way as:

$$\boldsymbol{\mathcal{T}} = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}; \boldsymbol{\mathcal{G}}]\!]. \tag{4}$$

Each R_ℓ is uniquely defined, and corresponds to the rank of the linear operator associated with the ℓth matrix unfolding of $\boldsymbol{\mathcal{T}}$, $\mathrm{unfold}_\ell \boldsymbol{\mathcal{T}}$. The point is that the Tucker decomposition is not unique, for it is defined up to three invertible matrices $\{\boldsymbol{M}_1, \boldsymbol{M}_2, \boldsymbol{M}_3\}$, because:

$$[\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}; \boldsymbol{\mathcal{G}}]\!] = [\![\boldsymbol{A}\boldsymbol{M}_1^{-1}, \boldsymbol{B}\boldsymbol{M}_2^{-1}, \boldsymbol{C}\boldsymbol{M}_3^{-1}; \boldsymbol{\mathcal{G}}']\!]$$

if $\boldsymbol{\mathcal{G}}' = [\![\boldsymbol{M}_1, \boldsymbol{M}_2, \boldsymbol{M}_3; \boldsymbol{\mathcal{G}}]\!]$. Yet, as in the SVD of matrices, it is possible to impose that the core tensor is obtained via semi-unitary transforms $\{\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}\}$ as:

$$\boldsymbol{\mathcal{T}} = [\![\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}; \boldsymbol{\mathcal{G}}]\!], \tag{5}$$

where $\boldsymbol{U}^{\mathsf{H}}\boldsymbol{U} = \boldsymbol{I}_{R_1}$, $\boldsymbol{V}^{\mathsf{H}}\boldsymbol{V} = \boldsymbol{I}_{R_2}$ and $\boldsymbol{W}^{\mathsf{H}}\boldsymbol{W} = \boldsymbol{I}_{R_3}$. Equation (5) defines the High-Order SVD (HOSVD), sometimes referred to as *multilinear SVD* of tensor $\boldsymbol{\mathcal{T}}$ [10].

a) Partial Tucker decomposition: We present next a variant of the Tucker decomposition for an Q-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times \cdots \times I_Q}$, with factor matrices $\boldsymbol{A}^{(q)} \in \mathbb{C}^{I_q \times R_q}$ whose $Q - Q_1$ last ones are equal to identity matrices $\boldsymbol{I}_{I_q}$ of order I_q, for $q = Q_1 + 1, \cdots, Q$. This so-called Tucker-(Q_1, Q) decomposition is compactly written as $[\![\boldsymbol{A}^{(1)}, \cdots, \boldsymbol{A}^{(Q_1)}, \boldsymbol{I}_{I_{Q_1+1}}, \cdots, \boldsymbol{I}_{I_Q}; \boldsymbol{\mathcal{G}}]\!]$, where the core tensor $\boldsymbol{\mathcal{G}}$ is of dimensions $R_1 \times \cdots \times R_{Q_1} \times I_{Q_1+1} \times \cdots \times I_Q$, which induces $R_q = I_q$ for $q = Q_1 + 1, \cdots, Q$. See [11] for more details. For instance, the standard Tucker2 and Tucker1 decompositions correspond to $(Q_1, Q) = (1, 3)$ and $(Q_1, Q) = (2, 3)$, respectively.

2) Exact CP decomposition, rank, uniqueness: Of course, there is no reason that tensor $\boldsymbol{\mathcal{G}}$ in Equation (5) be diagonal; this is clear by just looking at the number of degrees of freedom of both sides. On the other hand, if the unitary constraint of matrices $\{\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}\}$ is relaxed, it is possible to decompose any tensor $\boldsymbol{\mathcal{T}}$ as:

$$\boldsymbol{\mathcal{T}} = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}; \boldsymbol{\mathcal{S}}]\!], \tag{6}$$

where tensor $\boldsymbol{\mathcal{S}}$ is diagonal. This is known as the *Canonical Polyadic* (CP[1]) decomposition (CPD). Note that now, not only matrices $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}\}$ are not semi-unitary, but their number of columns, R, may exceed the number of their rows. The minimal value of R such that (6) holds *exactly* is called the *rank* of tensor $\boldsymbol{\mathcal{T}}$. The explicit writing of (6) in terms of array entries is:

$$\mathcal{T}_{i,j,k} = \sum_{r=1}^{R} A_{i,r} B_{j,r} C_{k,r}\ \sigma(r), \tag{7}$$

if $\sigma(r)$'s denote the diagonal entries of $\boldsymbol{\mathcal{S}}$. But a clearer writing of this CPD is provided in terms of *decomposable tensors* [9] as follows:

$$\boldsymbol{\mathcal{T}} = \sum_{r=1}^{R} \sigma_r \boldsymbol{\mathcal{D}}(r), \tag{8}$$

where $\boldsymbol{\mathcal{D}}(r) = \boldsymbol{a}(r) \otimes \boldsymbol{b}(r) \otimes \boldsymbol{c}(r)$, that is, $\mathcal{D}(r)_{i,j,k} = A_{i,r} B_{j,r} C_{k,r}$, $\|\boldsymbol{\mathcal{D}}(r)\| = 1$, and $\sigma_r > 0$. In fact, decomposable tensors may be seen as rank-1 tensors. This writing is clearer because it does not depend on the way tensors $\boldsymbol{\mathcal{D}}(r)$ are represented – and their representation is precisely rarely unique, which is *not related* to the uniqueness of (8). Note that the definition of tensor rank is pretty much the same as matrix rank. Hence, this CPD may be seen as another natural generalization of the matrix SVD to tensors, the former being HOSVD seen in the previous section. Because the acronym CP is sometimes used to mean "Completely Positive", it may sometimes be more suitable to call it *rank decomposition*, when the number of terms R is indeed minimal.

The uniqueness of the CPD (8) should not be confused with its array representation (7). In fact, the latter is never unique, since expressing a rank-1 tensor as the tensor product of N vectors is subject to $N - 1$ scaling indeterminacies. For instance, $\boldsymbol{\mathcal{D}} = \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c}$ can also be written as $\alpha\boldsymbol{a} \otimes \beta\boldsymbol{b} \otimes \gamma\boldsymbol{c}$ provided $\alpha\beta\gamma = 1$. This is precisely the difference between a tensor space and a product of linear spaces [9], [13], [14]. This is the reason why the wording of "essential uniqueness" is sometimes found in the literature; it expresses the fact that there are $N - 1$ scaling indeterminacies, and that the order of summation is subject to permutation because the addition is commutative.

With Definition (2), the CPD can be rewritten in matrix form. For a 3rd order rank-R tensor $\boldsymbol{\mathcal{T}} = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}; \boldsymbol{\mathcal{S}}]\!]$ for example, the three flattened representations of the CPD are [15]:

$$\mathrm{unfold}_1 \boldsymbol{\mathcal{T}} = \boldsymbol{A}\boldsymbol{S}(\boldsymbol{C} \odot \boldsymbol{B})^{\mathsf{T}}, \tag{9}$$

$$\mathrm{unfold}_2 \boldsymbol{\mathcal{T}} = \boldsymbol{B}\boldsymbol{S}(\boldsymbol{C} \odot \boldsymbol{A})^{\mathsf{T}}, \tag{10}$$

$$\mathrm{unfold}_3 \boldsymbol{\mathcal{T}} = \boldsymbol{C}\boldsymbol{S}(\boldsymbol{B} \odot \boldsymbol{A})^{\mathsf{T}}, \tag{11}$$

where $\boldsymbol{S}$ denotes the diagonal matrix with entries $\sigma(r) = \mathcal{S}_{r,r,r}$, $\forall 1 \leq r \leq R$.

One of the major interests in the CPD (8) lies in its uniqueness. In particular, it allows to relax the orthogonality constraint (necessary in the case of matrices to achieve some

[1]The acronym CP also stands for Candecomp/Parafac in the literature, due to two contributions [1], [2] in which this tensor decomposition has been independently rediscovered, years after the original publication [12].

form of uniqueness), which is often incompatible with physical reality. More precisely, several uniqueness conditions have been derived in the literature. We shall quote two of them. A sufficient condition is that [16]–[18]:

$$2R \leq \text{krank}\{\boldsymbol{A}\} + \text{krank}\{\boldsymbol{B}\} + \text{krank}\{\boldsymbol{C}\} - 2, \quad (12)$$

where $\text{krank}\{\boldsymbol{A}\}$ denotes the so-called *Kruskal rank* of $\boldsymbol{A}$ [9], [16]; by definition[2], the $krank$ is always smaller than or equal to the usual matrix rank. If entries of $\boldsymbol{A}$ are seen as randomly drawn according to an absolutely continuous distribution, $\text{krank}\{\boldsymbol{A}\}$ can be replaced by the smallest of its two dimensions. In the latter case, the condition is referred to as *generic* [19], [20].

A second generic condition ensuring uniqueness of the CPD is given by the bound of the so-called *expected rank* [21]–[23]:

$$R \leq \frac{R_1 R_2 R_3}{R_1 + R_2 + R_3 - 2} - 1, \quad (13)$$

where R_ℓ are the multilinear ranks defined in Section III-A1. To sum up, the exact CP decomposition (8) is almost surely unique if the tensor rank is not too large.

3) Approximate CPD, existence, uniqueness: Things become more complicated when it is thought to *approximate* a tensor by another one of lower rank. To fix the ideas, suppose that our objective is to minimize:

$$\Upsilon = \left\| \boldsymbol{\mathcal{T}} - \sum_{r=1}^{R_o} \sigma(r)\ \boldsymbol{a}(r) \otimes \boldsymbol{b}(r) \otimes \boldsymbol{c}(r) \right\|_F^2,$$

where $\sigma(r) > 0$ and $||\boldsymbol{a}(r)|| = ||\boldsymbol{b}(r)|| = ||\boldsymbol{c}(r)|| = 1$, $\forall r, 1 \leq r \leq R_o$. Unfortunately, it has been pointed out in *e.g.* [24] that this objective function may not always have a minimum (but only an infimum), which means that this low-rank approximation problem is ill-posed; there are ways to circumvent the difficulty, for instance by imposing extraneous angular constraints [25], which is meaningful in the context of DOA estimation for instance.

On the other hand, if all objects are real nonnegative, then the problem becomes well-posed [26]. In the latter case, it is important to recall that imposing a nonnegative constraint generally increases the rank [9], even for matrices. This being said, it has been proved in [22] that the best nonnegative low-rank tensor approximation of a nonnegative tensor is almost surely unique. One can even look at the conditions under which the best low-rank approximation admits a unique CPD [27], but this is more involved.

To conclude, the *best low-rank approximate* of a tensor does not always exist, and this is still often ignored in the literature.

4) Tensor-Train Decomposition (TTD) and TT-ranks: A Q-order tensor of size $I_1 \times \ldots \times I_Q$ admits a decomposition into a train of tensors [28] if

$$\boldsymbol{\mathcal{X}} = \boldsymbol{G}^{(1)} \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{G}}^{(2)} \overset{1}{\underset{3}{\bullet}} \boldsymbol{\mathcal{G}}^{(3)} \overset{1}{\underset{4}{\bullet}} \cdots \overset{1}{\underset{Q-1}{\bullet}} \boldsymbol{\mathcal{G}}^{(Q-1)} \overset{1}{\underset{Q}{\bullet}} \boldsymbol{G}^{(Q)}, \quad (14)$$

where the TT-cores $\boldsymbol{G}^{(1)}$, $\boldsymbol{\mathcal{G}}^{(q)} (2 \leq q \leq Q-1)$, and $\boldsymbol{G}^{(Q)}$ are, respectively, of size $I_1 \times R_1$, $R_{q-1} \times I_q \times R_q$ and $R_{Q-1} \times I_Q$. The $Q-1$ dimensions $\{R_1, \ldots, R_{Q-1}\}$ are referred to as the *TT-ranks* with boundary conditions $R_0 = R_Q = 1$.

The idea behind the TTD, is to transform a high Q-order tensor into a set of much lower 3-order tensors, which allows to break the "curse of dimensionality" [29]. Indeed, just like the CPD, the storage cost of the TTD scales linearly with the order Q. A graph-based representation of the TTD of a Q-order tensor $\boldsymbol{\mathcal{X}}$ is given in Fig. 1. It is straightforward to see that the TTD

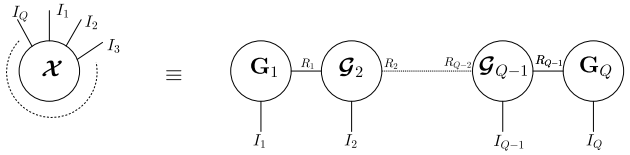

Fig. 1. TT decomposition of a Q-order tensor

is not unique since [30]

$$[\boldsymbol{\mathcal{X}}]_{i_1,\cdots,i_Q} = \boldsymbol{a}_1(i_1)^\mathsf{T} \boldsymbol{A}_2(i_2) \cdots \boldsymbol{A}_{Q-1}(i_{Q-1}) \boldsymbol{a}_Q(i_Q) \quad (15)$$

where

$$\boldsymbol{a}_1(i_1) = \boldsymbol{M}_1 [\boldsymbol{G}^{(1)}]^\mathsf{T}_{i_1,:} \quad : \quad R_1 \times 1, \quad (16)$$

$$\boldsymbol{a}_Q(i_Q) = \boldsymbol{M}_{Q-1}^{-1} [\boldsymbol{G}^{(Q)}]_{:,i_Q} \quad : \quad R_{Q-1} \times 1, \quad (17)$$

$$\boldsymbol{A}_q(i_q) = \boldsymbol{M}_{q-1}^{-\mathsf{T}} [\boldsymbol{\mathcal{G}}^{(q)}]_{:,i_q,:} \boldsymbol{M}_q \quad : \quad R_{q-1} \times R_q, \quad (18)$$

with $\boldsymbol{M}_q$, an invertible matrix of size $R_q \times R_q$. In [30], it is shown that, when we apply decomposition algorithms such as TT-SVD or TT-HSVD, the matrices $\boldsymbol{M}_q$ are *change-of-basis* matrices, that are mainly linked to the estimation of dominant subspaces by the SVD.

5) Model equivalence between TTD and CPD: An interesting model equivalence property between a Q-order CPD of rank-R and a train of $(Q-2)$ 3-order CPD(s) of rank R and two matrices is briefly discussed in [28]. We can summarize this fundamental algebraic relation according to the following result.

Theorem 1. *If the tensor $\boldsymbol{\mathcal{X}}$ follows a Q-order CPD of rank-R with factors $\boldsymbol{P}_q$, then a TTD can be given by:*

$$\boldsymbol{G}_1 = \boldsymbol{P}_1, \quad (19)$$

$$\boldsymbol{\mathcal{G}}_q = \boldsymbol{\mathcal{I}}_{3,R} \overset{2}{\bullet} \boldsymbol{P}_q \quad \textit{(3-order CPD), where } 2 \leq q \leq Q-1, \quad (20)$$

$$\boldsymbol{G}_Q = \boldsymbol{P}_Q^\mathsf{T}, \quad (21)$$

and the TT-ranks are all equal to the canonical rank R.

Proof. The TTD of the Q-order identity tensor $\boldsymbol{\mathcal{I}}_{Q,R}$ is given by

$$\boldsymbol{\mathcal{I}}_{Q,R} = \boldsymbol{I}_R \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{I}}_{3,R} \overset{1}{\underset{3}{\bullet}} \cdots \overset{1}{\underset{Q-1}{\bullet}} \boldsymbol{\mathcal{I}}_{3,R} \overset{1}{\underset{Q}{\bullet}} \boldsymbol{I}_R. \quad (22)$$

Substituting the above TTD into the CPD, we get

$$\boldsymbol{\mathcal{X}} = (\boldsymbol{I}_R \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{I}}_{3,R} \overset{1}{\underset{3}{\bullet}} \cdots \overset{1}{\underset{Q}{\bullet}} \boldsymbol{I}_R) \overset{\bullet}{\underset{1}{}} \boldsymbol{P}_1 \ldots \overset{\bullet}{\underset{Q}{}} \boldsymbol{P}_Q \quad (23)$$

$$(24)$$

[2]A matrix $\boldsymbol{M}$ has Kruskal rank R if *any* subset of R columns forms a full rank matrix. Remember that $\boldsymbol{M}$ has rank R if there exists *at least one* subset of R columns forming a full rank matrix.

by expressing the entries of $\boldsymbol{\mathcal{X}}$ and reorganizing them, we can equivalently write

$$\boldsymbol{\mathcal{X}} = (\boldsymbol{P}_1 \boldsymbol{I}_R) \underset{2}{\overset{1}{\bullet}} (\boldsymbol{\mathcal{I}}_{3,R} \underset{2}{\bullet} \boldsymbol{P}_2) \underset{3}{\overset{1}{\bullet}} \cdots \underset{Q}{\overset{1}{\bullet}} (\boldsymbol{I}_R \boldsymbol{P}_Q^{\mathsf{T}}). \tag{25}$$

By identifying the TT-cores in (25) with those in (14), we can deduce the relations (19), (20) and (21). Then, it is straightforward to conclude that the TT-ranks are all identical and equal to the canonical rank R. □

Thus, conditionally to the knowledge of the TT-cores, it is theoretically possible to recover the factors of the CPD by a one-to-one methodology. In Section III-B4, we present the so called JIRAFE framework, used for that aim.

6) Structured Tensor-Train models: In this subsection, we present a particular class of TT models, the so-called structured tensor-train (STT) models. These models are composed of a train of third-order tensors, each tensor being represented by means of a tensor decomposition like CP, Tucker or generalized Tucker.

