Double Coupled Canonical Polyadic Decomposition for Joint Blind Source Separation

Xiao-Feng Gong, *Member, IEEE*, Qiu-Hua Lin, *Member, IEEE*, Feng-Yu Cong, *Senior Member, IEEE*, and Lieven De Lathauwer, *Fellow, IEEE*

***Abstract* — Joint blind source separation (J-BSS) is an emerging data-driven technique for multi-set data-fusion. In this paper, J-BSS is addressed from a tensorial perspective. We show how, by using second-order multi-set statistics in J-BSS, a specific double coupled canonical polyadic decomposition (DC-CPD) problem can be formulated. We propose an algebraic DC-CPD algorithm based on a coupled rank-1 detection mapping. This algorithm converts a possibly underdetermined DC-CPD to a set of overdetermined CPDs. The latter can be solved algebraically via a generalized eigenvalue decomposition based scheme. Therefore, this algorithm is deterministic and returns the exact solution in the noiseless case. In the noisy case, it can be used to effectively initialize optimization based DC-CPD algorithms. In addition, we obtain the deterministic and generic uniqueness conditions for DC-CPD, which are shown to be more relaxed than their CPD counterparts. Experiment results are given to illustrate the superiority of DC-CPD over standard CPD with regards to uniqueness and accuracy.**

***Index Terms* — Joint blind source separation, Tensor, Coupled canonical polyadic decomposition.**

I. Introduction

CANONICAL polyadic decomposition (CPD) decomposes a higher-order tensor into a minimal number of rank-1 terms. Compared with its matrix counterpart, CPD is essentially unique under mild conditions, without the need to exploit prior knowledge or impose particular constraints, and this nice identifiability property makes it a fundamental tool that is of particular interest for blind source separation (BSS). CPD has indeed been successfully adopted in many BSS applications [1]–[8] including telecommunications, array signal processing, speech separation, and biomedical data analysis. Numerous CPD algorithms have been proposed in the literature [9]–[16], and its identifiability has been extensively studied, including its ability to deal with underdetermined BSS problems [6], [10]–[13], [17]–[22]. The incorporation of various constraints (e.g. nonnegativity, source independence, Vandermonde structure etc.) has also been investigated [23]–[28], aiming at improved performance in a variety of applications.

Recently, joint BSS (J-BSS) of multi-set signals has been considered in a number of applications such as multi-subject / multi-modal biomedical data fusion and BSS of transformed signals at multiple frequency bins for convolutive mixtures. These approaches usually assume dependence across datasets (inter-set dependence) and independence of latent sources within a dataset (intra-set independence), aiming at BSS of each individual dataset as well as indication of correspondences among decomposed components. A number of J-BSS algorithms have been proposed, e.g. joint and parallel independent component analysis (ICA) [29], independent vector analysis (IVA) [30]–[32], multi-set canonical correlation analysis (M-CCA) [33]–[35] and generalized joint diagonalization (GJD) [36]–[39]. We refer to [40] and references therein for an overview.

There are also tensor-based results that concern multi-set problems [41]–[59]. The concept of linked mode parallel factor analysis originated in [41]. Tensor probabilistic ICA (T-PICA) [42] has been extended and applied to multi-set fMRI via group T-PICA [43] and linked T-PICA [44], which assume shared source and shared mixing matrix, respectively. A simple case of coupled CPD (C-CPD) over two tensors has been discussed in [47]–[49] with applications to blind deconvolution of MIMO systems, joint EEG-MEG analysis, and array processing, respectively. Linked CPD with partially shared factors is addressed in [50]. A comprehensive study of uniqueness and an algebraic algorithm for the computation of C-CPD and coupled rank-($L_{r,n}$, $L_{r,n}$, 1) block term decomposition with one coupling factor matrix are provided in [51] and [52], respectively. Several applications of C-CPD in signal and array processing are discussed in [45], [46], [53]–[55]. Soft or flexible coupling has been considered as a way to deal with nonidentical but similar coupling factor matrices [56]–[58]. Recently, C-CPD with double coupling structure has been considered in [59].

However, the C-CPD approaches mentioned above are not specifically devised for J-BSS problems. More precisely, the

Manuscript received XXXX XX, XXXX. This research is funded by: (1) National natural science foundation of China (nos. 61671106, 61331019, 61379012, 81471742); (2) Scientific Research Fund of Liaoning Education Department (no. L2014016); (3) Fundamental Research Funds for the Central Universities (nos. DUT16QY07). (4) China Scholarship Council; (5) Research Council KU Leuven: C1 project C16/15/059-nD, CoE PFV/10/002 (OPTEC). (6) FWO: project G.0830.14N, G.0881.14N; (7) the Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO II, Dynamical systems, control and optimization, 2012-2017); (8) EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no. 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

Xiao-Feng Gong and Qiu-Hua Lin are with the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China, 116024. (e-mails: xfgong@dlut.edu.cn, qhlin@dlut.edu.cn).

Feng-Yu Cong is with the Department of Biomedical Engineering, Dalian University of Technology, Dalian, China, 116024 (e-mail: cong@dlut.edu.cn).

Lieven De Lathauwer is with the STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics; Department of Electrical Engineering (ESAT), KU Leuven, BE-3001 Leuven, Belgium; and Group Science, Engineering and Technology, KU Leuven, Kulak 8500 Kortrijk, Belgium (e-mail: Lieven.DeLathauwer@kuleuven.be).

C-CPD approaches in [47]–[49] are limited to the case of two coupled tensor datasets. The approaches in [50]–[58] based on either hard or soft coupling structure are mainly used to fuse multiple datasets which have tensor form, while the general multi-set data formulation in J-BSS is in matrix form. The T-PICA variants [42]–[44] assume that either the mixing matrices or the sources for different datasets are the same, and thus may not work well in applications where this crucial assumption is not satisfied. In fact, as will be shown later, by using second-order statistics of the multi-set signal, we can obtain a set of tensors admitting a specific double coupled CPD (DC-CPD) formulation with coupling in two modes, possibly with a conjugation structure for complex-valued problems (an illustration is given in Fig. 1). The coupling factor matrices of this specific DC-CPD may not have full column rank. We note that the GJD formulation in [36]–[39] is similar to DC-CPD, but is limited to the case where the coupling factor matrices have full column rank. As such, to the best of our knowledge, the specific but important DC-CPD problem has not been thoroughly studied in the literature. It will be addressed in this paper, both from a theoretical and a practical perspective.

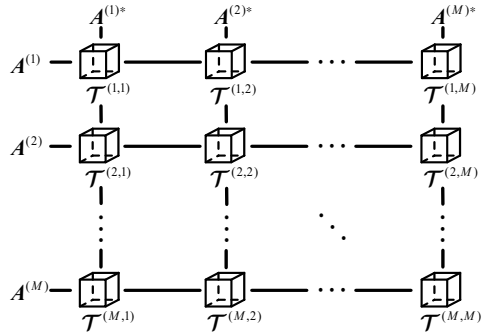


Fig. 1. Illustration of DC-CPD. The tensors are placed at different nodes of a grid according to their indices. The tensor at node (m, n) admits $\boldsymbol{\mathcal{T}}^{(m,n)} = \llbracket \boldsymbol{A}^{(m)}, \boldsymbol{A}^{(n)*}, \boldsymbol{C}^{(m,n)} \rrbracket_R$, m, n = 1,…,M. Tensors in the mth column of the grid are coupled in the first mode by $\boldsymbol{A}^{(m)}$. Tensors in the mth row of the grid are coupled in the second mode by $\boldsymbol{A}^{(m)*}$.

In Section II, we explain how second-order statistics can be used to formulate a DC-CPD from the matrix data in J-BSS. In Section III, we present an algebraic algorithm based on a new coupled rank-1 detection mapping. We will show that it is possible to compute an exact DC-CPD by means of conventional linear algebra, essentially by solving overdetermined sets of linear equations, low-rank matrix approximation and generalized eigenvalue decomposition (GEVD). In addition, we obtain deterministic and generic uniqueness conditions for DC-CPD. In Section IV, experiment results are provided to demonstrate the performance of DC-CPD, in comparison with CPD. Section V concludes this paper.

Notation: Vectors, matrices and tensors are denoted by lower-case boldface, uppercase boldface and uppercase calligraphic letters, respectively. The rth column vector and the (i, j)th entry of $\boldsymbol{A}$ are denoted by $\boldsymbol{a}_r$ and a_{ij}, respectively. The identity matrix, all-zero matrix and all-zero vectors are denoted by $\boldsymbol{I}_M \in \mathbb{C}^{M\times M}$, $\boldsymbol{0}_{M,N} \in \mathbb{C}^{M\times N}$, $\boldsymbol{0}_M \in \mathbb{C}^M$, respectively. The subscripts are omitted if there is no ambiguity. The null space of a matrix $\boldsymbol{M}$ is denoted as $\ker(\boldsymbol{M})$. The dimensionality of a vector space $\mathfrak{I}$ is denoted as $\dim(\mathfrak{I})$. Transpose, conjugate, conjugated transpose, Moore-Penrose pseudo-inverse, Frobenius norm and matrix determinant are denoted as $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, $(\cdot)^\dagger$, $\|\cdot\|_F$, and $|\cdot|$, respectively.

The symbols '$\otimes$', '$*$', '$\odot$' and '$\otimes$' denote Kronecker product, Hadamard product, Khatri-Rao product, and tensor outer product, respectively, defined as follows:

$$\boldsymbol{A}\otimes\boldsymbol{B} \triangleq \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \boldsymbol{A}*\boldsymbol{B} \triangleq \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots \\ a_{21}b_{21} & a_{22}b_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix},$$

$$\boldsymbol{A}\odot\boldsymbol{B} \triangleq [\boldsymbol{a}_1\otimes\boldsymbol{b}_1, \boldsymbol{a}_2\otimes\boldsymbol{b}_2, \cdots], (\boldsymbol{a}\otimes\boldsymbol{b}\otimes\boldsymbol{c})_{i,j,k} \triangleq a_i b_j c_k.$$

We denote $\bigodot_{u\in\Omega}\boldsymbol{B}^{(u)}$ and $\bigotimes_{u\in\Omega}\boldsymbol{B}^{(u)}$ as the Khatri-Rao product and Kronecker product of all the matrices $\boldsymbol{B}^{(u)}$ with $u\in\Omega$, respectively. Mathematical expectation is denoted as $\mathrm{E}\{\cdot\}$. MATLAB notation will be used to denote submatrices of a tensor. For instance, we use $(\boldsymbol{\mathcal{T}})_{(:,:,k)}$ or $\boldsymbol{T}_{(:,:,k)}$ to denote the frontal slice of tensor $\boldsymbol{\mathcal{T}}$ obtained by fixing the third index to k. A polyadic decomposition (PD) of $\boldsymbol{\mathcal{T}}$ expresses $\boldsymbol{\mathcal{T}}$ as the sum of rank-1 terms:

$$\boldsymbol{\mathcal{T}} = \llbracket \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \rrbracket_R \triangleq \sum_{r=1}^{R} \boldsymbol{a}_r \otimes \boldsymbol{b}_r \otimes \boldsymbol{c}_r \in \mathbb{C}^{I\times J\times K}, \qquad (1)$$

where $\boldsymbol{A} \triangleq [\boldsymbol{a}_1,...,\boldsymbol{a}_R] \in \mathbb{C}^{I\times R}$, $\boldsymbol{B} \triangleq [\boldsymbol{b}_1,...,\boldsymbol{b}_R] \in \mathbb{C}^{J\times R}$, and $\boldsymbol{C} \triangleq [\boldsymbol{c}_1,...,\boldsymbol{c}_R] \in \mathbb{C}^{K\times R}$. We call (1) a canonical PD (CPD) if R is minimal.

For an Nth-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{C}^{I_1\times\cdots\times I_N}$, $\mathrm{vec}(\boldsymbol{\mathcal{T}}) \in \mathbb{C}^{I_1\cdots I_N}$ denotes the vector representation of $\boldsymbol{\mathcal{T}}$ defined by $[\mathrm{vec}(\boldsymbol{\mathcal{T}})]_{\tilde{i}} \triangleq t_{i_1,...,i_N}$, with $\tilde{i} = \sum_{n=1}^{N}(i_n - 1)\prod_{m=1}^{N-n} I_m + 1$, while $\mathrm{unvec}(\cdot)$ performs the inverse. The mode-i matrix representation of a third-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{C}^{I\times J\times K}$ is denoted as $\boldsymbol{T}_i$, i = 1, 2, 3, and defined by:

$$(\boldsymbol{T}_1)_{(j-1)K+k,i} = (\boldsymbol{T}_2)_{(i-1)K+k,j} = (\boldsymbol{T}_3)_{(i-1)J+j,k} = t_{i,j,k}.$$

We define $\mathrm{Ten}(\boldsymbol{T}, [I_1,...,I_{K-1},J]) = \boldsymbol{\mathcal{T}}$ as the operation to reshape a matrix $\boldsymbol{T} \in \mathbb{C}^{\prod_{k=1}^{K-1} I_k \times J}$ into a Kth-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{C}^{I_1\times\cdots\times I_{K-1}\times J}$ such that:

$$t_{i_1,...,i_{K-1},j} = (\boldsymbol{T})_{\tilde{i},j}, \text{ where } \tilde{i} \triangleq \sum_{k=1}^{K-1}(i_k - 1)\prod_{m=1}^{K-k-1} I_m + 1.$$

The operator $\mathrm{perm}_{\boldsymbol{p}}(\boldsymbol{\mathcal{T}})$ permutes the index of $\boldsymbol{\mathcal{T}}$ according to a permutation vector $\boldsymbol{p}$. For instance, $\mathrm{perm}_{(2,1,3)}(\boldsymbol{\mathcal{T}})$ permutes the first and second indices of $\boldsymbol{\mathcal{T}}$. Concatenation of tensors $\boldsymbol{\mathcal{T}}_1 \in \mathbb{C}^{I\times J\times K_1}, ..., \boldsymbol{\mathcal{T}}_N \in \mathbb{C}^{I\times J\times K_N}$ in the last mode is denoted as $\boldsymbol{\mathcal{T}} \triangleq \mathrm{cat}(\boldsymbol{\mathcal{T}}_1,...,\boldsymbol{\mathcal{T}}_N) \in \mathbb{C}^{I\times J\times(K_1+\cdots+K_N)}$.

The number of k-permutations from a set of m elements is defined as $\mathrm{P}_m^k \triangleq m\cdot(m-1)\cdots(m-k+1)$, and the number of k-combinations from a set of m elements is defined as $\mathrm{C}_m^k \triangleq \mathrm{P}_m^k / \mathrm{P}_k^k$. The cardinality of a set is denoted as $\mathrm{card}(\cdot)$.

II. Problem Formulation

J-BSS usually assumes the following multi-set data model:

$$\boldsymbol{x}^{(m)}(t) = \boldsymbol{A}^{(m)}\boldsymbol{s}^{(m)}(t), \qquad m = 1,...,M, \qquad (2)$$

where $\boldsymbol{x}^{(m)}(t) \in \mathbb{C}^N$, $\boldsymbol{s}^{(m)}(t) \in \mathbb{C}^R$ denote the mixture and source at time instant t, respectively, and $\boldsymbol{A}^{(m)} \in \mathbb{C}^{N\times R}$ is the mixing

matrix for the mth dataset. In practice, there is also a noise term. Here we omit it for convenience. N, R and M are the number of observations, sources, and datasets respectively. J-BSS aims at estimating the mixing matrices $\boldsymbol{A}^{(m)}$ and/or sources $\boldsymbol{s}^{(m)}(t)$ for $m = 1,\ldots,M$. In particular, if only the mixing matrices are of interest, the problem is called joint blind identification (J-BI). We say a J-BSS is underdetermined if $N < R$, and is overdetermined if $N \geq R$.