For a Q-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times I_2 \cdots \times I_Q}$, an STT model can be written as:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{T}}^{(1)} \underset{3}{\overset{1}{\bullet}} \boldsymbol{\mathcal{T}}^{(2)} \underset{4}{\overset{1}{\bullet}} \cdots \underset{Q-1}{\overset{1}{\bullet}} \boldsymbol{\mathcal{T}}^{(Q-2)}, \tag{26}$$

with $\boldsymbol{\mathcal{T}}^{(1)} \in \mathbb{C}^{I_1 \times I_2 \times R_2}$, $\boldsymbol{\mathcal{T}}^{(Q-2)} \in \mathbb{C}^{R_{2Q-6} \times I_{Q-1} \times I_Q}$, and $\boldsymbol{\mathcal{T}}^{(q)} \in \mathbb{C}^{R_{2q-2} \times I_{q+1} \times R_{2q}}$ for $2 \leq q \leq Q-3$.

For instance, in the case of a fifth-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times I_2 \cdots \times I_5}$, let us assume that $\boldsymbol{\mathcal{T}}^{(q)}$ satisfies a Tucker-(1,3) decomposition for $q \in \{1, 2\}$ while $\boldsymbol{\mathcal{T}}^{(3)}$ satisfies a Tucker-(2,3) decomposition such that [11]:

$$\boldsymbol{\mathcal{T}}^{(1)} = [\![\boldsymbol{A}^{(1)}, \boldsymbol{I}_{I_2}, \boldsymbol{I}_{R_2}; \boldsymbol{\mathcal{G}}^{(1)}]\!] \in \mathbb{C}^{I_1 \times I_2 \times R_2} \tag{27}$$

$$\boldsymbol{\mathcal{T}}^{(2)} = [\![\boldsymbol{A}^{(2)}, \boldsymbol{I}_{I_3}, \boldsymbol{I}_{R_4}; \boldsymbol{\mathcal{G}}^{(2)}]\!] \in \mathbb{C}^{R_2 \times I_3 \times R_4} \tag{28}$$

$$\boldsymbol{\mathcal{T}}^{(3)} = [\![\boldsymbol{A}^{(3)}, \boldsymbol{I}_{I_4}, \boldsymbol{A}^{(4)}; \boldsymbol{\mathcal{G}}^{(3)}]\!] \in \mathbb{C}^{R_4 \times I_4 \times I_5} \tag{29}$$

with the core tensors $\boldsymbol{\mathcal{G}}^{(1)} \in \mathbb{C}^{R_1 \times I_2 \times R_2}$, $\boldsymbol{\mathcal{G}}^{(2)} \in \mathbb{C}^{R_3 \times I_3 \times R_4}$, $\boldsymbol{\mathcal{G}}^{(3)} \in \mathbb{C}^{R_5 \times I_4 \times R_6}$, and the matrix factors $\boldsymbol{A}^{(1)} \in \mathbb{C}^{I_1 \times R_1}$, $\boldsymbol{A}^{(2)} \in \mathbb{C}^{R_2 \times R_3}$, $\boldsymbol{A}^{(3)} \in \mathbb{C}^{R_4 \times R_5}$, $\boldsymbol{A}^{(4)} \in \mathbb{C}^{I_5 \times R_6}$. We then obtain a structured Tucker train of order five, abbreviated as STuT(5), which can be written by means of the following scalar equation:

$$\begin{aligned} x_{i_1,i_2,i_3,i_4,i_5} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \cdots \sum_{r_6=1}^{R_6} & a^{(1)}_{i_1,r_1} g^{(1)}_{r_1,i_2,r_2} a^{(2)}_{r_2,r_3} g^{(2)}_{r_3,i_3,r_4} \\ & a^{(3)}_{r_4,r_5} g^{(3)}_{r_5,i_4,r_6} a^{(4)}_{i_5,r_6}, \end{aligned} \tag{30}$$

or compactly as:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{T}}^{(1)} \underset{3}{\overset{1}{\bullet}} \boldsymbol{\mathcal{T}}^{(2)} \underset{4}{\overset{1}{\bullet}} \boldsymbol{\mathcal{T}}^{(3)}. \tag{31}$$

Comparing eqs. (30)-(31) of the STuT(5) model with eq. (14) of the TT model with $Q = 5$, we can conclude that the STuT model corresponds to a TT model for which each tensor of the train satisfies a Tucker-(1,3) or Tucker-(2,3) decomposition.

Such an STuT model was derived for the fourth-order tensor of received signals in a cooperative wireless MIMO communication system [31], and then generalized to a Q-order tensor in the case of multi-hop MIMO relay systems [32]. This model can also be interpreted as a nested Tucker train (NTT) model, with two successive tensors sharing a common factor.

Thus, in the case of a fourth-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times I_2 \times I_3 \times I_4}$, as illustrated by means of Fig. 2, with $\boldsymbol{A}^{(1)} \in \mathbb{C}^{I_1 \times R_1}$, $\boldsymbol{\mathcal{G}}^{(1)} \in \mathbb{C}^{R_1 \times I_2 \times R_2}$, $\boldsymbol{A}^{(2)} \in \mathbb{C}^{R_2 \times R_3}$, $\boldsymbol{\mathcal{G}}^{(2)} \in \mathbb{C}^{R_3 \times I_3 \times R_4}$, $\boldsymbol{A}^{(3)} \in \mathbb{C}^{I_4 \times R_4}$, the NTT(4) model can be described by means of the following scalar equation:

$$x_{i_1,i_2,i_3,i_4} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} \sum_{r_4=1}^{R_4} a^{(1)}_{i_1,r_1} g^{(1)}_{r_1,i_2,r_2} a^{(2)}_{r_2,r_3} g^{(2)}_{r_3,i_3,r_4} a^{(3)}_{i_4,r_4}. \tag{32}$$

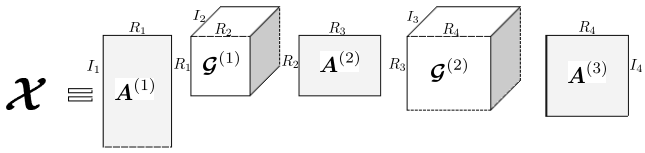

Fig. 2. NTT(4) model for a 4-order tensor

Let us define the following third-order Tucker models:

$$\begin{aligned} \boldsymbol{\mathcal{T}}^{(1)} &= [\![\boldsymbol{A}^{(1)}, \boldsymbol{I}_{I_2}, \boldsymbol{I}_{R_2}; \boldsymbol{\mathcal{G}}^{(1)}]\!] && (33) \\ &= \boldsymbol{\mathcal{G}}^{(1)} \underset{1}{\bullet} \boldsymbol{A}^{(1)} \in \mathbb{C}^{I_1 \times I_2 \times R_2} && (34) \\ \boldsymbol{\mathcal{T}}^{(2)} &= [\![\boldsymbol{A}^{(2)}, \boldsymbol{I}_{I_3}, \boldsymbol{A}^{(3)}; \boldsymbol{\mathcal{G}}^{(2)}]\!] && (35) \\ &= \boldsymbol{\mathcal{G}}^{(2)} \underset{1}{\bullet} \boldsymbol{A}^{(2)} \underset{3}{\bullet} \boldsymbol{A}^{(3)} \in \mathbb{C}^{R_2 \times I_3 \times I_4} && (36) \\ \boldsymbol{\mathcal{T}}^{(3)} &= [\![\boldsymbol{A}^{(1)}, \boldsymbol{I}_{I_2}, \boldsymbol{A}^{(2)}; \boldsymbol{\mathcal{G}}^{(1)}]\!] && (37) \\ &= \boldsymbol{\mathcal{G}}^{(1)} \underset{1}{\bullet} \boldsymbol{A}^{(1)} \underset{3}{\bullet} \boldsymbol{A}^{(2)\mathsf{T}} \in \mathbb{C}^{I_1 \times I_2 \times R_3} && (38) \\ \boldsymbol{\mathcal{T}}^{(4)} &= [\![\boldsymbol{I}_{R_3}, \boldsymbol{I}_{I_3}, \boldsymbol{A}^{(3)}; \boldsymbol{\mathcal{G}}^{(2)}]\!] && (39) \\ &= \boldsymbol{\mathcal{G}}^{(2)} \underset{3}{\bullet} \boldsymbol{A}^{(3)} \in \mathbb{C}^{R_3 \times I_3 \times I_4}. && (40) \end{aligned}$$

These Tucker models can be written in an element-wise form as:

$$t^{(1)}_{i_1,i_2,r_2} = \sum_{r_1=1}^{R_1} g^{(1)}_{r_1,i_2,r_2} a^{(1)}_{i_1,r_1} \tag{41}$$

$$t^{(2)}_{r_2,i_3,i_4} = \sum_{r_3=1}^{R_3} \sum_{r_4=1}^{R_4} g^{(2)}_{r_3,i_3,r_4} a^{(2)}_{r_2,r_3} a^{(3)}_{i_4,r_4} \tag{42}$$

$$t^{(3)}_{i_1,i_2,r_3} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} g^{(1)}_{r_1,i_2,r_2} a^{(1)}_{i_1,r_1} a^{(2)}_{r_2,r_3} \tag{43}$$

$$t^{(4)}_{r_3,i_3,i_4} = \sum_{r_4=1}^{R_4} g^{(2)}_{r_3,i_3,r_4} a^{(3)}_{i_4,r_4}. \tag{44}$$

The NTT(4) model (32) can be viewed as the nesting of the third-order Tucker models $\boldsymbol{\mathcal{T}}^{(3)}$ and $\boldsymbol{\mathcal{T}}^{(2)}$ which share the matrix factor $\boldsymbol{A}^{(2)}$. It can also be interpreted as a contraction of a Tucker-(1,3) model with a Tucker-(2,3) model, along their common mode:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{T}}^{(1)} \underset{3}{\overset{1}{\bullet}} \boldsymbol{\mathcal{T}}^{(2)} = \boldsymbol{\mathcal{T}}^{(3)} \underset{3}{\overset{1}{\bullet}} \boldsymbol{\mathcal{T}}^{(4)}. \tag{45}$$

These contraction operations correspond to summing the entries of the third-order tensors $\boldsymbol{\mathcal{T}}^{(1)}$ and $\boldsymbol{\mathcal{T}}^{(2)}$, or $\boldsymbol{\mathcal{T}}^{(3)}$ and $\boldsymbol{\mathcal{T}}^{(4)}$, along their common mode r_2, or r_3, as follows:

$$x_{i_1,i_2,i_3,i_4} = \sum_{r_2=1}^{R_2} t^{(1)}_{i_1,i_2,r_2} t^{(2)}_{r_2,i_3,i_4} \tag{46}$$

$$= \sum_{r_3=1}^{R_3} t^{(3)}_{i_1,i_2,r_3} t^{(4)}_{r_3,i_3,i_4}. \tag{47}$$

Defining the fourth-order tensor $\boldsymbol{\mathcal{C}} \in \mathbb{C}^{R_1 \times I_2 \times I_3 \times R_4}$ such that:

$$c_{r_1,i_2,i_3,r_4} = \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} g^{(1)}_{r_1,i_2,r_2} a^{(2)}_{r_2,r_3} g^{(2)}_{r_3,i_3,r_4}, \tag{48}$$

the NTT(4) model (32) can also be interpreted as the following Tucker-(2,4) model [33]:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{C}} \underset{1}{\bullet} \boldsymbol{A}^{(1)} \underset{4}{\bullet} \boldsymbol{A}^{(3)}, \tag{49}$$

or in scalar form:

$$x_{i_1,i_2,i_3,i_4} = \sum_{r_1=1}^{R_1} \sum_{r_4=1}^{R_4} c_{r_1,i_2,i_3,r_4} a^{(1)}_{i_1,r_1} a^{(3)}_{i_4,r_4}. \tag{50}$$

The nested CP model introduced in [34], and exploited in [35] in the context of MIMO relay systems, is a special case of the nested Tucker model (32), defined as follows:

$$x_{i_1,i_2,i_3,i_4} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} a^{(1)}_{i_1,r_1} g^{(1)}_{i_2,r_1} a^{(2)}_{r_1,r_2} g^{(2)}_{i_3,r_2} a^{(3)}_{i_4,r_2}, \tag{51}$$

with $\boldsymbol{A}^{(1)} \in \mathbb{C}^{I_1 \times R_1}, \boldsymbol{G}^{(1)} \in \mathbb{C}^{I_2 \times R_1}, \boldsymbol{A}^{(2)} \in \mathbb{C}^{R_1 \times R_2}, \boldsymbol{G}^{(2)} \in \mathbb{C}^{I_3 \times R_2}, \boldsymbol{A}^{(3)} \in \mathbb{C}^{I_4 \times R_2}$. The nested CP model is illustrated in Fig. 3 for a fourth-order tensor.

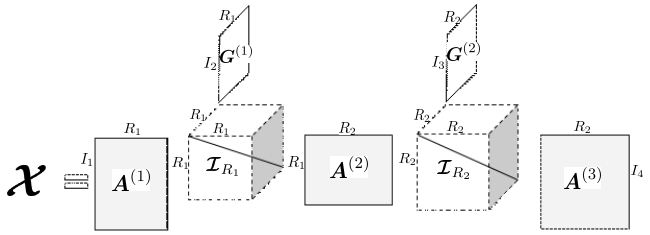

Fig. 3. Nested CP model for a 4-order tensor

This nested CP model can be deduced from the nested Tucker model (32) using the following correspondences:

$$(r_1, r_2, r_3, r_4) \leftrightarrow (r_1, r_1, r_2, r_2) \tag{52}$$

$$(\boldsymbol{A}^{(1)}, \boldsymbol{\mathcal{G}}^{(1)}, \boldsymbol{A}^{(2)}, \boldsymbol{\mathcal{G}}^{(2)}, \boldsymbol{A}^{(3)}) \leftrightarrow (\boldsymbol{A}^{(1)}, \boldsymbol{G}^{(1)}, \boldsymbol{A}^{(2)}, \boldsymbol{G}^{(2)}, \boldsymbol{A}^{(3)}). \tag{53}$$

More generally, one can define a STT model (26) for which each elementary tensor of the train has a CP decomposition. Thus, for a Q-order tensor $\boldsymbol{\mathcal{X}}$, Eq. (26) then becomes a structured CP train (SCPT) such that:

$$\begin{aligned}
\boldsymbol{\mathcal{T}}^{(1)} &= [\![\boldsymbol{A}^{(1)}, \boldsymbol{G}^{(1)}, \boldsymbol{A}^{(2)^{\mathsf{T}}}; \boldsymbol{\mathcal{I}}_{R_1}]\!] \in \mathbb{C}^{I_1 \times I_2 \times R_2} \\
\boldsymbol{\mathcal{T}}^{(q)} &= [\![\boldsymbol{A}^{(q)}, \boldsymbol{G}^{(q)}, \boldsymbol{A}^{(q+1)^{\mathsf{T}}}; \boldsymbol{\mathcal{I}}_{R_q}]\!] \\
&\in \mathbb{C}^{R_{q-1} \times I_{q+1} \times R_{q+1}} \text{ for } 2 \le q \le Q-3 \\
\boldsymbol{\mathcal{T}}^{(Q-2)} &= [\![\boldsymbol{A}^{(Q-2)}, \boldsymbol{G}^{(Q-2)}, \boldsymbol{A}^{(Q-1)}; \boldsymbol{\mathcal{I}}_{R_{Q-2}}]\!] \\
&\in \mathbb{C}^{R_{Q-2} \times I_{Q-1} \times I_Q},
\end{aligned}$$

with the matrix factors $\boldsymbol{A}^{(1)} \in \mathbb{C}^{I_1 \times R_1}$, $\boldsymbol{A}^{(q)} \in \mathbb{C}^{R_{q-1} \times R_q}$, for $2 \le q \le Q-1$, $\boldsymbol{A}^{(Q-1)} \in \mathbb{C}^{I_Q \times R_{Q-2}}$, and $\boldsymbol{G}^{(q)} \in \mathbb{C}^{I_{q+1} \times R_q}$, for $1 \le q \le Q-2$. The SCPT model of order Q can also be written in the following scalar form:

$$\begin{aligned}
x_{i_1,\cdots,i_Q} = \sum_{r_1=1}^{R_1} \cdots \sum_{r_{Q-2}=1}^{R_{Q-2}} & a^{(1)}_{i_1,r_1} g^{(1)}_{i_2,r_1} a^{(2)}_{r_1,r_2} g^{(2)}_{i_3,r_2} \\
& a^{(3)}_{r_2,r_3} g^{(3)}_{i_4,r_3} \cdots g^{(Q-2)}_{i_{Q-1},r_{Q-2}} a^{(Q-1)}_{i_Q,r_{Q-2}}.
\end{aligned}$$

B. Algorithms

1) Alternating least squares algorithm: There are many algorithms to compute a CP decomposition. By far, the most used one is the alternating least squares (ALS). ALS was proposed in [1], [2], and is considered today as the "workhorse" for CPD computation. It is a simple algorithm that fixes, iteratively, all but one factor, which is then updated by solving a linear least squares problem. In Algorithm 1, we present the ALS algorithm for a 3-order tensor $\boldsymbol{\mathcal{T}} \approx [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}; \boldsymbol{\mathcal{I}}_{3,R}]\!]$. The generalization for a Q-order case is straightforward.

Algorithm 1 ALS algorithm

Input: 3-order rank-R tensor $\boldsymbol{\mathcal{T}}$, `CritStop`
Output: Estimated CPD factors: $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$

1: **Initialize** $\boldsymbol{B}$ and $\boldsymbol{C}$
2: **while** `CritStop` and maximum iterations are not reached **do**
3: $\quad \mathbf{A} = \text{unfold}_1 \boldsymbol{\mathcal{T}} \cdot (\boldsymbol{C} \odot \boldsymbol{B})^{\mathsf{T}\dagger}$
4: $\quad \mathbf{B} = \text{unfold}_2 \boldsymbol{\mathcal{T}} \cdot (\boldsymbol{C} \odot \boldsymbol{A})^{\mathsf{T}\dagger}$
5: $\quad \mathbf{C} = \text{unfold}_3 \boldsymbol{\mathcal{T}} \cdot (\boldsymbol{B} \odot \boldsymbol{A})^{\mathsf{T}\dagger}$
6: **end while**

Generally `CritStop` is based on the evaluation of the fitting error $\|\boldsymbol{\mathcal{T}} - [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}; \boldsymbol{\mathcal{I}}_{3,R}]\!]\|_F$.

2) High-order SVD algorithm: Based on the definition of the HOSVD given in (5), an algorithm to compute the HOSVD is presented in Algorithm 2. This algorithm was proposed in [10], and it relies on the computation of the left singular vectors of the unfoldings $\text{unfold}_q \boldsymbol{\mathcal{T}}$. The truncated HOSVD has a moderate computing cost, and is not necessarily optimal, but it gives a good approximation, which can be used as initialization for ALS-like algorithms. For ease of presentation, we derive the method only for a 3-order tensor, but the generalization to a Q-order tensor is straightforward.

Algorithm 2 HOSVD algorithm

Input: 3-order tensor $\boldsymbol{\mathcal{T}}$, multilinear rank $\{R_1, R_2, R_3\}$
Output: $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{W}$ and $\boldsymbol{\mathcal{G}}$

1: $\boldsymbol{U} \leftarrow R_1$ first left singular vectors of $\text{unfold}_1\boldsymbol{\mathcal{T}}$
2: $\boldsymbol{V} \leftarrow R_2$ first left singular vectors of $\text{unfold}_2\boldsymbol{\mathcal{T}}$
3: $\boldsymbol{W} \leftarrow R_3$ first left singular vectors of $\text{unfold}_3\boldsymbol{\mathcal{T}}$
4: $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{T}} \underset{1}{\bullet} \boldsymbol{U}^{\mathsf{H}} \underset{2}{\bullet} \boldsymbol{V}^{\mathsf{H}} \underset{3}{\bullet} \boldsymbol{W}^{\mathsf{H}}$

3) TTD computation with the TT-SVD algorithm: The TT-SVD algorithm has been introduced in [28]. This algorithm minimizes in a sequential way the following LS criterion:

$$\psi(\boldsymbol{\mathcal{X}}) = \left\| \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{A}}_1 \underset{2}{\overset{1}{\bullet}} \boldsymbol{\mathcal{A}}_2 \underset{3}{\overset{1}{\bullet}} \cdots \underset{Q-1}{\overset{1}{\bullet}} \boldsymbol{\mathcal{A}}_{Q-1} \underset{Q}{\overset{1}{\bullet}} \boldsymbol{\mathcal{A}}_Q \right\|_F^2. \quad (54)$$

In Fig. 4, we present in a schematic way the TT-SVD algorithm applied on a 4-order tensor $\boldsymbol{\mathcal{X}}$. As illustrated, the TT-SVD computes in a sequential way $Q-1$ SVDs on matrix reshapings of the Q-order original tensor. Note that at each step, we apply the truncated SVD on the reshaped matrices $\boldsymbol{V}_{(2)}^{(q)}$ of size $(R_q I_{q+1}) \times (I_{q+2} \cdots I_Q)$ to recover matrices $\boldsymbol{U}^{(q+1)}$ and $\boldsymbol{V}^{(q+1)}$, this latter containing the product of the pseudo-diagonal singular values matrix and the right singular vectors matrix.

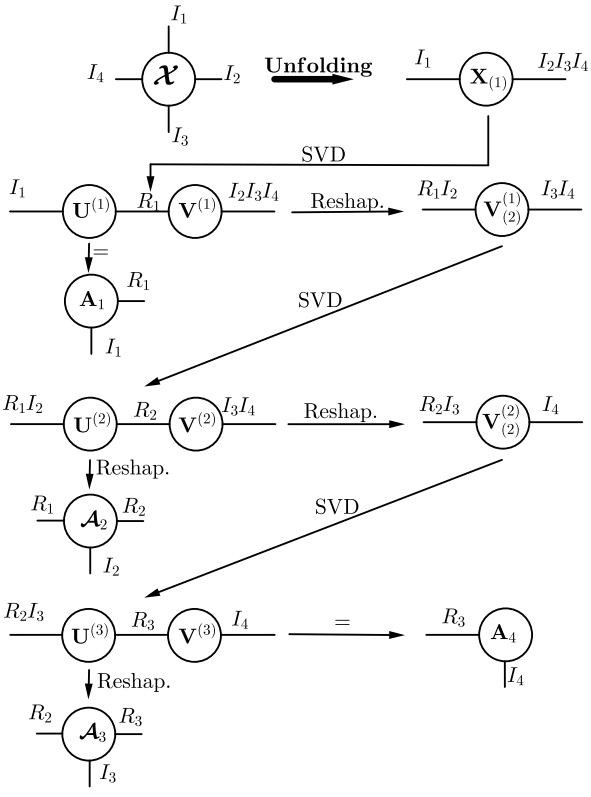

Fig. 4. TT-SVD algorithm applied on a 4-order tensor

It is worth noting that the TT-SVD in its current state cannot be parallelized, which is a problem when we deal with very high order tensors. An alternative method to compute the TTD is to consider different reshapings than those considered in the TT-SVD. Indeed, the recently proposed TT-HSVD [30] algorithm, for Tensor-Train Hierarchical SVD, is a hierarchical algorithm for TTD, that suggests to combine more than one mode in each dimension of the reshaped matrices, *i.e.*, instead of considering the matrix $\mathbf{X}_{(1)}$ of size $I_1 \times (I_2 I_3 I_4)$, we can for example use as first unfolding the matrix $\mathbf{X}_{(2)}$ of size $(I_1 I_2) \times (I_3 I_4)$. Note that using this strategy allows to parallelize the decomposition, after each SVD, across several processors, which suits very well high order tensors. Moreover, the choice of the best reshaping strategy when the order is very high is discussed in [30] in terms of the algorithmic complexity. Indeed, [30] shows that reshapings with the "most square " matrix (*i.e.*, matrices with more balanced row and column dimensions), leads to the lower computational complexity.
Remark that from an estimation point of view, the TT-SVD and the TT-HSVD algorithms suffer from the lack of uniqueness of the TT-cores described in Section III-A4. Indeed, as the latent matrices $\{\boldsymbol{M}_1, \cdots, \boldsymbol{M}_{Q-1}\}$ are unknown, the true TT-cores remain unknown. In the next section, we propose a methodology to solve this problem in the important context where the observed tensor follows a Q-order CPD of rank R.

4) JIRAFE principle for CPD: In this section, we present a TT-based methodology [36], JIRAFE, for high-order CPD factors retrieval. The acronym JIRAFE stands for "Joint dImensionality Reduction And Factors rEtrieval" and the method is described in [36], [37] by the following two-step procedure:

1) Reduce the dimensionality of the original factor retrieval problem by breaking the difficult multidimensional optimization problem into a collection of simpler optimization problems on small-order tensors. This step is performed using the TT-SVD algorithm.
2) Design a factor retrieval strategy by exploiting (or not) the coupled existing structure between the 1st and 3rd factors of two consecutive TT-cores. Here, the goal is to minimize a sum of coupled least-square (LS) criteria.

In [36], the structures of the TT-cores associated with the TT-SVD algorithm are described and a family of estimation algorithms, is proposed. This methodology is based on the following result:

Theorem 2. *If the data tensor follows a Q-order CPD of rank R parametrized by Q full column rank factors $\{\boldsymbol{P}_1, \ldots, \boldsymbol{P}_Q\}$. The TT-SVD algorithm recovers the TT-cores such that:*

$$\mathbf{A}_1 = \boldsymbol{P}_1 \boldsymbol{M}_1^{-1}, \quad (55)$$

$$\boldsymbol{\mathcal{A}}_q = \boldsymbol{\mathcal{I}}_{3,R} \underset{1}{\bullet} \boldsymbol{M}_{q-1} \underset{2}{\bullet} \boldsymbol{P}_q \underset{3}{\bullet} \boldsymbol{M}_q^{-\mathsf{T}}, \text{ where } 2 \leq q \leq Q-1 \quad (56)$$

$$\mathbf{A}_Q = \boldsymbol{M}_{Q-1} \boldsymbol{P}_Q^{\mathsf{T}}, \quad (57)$$

where $\boldsymbol{M}_q$ is a nonsingular $R \times R$ change of basis matrix.

This means that if a Q-order tensor admits a rank-R CPD, then its TTD involves a train of $(Q-2)$ 3-order CPD(s) and has all identical TT-ranks such as $R_1 = \ldots = R_{Q-1} = R$. The factors can be derived straightforwardly from the TT-cores up to two change-of-basis matrices.

Remark 1. *Note that each TT-core for $2 \leq q \leq Q-1$ follows a CPD coupled with its adjacent TT-cores (see Fig. 5).*

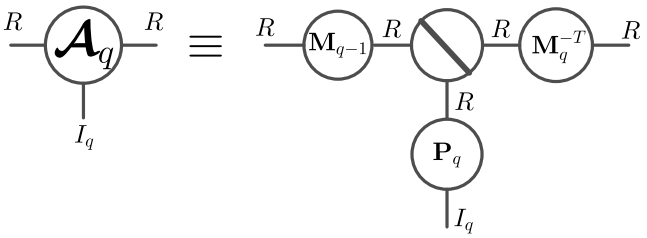

Fig. 5. 3-order CPD of the q-th TT-core

The above theorem and the remark allow us to derive the factor retrieval scheme, called JIRAFE and presented in Algorithm 3 as a pseudo code which minimizes the following criterion:

$$\min_{\mathbb{M},\mathbb{P}} \Big\{ \|\boldsymbol{A}_1 - \boldsymbol{P}_1 \boldsymbol{M}_1^{-1}\|_F + \|\boldsymbol{A}_Q - \boldsymbol{M}_{Q-1}\boldsymbol{P}_Q\|_F \tag{58}$$

$$+ \sum_{q=2}^{Q-2} \phi(\boldsymbol{\mathcal{A}}_q) \Big\}, \tag{59}$$

where $\mathbb{M} = \{\boldsymbol{M}_1, \cdots, \boldsymbol{M}_Q\}$ and $\mathbb{P} = \{\boldsymbol{P}_1, \cdots, \boldsymbol{P}_Q\}$ and

$$\phi(\boldsymbol{\mathcal{A}}_q) = \|\boldsymbol{\mathcal{A}}_q - \boldsymbol{\mathcal{I}}_{3,R} \underset{1}{\bullet} \boldsymbol{M}_{q-1} \underset{2}{\bullet} \boldsymbol{P}_q \underset{3}{\bullet} \boldsymbol{M}_q^{-\mathsf{T}}\|_F^2. \tag{60}$$

Algorithm 3 JIRAFE for CPD

Input: Q-order rank-R tensor $\boldsymbol{\mathcal{X}}$

Output: Estimated CPD factors: $\boldsymbol{P}_1, \cdots, \boldsymbol{P}_Q$.

1: Model reduction: Solve criterion (54) using TT-SVD or TT-HSVD, *i.e.*

$$\min_{\boldsymbol{A}_1, \boldsymbol{\mathcal{A}}_2, \ldots, \boldsymbol{\mathcal{A}}_{Q-1}, \boldsymbol{A}_Q} \psi(\boldsymbol{\mathcal{X}})$$

2: Factor retrieval:

$$\min_{\boldsymbol{M}_1, \boldsymbol{P}_2, \boldsymbol{M}_2} \phi(\boldsymbol{\mathcal{A}}_2).$$

3: **for** $k = 3 \cdots Q-1$ **do**

4: $\quad \min_{\boldsymbol{M}_q, \boldsymbol{P}_q} \phi(\boldsymbol{\mathcal{A}}_q | \boldsymbol{M}_{q-1})$

5: **end for**

6: $\boldsymbol{P}_1 = \boldsymbol{A}_1 \boldsymbol{M}_1$, and $\boldsymbol{P}_Q = \boldsymbol{A}_Q^{\mathsf{T}} \boldsymbol{M}_{Q-1}^{-\mathsf{T}}$

In Algorithm 3, we denote by $\phi(\boldsymbol{\mathcal{A}}|\boldsymbol{M})$ the criterion $\phi(\boldsymbol{\mathcal{A}})$ when factor $\boldsymbol{M}$ has been previously estimated. The JIRAFE approach breaks this delicate optimization problem into a set of Q low-dimensional and coupled LS criterion. Indeed, the JIRAFE methodology is essentially based on a single tridimensional LS optimization, $Q-3$ bidimensional LS optimizations and the computation of two inverse $R \times R$ matrices. It is worth noting that any processing applicable to CPDs can be done at the level of the estimated 3-order CPD TT-cores, using the JIRAFE principle, with a much lower complexity. Moreover, the ambiguities of the CPD-Train model, presented in Theorem 2, are the same as the ones of the CPD. Indeed, in [36], a proof of the following result is given:

Remark 2. *Just like the CPD, the CPD-Train model, given in Theorem 2, is characterized by the following indeterminacies:*

1) *a unique column permutation matrix denoted by* $\boldsymbol{\Pi}$*,*
2) *diagonal scaling matrices satisfying the following relation:*

$$\boldsymbol{\Lambda}_1 \boldsymbol{\Lambda}_2 \boldsymbol{\Lambda}_3 \cdots \boldsymbol{\Lambda}_{Q-1} \boldsymbol{\Lambda}_Q = \boldsymbol{I}_R \tag{61}$$

where $\boldsymbol{\Lambda}_k$ *is the scaling ambiguity for the* k*-mode factor* $\boldsymbol{P}_k$*.*

5) Least squares Kronecker factorization: Consider the following minimization problem

$$\min_{\boldsymbol{A},\boldsymbol{B}} \|\boldsymbol{X} - \boldsymbol{A} \boxtimes \boldsymbol{B}\|_F, \tag{62}$$

where $\boldsymbol{A} \in \mathbb{C}^{I_2 \times R_2}$, $\boldsymbol{B} \in \mathbb{C}^{I_1 \times R_1}$ and $\boldsymbol{X} = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{V} \in \mathbb{C}^{I_1 I_2 \times R_1 R_2}$, and $\boldsymbol{V}$ represents a zero-mean uncorrelated noise term. The solution of the problem in Eq. (62), is based on a rank-one matrix approximation (via SVD) of $\overline{\boldsymbol{X}}$ (a permuted version of $\boldsymbol{X}$, the construction of which was proposed in [38]). The problem in (62) becomes

$$\min_{\boldsymbol{a},\boldsymbol{b}} \left\|\overline{\boldsymbol{X}} - \boldsymbol{b} \otimes \boldsymbol{a}\right\|_F, \tag{63}$$

meaning to find the nearest rank-one matrix to $\overline{\boldsymbol{X}}$, where $\boldsymbol{a} = \text{vec}(\boldsymbol{A}) \in \mathbb{C}^{I_2 R_2 \times 1}$ and $\boldsymbol{b} = \text{vec}(\boldsymbol{B}) \in \mathbb{C}^{I_1 R_1 \times 1}$. In [39], the authors proposed a solution generalizing [38] to a Kronecker product involving N factor matrices. Let us consider the case $N = 3$, usually encountered in practice. The problem then becomes

$$\min_{\boldsymbol{A},\boldsymbol{B}} \|\boldsymbol{X} - \boldsymbol{A} \boxtimes \boldsymbol{B} \boxtimes \boldsymbol{C}\|_F, \tag{64}$$

where $\boldsymbol{A} \in \mathbb{C}^{I_3 \times R_3}$, $\boldsymbol{B} \in \mathbb{C}^{I_2 \times R_2}$ and $\boldsymbol{C} \in \mathbb{C}^{I_1 \times R_1}$. The problem in (64) now becomes

$$\min_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}} \left\|\overline{\boldsymbol{\mathcal{X}}} - \boldsymbol{c} \otimes \boldsymbol{b} \otimes \boldsymbol{a}\right\|_F, \tag{65}$$

where $\boldsymbol{a} = \text{vec}(\boldsymbol{A}) \in \mathbb{C}^{I_3 R_3 \times 1}$, $\boldsymbol{b} = \text{vec}(\boldsymbol{B}) \in \mathbb{C}^{I_2 R_2 \times 1}$, $\boldsymbol{c} = \text{vec}(\boldsymbol{C}) \in \mathbb{C}^{I_1 R_1 \times 1}$. We have that $\overline{\boldsymbol{x}} = \text{vec}(\overline{\boldsymbol{X}})$ and $\overline{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{T}}\{\overline{\boldsymbol{x}}\} \in \mathbb{C}^{I_1 R_1 \times I_2 R_2 \times I_3 R_3}$, where the operator $\boldsymbol{\mathcal{T}}\{\cdot\}$ maps the elements of $\overline{\boldsymbol{x}}$ into $\overline{\boldsymbol{\mathcal{X}}}$, as follows

$$\overline{\boldsymbol{x}}_{q_1 + (q_2-1)Q_1 + (q_3-1)Q_1 Q_2} \xrightarrow[\mathcal{T}\{\cdot\}]{} \overline{\mathcal{X}}_{q_1,q_2,q_3} \tag{66}$$

where $q_i = \{1, \ldots, Q_i\}$ and $Q_i = I_i R_i$, with $i = \{1, 2, 3\}$. Otherwise stated, we have $\overline{\boldsymbol{\mathcal{X}}} = \text{reshape}\,(\overline{\boldsymbol{x}}, Q_1, Q_2, Q_3)$. Hence, finding the matrix triplet $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}\}$ that solves (64) is equivalent to finding the vector triplet $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ that solves (65), i.e., the solution of a Kronecker approximation problem can be recast as the solution to a rank-one tensor approximation problem, for which effective algorithms exist in the literature (see, e.g., [40]–[42]).