With the multi-set data model (2), we additionally assume that $s_r^{(m)}(t)$ and $s_u^{(n)}(t)$ are independent for $1 \leq r \neq u \leq R$ (intra-set independence) and dependent for $1 \leq r = u \leq R$ (inter-set dependence), regardless of the value of m and n. The idea is now to obtain a set of tensors by stacking second-order cross-covariance matrices. We assume that the sources are temporally nonstationary with zero mean and unit variance. Then we obtain a set of cross-covariance tensors as:

$$\begin{aligned}\left(\boldsymbol{\mathcal{T}}^{(m,n)}\right)_{(:,:,t)} &= \mathrm{E}\{\boldsymbol{x}^{(m)}(t)[\boldsymbol{x}^{(n)}(t)]^H\} \\ &= \boldsymbol{A}^{(m)}\,\mathrm{E}\{\boldsymbol{s}^{(m)}(t)[\boldsymbol{s}^{(n)}(t)]^H\}\boldsymbol{A}^{(n)H},\end{aligned} \tag{3}$$

where $\boldsymbol{\mathcal{T}}^{(m,n)} \in \mathbb{C}^{N\times N\times T}$, $1 \leq m,n \leq M$. We note that the frontal slice $(\boldsymbol{\mathcal{T}}^{(m,n)})_{(:,:,t)}$ is the cross-covariance matrix between the mth and nth datasets at time instant t, and that T denotes the number of time frames for which such a cross-covariance is computed. Observing that $\mathrm{E}\{\boldsymbol{s}^{(m)}(t)\cdot[\boldsymbol{s}^{(n)}(t)]^H\}$ is diagonal under the assumption of intra-set independence and inter-set dependence, we can rewrite (3) as follows:

$$\boldsymbol{\mathcal{T}}^{(m,n)} = \sum\nolimits_{r=1}^{R} \boldsymbol{a}_r^{(m)} \otimes \boldsymbol{a}_r^{(n)*} \otimes \boldsymbol{c}_r^{(m,n)} = \left[\!\left[\boldsymbol{A}^{(m)}, \boldsymbol{A}^{(n)*}, \boldsymbol{C}^{(m,n)} \right]\!\right]_R, \tag{4}$$

where $\boldsymbol{c}_r^{(m,n)}(t) = \mathrm{E}\{s_r^{(m)}(t)[s_r^{(n)}(t)]^H\}$, and $\boldsymbol{C}^{(m,n)} \triangleq [\boldsymbol{c}_1^{(m,n)},\ldots,\boldsymbol{c}_R^{(m,n)}]$. We see that for fixed m and n, (4) is by itself a PD. For varying m and n, each PD shares factor matrices with other PDs in the first two modes, which we denote as "double coupling". More precisely, if we place the tensors $\boldsymbol{\mathcal{T}}^{(m,n)}$ at different nodes of a grid according to their indices (see Fig. 1), we can see that all the tensors in the same "column" (e.g. the mth column) are coupled in the first mode (by $\boldsymbol{A}^{(m)}$), and all the tensors in the same "row" (e.g. the nth row) are coupled in the second mode (by $\boldsymbol{A}^{(n)*}$). For instance, for $M = 3$ we note that $\boldsymbol{\mathcal{T}}^{(1,1)}, \boldsymbol{\mathcal{T}}^{(1,2)}$, and $\boldsymbol{\mathcal{T}}^{(1,3)}$ are coupled in the first mode by factor matrix $\boldsymbol{A}^{(1)}$. Similarly, $\boldsymbol{\mathcal{T}}^{(1,3)}, \boldsymbol{\mathcal{T}}^{(2,3)}$, and $\boldsymbol{\mathcal{T}}^{(3,3)}$ are coupled in the second mode by $\boldsymbol{A}^{(3)*}$. We define the set of rank-1 terms $\{\boldsymbol{a}_r^{(m)} \otimes \boldsymbol{a}_r^{(n)*} \otimes \boldsymbol{c}_r^{(m,n)}, m,n = 1,\ldots,M\}$ as a double coupled rank-1 term, and the set of tensors $\{\boldsymbol{\mathcal{T}}^{(m,n)}, m,n = 1,\ldots,M\}$ is said to admit an R-term double coupled PD. If the number of double coupled rank-1 terms R in (4) is minimal, then the double coupled PD is denoted as double coupled canonical PD (DC-CPD). Our goal is to decompose the tensors $\boldsymbol{\mathcal{T}}^{(m,n)}$ into a minimal number of double coupled rank-1 terms, to recover the mixing matrices $\boldsymbol{A}^{(m)}$.

So far, we have shown how, via the computation of second-order statistics, the stochastic multi-set signals in (2) are transformed into a set of tensors, the joint decomposition of which takes the form of a DC-CPD (4). This method fits in the tensorization framework proposed in [60]. In line with the discussion in [60], the inherent intra-set independence and inter-set dependence of sources are key in the conversion, and have in fact already been used in existing practical multi-set signal processing applications, not necessarily in the J-BSS framework sensu stricto. For example, for linear convolutive mixtures that are transformed to the frequency domain, the cross-correlations of components at neighboring frequency bins are intensively exploited to solve the permutation misalignment problem [30], [39], [47], [63]. The above assumptions on multi-set statistics are also used in group analysis of biomedical data [29], [31]–[39] where the multiple datasets may refer to data of multiple modalities (e.g. fMRI, MEG, EEG) collected from a single subject under equal conditions, or data collected from multiple subjects, possibly using a single modality but acquired under the same conditions.

Remark 1: Note that the DC-CPD (4) derived from (3) has a conjugated symmetric structure: $\boldsymbol{\mathcal{T}}^{(m,n)} = \mathrm{perm}_{(2,1,3)}(\boldsymbol{\mathcal{T}}^{(n,m)*})$. In cases where this symmetry is not readily present, we are able to create it by tensor concatenation. Let us denote the tensors in an asymmetric DC-CPD as $\boldsymbol{\mathcal{T}}'^{(m,n)} = \left[\!\left[\boldsymbol{A}^{(m)}, \boldsymbol{A}^{(n)*}, \boldsymbol{C}'^{(m,n)} \right]\!\right]_R$, , and group these tensors into M sets $\Upsilon'^{(1)},\ldots,\Upsilon'^{(M)}$. Each set $\Upsilon'^{(m)}$ contains $2M$ tensors that have a common factor matrix $\boldsymbol{A}^{(m)}$, for fixed m:

$$\begin{aligned}\Upsilon'^{(m)} \triangleq \{&\boldsymbol{\mathcal{T}}'^{(m,1)}, \mathrm{perm}_{(2,1,3)}(\boldsymbol{\mathcal{T}}'^{(1,m)*}), \\ &\ldots, \boldsymbol{\mathcal{T}}'^{(m,M)}, \mathrm{perm}_{(2,1,3)}(\boldsymbol{\mathcal{T}}'^{(M,m)*})\}.\end{aligned} \tag{5}$$

The $2M$ tensors in $\Upsilon'^{(m)}$ can be further grouped into M pairs: $\{\boldsymbol{\mathcal{T}}'^{(m,n)}, \mathrm{perm}_{(2,1,3)}(\boldsymbol{\mathcal{T}}'^{(n,m)*})\}$, with fixed m and varying n. We concatenate each pair of tensors along the third mode into a new tensor that admits a PD, as:

$$\boldsymbol{\mathcal{T}}^{(m,n)} \triangleq \begin{cases} \mathrm{cat}\left(\boldsymbol{\mathcal{T}}'^{(m,n)}, \mathrm{perm}_{(2,1,3)}(\boldsymbol{\mathcal{T}}'^{(n,m)*})\right) \\ = \left[\!\left[\boldsymbol{A}^{(m)}, \boldsymbol{A}^{(n)*}, [\boldsymbol{C}'^{(m,n)}, \boldsymbol{C}'^{(n,m)*}] \right]\!\right]_R \end{cases} \quad m \leq n, \\ \begin{cases} \mathrm{cat}\left(\mathrm{perm}_{(2,1,3)}(\boldsymbol{\mathcal{T}}'^{(n,m)*}), \boldsymbol{\mathcal{T}}'^{(m,n)}\right) \\ = \left[\!\left[\boldsymbol{A}^{(m)}, \boldsymbol{A}^{(n)*}, [\boldsymbol{C}'^{(n,m)}, \boldsymbol{C}'^{(m,n)*}] \right]\!\right]_R \end{cases} \quad m > n. \tag{6}$$

The new set of tensors constructed by (6) admits a DC-CPD that has the conjugated symmetry:

$$\begin{cases} \boldsymbol{\mathcal{T}}^{(m,n)} = \left[\!\left[\boldsymbol{A}^{(m)}, \boldsymbol{A}^{(n)*}, \boldsymbol{C}^{(m,n)} \right]\!\right]_R, \\ \text{where } \boldsymbol{\mathcal{T}}^{(m,n)} = \mathrm{perm}_{(2,1,3)}(\boldsymbol{\mathcal{T}}^{(n,m)*}). \end{cases} \tag{7}$$

The concatenated tensors $\boldsymbol{\mathcal{T}}^{(m,n)}$ have the advantage over $\boldsymbol{\mathcal{T}}'^{(m,n)}$ that their third dimension is larger, which will make it easier to satisfy the full column rank assumption of the factor matrix in the third mode, as imposed by the derivation in the next section.