6) Least squares Khatri-Rao factorization: Consider the following minimization problem

$$\min_{\boldsymbol{A},\boldsymbol{B}} \|\boldsymbol{X} - \boldsymbol{A} \odot \boldsymbol{B}\|_F, \tag{67}$$

where $\boldsymbol{A} \in \mathbb{C}^{I \times R}$, $\boldsymbol{B} \in \mathbb{C}^{J \times R}$ and $\boldsymbol{X} = \boldsymbol{A} \odot \boldsymbol{B} + \boldsymbol{V} \in \mathbb{C}^{JI \times R}$, and $\boldsymbol{V}$ represents the zero-mean uncorrelated noise term. Note that problem (67) can be rewritten as

$$\min_{\boldsymbol{a}_r,\boldsymbol{b}_r} \sum_{i=1}^{R} \|\boldsymbol{x}_r - \boldsymbol{a}_r \boxtimes \boldsymbol{b}_r\|_2, \tag{68}$$

or, equivalently,

$$\min_{\boldsymbol{a}_r,\boldsymbol{b}_r} \sum_{i=1}^{R} \left\| \bar{\boldsymbol{X}}_r - \boldsymbol{b}_r \otimes \boldsymbol{a}_r \right\|_F , \tag{69}$$

where $\bar{\boldsymbol{X}}_r \in \mathbb{C}^{J\times I}$. Defining $\boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^{\mathrm{H}}$ as the singular value decomposition (SVD) of $\boldsymbol{X}_r$, estimates for $\boldsymbol{a}_r \in \mathbb{C}^I$ and $\boldsymbol{b}_r \in \mathbb{C}^J$, $(r = 1, \ldots, R)$ can be obtained by truncating the SVD of $\boldsymbol{X}_r$ to its dominant eigenmodes, i.e., [43]

$$\hat{\boldsymbol{b}}_r = \sqrt{\sigma_r^{(1)}}\boldsymbol{u}_r^{(1)} \quad \text{and} \quad \hat{\boldsymbol{a}}_r^{(1)} = \sqrt{\sigma_r^{(1)}}\boldsymbol{v}_r^{(1)*}, \tag{70}$$

where $\boldsymbol{u}_r^{(1)} \in \mathbb{C}^J$ and $\boldsymbol{v}_r^{(1)} \in \mathbb{C}^I$ are the dominant first left and right singular vectors of $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$, respectively, while $\sigma_r^{(1)}$ denotes the largest singular value of $\bar{\boldsymbol{X}}_r$. Hence, estimates of the full matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ that minimize (67) are obtained by repeating such a rank-1 approximation problem R times in parallel, one for each column $\boldsymbol{x}_r$, $r = 1, \ldots, R$.

IV. Tensor-based DOA estimation with sensor arrays

A fundamental problem in array processing is the estimation of the DOAs for multiple sources impinging on an array of sensors. Assume that P narrow-band far-field sources are impinging on an array of L identical sensors ($P < L$, in general). The direction of arrival of a source p in a Cartesian coordinate system $OXYZ$ associated with the array is given by the unitary vector $\boldsymbol{k}_p = \begin{bmatrix}\sin\theta_p\cos\phi_p & \sin\theta_p\sin\phi_p & \cos\theta_p\end{bmatrix}^{\mathsf{T}}$, where θ and ϕ are the elevation and azimuth angles, respectively, as illustrated on Fig.6.

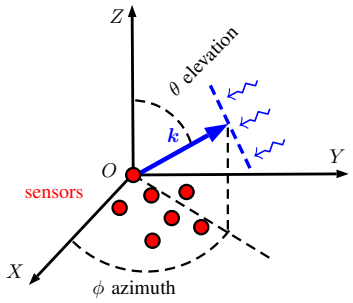

Fig. 6. DOA estimation set-up

With these notations, a snapshot of the array at time instant t is given by:

$$\boldsymbol{y}(t) = \sum_{p=1}^{P} \boldsymbol{a}(\boldsymbol{k}_p)s_p(t) + \boldsymbol{b}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}(t), \tag{71}$$

where $\boldsymbol{y}(t)$ is the $(L \times 1)$ output vector of the array, $\boldsymbol{A} = [\boldsymbol{a}(\boldsymbol{k}_1), \ldots, \boldsymbol{a}(\boldsymbol{k}_P)]$ is the $(L \times P)$ steering vectors matrix of the P sources, $\boldsymbol{x}(t) = [s_1(t), \ldots, s_P(t)]^{\mathsf{T}}$ is the $(P\times 1)$ source signals vector at time instant t, and $\boldsymbol{b}(t)$ is a $(L \times 1)$ vector that accounts for the spatially and temporally white Gaussian noise on the array sensors. If K snapshots $t_1, t_2, \ldots, t_K$ are considered, the output of the array can be expressed in the matrix form as:

$$\begin{aligned}\boldsymbol{Y} =& [\boldsymbol{y}(t_1), \ldots, \boldsymbol{y}(t_K)] \\ =& \boldsymbol{A}(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_P)[\boldsymbol{x}(t_1), \ldots, \boldsymbol{x}(t_K)] + \boldsymbol{B} \\ =& \boldsymbol{A}(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_P)\boldsymbol{S}^{\mathsf{T}} + \boldsymbol{B}\end{aligned} \tag{72}$$

with $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_P]$, a $(K \times P)$ matrix gathering on its columns the K time samples for the P sources, $\boldsymbol{s}_p = [s_p(t_1), \ldots, s_p(t_K)]^{\mathsf{T}}$, with $p = 1, \ldots, P$ and $\boldsymbol{B} = [\boldsymbol{b}(t_1), \ldots, \boldsymbol{b}(t_K)]$, a $(L \times K)$ noise matrix. The DOA estimation problem consists in finding $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_P$ from the data $\boldsymbol{Y}$.

The use of multilinear algebra to solve this problem is inspired from the principle of ESPRIT algorithm introduced by Roy *et al.* in [44], [45]. The idea is to exploit the invariances in the array output data in order to create multilinear structures. These invariances can be intrinsic to the physics of the acquired signals (*i.e.* polarization), to the acquisition setup (*i.e.,* spatially shifted subarrays), or artificially created (*i.e.* matricization / time-frequency transform of 1D signals).

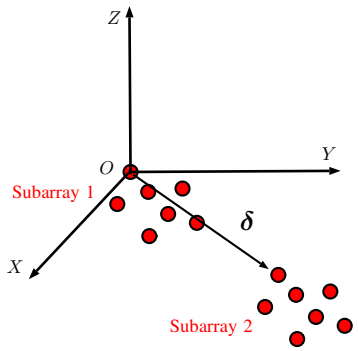

Fig. 7. Acquisition set-up for ESPRIT

The main idea of ESPRIT, when applied to DOA estimation is to employ two identical subarrays, the second one being spatially shifted compared to the first one by a known displacement vector $\boldsymbol{\delta}$, as illustrated in Fig. 7. The outputs of the first and the second subarrays, denoted by $\boldsymbol{y}_1(t)$ and $\boldsymbol{y}_2(t)$, respectively, can then be expressed as:

$$\boldsymbol{y}_1(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}_1(t), \tag{73}$$

$$\boldsymbol{y}_2(t) = \boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{x}(t) + \boldsymbol{b}_2(t), \tag{74}$$

with $\boldsymbol{\Phi}$, a diagonal matrix having on its diagonals the phase-shifts between the two arrays for the P sources:

$$\boldsymbol{\Phi} = \begin{bmatrix} e^{j\frac{2\pi}{\lambda_1}\boldsymbol{k}_1^{\mathsf{T}}\boldsymbol{\delta}} & & \\ & \ddots & \\ & & e^{j\frac{2\pi}{\lambda_P}\boldsymbol{k}_P^{\mathsf{T}}\boldsymbol{\delta}} \end{bmatrix}.$$

In the expression of $\boldsymbol{\Phi}$, λ_p denotes the pth source wavelength. Thus, the DOA estimation problem comes down at estimating the $\boldsymbol{\Phi}$ matrix. To this end, the covariance matrix $\boldsymbol{R}_{\boldsymbol{yy}}$ of the entire array:

$$\boldsymbol{y}(t) = \begin{bmatrix} \boldsymbol{y}_1(t) \\ \boldsymbol{y}_2(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{A}\boldsymbol{\Phi} \end{bmatrix} \boldsymbol{x}(t) + \begin{bmatrix} \boldsymbol{b}_1(t) \\ \boldsymbol{b}_2(t) \end{bmatrix}$$

is first computed as:

$$\boldsymbol{R}_{\boldsymbol{yy}} = E\{\boldsymbol{y}(t)\boldsymbol{y}^{\mathsf{H}}(t)\} \approx \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{y}(t_k)\boldsymbol{y}^{\mathsf{H}}(t_k).$$

Then, we use the fact that the first P eigenvectors $[\boldsymbol{E}_1^{\mathsf{T}}, \boldsymbol{E}_2^{\mathsf{T}}]^{\mathsf{T}}$ of $\boldsymbol{R}_{\boldsymbol{yy}}$ span the same subspace as the source steering vectors, *i.e,*

$$\begin{bmatrix} \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{A}\boldsymbol{\Phi} \end{bmatrix} \boldsymbol{T} = \begin{bmatrix} \boldsymbol{A}\boldsymbol{T} \\ \boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{T} \end{bmatrix}, \tag{75}$$

where $\boldsymbol{T}$ is a nonsingular matrix. As $\text{span}\{\boldsymbol{A}\} = \text{span}\{\boldsymbol{E}_1\} = \text{span}\{\boldsymbol{E}_2\}$, there exist a nonsingular matrix $\boldsymbol{\Psi}$ such that

$$\boldsymbol{E}_1\boldsymbol{\Psi} = \boldsymbol{E}_2. \tag{76}$$

From (75) and (76), it can be shown, after some algebraic manipulations (see *e.g.,* [45]), that $\boldsymbol{\Phi}$ can be directly obtained by the diagonalization of the matrix $\boldsymbol{\Psi} = \boldsymbol{E}_1^{\dagger}\boldsymbol{E}_2$. However, this *least-squares* (LS) estimate $\boldsymbol{\Psi}$ is biased because it assumes implicitly that $\boldsymbol{E}_1$ is perfectly known and the error is to be attributed only to $\boldsymbol{E}_2$. In practice, $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ are equally "noisy" as they are both estimates of $\text{span}\{\boldsymbol{A}\}$. A more appropriate criterion that takes into account noise on both $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ is the *total least-squares* (TLS) criterion. It can be formulated as finding two minimum Frobenius norm residual matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, and a matrix $\boldsymbol{\Psi}$ satisfying $(\boldsymbol{E}_1 + \boldsymbol{R}_1)\boldsymbol{\Psi} = \boldsymbol{E}_2 + \boldsymbol{R}_2$. If we denote $\boldsymbol{B} = \boldsymbol{E}_1 + \boldsymbol{R}_1$, the TLS-ESPRIT problem can be expressed as the following minimization problem:

$$\min_{\boldsymbol{\Psi},\boldsymbol{B}} \left\| \begin{bmatrix} \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \end{bmatrix} - \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{B}\boldsymbol{\Psi} \end{bmatrix} \right\|_F^2. \tag{77}$$

The solution of (77) is obtained by the eigenvalue decomposition of the $2P \times 2P$ matrix $[\boldsymbol{E}_1|\boldsymbol{E}_2]^{\mathsf{H}}[\boldsymbol{E}_1|\boldsymbol{E}_2]$.

The ESPRIT algorithm presents several advantages over the "classical" DOA estimation methods such as the beamforming or MUSIC [46], [47]. For example, it requires no calibration step and parameters estimates are obtained directly via two well- conditioned eigenproblems. Thus ESPRIT avoids the parameter grid search, inherent to MUSIC-like algorithms. However, in order to use ESPRIT one needs to estimate the data covariance matrix, which can be a difficult task in the case of highly correlated sources. Another major drawback of ESPRIT is the fact that it cannot handle more that two displaced subarrays.

A. CPD-based array processing

To overcome these drawbacks, a DOA estimation approach was introduced in [5], capable of handling N arbitrarily displaced subarray, as illustrated in Fig. 8.

It is also the first time that a tensor-based algorithm is proposed for the DOA estimation problem. If $\boldsymbol{y}_1(t), \ldots, \boldsymbol{y}_N(t)$ are the outputs of the N subarrays and using notations similar to those used for ESPRIT, we have:

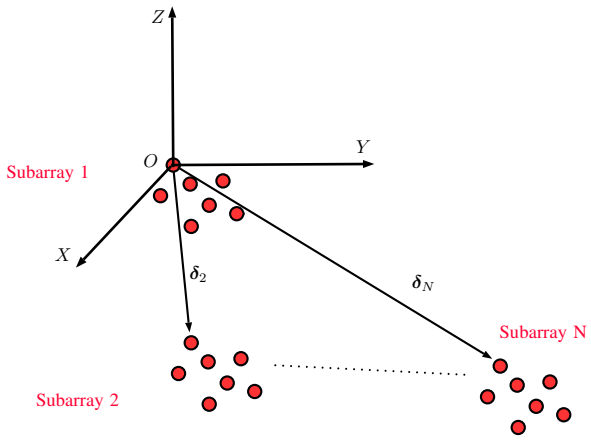

Fig. 8. Acquisition set-up for the CP approach

$$\boldsymbol{y}_1(t) = \sum_{p=1}^{P}\boldsymbol{a}(\boldsymbol{k}_p)s_p(t) + \boldsymbol{b}_1(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}_1(t),$$

$$\boldsymbol{y}_2(t) = \sum_{p=1}^{P}\boldsymbol{a}(\boldsymbol{k}_p)\mathrm{e}^{j\frac{2\pi}{\lambda_p}\boldsymbol{k}_p^{\mathsf{T}}\boldsymbol{\delta}_2}s_p(t) + \boldsymbol{b}_2(t) = \boldsymbol{A}\boldsymbol{\Phi}_2\boldsymbol{x}(t) + \boldsymbol{b}_2(t),$$

$$\vdots$$

$$\boldsymbol{y}_N(t) = \sum_{p=1}^{P}\boldsymbol{a}(\boldsymbol{k}_p)\mathrm{e}^{j\frac{2\pi}{\lambda_p}\boldsymbol{k}_p^{\mathsf{T}}\boldsymbol{\delta}_N}s_p(t) + \boldsymbol{b}_2(t) = \boldsymbol{A}\boldsymbol{\Phi}_N\boldsymbol{x}(t) + \boldsymbol{b}_N(t),$$

where,

$$\boldsymbol{\Phi}_n = \begin{bmatrix} \mathrm{e}^{j\frac{2\pi}{\lambda_1}\boldsymbol{k}_1{}^{\mathsf{T}}\boldsymbol{\delta}_n} & & \\ & \ddots & \\ & & \mathrm{e}^{j\frac{2\pi}{\lambda_P}\boldsymbol{k}_P{}^{\mathsf{T}}\boldsymbol{\delta}_n} \end{bmatrix}, \quad n = 2, \ldots, N.$$

If K snapshots are considered, the output of the entire array can be expressed as the $(NL \times K)$ matrix $\boldsymbol{Y}$:

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_1\boldsymbol{\Phi}_2 \\ \vdots \\ \boldsymbol{A}_1\boldsymbol{\Phi}_N \end{bmatrix} \boldsymbol{S}^{\mathsf{T}} + \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \\ \vdots \\ \boldsymbol{B}_N \end{bmatrix} = (\boldsymbol{A} \odot \boldsymbol{D})\boldsymbol{S}^{\mathsf{T}} + \boldsymbol{B}, \tag{78}$$

where $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_P]$ is a $(K \times P)$ source matrix, $\boldsymbol{A} = [\boldsymbol{a}(\boldsymbol{k}_1), \ldots, \boldsymbol{a}(\boldsymbol{k}_P)]$ is the $(L \times P)$ steering vector matrix and $\boldsymbol{D} = [\boldsymbol{d}(\boldsymbol{k}_1), \ldots, \boldsymbol{d}(\boldsymbol{k}_P)] = \begin{bmatrix} \text{diag}\{\boldsymbol{I}_P\}^{\mathsf{T}} \\ \text{diag}\{\boldsymbol{\Phi}_2\}^{\mathsf{T}} \\ \vdots \\ \text{diag}\{\boldsymbol{\Phi}_N\}^{\mathsf{T}} \end{bmatrix}$ is a $(N \times P)$ matrix regrouping on its rows the diagonal elements of $\boldsymbol{\Phi}_n$, $\boldsymbol{B}$ is a noise matrix. In the light of the concepts introduced in Sections III-A2 and III-A3, equation (78) expresses a noisy CP decomposition of the sensor array output. Using the tensor

formalism, equation (78) can be expressed as a $(L \times K \times N)$ tensor $\boldsymbol{\mathcal{Y}}$:

$$\boldsymbol{\mathcal{Y}} = \sum_{p=1}^{P} \mathbf{a}_p \otimes \boldsymbol{s}_p \otimes \boldsymbol{d}_p + \boldsymbol{\mathcal{B}} = [\![\boldsymbol{A}, \boldsymbol{S}, \boldsymbol{D}; \boldsymbol{\mathcal{I}}_{3,P}]\!] + \boldsymbol{\mathcal{B}}. \quad (79)$$

The tensor form (79) clearly highlights the three *diversities / invariances* of the array output: *sensors - time - subarrays*. To estimate the loading matrices $\boldsymbol{A}, \boldsymbol{S}, \boldsymbol{D}$, and consequently the source DOAs, an approximate CP decomposition of the data tensor $\boldsymbol{\mathcal{Y}}$ must be performed, using one of the many existing algorithms. Unlike the previous presented methods, the CP approach applies directly to the "raw" data and does not require the estimation of the second order statistics of the array output.

In [25], the ill-posedness of the tensor approximation problem induced by (79) is solved thanks to a constraint imposing that sources must be different, in the sense that either they should stem from different directions (e.g. if they are fully correlated like multipaths), or they should not be fully correlated, or both.