Because of the conjugated symmetry in DC-CPD, it suffices to consider the following set $\Upsilon^{(m)}$ of tensors to take into account all occurrences of $\boldsymbol{A}^{(m)}$, for fixed m:

$$\Upsilon^{(m)} = \left\{ \boldsymbol{\mathcal{T}}^{(m,n)} = \left[\!\left[\boldsymbol{A}^{(m)}, \boldsymbol{A}^{(n)*}, \boldsymbol{C}^{(m,n)} \right]\!\right]_R, n = 1,\ldots,M \right\}. \tag{8}$$

Note that in (8), seen as C-CPD, occurrences of $\boldsymbol{A}^{(m)*}$ in the second mode are taken care of automatically because of the conjugated symmetry.

Remark 2: We note that DC-CPD (7) differs from existing

C-CPD works in the following aspects: (i) DC-CPD (7) has a double coupling structure in the first two modes while C-CPD in [45]–[58] consider coupling in a single mode; (ii) the common factor matrix in DC-CPD (7) corresponds to the mixing matrix in J-BSS (2), and does not have full column rank if the J-BSS problem is underdetermined. The existing C-CPD works all assume that the common factor matrix has full column rank. In [59] it is assumed that at least one of the common factor matrices in a DC-CPD has full column rank. In the following context, we say a DC-CPD derived from an over-determined J-BSS ($N \geq R$) is an overdetermined DC-CPD. On the other hand, a DC-CPD is said to be underdetermined if it corresponds to an underdetermined J-BSS ($N < R$).

III. Algebraic Algorithm

In this section, we introduce an algebraic algorithm for DC-CPD. We formally call an algorithm algebraic if it relies only on arithmetic operations, overdetermined sets of linear equations, matrix singular value decomposition (SVD) and generalized eigenvalue decomposition (GEVD). In particular, an algebraic computation does not involve a numerical optimization in which the cost function may have multiple local optima, or of which the convergence is not guaranteed.

A. Overall approach

The algebraic DC-CPD algorithm fits in the framework of the algebraic algorithms in [10], [12], [13], [19], [20], [51], [52], which, by the use of a rank-1 or rank-R detection mapping, convert a possibly underdetermined (coupled) CPD into an overdetermined CPD. The latter can be solved exactly by means of a GEVD. Here we start from the formulation (8). The derivation follows a similar line of thought and consists of the following main steps:

(i) We will extend the rank-1 detection mapping [10], [12], [13], [19], [20] to a coupled rank-1 detection mapping, and use it to obtain a matrix $\boldsymbol{\Gamma}^{(m,n_1,\ldots,n_K)}$, for any K tensors in $\Upsilon^{(m)}$, e.g. $\boldsymbol{\mathcal{T}}^{(m,n_1)},\ldots,\boldsymbol{\mathcal{T}}^{(m,n_K)}$ that have a common factor matrix $\boldsymbol{A}^{(m)}$, $m = 1,\ldots,M$, $1 \leq n_1 < \cdots < n_K \leq M$, $K = 2,\ldots,M$.

(ii) In the null space of $\boldsymbol{\Gamma}^{(m,n_1,\ldots,n_K)}$, we will find a set of basis vectors that can be reshaped into a tensor $\boldsymbol{\mathcal{W}}^{(m,n_1,\ldots,n_K)}$. This tensor admits an overdetermined CPD, involving $\boldsymbol{B}^{(m,n_1)},\ldots,\boldsymbol{B}^{(m,n_K)}$ as factor matrices, which are related to the inverse of $\boldsymbol{C}^{(m,n_1)},\ldots,\boldsymbol{C}^{(m,n_K)}$.

(iii) We will perform steps (i) and (ii) for all possible K-combinations of tensors in $\Upsilon^{(m)}$ (8) to construct a set of tensors $\boldsymbol{\mathcal{W}}^{(m,n_1,\ldots,n_K)}$. For varying indices m, $n_1,\cdots,n_K$, these tensors overall admit an overdetermined C-CPD with coupling in multiple modes.

(iv) We will compute the overdetermined C-CPD to obtain the factor matrices $\boldsymbol{B}^{(m,n)}$, m, $n = 1,\ldots,M$. This can be done algebraically by means of a GEVD. It can also be done by optimization based algorithms such as those in [38], [39]. The matrix $\boldsymbol{C}^{(m,n)}$ can be calculated as the inverse of $\boldsymbol{B}^{(m,n)}$ up to trivial indeterminancies. Once $\boldsymbol{C}^{(m,n)}$ is known, the remaining factor matrices $\boldsymbol{A}^{(m)}$ can be recovered using the DC-CPD structure of the data tensors.

B. Assumptions

We assume both $\boldsymbol{A}^{(m)} \odot \boldsymbol{A}^{(n)*}$ and $\boldsymbol{C}^{(m,n)}$ have full column rank, for all the indices m, n. Together these assumptions imply that the rank of tensor $\boldsymbol{\mathcal{T}}^{(m,n)}$ is equal to the rank of the mode-3 matrix representation $\boldsymbol{T}_3^{(m,n)}$. This is of practical use in the sense that the number of sources can be determined by checking the number of significant singular values of $\boldsymbol{T}_3^{(m,n)}$. We note that DC-CPD (8) is not unique if $\boldsymbol{A}^{(m)} \odot \boldsymbol{A}^{(n)*}$ does not have full column rank. As a matter of fact, in this case a decomposition in a smaller number of terms is possible. For instance, if $\boldsymbol{a}_R^{(m)} \otimes \boldsymbol{a}_R^{(n)*} = \sum_{r=1}^{R-1} \alpha_r^{(m,n)} \boldsymbol{a}_r^{(m)} \otimes \boldsymbol{a}_r^{(n)*}$, then $\boldsymbol{\mathcal{T}}^{(m,n)} = \sum_{r=1}^{R-1} \boldsymbol{a}_r^{(m)} \otimes \boldsymbol{a}_r^{(n)*} \otimes (\boldsymbol{c}_r^{(m,n)} + \alpha_r^{(m,n)} \boldsymbol{c}_R^{(m,n)})$. On the other hand, if $\boldsymbol{C}^{(m,n)}$ does not have full column rank, decomposition (4) may still be unique (for instance, algebraic algorithms for CPD have been derived for cases where none of the factor matrices has full column rank [12], [13]), while the rank of tensor $\boldsymbol{\mathcal{T}}^{(m,n)}$ is not equal to the rank of the mode-3 matrix representation $\boldsymbol{T}_3^{(m,n)}$. Here we do not consider this more complicated case. In addition, we make the notational assumption that the factor matrices $\boldsymbol{C}^{(m,n)}$ in the third mode have size $R \times R$. In practice, this can always be achieved by a classical dimensionality reduction step: if the columns of a matrix $\boldsymbol{U}^{(m,n)}$ form an orthonormal basis of the row space of $\boldsymbol{T}_3^{(m,n)}$, then $\boldsymbol{T}_3^{(m,n)}\boldsymbol{U}^{(m,n)}$ has R columns, tensor $\mathrm{Ten}(\boldsymbol{T}_3^{(m,n)}\boldsymbol{U}^{(m,n)}, I, I, R)$ has reduced dimensionality R in the third mode, and its factor matrices are $\boldsymbol{A}^{(m)}$, $\boldsymbol{A}^{(n)*}$, and $\boldsymbol{U}^{(m,n)T}\boldsymbol{C}^{(m,n)} \in \mathbb{C}^{R \times R}$.

C. Coupled rank-1 detection

We give the following two definitions and a theorem.