In [48], an approach that generalizes the results given in [5] to an array that presents an arbitrary number of spatial invariances was proposed. Consider a subarray composed of L_1 isotropic identical sensors indexed by $l_1 = 1, \ldots, L_1$. Consider then, L_2 identical replicas of this subarray, spatially translated to arbitrary locations. The L_2 distinct copies of the original subarray, indexed by $l_2 = 1, \ldots, L_2$, can now be seen as *subarrays* that together constitute a larger (*higher-level*) array. This proposed array structure can be further generalized by considering an additional level, composed of L_3 translated replicas of the previous sensor subarrays, indexed by $l_3 = 1, \ldots, L_3$. Let us generalize this scheme to a total of N such hierarchical levels, with the "highest" level consisting of L_N subarrays indexed by $l_N = 1, \ldots, L_N$. It is worth noting that two different subarrays at a given level n need not be disjoint, *i.e.* they may have in common subarrays/sensors of the previous level $(n-1)$. However, if all subarrays at each level are disjoint, then the entire array will contain a total number of $L = L_1 L_2 \ldots L_N$ identical sensors. Fig. 9 illustrates a three-level array of co-planar sensors.

Consider first a narrow-band plane wave impinging on the array as described in Section **??** Let us denote by $a_{l_1 l_2 \ldots l_N}$ its phase factor at the sensor indexed by $l_1, l_2, \ldots, l_N$ at the N various levels of the array. Define $\boldsymbol{d}_{l_n}^{(n)} = [x_{l_n}^{(n)} \; y_{l_n}^{(n)} \; z_{l_n}^{(n)}]^{\mathsf{T}}$, with $n = 1, \ldots, N$. With the notation introduced above, the spatial phase factor is given by:

$$a_{l_1 l_2 \ldots l_N}(\boldsymbol{k}) = \exp\{ j \frac{2\pi}{\lambda} \sum_{n=1}^{N} \boldsymbol{k}^{\mathsf{T}} \boldsymbol{d}_{l_n}^{(n)} \} = \prod_{n=1}^{N} \exp\{ j \frac{2\pi}{\lambda} \boldsymbol{k}^{\mathsf{T}} \boldsymbol{d}_{l_n}^{(n)} \}. \quad (80)$$

Thus, the array manifold for the entire sensor-array is

$$\boldsymbol{a}(\boldsymbol{k}) = \boldsymbol{a}_1(\boldsymbol{k}) \boxtimes \ldots \boxtimes \boldsymbol{a}_N(\boldsymbol{k}), \quad (81)$$

with

$$\boldsymbol{a}_n(\boldsymbol{k}) = \begin{bmatrix} e^{j(2\pi/\lambda)\boldsymbol{k}^{\mathsf{T}} \boldsymbol{d}_1^{(n)}} \\ \vdots \\ e^{j(2\pi/\lambda)\boldsymbol{k}^{\mathsf{T}} \boldsymbol{d}_{L_n}^{(n)}} \end{bmatrix} \quad (82)$$

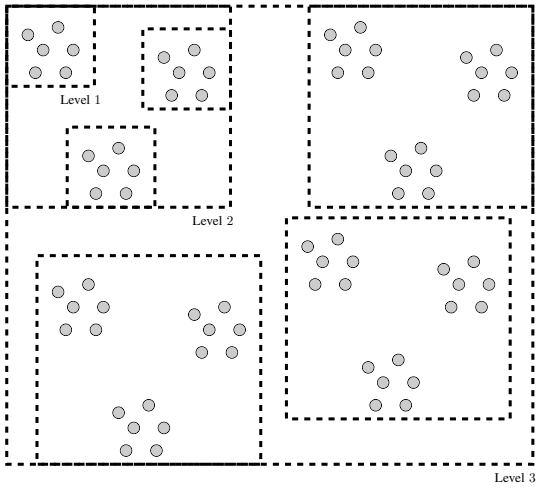

Fig. 9. A multi-scale planar array with three hierarchical levels.

an $L_n \times 1$ vector, $\forall n = 1, \ldots, N$.

Next, consider P narrow-band plane-waves, having traveled through a nonconductive homogeneous isotropic medium, impinging upon the array from directions $\boldsymbol{k}_p$, with $p = 1, \ldots, P$. Denote by $s_p(t)$ the time signal emitted by the pth narrow-band source. Then, the output at time t of the entire sensor-array can be expressed as an $L \times 1$ vector,

$$\boldsymbol{z}(t) = \sum_{p=1}^{P} \big(\boldsymbol{a}_1(\boldsymbol{k}_p) \boxtimes \ldots \boxtimes \boldsymbol{a}_N(\boldsymbol{k}_p) \big) s_p(t) + \boldsymbol{n}(t), \quad (83)$$

where $\boldsymbol{n}(t)$ is a complex-valued zero-mean additive white noise. Let us assume that we have at our disposal K snapshots at time instants, $t_1, t_2, \ldots, t_K$. Define the following $L_n \times P$ matrices:

$$\begin{aligned} \boldsymbol{A}_1 &= \big[\boldsymbol{a}_1(\boldsymbol{k}_1), \ldots, \boldsymbol{a}_1(\boldsymbol{k}_P) \big] \quad (84) \\ &\vdots \\ \boldsymbol{A}_N &= \big[\boldsymbol{a}_N(\boldsymbol{k}_1), \ldots, \boldsymbol{a}_N(\boldsymbol{k}_P) \big] , \quad (85) \end{aligned}$$

and the $K \times P$ matrix:

$$\boldsymbol{S} = \begin{bmatrix} s_1(t_1) & s_2(t_1) & \ldots & s_P(t_1) \\ s_1(t_2) & s_2(t_2) & \ldots & s_P(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ s_1(t_K) & s_2(t_K) & \ldots & s_P(t_K) \end{bmatrix} = \big[\boldsymbol{s}_1, \; \boldsymbol{s}_2, \ldots, \boldsymbol{s}_P \big] . \quad (86)$$

The collection of K snapshots of the array can then be organized into an $L \times K$ data matrix as

$$\boldsymbol{Z} = [\boldsymbol{z}(t_1), \ldots, \boldsymbol{z}(t_K)] = \big(\boldsymbol{A}_1 \odot \cdots \odot \boldsymbol{A}_N \big) \boldsymbol{S}^{\mathsf{T}} + \boldsymbol{N}, \quad (87)$$

where $\boldsymbol{N}$ is a $(L \times K)$ complex-valued matrix, modeling the noise on the entire array for all K temporal snapshots. Equation (87) reveals a $(N+1)$-dimensional CPD structure. In the case where only one time-sample is available, *i.e.* matrix $\boldsymbol{S}$ is a $1 \times P$ vector, the data model given by (87) becomes

$$\boldsymbol{z} = \big(\boldsymbol{A}_1 \odot \cdots \odot \boldsymbol{A}_N \big) \boldsymbol{s} + \boldsymbol{n}, \quad (88)$$

with $\boldsymbol{z} = \boldsymbol{z}(t_1)$, $\boldsymbol{s} = \boldsymbol{s}(t_1) = \big(\boldsymbol{S}(1,:)\big)^{\mathsf{T}}$ and $\boldsymbol{n} = \boldsymbol{N}(:,1)$. In the definitions above, we used the Matlab notations for columns and rows selection operators. Equation (88) is a vectorized representation of a N-dimensional CP data model. It is worth noting that if only one snapshot of the array is available, the $N+1$ CP model degenerates into an N-dimensional one. A discussion on the identifiability of these CPD models, for different scenarios can be found in [48].

For this multi-scale invariance array, a two stages DOA estimation procedure was also proposed in [48]. The first stage estimates the N steering vectors $\boldsymbol{a}_n(\boldsymbol{k}_p)$ $(n = 1, \ldots, N)$ for each of the P sources $(p = 1, \ldots, P)$, by exploiting the CPD structure (87) of the collected data. The second stage estimates the sources' direction-cosines $\boldsymbol{k}_p$, $p = 1, \ldots, P$ from the steering vectors obtained at the previous stage. To this end, the DOA estimation can be formulated as an optimization problem, and a sequential procedure that exploits all the available information from the source's steering vectors encompassing all scale levels, is adopted.

Define the following cost functions:

$$\mathcal{J}_n(\boldsymbol{k}_p) = \|\hat{\boldsymbol{a}}_n^p - \boldsymbol{a}_n(\boldsymbol{k}_p)\|_2^2, \text{ with } n = 1, \ldots, N, \qquad (89)$$

where $\hat{\boldsymbol{a}}_n^p$ denotes the estimated steering vector at the nth level for the pth source. Estimating the DOAs for the pth source comes down to minimising the following criterion:

$$\mathcal{I}_N(\boldsymbol{k}_p) = \sum_{n=1}^{N} \mathcal{J}_n(\boldsymbol{k}_p). \qquad (90)$$

This function is non-convex and highly non-linear with respect to the direction-cosines; hence, a direct local optimization procedure would fail in most cases. In [48], a sequential strategy was adopted to minimize $\mathcal{I}_N(\boldsymbol{k}_p)$, progressing from one level to the next higher level, using an iterative refinement of the direction-cosine estimates within each level. The method is based on the fact that, when noise-free, the N cost-functions in (89) have the same global minimum.

Assume that the level-1 subarrays' inter-sensor separations do not exceed half a wavelength. This assumption is essential to obtaining a set of high-variance but unambiguous direction-cosine estimates. On the contrary, to achieve a practical advantage, it is important that the spatial displacement between any two subarrays of the highest level exceeds $\lambda/2$, where λ is the wavelength. This will produce estimates of lower variance but with cyclic ambiguity for the same set of direction-cosines. On the other hand, under the first assumption, the $\mathcal{J}_1(\boldsymbol{k}_p)$ function is unimodal on the support region of the DOAs. Therefore, any local optimization procedure should converge towards the global minimum for the criterion. Thus, we obtain another set of estimates, now of high-variance but with no cyclic ambiguity, for the DOAs, to be denoted by $\boldsymbol{k}_{p,1}^*$ with $p = 1, \ldots, P$. These estimates will subsequently be used, in a second step, as the initial point for the minimization of $\mathcal{I}_2(\boldsymbol{k}_p) = \mathcal{J}_1(\boldsymbol{k}_p) + \mathcal{J}_2(\boldsymbol{k}_p)$. As no assumption is made on the distances between the level-2 subarrays, $\mathcal{I}_2(\boldsymbol{k}_p)$ may present more than one local minimum. Hence, a good initial estimate is crucial for the optimization procedure. The estimates obtained by the minimization of $\mathcal{I}_2(\boldsymbol{k}_p)$, denoted by $\boldsymbol{k}_{p,2}^*$, are then used for the minimization of $\mathcal{I}_3(\boldsymbol{k}_p) = \sum_{n=1}^{3} \mathcal{J}_n(\boldsymbol{k}_p)$, and so on, until the final estimates are obtained by the minimization of $\mathcal{I}_N(\boldsymbol{k}_p)$. The proposed algorithm can be summarized as follows:

First Stage:

- Estimate $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N$ by CP decomposition of the data $\boldsymbol{\mathcal{Z}}$ or $\boldsymbol{z}$ (see eq. (87) or (88)).

Second Stage:

- For $p = 1, \ldots, P$ and for $n = 1, \ldots, N$, compute

$$\boldsymbol{k}_{p,n}^* = \underset{\boldsymbol{k}_p}{\operatorname{argmin}}\, \mathcal{I}_n(\boldsymbol{k}_p) = \underset{\boldsymbol{k}_p}{\operatorname{argmin}} \sum_{i=1}^{n} \mathcal{J}_i(\boldsymbol{k}_p). \qquad (91)$$

- *Output:* The estimated parameters for the P sources: $\hat{\boldsymbol{k}}_p = (\hat{u}_p, \hat{v}_p, \hat{w}_p) = \boldsymbol{k}_{p,N}^*$ with $p = 1, \ldots, P$.

Several CPD-based DOA estimation approaches, based on other diversity schemes have been proposed in the literature. In [49], polarization of electromagnetic waves was exploited as an additional diversity, by using vector-sensor arrays capable of capturing the six components of the electromagnetic field. A coupled CPD approach for DOA estimation, using the multiple baselines in sparse arrays was introduced in [50]. High-order statistics (cumulants) were also used to create diversity in CPD based direction finding algorithms [51].

While CPD remains the most commonly used decomposition in tensor-based array processing, other tensor decompositions made their way in. For example, the HOSVD has been used to develop multidimensional versions of the popular ESPRIT and MUSIC algorithms (see *e.g.,* [52], [53]).

V. Tensor-based multidimensional harmonic retrieval

Multidimensional Harmonic Retrieval (MHR) [6], [54] is a classical signal processing problem that has found several applications in spectroscopy [55], radar communications [56], sensor array processing [57], [58], to mention a few. The Multidimensional Harmonic (MH) model can be viewed as the tensor-based generalization of the one-dimensional harmonic one, resulting from the sampling process over a multidimensional regular grid. As a consequence, the Q-dimensional harmonic model needs the estimation of a large number (QR) of angular-frequencies of interest. We can easily note that the number of unknown parameters and the order of the associated data tensor grow with Q. Moreover, it is likely that the joint exploitation of multi-diversity/modality sensing technologies for data fusion [59]–[61] further increases the data tensor order. This trend is usually called the "curse of dimensionality" [62]–[64] and the challenge here is to reformulate a high-order tensor as a set of low-order tensors over a graph. In this context, we observe an increasing interest for the tensor network theory (see [62] and references therein). Tensor network provides a useful and rigorous graphical representation of a high-order tensor into a factor graph where the nodes are low-order tensors, called cores, and the edges encode their dependencies, *i.e.*, their common dimensions, often called "rank". In addition, tensor network allows to perform scalable/distributed computations over the cores [62]. In the tensor network framework, Hierarchical/tree Tucker [29], [65] and Tensor Train (TT) [66] are two popular representations of a high-order tensor into a graph-connected

low-order (at most 3) tensors. In this section, we focus our effort on the TT formalism for its simplicity and compactness in terms of storage cost. Unlike the hierarchical Tucker model, TT is exploited in many practical and important contexts as, for instance, tensor completion [67], blind source separation [68], and machine learning [4], to mention a few. In the context of the MHR problem, this strategy has at least two advantages. First, it is well-known that the convergence of the Alternating Least Squares (ALS) algorithm becomes more and more difficult when the order increases [69]–[71]. To deal with this problem, applying ALS on lower-order tensors is preferable. The second argument is to exploit some latent coupling properties between the cores [4], [72] to propose new efficient estimators.

The Maximum Likelihood estimator [73], [74] is the optimal choice from an estimation point of view, since it is statistically efficient, *i.e.*, its Mean Squared Error (MSE) reaches the Cramér-Rao Bound (CRB) [75] in the presence of noise. The main drawback of the maximum likelihood estimator is its prohibitive complexity cost. This limitation is particularly severe in the context of a high-order data tensor. To overcome this problem, several low-complexity methods can be found in the literature. These methods may not reach, sometimes, the CRB, but they provide a significant gain in terms of the computational cost compared to the maximum likelihood estimator. There are essentially two main families of methods. The first one is based on the factorization of the data to estimate the well-known signal/noise subspace such as the Estimation of Signal Parameters *via* Rotational Invariance Techniques (ESPRIT) [45], the ND-ESPRIT [76], the Improved Multidimensional Folding technique [77], and the CP-VDM [78]. The second one is based on the uniqueness property of the CPD. Indeed, factorizing the data tensor thanks to the ALS algorithm [79] allows an identification of the unknown parameters by Vandermonde-based rectification of the factor matrices.

A. Generalized Vandermonde Canonical Polyadic Decomposition

The multidimensional harmonic (MH) model assumes that the measurements can be modeled as the superposition of R undamped exponentials sampled on a Q-dimensional grid according to [6]

$$[\boldsymbol{\mathcal{X}}]_{i_1\ldots i_Q} = \sum_{r=1}^{R} \alpha_r \prod_{q=1}^{Q} z_{r,q}^{i_q-1}, \quad 1 \leq i_q \leq I_q \tag{92}$$

in which the r-th complex amplitude is denoted by α_r and the pole is defined by $z_{r,q} = e^{j\omega_{r,q}}$ where $\omega_{r,q}$ is the r-th angular-frequency along the q-th dimension, and we have $\mathbf{z}_q = \begin{bmatrix} z_{1,q} & z_{2,q} & \ldots & z_{R,q} \end{bmatrix}^{\mathsf{T}}$. Note that the tensor $\boldsymbol{\mathcal{X}}$ is expressed as the linear combination of M rank-1 tensors, each of size $I_1 \times \ldots \times I_Q$ (the size of the grid), and follows a generalized Vandermonde CPD [80]:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{A}} \underset{1}{\bullet} \boldsymbol{V}_1 \underset{2}{\bullet} \ldots \underset{Q}{\bullet} \boldsymbol{V}_Q \tag{93}$$

where $\boldsymbol{\mathcal{A}}$ is a $R \times \ldots \times R$ diagonal tensor with $[\boldsymbol{\mathcal{A}}]_{r,\ldots,r} = \alpha_r$ and

$$\boldsymbol{V}_q = \begin{bmatrix} \boldsymbol{v}(z_{1,q}) & \ldots & \boldsymbol{v}(z_{R,q}) \end{bmatrix}$$

is a $I_q \times R$ rank-R Vandermonde matrix, where

$$\boldsymbol{v}(z_{r,q}) = \begin{bmatrix} 1 & z_{r,q} & z_{r,q}^2 & \ldots & z_{r,q}^{I_q-1} \end{bmatrix}^{\mathsf{T}}.$$

We define a noisy MH tensor model of order Q as:

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} + \sigma\boldsymbol{\mathcal{E}}, \tag{94}$$

where $\sigma\boldsymbol{\mathcal{E}}$ is the noise tensor, σ is a positive real scalar, and each entry $[\boldsymbol{\mathcal{E}}]_{i_1\ldots i_Q}$ follows an *i.i.d.* circular Gaussian distribution $\mathcal{CN}(0,1)$, and $\boldsymbol{\mathcal{X}}$ has a canonical rank equal to R. The reader may be referred to [76] for the case of damped signals, which is not addressed in the present article.

B. Algorithms

1) Vandermonde rectification of the ALS algorithm:

a) Limit of the ALS algorithm for structured CPD: The CPD of any order-Q rank-R tensor $\boldsymbol{\mathcal{X}}$ involves the estimation of Q factors $\boldsymbol{V}_q$ of size $I_q \times R$. As pointed out above, in the context of the MD-harmonic model, the factors $\boldsymbol{V}_q$ of the CPD are Vandermonde matrices. Consider, $\boldsymbol{Y}_q$, the q-th mode unfolding [9] of tensor $\boldsymbol{\mathcal{Y}}$, at the k-th iteration with $1 \leq k \leq M$, M denoting the maximal number of iterations. The ALS algorithm solves alternatively for each of the Q dimensions the minimization problem [81], [82]: $\min_{\boldsymbol{V}_q} ||\boldsymbol{Y}_q - \boldsymbol{V}_q\boldsymbol{S}_q||^2$ where $\boldsymbol{S}_q^{\mathsf{T}} = \boldsymbol{V}_Q \odot \ldots \odot \boldsymbol{V}_{q+1} \odot \boldsymbol{V}_{q-1} \odot \ldots \odot \boldsymbol{V}_1$. It aims at approximating tensor $\boldsymbol{\mathcal{Y}}$ by a tensor of rank R, hopefully close to $\boldsymbol{\mathcal{X}}$. The LS solution conditionally to matrix $\boldsymbol{S}_q$ is given by $\boldsymbol{V}_q = \boldsymbol{Y}_q\boldsymbol{S}_q^{\dagger}$ where $\dagger$ stands for the pseudo-inverse. Now, remark that there is no reason that the above LS criterion promotes the Vandermonde structure in the estimated factors in the presence of noise. In other words, ignoring the structure in the CPD leads to estimate an excessive number of free parameters. This mismatched model dramatically decreases the estimation performance [83]. Hence there is a need to rectify the ALS algorithm to take into account the factor structure.

b) Rectified ALS (RecALS): The RecALS algorithm belongs to the family of Lift-and-Project Algorithms [84], [85]. The optional lift step computes a low rank approximation and the projection step performs a rectification toward the desired structure. The RecALS algorithm is based on iterated projections and split LS criteria. Its algorithmic description is provided in Algorithm 4 for $Q = 3$. We insist that several iterations in the while loops are necessary, since restoring the structure generally increases the rank, and computing the low-rank approximation via truncated SVD generally destroys the structure. The most intuitive way to exploit the Vandermonde structure is called column-averaging. Let $\omega = \frac{1}{i}\angle z^i$ where $\angle$ stands for the angle function. Define the sets

$$\mathbb{J} = \left\{ \boldsymbol{v} = \frac{\boldsymbol{f}}{[\boldsymbol{f}]_1} : \boldsymbol{f} \in \mathbb{C}^I \right\} \tag{95}$$

$$\mathbb{A}_\ell = \left\{ \boldsymbol{v}(z = e^{j\bar{\omega}}) : \bar{\omega} = \frac{1}{\ell}\sum_{i=1}^{\ell} \frac{1}{i}\angle[\boldsymbol{f}]_{i+1} \right\} \tag{96}$$

Algorithm 4 Rectified ALS (RecALS)

Require: $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{V}_1, \boldsymbol{V}_2, \{\mathbb{V}_1, \ldots, \mathbb{V}_K\}$, `CritStop`
Ensure: $\{z_{1,q}, \ldots, z_{R,q}\}$ for $1 \leq q \leq 3$

while `CritStop` and maximum iterations are not reached **do**
  $\boldsymbol{V}_3 = \text{unfold}_3\boldsymbol{\mathcal{Y}} \cdot \left((\boldsymbol{V}_2 \odot \boldsymbol{V}_1)^{\mathsf{T}}\right)^{\dagger}$
  for $r = 1, \ldots, R$ **do**
    $\boldsymbol{v} := [\boldsymbol{V}_3]_r$
    while (`CritStop` is false) **do**
      $\boldsymbol{v} = \pi_{\mathbb{V}_K} \ldots \pi_{\mathbb{V}_1}(\boldsymbol{v})$
    end while
    $z_{r,3} = \min_z ||\boldsymbol{v}(z) - \boldsymbol{v}||^2$
  end for
  $\boldsymbol{V}_3 := \left[\boldsymbol{v}(z_{1,3}) \quad \ldots \quad \boldsymbol{v}(z_{R,3})\right]$
  $\boldsymbol{V}_2 = \text{unfold}_2\boldsymbol{\mathcal{Y}} \cdot \left((\boldsymbol{V}_3 \odot \boldsymbol{V}_1)^{\mathsf{T}}\right)^{\dagger}$
  for $r = 1, \ldots, R$ **do**
    $\boldsymbol{v} := [\boldsymbol{V}_2]_r$
    while (`CritStop` is false) **do**
      $\boldsymbol{v} = \pi_{\mathbb{V}_K} \ldots \pi_{\mathbb{V}_1}(\boldsymbol{v})$
    end while
    $z_{r,2} = \min_z ||\boldsymbol{v}(z) - \boldsymbol{v}||^2$
  end for
  $\boldsymbol{V}_2 := \left[\boldsymbol{v}(z_{1,2}) \quad \ldots \quad \boldsymbol{v}(z_{R,2})\right]$
  $\boldsymbol{V}_1 = \text{unfold}_1\boldsymbol{\mathcal{Y}} \cdot \left((\boldsymbol{V}_3 \odot \boldsymbol{V}_2)^{\mathsf{T}}\right)^{\dagger}$
  for $r = 1, \ldots, R$ **do**
    $\boldsymbol{v} := [\boldsymbol{V}_1]_r$
    while (`CritStop` is false) **do**
      $\boldsymbol{v} = \pi_{\mathbb{V}_K} \ldots \pi_{\mathbb{V}_1}(\boldsymbol{v})$
    end while
    $z_{r,1} = \min_z ||\boldsymbol{v}(z) - \boldsymbol{v}||^2$
  end for
  $\boldsymbol{V}_1 := \left[\boldsymbol{v}(z_{1,1}) \quad \ldots \quad \boldsymbol{v}(z_{R,1})\right]$
end while

where $1 \leq \ell \leq I-2$. This method exploits the Vandermonde structure in a heuristic way. So, the rectified strategy is to consider the iterated vector $\boldsymbol{f}(h) = (\pi_{\mathbb{A}_\ell}\pi_{\mathbb{J}})^h(\boldsymbol{f})$. Another straightforward method is the periodogram maximization. Under Gaussian noise and for a single tone $\alpha_r z_{r,q}^{i_q}$, the maximum likelihood estimator (MLE) is optimal and is given by the location of the maximal peak of the Fourier-periodogram [8], [86]. To increase the precision of the estimation, it is standard to use the well-known zero-padding technique at the price of an increase in computational cost. In [87], two strategies for Vandermonde-based rectification of the CP factors are provides but the RecALS principle is generic and flexible in the sense that any Vandermonde-based rectification methods can be used.

C. VTT-RecALS

In this section, the JIRAFE algorithm and the RecALS method are associated to solve the multidimensional harmonic retrieval problem. The VTT-RecALS estimator is based on the JIRAFE principle which is composed of two main steps.

1) The first one is the computation of the TTD of the initial tensor. By doing this, the initial Q-order tensor is broken down into Q graph-connected third-order tensors, called TT-cores. This dimensionality reduction is an efficient way to mitigate the "curse of dimensionality". To reach this goal, the TT-SVD [28] is used as a first step.
2) The second step is dedicated to the factorization of the TT-cores. Recall the main result given by Theorem 2, *i.e.*, if the initial tensor follows a Q-order CPD of rank R, then the TT-cores for $2 \leq q \leq Q-1$ follow coupled 3-order CPD of rank R. Consequently, the JIRAFE minimizes the following criterion over the physical quantities $\{\boldsymbol{V}_1, \ldots, \boldsymbol{V}_Q\}$ and over the latent quantities $\{\boldsymbol{M}_1, \ldots, \boldsymbol{M}_{Q-1}\}$:

$$\mathcal{C} = ||\boldsymbol{G}_1 - \boldsymbol{V}_1\boldsymbol{A}\boldsymbol{M}_1^{-1}||_F^2 + ||\boldsymbol{G}_Q - \boldsymbol{M}_{Q-1}\boldsymbol{V}_Q^{\mathsf{T}}||_F^2 \tag{97}$$

$$+ \sum_{q=2}^{Q-1} ||\boldsymbol{\mathcal{G}}_q - \boldsymbol{\mathcal{I}}_{3,R} \underset{1}{\bullet} \boldsymbol{M}_{q-1} \underset{2}{\bullet} \boldsymbol{V}_q \underset{3}{\bullet} \boldsymbol{M}_q^{-\mathsf{T}}||_F^2, \tag{98}$$

where $\boldsymbol{A}$ is a $R \times R$ diagonal matrix with $[\boldsymbol{A}]_{r,r} = \alpha_r$. The above cost function is the sum of coupled LS criteria. The aim of JIRAFE is to recover the original tensor factors using only the 3-order tensors $\boldsymbol{\mathcal{G}}_q$. Consequently, the JIRAFE approach adopts a straightforward sequential methodology, described in Fig. 10, to minimize the cost function $\mathcal{C}$.

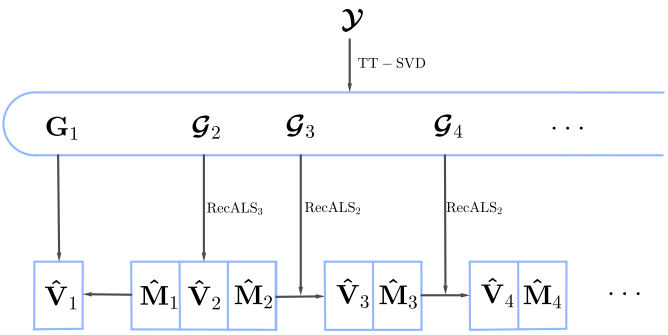

Fig. 10. VTT-RecALS representation

We present in Algorithm 5, the pseudo-code of the VTT-RecALS algorithm, where RecALS$_3$ denotes the RecALS applied to a 3-order tensor, while RecALS$_2$ denotes the RecALS applied to a 3-order tensor using the knowledge of one factor. The VTT-RecALS algorithm is actually divided into two parts. The first part is dedicated to dimensionality reduction, *i.e.*, breaking the dimensionality of the high Q-order tensor into a set of 3-order tensors using Theorem 2. The second part is dedicated to the factors retrieval from the TT-cores using the RecALS algorithm presented in the previous section. It is worth noting that the factors $\mathbf{V}_q$ are estimated up to a trivial (common) permutation ambiguity [88]. As noted in [37], since all the factors are estimated up to a unique column permutation matrix, the estimated angular-frequencies are automatically paired.

Algorithm 5 VTT-RecALS

Input: $\boldsymbol{\mathcal{Y}}$, R, `CritStop`

Output: Estimated parameters: $\{\mathbf{z}_1, \ldots, \mathbf{z}_Q\}$.

1: Dimensionality reduction:

$$[\boldsymbol{G}_1, \boldsymbol{\mathcal{G}}_2, \cdots, \boldsymbol{\mathcal{G}}_{Q-1}, \boldsymbol{G}_Q] = \text{TT-SVD}(\boldsymbol{\mathcal{X}}, R).$$

2: Factor retrieval:

3: For $q = 2$,

$$[\boldsymbol{M}_1, \boldsymbol{V}_2, \boldsymbol{M}_2, \mathbf{z}_2] = \text{RecALS}_3(\boldsymbol{\mathcal{G}}_2, R, \texttt{CritStop}).$$

4: **for** $q = 3 \cdots Q-1$ **do**

5: $\quad [\boldsymbol{V}_q, \boldsymbol{M}_q, \mathbf{z}_q] = \text{RecALS}_2(\boldsymbol{\mathcal{G}}_q, \boldsymbol{M}_{q-1}, M, \texttt{CritStop})$

6: **end for**

7: $\boldsymbol{V}_1 = \boldsymbol{G}_1 \boldsymbol{M}_1$, and $\boldsymbol{V}_Q = \boldsymbol{G}_Q^{\mathsf{T}} \boldsymbol{M}_{Q-1}^{-\mathsf{T}}$

VI. Tensor-based MIMO wireless communication systems

As it is now well known, the use of multiple transmit and receive antennas, corresponding to a MIMO channel, allows to significantly improve the performance of wireless communication systems [89]. Such MIMO systems exploit space diversity to combat channel fading. Each information symbol is transmitted through different paths between the transmitter and the receiver, each path being characterized by an attenuation modeled by means of a fading coefficient of the channel. These channel coefficients are generally assumed to be independent and identically distributed (i.i.d.), which corresponds to an i.i.d. Rayleigh flat fading channel. In the presence of a slow fading environment, the channel coefficients are also assumed to be constant during P time slots, the duration of transmission of N data streams.

Another way to improve the performance consists in using space-time (ST) or space-time-frequency (STF) codings to get space, time and frequency diversities, corresponding to redundancies of the information symbols in each of these domains [33], [90], [91]. One objective of MIMO communication systems is to maximize the combined diversity gain while ensuring the best transmission rate possible, which corresponds to the fundamental diversity-rate tradeoff. Since the pioneering work [7], wireless communications have been a privileged field of application for tensors. Indeed, tensor-based approaches in the context of MIMO communication systems allow:

- to take simultaneously into account different diversities like space, time, frequency or polarization, leading to high order tensor models for the signals received at destination;
- to design tensor coding like tensor space-time (TST) or tensor space-time-frequency (TSTF) codings;
- to develop semi-blind receivers for jointly estimating the information symbols and the channel, by exploiting uniqueness properties of tensor models, which is an advantage of tensor-based systems over matrix-based ones;
- to derive closed-form receivers under *a priori* knowledge of some transmitted symbols (semi-blind systems);
- to deal with point-to-point systems as well as multi-hop systems with relays.

A possible classification of MIMO wireless communication systems can be made according to:

- the considered channel access and multiplexing technology like CDMA (code-division multiple access), TDMA (time-division multiple access), FDMA (frequency-division multiplexing access), OFDMA (orthogonal frequency-division multiplexing access), or yet hybrid technologies like CDMA-OFDM or TDMA-FDMA;
- the assumptions on the channel: flat fading versus frequency-selective fading channel, according to whether all the frequency components of the transmitted signals are attenuated by the same fading or not; in this last case, the frequency dependency of the channel coefficients leads to a third-order MIMO channel tensor;
- the consideration or not of a sparsely scattering environment characterized by a multipath assumption with DoD (direction-of-departure) and DoA (direction-of-arrival) angles;
- the consideration or not of resource allocation under the form of matrix or tensor allocation;
- the type of coding (matrix/tensor), and the diversities taken into account (space-time (ST), space-time-frequency (STF) codings);
- the type of communication system (supervised versus semi-blind, depending on the use or not of a training sequence);
- the type of receiver (iterative versus closed-form);
- the type of optimization algorithm used at the receiver (ALS, Levenberg-Marquardt...);
- the type of communication: half-duplex (HD) versus full-duplex (FD) depending on the possibility or not of a simultaneous communication between two users, meaning that the communication can be in only one direction or in both directions, respectively;
- the existence or not of relays corresponding to multi-hop and single-hop communication systems, respectively;
- the relaying protocol used at the relays (amplify-and-forward (AF) versus decode-and-forward (DF) protocols). Note that the DF protocol means that the signals received at the relay are decoded and then re-encoded before to be forwarded by the relay, whereas in the case of the AF protocol, the received signals are only amplified and forwarded by the relay. When two end nodes are communicating with each other through a relay, via bidirectionnal links, one can distinguish one-way and two-way relaying depending on the way the data streams are exchanged between the end nodes via the relay. With two-way relaying, the end nodes transmit simultaneously their data to the relay, and then the relay combines this information before broadcasting them to the end nodes, which allows to increase the system throughput (in bits/channel use).

A specificity of tensor-based communication systems lies in the development of new tensor models through the design of new systems. For instance, in the case of point-to-point

systems, CONFAC (constrained factors) [92], generalized PARAFAC-Tucker (PARATUCK) [90], and nested PARAFAC models [34] were proposed. Most of these models can be viewed as constrained PARAFAC decompositions. See [11], [93] and [94] for overviews of such models and corresponding point-to-point MIMO systems. In the context of cooperative or relay-assisted communications, several papers have proposed tensor-based models and algorithms for the supervised channel estimation problem [95], [96], considering two-hop AF relays and two-way MIMO relaying [97]. Tensor-based semi-blind receivers for the uplink multiuser cooperative scenario were first proposed in [98] and [99]. For the two-hop MIMO AF relaying case, tensor modeling was exploited in several papers to derive semi-blind joint channel and symbol estimation algorithms [100], [35], [101]. More recently, tensor-based receivers have been developed for multi-hop relaying systems, by exploiting the nested Tucker and coupled nested Tucker models [31], [102].

In this section, we make a brief presentation of three point-to-point systems and of one relaying system to illustrate the tensor-based approach in the context of wireless communication systems.

A. Khatri-Rao space-time coding

Let us consider a MIMO wireless system with M transmit and K receive antennas. The information symbols to be transmitted by the M transmit antennas define a symbol matrix $\boldsymbol{S} \in \mathbb{C}^{N \times M}$, composed of N data streams, each one containing M symbols. These symbols are coded using a space-time coding matrix $\boldsymbol{C} \in \mathbb{C}^{P \times M}$ such that the coded symbols to be transmitted are given by:

$$\boldsymbol{U} = \boldsymbol{C} \odot \boldsymbol{S} \in \mathbb{C}^{PN \times M}, \tag{99}$$

or equivalently:

$$u_{(p-1)N+n,m} = c_{p,m} s_{n,m}. \tag{100}$$

The coding matrix $\boldsymbol{C}$ introduces a time diversity in the sense that every symbol $s_{n,m}$ is duplicated P times, with P corresponding to the number of time slots, and the channel being assumed to be constant during the transmission, i.e., for $p \in [1, P]$. The coding (99) can be viewed as a simplified version of the Khatri-Rao space-time (KRST) coding proposed in [103].