Definition 1: Matrices $\boldsymbol{X}^{(1)} \in \mathbb{C}^{N \times J_1},\ldots,\boldsymbol{X}^{(K)} \in \mathbb{C}^{N \times J_K}$ are said to be coupled rank-1 matrices if they are all rank-1 and have the same column space.

Definition 2: The coupled rank-1 detection mapping $\Phi_{(k_1,k_2)}$: $(\boldsymbol{X}^{(1)},\ldots,\boldsymbol{X}^{(K)}) \in \mathbb{C}^{N \times J_1} \times \cdots \times \mathbb{C}^{N \times J_K} \rightarrow \Phi_{(k_1,k_2)}(\boldsymbol{X}^{(1)},\ldots,\boldsymbol{X}^{(K)}) \in \mathbb{C}^{N \times \cdots \times N \times J_1 \times \cdots \times J_K}$ is defined by:

$$\left[\Phi_{(k_1,k_2)}(\boldsymbol{X}^{(1)},\ldots,\boldsymbol{X}^{(K)})\right]_{i_1 \ldots i_K, j_1 \ldots j_K} \triangleq \begin{vmatrix} x_{i_{k_1},j_{k_1}}^{(k_1)} & x_{i_{k_1},j_{k_2}}^{(k_2)} \\ x_{i_{k_2},j_{k_1}}^{(k_1)} & x_{i_{k_2},j_{k_2}}^{(k_2)} \end{vmatrix} \prod_{l \notin \{k_1,k_2\}} x_{i_l,j_l}^{(l)}, \quad (9)$$

with $1 \leq k_1 < k_2 \leq K$. The number of arguments of $\Phi_{(k_1,k_2)}$, K, is called the order, and the subscript (k_1, k_2) is called the index of the coupled rank-1 detection mapping, respectively.

Theorem 1: For K nonzero matrices $\boldsymbol{X}^{(k)} \in \mathbb{C}^{N \times J_k}$, $k = 1,\ldots,K$, consider the vector $\psi(\boldsymbol{X}^{(1)},\ldots,\boldsymbol{X}^{(K)}) \in \mathbb{C}^{\binom{K}{2} N^K \prod_{k=1}^K J_k}$ obtained via the coupled rank-1 detection mapping (9):

$$\psi\left(\boldsymbol{X}^{(1)},\ldots,\boldsymbol{X}^{(K)}\right) \triangleq \begin{bmatrix} \mathrm{vec}\left(\Phi_{(1,2)}(\boldsymbol{X}^{(1)},\ldots,\boldsymbol{X}^{(K)})\right) \\ \mathrm{vec}\left(\Phi_{(1,3)}(\boldsymbol{X}^{(1)},\ldots,\boldsymbol{X}^{(K)})\right) \\ \vdots \\ \mathrm{vec}\left(\Phi_{(K-1,K)}(\boldsymbol{X}^{(1)},\ldots,\boldsymbol{X}^{(K)})\right) \end{bmatrix}. \quad (10)$$

Then $\psi(\boldsymbol{X}^{(1)},\ldots,\boldsymbol{X}^{(K)})$ is a zero vector if and only if its arguments $\boldsymbol{X}^{(1)},\ldots,\boldsymbol{X}^{(K)}$ are coupled rank-1 matrices.

The proof of *Theorem 1* is given in the Appendix. We note that the above definitions and theorem generalize the rank-1

detection mapping and relevant results [10], [12], [13], [19], [20], [51], [52] to the coupled case.

Next we explain how to use the coupled rank-1 detection mapping to convert a possibly underdetermined DC-CPD into an overdetermined C-CPD. For clarity, we first explain the simplest case where $K = 2$, and then address the general case where $K \geq 2$.

D. The simplest case: K = 2

When $K = 2$, the coupled rank-1 detection mapping defined in (9) simplifies to:

$$\left[\Phi_{(1,2)}(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})\right]_{i_1,i_2,j_1,j_2} \triangleq \begin{vmatrix} x^{(1)}_{i_1,j_1} & x^{(2)}_{i_1,j_2} \\ x^{(1)}_{i_2,j_1} & x^{(2)}_{i_2,j_2} \end{vmatrix} = x^{(1)}_{i_1,j_1}x^{(2)}_{i_2,j_2} - x^{(1)}_{i_2,j_1}x^{(2)}_{i_1,j_2}. \quad (11)$$

The vector $\psi(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})$ defined in (10) is now equal to $\mathrm{vec}(\Phi_{(1,2)}(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}))$. From *Theorem 1* we know that $\Phi_{(1,2)}(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})$ is a zero tensor if and only if $\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}$ are coupled rank-1 matrices. We now briefly explain how (11) can be used for computing the C-CPD of two tensors $\boldsymbol{\mathcal{T}}^{(m,n_1)}$ and $\boldsymbol{\mathcal{T}}^{(m,n_2)}$ in $\Upsilon^{(m)}$.

First, we use the operation $\Phi_{(1,2)}(\cdot)$ on the sth and uth frontal slice of $\boldsymbol{\mathcal{T}}^{(m,n_1)},\boldsymbol{\mathcal{T}}^{(m,n_2)}$, respectively, $1 \leq s,u \leq R$:

$$\left[\Phi_{(1,2)}\left(\boldsymbol{T}^{(m,n_1)}_{(:,:,s)},\boldsymbol{T}^{(m,n_2)}_{(:,:,u)}\right)\right]_{i_1,i_2,j_1,j_2} = t^{(m,n_1)}_{i_1,j_1,s}t^{(m,n_2)}_{i_2,j_2,u} - t^{(m,n_1)}_{i_2,j_1,s}t^{(m,n_2)}_{i_1,j_2,u}. \quad (12)$$

Since $\Phi_{(1,2)}(\cdot)$ is bilinear, we have:

$$\begin{aligned}\psi\left(\boldsymbol{T}^{(m,n_1)}_{(:,:,s)},\boldsymbol{T}^{(m,n_2)}_{(:,:,u)}\right) &= \mathrm{vec}\left(\Phi_{(1,2)}\left(\boldsymbol{T}^{(m,n_1)}_{(:,:,s)},\boldsymbol{T}^{(m,n_2)}_{(:,:,u)}\right)\right) \\ &= \sum_{r,t=1}^{R} c^{(m,n_1)}_{s,t}c^{(m,n_2)}_{u,r}\psi[\boldsymbol{a}^{(m)}_t(\boldsymbol{a}^{(n_1)}_t)^H,\boldsymbol{a}^{(m)}_r(\boldsymbol{a}^{(n_2)}_r)^H] \\ &= \sum_{\substack{r,t=1 \\ r\neq t}}^{R} c^{(m,n_1)}_{s,t}c^{(m,n_2)}_{u,r}\psi[\boldsymbol{a}^{(m)}_t(\boldsymbol{a}^{(n_1)}_t)^H,\boldsymbol{a}^{(m)}_r(\boldsymbol{a}^{(n_2)}_r)^H].\end{aligned} \quad (13)$$

In (13) we have used the property that $\psi[\boldsymbol{a}^{(m)}_t(\boldsymbol{a}^{(n_1)}_t)^H,\boldsymbol{a}^{(m)}_t(\boldsymbol{a}^{(n_2)}_t)^H] = \boldsymbol{0}$. For fixed m, n_1, n_2, we assume that $\psi[\boldsymbol{a}^{(m)}_t(\boldsymbol{a}^{(n_1)}_t)^H,\boldsymbol{a}^{(m)}_r(\boldsymbol{a}^{(n_2)}_r)^H]$ are linearly independent, $t \neq r$. With a short derivation analogous to equation (2.9)–(2.13) in [10], we obtain the result that there exist precisely R linearly independent matrices $\boldsymbol{V}^{(m,n_1,n_2)}_l \in \mathbb{C}^{R\times R}$, $l = 1,\ldots,R$, such that:

$$\sum_{s,u=1}^{R} (\boldsymbol{V}^{(m,n_1,n_2)}_l)_{s,u}\psi(\boldsymbol{\mathcal{T}}^{(m,n_1)}_{:,:,s},\boldsymbol{\mathcal{T}}^{(m,n_2)}_{:,:,u}) = \boldsymbol{0}. \quad (14)$$

The matrices $\boldsymbol{B}^{(m,n_1)},\boldsymbol{B}^{(m,n_2)}$ jointly diagonalize $\boldsymbol{V}^{(m,n_1,n_2)}_l$:

$$\boldsymbol{V}^{(m,n_1,n_2)}_l = \boldsymbol{B}^{(m,n_1)} \cdot \boldsymbol{\Lambda}^{(m,n_1,n_2)}_l \cdot \boldsymbol{B}^{(m,n_2)T}, \quad (15)$$

where $\boldsymbol{\Lambda}^{(m,n_1,n_2)}_l$ is a diagonal matrix, $\boldsymbol{B}^{(m,n_k)} = \tilde{\boldsymbol{C}}^{(m,n_k)-T} \in \mathbb{C}^{R\times R}$, $\tilde{\boldsymbol{C}}^{(m,n_k)} \triangleq (\boldsymbol{C}^{(m,n_k)}\boldsymbol{D}^{(m,n)}_C\boldsymbol{\Pi})$, $k = 1,2$, $\boldsymbol{D}^{(m,n)}_C$ is a diagonal matrix, and $\boldsymbol{\Pi}$ is a common permutation matrix for $\boldsymbol{B}^{(m,n_1)},\boldsymbol{B}^{(m,n_2)}$.