The MIMO channel is assumed to be i.i.d. Rayleigh flat fading and defined as $\boldsymbol{H} \in \mathbb{C}^{K \times M}$, where $h_{k,m}$ is the complex fading coefficient between the m^{th} transmit antenna and the k^{th} receive antenna. The transmitter transmits the coded signals via the communication channel, so that the signal received by the k^{th} receive antenna, during the p^{th} time slot, and associated with the n^{th} data stream, can be written as:

$$\begin{aligned} x_{k,n,p} &= \sum_{m=1}^{M} h_{k,m} u_{(p-1)N+n,m} + b_{k,n,p} \\ &= \sum_{m=1}^{M} h_{k,m} s_{n,m} c_{p,m} + b_{k,n,p}. \end{aligned} \tag{101}$$

These received signals form a third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{K \times N \times P}$, which satisfies a CPD model, with rank M and factor matrices $(\boldsymbol{H}, \boldsymbol{S}, \boldsymbol{C})$, and $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{K \times N \times P}$ represents the additive noise tensor. These matrices can be estimated using the tri-ALS algorithm. Assuming the coding matrix known at the receiver, it is possible to use the bi-ALS algorithm for jointly estimating the channel and symbol matrices. Another solution consists in applying the Khatri-Rao factorisation (KRF) algorithm [43] after a LS estimation of the Khatri-Rao product $\boldsymbol{H} \odot \boldsymbol{S}$ as:

$$\widehat{\boldsymbol{H} \odot \boldsymbol{S}} = \boldsymbol{X}_{KN \times P} (\boldsymbol{C}^{\mathsf{T}})^{\dagger}, \tag{102}$$

where $(.)^{\dagger}$ denotes the matrix right inverse, and $\boldsymbol{X}_{KN \times P}$ represents a tall mode-3 matrix unfolding of the received signals tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{K \times N \times P}$ (see Section III-B6). It is worth noting that choosing a column-orthonormal coding matrix allows to simplify the right-inverse of $\boldsymbol{C}^{\mathsf{T}}$ as $(\boldsymbol{C}^{\mathsf{T}})^{\dagger} = \mathbf{C}^{*} (\mathbf{C}^{\mathsf{T}} \mathbf{C}^{*})^{-1} = \mathbf{C}^{*}$, with the necessary condition $M \leq P$ for uniqueness of this inverse.

B. Tensor space-time coding

We now present another MIMO system which can be viewed as a simplified version of the TST one [33] in the sense that no resource allocation matrix is considered. A tensor space-time coding $\boldsymbol{\mathcal{C}} \in \mathbb{C}^{M \times R \times P}$ is used to code the symbol matrix $\boldsymbol{S} \in \mathbb{C}^{N \times R}$, composed of N data streams, each one containing R symbols.

The coded signals to be transmitted are given by:

$$\boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{C}} \underset{2}{\bullet} \boldsymbol{S} \in \mathbb{C}^{M \times N \times P} \tag{103}$$

or equivalently:

$$u_{m,n,p} = \sum_{r=1}^{R} c_{m,r,p} s_{n,r}, \tag{104}$$

which can be viewed as a linear combination of the R information symbols belonging to the n^{th} data stream, with the coding coefficients $c_{m,r,p}$ as coefficients of the linear combination.

After transmitting the coded signals from the M transmit antennas via the channel $\boldsymbol{H} \in \mathbb{C}^{K \times M}$, the signals received at destination satisfy the following equation:

$$\begin{aligned} \boldsymbol{\mathcal{X}} &= \boldsymbol{\mathcal{U}} \underset{1}{\bullet} \boldsymbol{H} + \boldsymbol{\mathcal{B}} \in \mathbb{C}^{K \times N \times P} \\ &= \boldsymbol{\mathcal{C}} \underset{1}{\bullet} \boldsymbol{H} \underset{2}{\bullet} \boldsymbol{S} + \boldsymbol{\mathcal{B}} \end{aligned} \tag{105}$$

or in scalar form:

$$\begin{aligned} x_{k,n,p} &= \sum_{m=1}^{M} u_{m,n,p} h_{k,m} + b_{k,n,p} \\ &= \sum_{m=1}^{M} \sum_{r=1}^{R} c_{m,r,p} h_{k,m} s_{n,r} + b_{k,n,p}, \end{aligned} \tag{106}$$

which defines a third-order tensor satisfying a Tucker-(2,3) model [11], with the coding tensor $\boldsymbol{\mathcal{C}}$ as core tensor, and the factor matrices $(\boldsymbol{H}, \boldsymbol{S}, \boldsymbol{I}_P)$, where $\boldsymbol{I}_P$ is the identity matrix of order P. One factor matrix being an identity matrix, such

a Tucker model is unique up to two nonsingular matrices $(\boldsymbol{\Delta H}, \boldsymbol{\Delta S})$. If the coding tensor is assumed to be known at the receiver, the model ambiguity becomes a scalar factor related to a Kronecker product. This scalar ambiguity can be eliminated assuming the *a priori* knowledge of only one symbol, leading to a semi-blind Kronecker receiver [94].

Then, the channel and symbol matrices can be jointly estimated using the Kronecker factorisation (KF) algorithm (see section III-B5), after the LS estimation of the Kronecker product $\boldsymbol{H} \boxtimes \boldsymbol{S}$ as:

$$\widehat{\boldsymbol{H} \boxtimes \boldsymbol{S}} = \boldsymbol{X}_{KN\times P}\, \boldsymbol{C}^{\dagger}_{MR\times P}, \tag{107}$$

where $\boldsymbol{C}_{MR\times P}$ denotes a tall mode-3 matrix unfolding of the coding tensor. Choosing this coding tensor such that its matrix unfolding $\boldsymbol{C}_{MR\times P}$ is row-orthonormal allows to simplify the right-inverse as $\boldsymbol{C}^{\dagger}_{MR\times P} = \boldsymbol{C}^{*}_{P\times MR}\big[\boldsymbol{C}_{MR\times P}(\boldsymbol{C}^{*}_{P\times MR}\big]^{-1} = \boldsymbol{C}^{*}_{P\times MR}$, with the necessary condition $MR \leq P$ for uniqueness of this inverse.

Comparing the two presented MIMO systems which are simplified versions of the KRST and TST coding-based systems ([103] and [33], respectively), one can conclude that both systems lead to a third-order tensor for the received signals, with the same dimensions. This tensor satisfies a PARAFAC model, in the first case, while it satisfies a Tucker-(2,3) model in the second case, which induces different ambiguities. Another important difference is that with the TST coding, the number (R) of symbols per data stream can be different from the number (M) of transmit antennas, which is not the case with the KRST coding.

In a practical situation where the coding used at the transmitter is known at the receiver, one can draw the supplementary conclusion that both systems offer the possibility of jointly and semi-blindly estimating the channel and symbol matrices by means of a closed-form receiver. However, in order to eliminate model ambiguities, the Kronecker factorization algorithm used with TST coding presents the advantage over the Khatri-Rao factorization algorithm used with KRST coding, to require the *a priori* knowledge of only one symbol for the Kronecker model while the *a priori* knowledge of M symbols is needed for the PARAFAC model.

C. Tensor space-time-frequency coding

A more general and flexible tensor-based transceiver scheme for space, time and frequency transmit signalling was proposed in [90], referred to as tensor space-time-frequency (TSTF) coding. By relying on a new class of tensor models, namely, the generalized PARATUCK-(N_1, N) and the Tucker-(N_1, N) models, the TSTF system combines a fifth-order coding tensor with a fourth-order allocation tensor. It can also be seen as a generalization of three previous tensor-based ST/TST/STF coding schemes, offering new performance/complexity trade-offs and space, time and frequency allocation flexibility.

Herein, the transmission time is decomposed into P time-slots (data blocks) of N symbol periods, each one being composed of J chips. At each symbol period n of the p^{th} block, the transceiver transmits a linear combination of the n^{th} symbols of certain data streams, using a set of transmit antennas and of sub-carriers. The coding is carried out by means of a fifth-order code tensor $\boldsymbol{\mathcal{W}} \in \mathbb{C}^{M\times R\times F\times P\times J}$ whose dimensions are the numbers of transmit antennas (M), data streams (R), sub-carriers (F), time blocks (P), and chips (J). An allocation tensor $\boldsymbol{\mathcal{C}} \in \mathbb{C}^{M\times R\times F\times P}$ determines the transmit antennas and the sub-carriers that are used, as well as the data streams transmitted in each block p. As an example, $c_{m,r,f,p} = 1$ means that the data stream r is transmitted using the transmit antenna m, with the sub-carrier f, during the time block p.

The TSTF-coded signals to be transmitted are given by:

$$\boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{G}} \underset{1}{\bullet} \boldsymbol{S}, \quad \text{with } \boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{W}} \underset{\{m,r,f,p\}}{\star} \boldsymbol{\mathcal{C}}, \tag{108}$$

or, alternatively, in scalar form

$$u_{m,n,f,p,j} = \sum_{r=1}^{R} w_{m,r,f,p,j}\, s_{n,r} c_{m,r,f,p}, \tag{109}$$

where $\boldsymbol{\mathcal{G}} \in \mathbb{C}^{M\times R\times F\times P\times J}$ is the effective code-allocation tensor, given by the Hadamard product between the code tensor $\boldsymbol{\mathcal{W}} \in \mathbb{C}^{M\times R\times F\times P\times J}$ and the allocation tensor $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{M\times R\times F\times P}$ along their common modes, i.e.,

$$g_{m,r,f,p,j} = w_{m,r,f,p,j} c_{m,r,f,p}.$$

Assuming a flat Rayleigh fading propagation channel, the discrete-time baseband-equivalent model for the signal received at the k^{th} receive antenna during the j^{th} chip period of the n^{th} symbol period of the p^{th} block, and associated with the f^{th} sub-carrier, is given by:

$$x_{k,n,f,p,j} = \sum_{m=1}^{M}\sum_{r=1}^{R} h_{k,m,f}\, s_{n,r} w_{m,r,f,p,j} c_{m,r,f,p} + b_{k,n,f,p,j}.$$

The above equation can be written as the following generalized Tucker-(2,5) model [11]

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \underset{1}{\bullet} \boldsymbol{\mathcal{H}} \underset{2}{\bullet} \boldsymbol{S}. \tag{110}$$

Joint channel and symbol estimation is carried out by exploiting the following unfolding of the received signal tensor:

$$\boldsymbol{X}_{NK\times JPF} = (\boldsymbol{S} \boxtimes \boldsymbol{H}_{K\times MF})\boldsymbol{G}_{RMF\times JPF}, \tag{111}$$

where $\boldsymbol{G}_{RMF\times JPF}$ is a matrix unfolding of the code-allocation tensor defined in (108). By properly designing this tensor to ensure its right-invertibility, the channel and symbol matrices are estimated in closed form from a LS estimation of the Kronecker product

$$\widehat{\boldsymbol{S} \boxtimes \boldsymbol{H}}_{K\times MF} = \boldsymbol{X}_{NK\times JPF} \boldsymbol{G}^{\dagger}_{RMF\times JPF}. \tag{112}$$

The use of a pilot-assisted channel estimation and of a semi-blind alternating least squares (ALS) based receiver are discussed in [90].

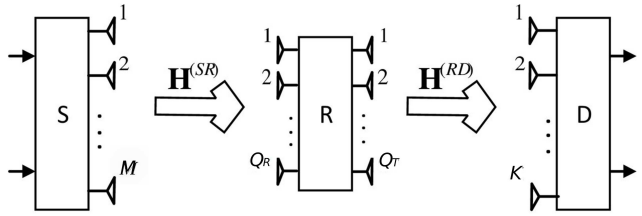

Fig. 11. Block diagram of a one-way two-hop MIMO relay system.

D. MIMO relaying systems with tensor coding

The benefits of tensor space-time coding (TSTC) have been extended to MIMO relaying systems in [31], where semi-blind receivers have been derived for the joint estimation of the source-relay and relay-destination channels, in addition to symbol estimation. Consider a one-way two-hop MIMO relaying communication system shown in Figure 11, where M, and K denote the numbers of antennas at the source (S), and destination (D) nodes, respectively, while Q_R and Q_T represent the numbers of receive and transmit antennas at the relay, respectively. The *source-relay* channel $\boldsymbol{H}^{(SR)} \in \mathbb{C}^{Q_R \times M}$ and the *relay-destination* channel $\boldsymbol{H}^{(RD)} \in \mathbb{C}^{K \times Q_T}$ are assumed to be flat fading and quasi-static, i.e. constant during the transmission protocol which is divided into two consecutive phases. Assume that the direct link between the source and destination is absent or negligible. During the first hop, the transmission is composed of N time-blocks associated with N symbol periods, R data streams being transmitted per time-block. The transmitted symbol matrix $\boldsymbol{S} \in \mathbb{C}^{N \times R}$ contains R data streams composed of N symbols each. During the time-block n, each antenna m of the source node transmits a combination of R information symbols $s_{n,r}, r = 1, ..., R$, to the relay, after a tensor space-time coding [31] by means of a third-order tensor $\boldsymbol{\mathcal{C}}^{(S)} \in \mathbb{C}^{M \times P \times R}$, which introduces space-time redundancies, since each symbol $s_{n,r}$ is repeated P times over the M transmit antennas, P being the source code length. During the second hop, the source remains silent and the relay uses a second three dimensional coding $\boldsymbol{\mathcal{C}}^{(R)} \in \mathbb{C}^{Q_T \times J \times Q_R}$ before forwarding the Q_R received signals to the destination using Q_T transmit antennas. Not that this second tensor coding consists in repeating the signals received at the relay J times, over the Q_T transmit antennas, J being the relay code length. The tensor $\boldsymbol{\mathcal{T}} \in \mathbb{C}^{Q_T \times J \times P \times N}$ of signals transmitted by the relay is defined as

$$t_{q_T,j,p,n} = \sum_{q_R=1}^{Q_R} \sum_{m=1}^{M} \sum_{r=1}^{R} c^{(R)}_{q_T,j,q_R} h^{(SR)}_{q_R,m} c^{(S)}_{m,p,r} s_{n,r}. \quad (113)$$

After transmission through the relay-destination channel, the noiseless signals received by K antennas at destination satisfies the following equation

$$\begin{aligned} x_{k,j,p,n} &= \sum_{q_T=1}^{Q_T} h^{(RD)}_{k,q_T} t_{q_T,j,p,n} \\ &= \sum_{q_T=1}^{Q_T} \sum_{q_R=1}^{Q_R} \sum_{m=1}^{M} \sum_{r=1}^{R} h^{(RD)}_{k,q_T} c^{(R)}_{q_T,j,q_R} h^{(SR)}_{q_R,m} c^{(S)}_{m,p,r} s_{n,r}. \end{aligned} \quad (114)$$

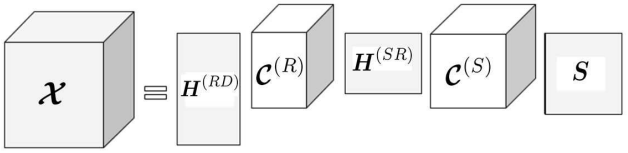

Fig. 12. NTT(4) model for a one-way two-hop MIMO relaying system with tensor coding.

Comparing (114) with (32), we can conclude that the tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{K \times J \times P \times N}$ of noiseless signals received at destination satisfies a NTT(4) model. Figure 12 illustrates the NTT(4) for the noiseless signal tensor $\boldsymbol{\mathcal{X}}$ received at the destination. We have to note that the four dimensions of the tensor $\boldsymbol{\mathcal{X}}$ are associated with four signal diversities (space k, relay code j, source code p, time n).

A special case of this model is the one presented in [35], [100], where the coding tensors $\boldsymbol{\mathcal{C}}^{(S)}$ and $\boldsymbol{\mathcal{C}}^{(R)}$ have a special structure where their frontal slices are diagonal matrices, with $M = R$ and $Q_R = Q_T$. Physically, this means that the source and the relay transmits a single data stream, in contrast to the NTT(4) model in (114) where R and Q_R signals are combined at the source and relay, respectively, before transmission. In this case, the noiseless received signal tensor at the destination node reduces to

$$x_{k,j,p,n} = \sum_{q_T=1}^{Q_T} \sum_{m=1}^{M} h^{(RD)}_{k,q_T} c^{(R)}_{q_T,j} h^{(SR)}_{q_R,m} c^{(S)}_{m,p} s_{n,r}, \quad (115)$$

which corresponds to a Nested CP model. This can be concluded by comparing (115) with (51).

Assuming that the coding tensors $\boldsymbol{\mathcal{C}}^{(S)}$ and $\boldsymbol{\mathcal{C}}^{(R)}$ are known at the destination node (i.e. by the receiver), semi-blind receivers for jointly estimating the symbol matrix ($\boldsymbol{S}$) and the individual channels ($\boldsymbol{H}^{(SR)}, \boldsymbol{H}^{(RD)}$) are proposed in [31]. We refer the interested reader to this reference for further details on the receiver algorithms.

We summarized in Table I the different tensor decompositions and algorithms used in the three applications presented in this manuscript.

VII. Conclusion

Collecting data from multisensor signal processing can be naturally interpreted in the multilinear algebra context. In this work, three important array signal processing-based applications are exposed in which the tensor-based framework has proven to be useful. Namely, the problems of interest are (i) array signal processing for localization, (ii) multidimensional harmonic retrieval and (iii) MIMO processing for wireless communications. In this overview paper, we present systematically in a synthetic way the necessary mathematical material to ensure a good understanding of the problems of interest.