The matrices $\boldsymbol{V}^{(m,n_1,n_2)}_l$ are obtained by solving the set of linear equations (14). We stack the matrices $\boldsymbol{V}^{(m,n_1,n_2)}_l$ along the third mode into a tensor: $(\boldsymbol{\mathcal{W}}^{(m,n_1,n_2)})_{:,:,l} \triangleq \boldsymbol{V}^{(m,n_1,n_2)}_l$. Then we know by (15) that, for fixed m, n_1, n_2, $\boldsymbol{\mathcal{W}}^{(m,n_1,n_2)}$ admits a third-order CPD:

$$\boldsymbol{\mathcal{W}}^{(m,n_1,n_2)} = \left[\!\left[\boldsymbol{B}^{(m,n_1)},\boldsymbol{B}^{(m,n_2)},\boldsymbol{F}^{(m,n_1,n_2)}\right]\!\right]_R, \quad (16)$$

in which $\boldsymbol{F}^{(m,n_1,n_2)} \in \mathbb{C}^{R\times R}$ holds the sth diagonal entry of $\boldsymbol{\Lambda}^{(m,n_1,n_2)}_l$ as its (l,s)th entry. Note that the factor matrices $\boldsymbol{B}^{(m,n_k)}$ and $\boldsymbol{F}^{(m,n_1,n_2)}$ have full column rank, $k = 1, 2$.

In (12)–(16), we have used the coupled rank-1 detection mapping (11) to convert the decomposition of a pair of coupled tensors into an asymmetric overdetermined third-order CPD. For the overall DC-CPD (8), we perform the same procedure for all pairs of coupled tensors and obtain a set of overdetermined CPDs. With varying indices (m, n_1, n_2), $m = 1,\ldots,M$, $1 \leq n_1 < n_2 \leq M$, these CPDs together form an overdetermined third-order DC-CPD with coupling in the first two modes.

As the new DC-CPD is overdetermined, it admits an algebraic solution. There are different ways to compute this solution. First, as all the factor matrices have full column rank, any of the CPDs (16) can be computed by GEVD [62]. This yields us the factor matrices $\boldsymbol{B}^{(m,n_1)}$ and $\boldsymbol{B}^{(m,n_2)}$ up to trivial scaling and permutation indeterminacies, for any choice of m, n_1 and n_2. Assume for instance that $\boldsymbol{B}^{(m,1)}$ has been found, then we can find all remaining $\boldsymbol{B}^{(m,n_2)}$ and $\boldsymbol{F}^{(m,1,n_2)}$ from the CPD of $\boldsymbol{\mathcal{W}}^{(m,1,n_2)}$ for fixed m and varying n_2. In fact, since in these CPDs the factor matrix $\boldsymbol{B}^{(m,1)}$ is already known, the other factor matrices follow from rank-1 approximation of the matricized columns of [22]:

$$\boldsymbol{W}^{(m,1,n_2)}_1\boldsymbol{B}^{(m,1)-T} = \boldsymbol{B}^{(m,n_2)} \odot \boldsymbol{F}^{(m,1,n_2)}. \quad (17)$$

Indeed, each column of $\boldsymbol{B}^{(m,n_2)} \odot \boldsymbol{F}^{(m,1,n_2)}$ is a vectorized rank-1 matrix:

$$\mathrm{unvec}(\boldsymbol{q}^{(m,1,n_2)}_{1,r}) = \boldsymbol{b}^{(m,n_2)}_r \cdot \boldsymbol{f}^{(m,1,n_2)T}_r, \quad r = 1,\ldots,R, \quad (18)$$

where $\boldsymbol{q}^{(m,1,n_2)}_{1,r} \in \mathbb{C}^{R^2}$ is the rth column of $\boldsymbol{W}^{(m,1,n_2)}_1\boldsymbol{B}^{(m,1)-T}$. Continuing this way, all factor matrices $\boldsymbol{B}^{(m,n)}$ can be determined, $m, n = 1,\ldots,M$.

Now that all the overdetermined CPDs (16) have been computed, we obtain $\tilde{\boldsymbol{C}}^{(m,n)} = \boldsymbol{C}^{(m,n)}\boldsymbol{D}^{(m,n)}_C\boldsymbol{\Pi} = \boldsymbol{B}^{(m,n)-T}$, and $\tilde{\boldsymbol{A}}^{(m)} \odot \tilde{\boldsymbol{A}}^{(n)*} = \boldsymbol{T}^{(m,n)}_3\boldsymbol{B}^{(m,n)}$, where $\tilde{\boldsymbol{A}}^{(m)} = \boldsymbol{A}^{(m)}\boldsymbol{D}^{(m)}_A\boldsymbol{\Pi}$ and $\tilde{\boldsymbol{A}}^{(n)} = \boldsymbol{A}^{(n)}\boldsymbol{D}^{(n)}_A\boldsymbol{\Pi}$ are estimates of $\boldsymbol{A}^{(m)}$ and $\boldsymbol{A}^{(n)}$ up to scaling and permutation ambiguities, $m, n = 1,\ldots,M$. Matrices $\boldsymbol{D}^{(m)}_A$ and $\boldsymbol{D}^{(n)}_A$ are diagonal matrices, and $\boldsymbol{D}^{(m)}_A\boldsymbol{D}^{(n)*}_A\boldsymbol{D}^{(m,n)}_C = \boldsymbol{I}_R$. We define $\boldsymbol{G}^{(m,n)}_r \triangleq \mathrm{unvec}(\boldsymbol{T}^{(m,n)}_3\boldsymbol{B}^{(m,n)})_{(:,r)}$, and collect these matrices in a matrix $\boldsymbol{G}_r \in \mathbb{C}^{N^2\times N^2}$ as follows:

$$\boldsymbol{G}_r = \begin{bmatrix} \boldsymbol{G}^{(1,1)}_r & \cdots & \boldsymbol{G}^{(1,M)}_r \\ \vdots & \ddots & \vdots \\ \boldsymbol{G}^{(M,1)}_r & \cdots & \boldsymbol{G}^{(M,M)}_r \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{a}}^{(1)}_r \\ \vdots \\ \tilde{\boldsymbol{a}}^{(M)}_r \end{bmatrix} \cdot [\tilde{\boldsymbol{a}}^{(1)H}_r,\ldots,\tilde{\boldsymbol{a}}^{(M)H}_r], \quad (19)$$

where $\tilde{\boldsymbol{a}}^{(m)}_r$ denotes the rth column of $\tilde{\boldsymbol{A}}^{(m)}$, $r = 1,\ldots,R$, $m = 1,\ldots,M$. We can calculate $[\tilde{\boldsymbol{a}}^{(1)H}_r,\ldots,\tilde{\boldsymbol{a}}^{(M)H}_r]^H$ as the dominant left singular vector of $\boldsymbol{G}_r$, or equivalently as the dominant right singular vector up to a conjugation.

The algebraic approach is guaranteed to return the exact solution in the exact case. When noise is present, conventional linear algebra operations in some of the steps, such as the solution to the set of linear equations and the matrix best rank-1 approximation, can by themselves be found in LS sense. The overall result, however, is not guaranteed to be optimal and is actually (mildly) sub-optimal in practice. To estimate the

optimal result, we can fit the DC-CPD in LS sense by an optimization algorithm. Several options are: (i) the framework of SDF is well-suited for the task; (ii) ALS is a specific type of optimization algorithm that can be used. The updating equations can be explicitly derived in analogy with the derivation in the supplementary materials for DC-CPD. Note that the algebraic approach can be used to efficiently initialize these optimization algorithms. The above discussion holds for both the overdetermined DC-CPD (16) and the original DC-CPD (7).

We note that the derivations here are analogous to the algebraic (coupled) CPD algorithms in the literature. In particular, the pairwise coupled rank-1 detection mapping $\Phi_{(1,2)}(\cdot)$ in (11) can be considered as a variant of the rank-1 detection mapping, originally proposed in [61] for ICA and subsequently adopted in [10] for computing CPD. The main results and derivations (12)–(16) follow a similar reasoning to those in [10] and [49]. The main difference between the proposed algorithm and existing algebraic CPD and C-CPD algorithms is that the proposed algorithm uses coupled rank-1 detection mapping while existing ones are based on rank-1 detection mapping.

The above derivation of the algebraic DC-CPD algorithm provides a constructive proof of the following theorem.

Theorem 2: Let $\boldsymbol{\mathcal{T}}^{(m,n_k)} = \left[\!\left[\boldsymbol{A}^{(m)}, \boldsymbol{A}^{(n_k)*}, \boldsymbol{C}^{(m,n_k)} \right]\!\right]_R$, $k = 1, 2$. We assume that $\boldsymbol{C}^{(m,n_k)}$ has full column rank for all $1 \le n_1 < n_2 \le M$, and $m = 1,...,M$, and that $\psi(\boldsymbol{a}_t^{(m)}(\boldsymbol{a}_t^{(n_1)})^H, \boldsymbol{a}_r^{(m)}(\boldsymbol{a}_r^{(n_2)})^H) = (\boldsymbol{a}_t^{(m)} \otimes \boldsymbol{a}_r^{(m)} - \boldsymbol{a}_r^{(m)} \otimes \boldsymbol{a}_t^{(m)}) \otimes \boldsymbol{a}_t^{(n_1)*} \otimes \boldsymbol{a}_r^{(n_2)*}$ are linearly independent for $1 \le t \ne r \le R$. Then we have:

(i) The DC-CPD consists of the sum of R double coupled rank-1 terms, i.e. the number of terms cannot be reduced.

(ii) The DC-CPD is unique.

(iii) The DC-CPD can be calculated algebraically.

Theorem 2 provides deterministic conditions for which DC-CPD is unique and for which DC-CPD can be calculated algebraically for $K = 2$. We also provide the generic value of the upper bound of R, denoted as $R_{\max}$. We call a property generic if it holds with probability one, when the parameters it involves are drawn from continuous probability densities. Generically, the matrix $\boldsymbol{C}^{(m,n)} \in \mathbb{C}^{R \times R}$ has full column rank. Hence, the generic version of the uniqueness condition in *Theorem 2* depends only on N. We give the generic value of $R_{\max}$ for $K = 2$, for a number of different N in Table I. The generic value of $R_{\max}$ of DC-CPD for $K = 3$ is also included in Table I, but will be addressed later. For comparison, we also list the generic value of $R_{\max}$ for CPD for $K = 2$ [10], and $K = 3$ [13], respectively.

The numerical values of $R_{\max}$ can be easily obtained using *Theorem 2* and Fisher's lemma (Corollary in p.10 of [66]). Fisher's lemma implies that, for fixed N and R, DC-CPD is generically unique if we can find one example for which the decomposition is unique. Hence, one only needs to check if the vectors $\psi(\boldsymbol{a}_t^{(m)}(\boldsymbol{a}_t^{(n_1)})^H, \boldsymbol{a}_r^{(m)}(\boldsymbol{a}_r^{(n_2)})^H)$ are linearly independent for $1 \le t \ne r \le R$, for a set of randomly generated factor matrices $\boldsymbol{A}^{(1)},...,\boldsymbol{A}^{(M)}$.

We summarize the algebraic algorithm for $K = 2$ in Table II.

TABLE I
GENERIC VALUE OF $R_{\max}$ OF DC-CPD AND CPD ($K = 2, 3$)

N	2	3	4	5	6
DC-CPD ($K = 2$)	2	5	10	16	23
DC-CPD ($K = 3$)	2	6	11	18	27
CPD ($K = 2$)	2	4	9	14	21
CPD ($K = 3$)	2	4	9	16	24

TABLE II
DC-CPD: ALGEBRAIC ALGORITHM FOR $K = 2$

Input: $\{\boldsymbol{\mathcal{T}}^{(m,n)} \in \mathbb{C}^{N \times N \times R} | 1 \le m, n \le M\}$ admitting DC-CPD (7).

1: Group $\boldsymbol{\mathcal{T}}^{(m,n)}$ into M sets $\Upsilon^{(m)}$ by (8).

2: For each pair of tensors $\boldsymbol{\mathcal{T}}^{(m,n_1)}, \boldsymbol{\mathcal{T}}^{(m,n_2)} \in \Upsilon^{(m)}$, perform operation (13) defined through second-order coupled rank-1 detection mapping (12), for all the frontal slices $\boldsymbol{\mathcal{T}}_{:,:,s}^{(m,n_1)}, \boldsymbol{\mathcal{T}}_{:,:,u}^{(m,n_2)}$, to obtain $\psi(\boldsymbol{\mathcal{T}}_{:,:,s}^{(m,n_1)}, \boldsymbol{\mathcal{T}}_{:,:,u}^{(m,n_2)})$, $1 \le s, u \le R$.

3: Solve the set of linear equations (14) to obtain $\boldsymbol{V}_l^{(m,n_1,n_2)}$, $l = 1,...,R$. Stack $\boldsymbol{V}_l^{(m,n_1,n_2)}$ with varying l into a tensor $\boldsymbol{\mathcal{W}}^{(m,n_1,n_2)}$.

4: For all the pairs of tensors in $\Upsilon^{(m)}$ and $m = 1,...,M$, the tensors $\boldsymbol{\mathcal{W}}^{(m,n_1,n_2)}$ together admit the overdetermined DC-CPD (16). Solve (16) to compute the matrices $\boldsymbol{B}^{(m,n)}$. Then we have $\tilde{\boldsymbol{C}}^{(m,n)} = \boldsymbol{B}^{(m,n)-T}$ and $\tilde{\boldsymbol{A}}^{(m)} \odot \tilde{\boldsymbol{A}}^{(n)*} = \boldsymbol{T}_3^{(m,n)} \boldsymbol{B}^{(m,n)}$.

5: Compute matrix $\boldsymbol{G}_r$ by (19). Then the rth column of the factor matrices $\tilde{\boldsymbol{A}}^{(m)}$ can be computed as the left dominant singular vector of $\boldsymbol{G}_r$, $m = 1,...,M$, $r = 1,...,R$.

Output: Estimates of the factor matrices $\tilde{\boldsymbol{A}}^{(m)}$ and $\tilde{\boldsymbol{C}}^{(m,n)}$, $1 \le m, n \le M$.

E. The general case: $K \in [2, M]$

First, we choose a constant $K \in [2, M]$, which denotes the order of the coupled rank-1 detection mapping. A larger K generally relaxes the uniqueness condition, but also increases the computational complexity, as will be explained later. In practice one may start with $K = 2$, and increase its value if the number of sources that can be handled is not sufficient. When K is chosen, we select K tensors $\boldsymbol{\mathcal{T}}^{(m,n_1)},...,\boldsymbol{\mathcal{T}}^{(m,n_K)}$ from $\Upsilon^{(m)}$ (8), $1 \le n_1 < \cdots < n_K \le M$. There are C_M^K such K-combinations and we perform for each K-combination the following steps, from the next paragraph until equation (24).