References

[1] R. A. Harshman, "Foundations of the Parafac procedure: Models and conditions for an explanatory multimodal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

Application	Tensor decompositions	Algorithms
DOA	Canonical Polyadic Decomposition (CPD)	Alternating Least Squares (ALS)
MHR	Generalized Vandermonde CPD	Rectified ALS (RecALS), Vandermonde Tensor-Train RecALS (VTT-RecALS)
MIMO	CPD, Tucker-(N_1, N), Nested Tucker Train (NTT)	Kronecker factorization (KF), Khatri-Rao factorization (KRF)

TABLE I
A SUMMARY OF THE TENSOR DECOMPOSITIONS AND ALGORITHMS USED IN THE APPLICATIONS.

[2] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via N-way generalization of Eckart-Young decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, Sep. 1970.

[3] E. Acar and B. Yener, "Unsupervised multiway data analysis: A literature survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 1, pp. 6–20, 2008.

[4] N. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, pp. 3551 – 3582, 2017.

[5] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. on Signal Processing*, vol. 48, no. 8, pp. 2377–2388, 2000.

[6] T. Jiang, N. D. Sidiropoulos, and J. M. F. T. Berge, "Almost-sure identifiability of multidimensional harmonic retrieval," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1849–1859, 2001.

[7] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. on Signal Processing*, vol. 48, no. 3, pp. 810–823, 2000.

[8] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for mimo radar," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5693–5705, 2010.

[9] P. Comon, "Tensors: a brief introduction," *IEEE Sig. Proc. Magazine*, vol. 31, no. 3, pp. 44–53, May 2014, special issue on BSS.

[10] L. De Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Ana. Appl.*, vol. 21, no. 4, pp. 1253–1278, Apr. 2000.

[11] G. Favier and A. de Almeida, "Overview of constrained PARAFAC models," *EURASIP J. on Advances in Signal Processing*, vol. 5, pp. 1–41, 2014.

[12] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *J. Math. and Phys.*, vol. 6, no. 1, pp. 165–189, 1927.

[13] J. M. Landsberg, *Tensors: Geometry and Applications*, ser. Graduate Studies in Mathematics. AMS publ., 2012, vol. 128.

[14] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, ser. Series in Computational Mathematics. Berlin, Heidelberg: Springer, 2012.

[15] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, Sep. 2009.

[16] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions," *Linear Algebra and Applications*, vol. 18, pp. 95–138, 1977.

[17] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *Jour. Chemo.*, vol. 14, pp. 229–239, 2000.

[18] A. Stegeman and N. Sidiropoulos, "On Kruskal's uniqueness condition for the CP decomposition," *Lin. Alg. Appl.*, vol. 420, no. 2-3, pp. 540–552, Jan. 2007.

[19] P. Comon, B. Mourrain, L.-H. Lim, and G. Golub, "Genericity and rank deficiency of high order symmetric tensors," in *ICASSP'06*, Toulouse, May 14-19 2006.

[20] P. Comon, J. M. F. T. Berge, L. De Lathauwer, and J. Castaing, "Generic and typical ranks of multi-way arrays," *Linear Algebra Appl.*, vol. 430, no. 11–12, pp. 2997–3007, Jun. 2009.

[21] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, "An algorithm for generic and low-rank specific identifiability of complex tensors," *SIAM J. matrix Ana. Appl.*, vol. 35, no. 4, pp. 1265–1287, 2014.

[22] Y. Qi, P. Comon, and L. H. Lim, "Uniqueness of nonnegative tensor approximations," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 2170–2183, Apr. 2016, arXiv:1410.8129.

[23] P. Comon, X. Luciani, and A. L. F. De Almeida, "Tensor decompositions, alternating least squares and other tales," *Jour. Chemometrics*, vol. 23, no. 7-8, pp. 393–405, Aug. 2009.

[24] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1260–1280, Feb. 2014.

[25] S. Sahnoun and P. Comon, "Joint source estimation and localization," *IEEE Trans. Sig. Proc.*, vol. 63, no. 10, pp. 2485–2495, May 2015, hal-01005352.

[26] L.-H. Lim and P. Comon, "Nonnegative approximations of nonnegative tensors," *Jour. Chemometrics*, vol. 23, pp. 432–441, Aug. 2009.

[27] Y. Qi, P. Comon, and L.-H. Lim, "Semialgebraic geometry of nonnegative tensor rank," *SIAM J. Matrix Ana. Appl.*, vol. 37, no. 4, pp. 1556–1580, Nov. 2016, hal-01763832.

[28] I. V. Oseledets, "Tensor-train decomposition," *SIAM J. Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, 2011.

[29] I. V. Oseledets and E. E. Tyrtyshnikov, "Breaking the curse of dimensionality, or how to use SVD in many dimensions," *SIAM J. Scientific Computing*, vol. 31, pp. 3744–3759, 2009.

[30] Y. Zniyed, R. Boyer, A. L. de Almeida, and G. Favier, "A TT-based hierarchical framework for decomposing high-order tensors," *SIAM Journal on Scientific Computing*, vol. 42, no. 2, pp. A822–A848, 2020.

[31] G. Favier, C. Fernandes, and A. de Almeida, "Nested Tucker tensor decomposition with application to MIMO relay systems using tensor space-time coding (TSTC)," *Elsevier Signal Processing*, vol. 128, no. 4, pp. 318–331, 2016.

[32] D. Rocha, G. Favier, and C. Fernandes, "Closed-form receiver for multi-hop MIMO relay systems with tensor space-time coding," *Journal of Communication and Information Systems*, vol. 34, no. 1, pp. 50–54, 2019.

[33] G. Favier, M. da Costa, A. de Almeida, and J. Romano, "Tensor space-time (TST) coding for (MIMO) wireless communication systems," *Elsevier Signal Processing*, vol. 92, no. 4, pp. 1079–1092, 2012.

[34] A. de Almeida and G. Favier, "Double Khatri-Rao space-time-frequency coding using semi-blind PARAFAC based receiver," *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 471–474, 2013.

[35] L. Ximenes, G. Favier, and A. de Almeida, "Closed-form semi-blind receiver for MIMO relay systems using double Khatri-Rao space-time coding," *IEEE Signal Processing Letters*, vol. 23, no. 3, pp. 316–320, 2016.

[36] Y. Zniyed, R. Boyer, A. L. de Almeida, and G. Favier, "High-order tensor estimation via trains of coupled third-order CP and Tucker decompositions," *Linear Algebra and its Applications*, vol. 588, pp. 304–337, 2020.

[37] ——, "High-order CPD estimation with dimensionality reduction using a tensor train model," in *Proc. of the 26th European Signal Processing Conference*, 2018.

[38] C. F. Van Loan and N. Pitsianis, *Approximation with Kronecker Products*. Dordrecht: Springer Netherlands, 1993, pp. 293–314.

[39] K. K. Wu, Y. Yam, H. Meng, and M. Mesbahi, "Kronecker Product Approximation with Multiple Factor Matrices via the Tensor Product Algorithm," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016)*, 2016, pp. 004 277–004 282.

[40] T. Zhang and G. H. Golub, "Rank-one approximation to high order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 23, no. 2, pp. 534–550, 2001.

[41] E. Kofidis and P. A. Regalia, "On the best rank-1 approximation of higher-order supersymmetric tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 23, no. 3, pp. 863–884, 2002.

[42] A. P. da Silva, P. Comon, and A. L. F. de Almeida, "A finite algorithm to compute rank-1 tensor approximations," *IEEE Signal Processing Letters*, vol. 23, no. 7, pp. 959–963, July 2016.

[43] A. Y. Kibangou and G. Favier, "Non-iterative solution for PARAFAC with a Toeplitz matrix factor," in *Proc. EUSIPCO*, Glasgow, Scotland, Aug. 2009.

[44] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT–a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE*

Transactions on Acoustics, Speech, and Signal Processing, vol. 34, no. 5, pp. 1340–1342, 1986.

[45] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.

[46] G. Bienvenu and L. Kopp, "Principe de la goniométrie passive adaptative," in *1979 7ème Colloque sur le traitement du signal et des images (GRETSI)*. Groupe d'Etudes du Traitement du Signal et des Images, 1979.

[47] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," *Stanford Univ., Stanford, CA*, 1981.

[48] S. Miron, Y. Song, D. Brie, and K. T. Wong, "Multilinear direction finding for sensor-array with multiple scales of invariance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 2057–2070, 2015.

[49] X. Guo, S. Miron, D. Brie, S. Zhu, and X. Liao, "A CANDECOMP/PARAFAC perspective on uniqueness of DOA estimation using a vector sensor array," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3475–3481, 2011.

[50] M. Sørensen and L. De Lathauwer, "Multiple invariance ESPRIT for nonuniform linear arrays: A coupled canonical polyadic decomposition approach," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3693–3704, 2016.

[51] J. Liang, S. Yang, J. Zhang, L. Gao, and F. Zhao, "4D near-field source localization using cumulant," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–10, 2007.

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

[53] M. Boizard, G. Ginolhac, F. Pascal, S. Miron, and P. Forster, "Numerical performance of a tensor MUSIC algorithm based on HOSVD for a mixture of polarized sources," in *2013 21st European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, September 2013.

[54] N. D. Sidiropoulos, "Generalizing carathéodory's uniqueness of harmonic parameterization to n dimensions," *IEEE Transactions on Information Theory*, vol. 47, pp. 1687–1690, 2001.

[55] Y. Li, J. Razavilar, and K. Liu, "A high-resolution technique for multidimensional NMR spectroscopie," *IEEE Transactions on Biomedical Engineering*, vol. 45, pp. 78–86, 1998.

[56] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radars," *IEEE Transactions on Signal Processing*, vol. 58, pp. 5693–57 057, 2010.

[57] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Transactions on Signal Processing*, vol. 48, pp. 2377–2388, 2000.

[58] M. D. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2-d angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT," *IEEE Transactions on Signal Processing*, vol. 44, pp. 316–328, 1996.

[59] E. Acar, R. Bro, and A. Smilde, "Data fusion in metabolomics using coupled matrix and tensor factorizations," *Proceedings of the IEEE*, vol. 103, pp. 1602–1620, 2015.

[60] R. C. Farias, J. Cohen, and P. Comon, "Exploring multimodal data fusion through joint decompositions with flexible couplings," *IEEE Transactions on Signal Processing*, vol. 64, pp. 4830–4844, 2016.

[61] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges and prospects," *Proceedings of the IEEE*, vol. 103, pp. 1449–1477, 2015.

[62] A. Cichocki, "Era of big data processing: A new approach via tensor networks and tensor decompositions," *International Workshop on Smart Info-Media Systems in Asia*, 2013.

[63] A. Phan, A. Cichocki, A. Uschmajew, P. Tichavsky, G. Luta, and D. Mandic, "Tensor networks for latent variable analysis. part i: Algorithms for tensor train decomposition," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[64] N. Vervliet, O. Debals, L. Sorber, and L. De Lathauwer, "Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis," *SIAM J. Scientific Computing*, vol. 31, pp. 71–79, 2014.

[65] W. Hackbusch and S. Kuhn, "A new scheme for the tensor representation," *Journal of Fourier Analysis and Applications*, vol. 15, pp. 706–722, 2009.

[66] I. V. Oseledets, "Tensor-train decomposition," *SIAM Journal on Scientific Computing*, vol. 33, pp. 2295–2317, 2011.

[67] D. Kressner, M. Steinlechner, and B. Vandereycken, "Low-rank tensor completion by riemannian optimization," *BIT Numerical Mathematics*, vol. 54, pp. 447–468, 2014.

[68] M. Bousse, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Transactions on Signal Processing*, vol. 65, pp. 346–358, 2016.

[69] F. Roemer and M. Haardt, "A closed-form solution for parallel factor (PARAFAC) analysis," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008.

[70] R. Bro, "PARAFAC: Tutorial and applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, pp. 149–171, 1997.

[71] N. Li, S. Kindermann, and C. Navasca, "Some convergence results on the regularized alternating least-squares method for tensor decomposition," *Linear Algebra and its Applications*, vol. 438, pp. 796–812, 2013.

[72] A. Cichocki, I. O. N. Lee, A.-H. Phan, Q. Zhao, and D. Mandic, "Low-rank tensor networks for dimensionality reduction and large-scale optimization problems: Perspectives and challenges," *Foundations and Trends in Machine Learning*, vol. 9, pp. 249–420, 2016.

[73] P. M. Kroonenberg and J. de Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms," *Psychometrika*, vol. 45, pp. 69–97, 1980.

[74] M. Clark and L. Scharf, "Two-dimensional modal analysis based on maximum likelihood," *IEEE Transactions on Signal Processing*, vol. 42, pp. 1443–1452, 1994.

[75] R. Boyer, "Deterministic asymptotic Cramér-Rao bound for the multidimensional harmonic model," *Signal Processing*, vol. 88, pp. 2869–2877, 2008.

[76] S. Sahnoun, K. Usevich, and P. Comon, "Multidimensional ESPRIT for damped and undamped signals: Algorithm, computations and perturbation analysis," *IEEE Trans. Sig. Proc.*, vol. 65, no. 22, pp. 5897–5910, Nov. 2017, hal-01360438.

[77] J. Liu and X. Liu, "An eigenvector-based approach for multidimensional frequency estimation with improved identifiability," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4543–4556, 2006.

[78] M. Sorensen and L. De Lathauwer, "Blind signal separation via tensor decomposition with Vandermonde factor: Canonical polyadic decomposition," *IEEE Transactions on Signal Processing*, vol. 61, pp. 5507–5519, 2013.

[79] R. Bro, N. D. Sidiropoulos, and G. B. Giannakis, "A fast least squares algorithm for separating trilinear mixtures," *Int. Workshop on Independent Component Analysis and Blind Separationg*, 1999.

[80] J. M. Papy, L. De Lathauwer, and S. V. Huffel, "Exponential data fitting using multilinear algebra: the single-channel and multi-channel case," *Wiley Online Library*, vol. 12, pp. 809–826, 2005.

[81] P. Kroonenberg and J. Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms," *Psychometrika*, vol. 45, no. 1, pp. 69–97, 1980.

[82] P. Comon, X. Luciani, and A. L. D. Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 393–405, 2009.

[83] J. H. Goulart, M. Boizard, R. Boyer, G. Favier, and P. Comon, "Tensor CP decomposition with structured factor matrices: Algorithms and performance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 757–769, 2016.

[84] I. Markovsky, "Low rank approximation: algorithms, implementation, applications," *Springer Science & Business Media*, 2011.

[85] J. P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in hilbert spaces," *Advances in order restricted statistical inference, Springer*, pp. 28–47, 1986.

[86] D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Transactions on Information Theory*, vol. 20, no. 5, pp. 591–598, 1974.

[87] R. Boyer and P. Comon, "Rectified ALS algorithm for multidimensional harmonic retrieval," in *Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2016.

[88] M. Sørensen and L. De Lathauwer, "New uniqueness conditions for the canonical polyadic decomposition of third-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 36, pp. 1381–1403, 2015.

[89] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Sig. Proc.*, vol. 49, no. 5, pp. 1073–1096, 2003.

[90] G. Favier and A. de Almeida, "Tensor space-time-frequency coding with semi-blind receivers for MIMO wireless communication systems," *IEEE Trans. Signal Processing*, vol. 62, no. 22, pp. 5987–6002, 2014.

[91] K. Liu, J. P. C. da Costa, H. So, and A. L. F. de Almeida, "Semi-blind receivers for joint symbol and channel estimation in space-time-frequency MIMO-OFDM systems," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5444–5457, 2013.

[92] A. de Almeida, G. Favier, and J. Mota, "A constrained factor decomposition with application to MIMO antenna systems," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2429–2442, 2008.

[93] ——, "PARAFAC-based unified tensor modeling for wireless communication systems with application to blind multiuser equalization," *Signal Processing*, vol. 87, no. 2, pp. 337–351, 20O7.

[94] M. da Costa, G. Favier, and J. Romano, "Tensor modelling of MIMO communication systems with performance analysis and kronecker receivers," *Elsevier Signal Processing*, vol. 145, no. 4, pp. 304–316, 2018.

[95] Y. Rong, M. R. A. Khandaker, and Y. Xiang, "Channel estimation of dual-hop MIMO relay system via parallel factor analysis," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2224–2233, June 2012.

[96] I. V. Cavalcante, A. L. F. de Almeida, and M. Haardt, "Tensor-based Approach to Channel Estimation in Amplify-and-Forward MIMO Relaying Systems," in *Proc. IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM 2014)*, 2014, pp. 445–448.

[97] F. Roemer and M. Haardt, "Tensor-Based Channel Estimation and Iterative Refinements for Two-Way Relaying with Multiple Antennas and Spatial Reuse," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5720–5735, Nov. 2010.

[98] C. A. R. Fernandes, A. L. F. de Almeida, and D. B. da Costa, "Unified tensor modeling for blind receivers in multiuser uplink cooperative systems," *IEEE Signal Processing Letters*, vol. 19, no. 5, pp. 247–250, 2012.

[99] A. L. F. de Almeida, C. A. Fernandes, and D. B. da Costa, "Multiuser Detection for Uplink DS-CDMA Amplify-and-Forward Relaying Systems," *IEEE Signal Processing Letters*, vol. 20, no. 7, pp. 697–700, 2013.

[100] L. R. Ximenes, G. Favier, and A. L. F. de Almeida, "Semi-Blind Receivers for Non-Regenerative Cooperative MIMO Communications Based on Nested PARAFAC Modeling," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4985–4998, 2015.

[101] B. Sokal, A. L. de Almeida, and M. Haardt, "Semi-blind receivers for MIMO multi-relaying systems via rank-one tensor approximations," *Signal Processing*, vol. 166, p. 107254, 2020.

[102] D. Rocha, C. Fernandes, and G. Favier, "MIMO multi-relay systems with tensor space-time coding based on coupled nested tucker decomposition," *Digital Signal Processing*, vol. 89, no. 3, pp. 170–185, 2019.

[103] N. Sidiropoulos and R. Budampati, "Khatri-Rao space-time codes," *IEEE Trans. on Signal Processing*, vol. 50, no. 10, pp. 2396–2407, 2002.