For each K-combination $\boldsymbol{\mathcal{T}}^{(m,n_1)},...,\boldsymbol{\mathcal{T}}^{(m,n_K)} \in \Upsilon^{(m)}$, we perform the operation (10) on all the possible combinations of the frontal slices of $\boldsymbol{\mathcal{T}}^{(m,n_1)},...,\boldsymbol{\mathcal{T}}^{(m,n_K)}$ to obtain a (K+1)th-order tensor $\boldsymbol{\mathcal{G}}^{(m,n_1,...,n_K)} \in \mathbb{C}^{\mathrm{C}_K^2 N^{2K} \times R \times \cdots \times R}$:

$$\left[\boldsymbol{\mathcal{G}}^{(m,n_1,...,n_K)}\right]_{(:,r_1,...,r_K)} \triangleq \psi\left(\boldsymbol{\mathcal{T}}_{(:,:,r_1)}^{(m,n_1)},...,\boldsymbol{\mathcal{T}}_{(:,:,r_K)}^{(m,n_K)}\right), \tag{20}$$

where $1 \le r_k \le R$, $1 \le k \le K$. Note that operation (20) implicitly uses the Kth-order coupled rank-1 detection mapping (9). We reshape $\boldsymbol{\mathcal{G}}^{(m,n_1,...,n_K)}$ into a matrix $\boldsymbol{\Gamma}^{(m,n_1,...,n_K)} \in \mathbb{C}^{\mathrm{C}_K^2 N^{2K} \times R^K}$ as:

$$\left[\boldsymbol{\Gamma}^{(m,n_1,...,n_K)}\right]_{(l,:)} \triangleq \mathrm{vec}\left(\left[\boldsymbol{\mathcal{G}}^{(m,n_1,...,n_K)}\right]_{(l,:,...,:)}\right)^T, \tag{21}$$

where $1 \le l \le \mathrm{C}_K^2 N^{2K}$. Due to the multilinearity of ψ, and by making use of the property that $\psi(\boldsymbol{a}_{r_1}^{(m)}\boldsymbol{a}_{r_1}^{(n_1)H},...,\boldsymbol{a}_{r_K}^{(m)}\boldsymbol{a}_{r_K}^{(n_K)H}) = \boldsymbol{0}$ when $r_1 = \cdots = r_K$ (*Theorem 1*), we have the following result after a short derivation given in the supplementary materials:

$$\boldsymbol{\Gamma}^{(m,n_1,...,n_K)} = \boldsymbol{\Phi}_{n_1,...,n_K}^{(m)} \left(\boldsymbol{P}_{n_1,...,n_K}^{(m)}\right)^T, \tag{22}$$

where matrices $\boldsymbol{P}^{(m)}_{m,\dots,n_K} \in \mathbb{C}^{R^K \times (R^K - R)}$ and $\boldsymbol{\Phi}^{(m)}_{m,\dots,n_K} \in \mathbb{C}^{C_K^2 N^{2K} \times (R^K - R)}$ have $\boldsymbol{c}^{(m,n_1)}_{r_1} \otimes \cdots \otimes \boldsymbol{c}^{(m,n_K)}_{r_K}$ and $\psi(\boldsymbol{a}^{(m)}_{r_1}\boldsymbol{a}^{(n_1)H}_{r_1},\dots,\boldsymbol{a}^{(m)}_{r_K}\boldsymbol{a}^{(n_K)H}_{r_K})$ as their columns, respectively. These columns are indexed by tuples $(r_1,\dots,r_K) \in \Theta^{(K)}$, where $\Theta^{(K)}$ is a set of tuples $(r_1,\dots,r_K)$, $1 \le r_1 \le \cdots \le r_K \le R$, in which each tuple contains at least two distinct elements.

Next, we give a theorem implying that the basis vectors in the null space of $\boldsymbol{\Gamma}^{(m,n_1,\dots,n_K)}$ can be reshaped into a tensor that admits an overdetermined (K+1)th-order CPD.

Theorem 3: Let $\boldsymbol{\mathcal{T}}^{(m,n_k)} = \left[\!\left[\boldsymbol{A}^{(m)}, \boldsymbol{A}^{(n_k)*}, \boldsymbol{C}^{(m,n_k)} \right]\!\right]_R \in \Upsilon^{(m)}$, k = 1,…,K, $1 \le n_1 < \cdots < n_K \le M$, and assume that $\boldsymbol{\Phi}^{(m)}_{n_1,\dots,n_K}$ in (22) and $\boldsymbol{C}^{(m,n_k)}$ for all values of k have full column rank for fixed $(m,n_1,\dots,n_K)$, then we have:

(i) $\ker(\boldsymbol{\Gamma}^{(m,n_1,\dots,n_K)}) = \ker((\boldsymbol{P}^{(m)}_{n_1,\dots,n_K})^T)$.

(ii) $\dim(\ker(\boldsymbol{\Gamma}^{(m,n_1,\dots,n_K)})) = R$.

(iii) the basis vectors in $\ker(\boldsymbol{\Gamma}^{(m,n_1,\dots,n_K)})$, denoted as $\boldsymbol{w}^{(m,n_1,\dots,n_k)}_r \in \mathbb{C}^{R^K}$, r = 1,…R, can be written as linear combinations of the vectors $\boldsymbol{b}^{(m,n_1)}_u \otimes \cdots \otimes \boldsymbol{b}^{(m,n_K)}_u$, u = 1,…,R:

$$\boldsymbol{w}^{(m,n_1,\dots,n_K)}_r = \sum\nolimits_{u=1}^{R} f^{(m,n_1,\dots,n_K)}_{r,u} \cdot \boldsymbol{b}^{(m,n_1)}_u \otimes \cdots \otimes \boldsymbol{b}^{(m,n_K)}_u, \qquad (23)$$

where $[\boldsymbol{b}^{(m,n_k)}_1,\dots,\boldsymbol{b}^{(m,n_k)}_R] = \boldsymbol{B}^{(m,n_k)} = \tilde{\boldsymbol{C}}^{(m,n_k)-T} \in \mathbb{C}^{R \times R}$, $\tilde{\boldsymbol{C}}^{(m,n_k)} \triangleq (\boldsymbol{C}^{(m,n_k)}\boldsymbol{D}^{(m,n)}_C\boldsymbol{\Pi})$, $\boldsymbol{D}^{(m,n)}_C$ is a diagonal matrix, and $\boldsymbol{\Pi}$ is a common permutation matrix for all $\boldsymbol{B}^{(m,n_k)}$, k = 1,…,K.

We give the proof of this theorem in the Appendix. We note that this theorem provides the following key results for the algebraic algorithm: (i) the dimension of the null space of $\boldsymbol{\Gamma}^{(m,n_1,\dots,n_K)}$ reveals the number of sources; (ii) via (23), the basis vectors in the null space of $\boldsymbol{\Gamma}^{(m,n_1,\dots,n_K)}$ are explicitly linked to the inverse of $\boldsymbol{C}^{(m,n_k)T}$ up to trivial indeterminacies. Therefore, solving the problem (23) yields estimates of the factor matrices $\boldsymbol{C}^{(m,n_k)}$. Subsequently estimates of the factor matrices $\boldsymbol{A}^{(m)}$ can be obtained.

We denote $\boldsymbol{W}^{(m,n_1,\dots,n_K)} \triangleq [\boldsymbol{w}^{(m,n_1,\dots,n_K)}_1,\dots,\boldsymbol{w}^{(m,n_1,\dots,n_K)}_R] \in \mathbb{C}^{R^K \times R}$, and $\boldsymbol{\mathcal{W}}^{(m,n_1,\dots,n_K)} \triangleq \mathrm{Ten}(\boldsymbol{W}^{(m,n_1,\dots,n_K)},[R,R,\dots,R]) \in \mathbb{C}^{R \times \cdots \times R}$. According to (23), $\boldsymbol{\mathcal{W}}^{(m,n_1,\dots,n_K)}$ admits the following (K+1)th-order CPD for fixed m, n_1,…,n_k:

$$\boldsymbol{\mathcal{W}}^{(m,n_1,\dots,n_K)} = \left[\!\left[\boldsymbol{B}^{(m,n_1)},\dots,\boldsymbol{B}^{(m,n_K)},\boldsymbol{F}^{(m,n_1,\dots,n_K)} \right]\!\right]_R. \qquad (24)$$

where all the factor matrices $\boldsymbol{B}^{(m,n_k)}, \boldsymbol{F}^{(m,n_1,\dots,n_K)}$ have full column rank, k = 1,…,K. Hence, for all the possible values of $1 \le m \le M$ and $1 \le n_1 < \cdots < n_K \le M$, (24) is a ($K$+1)th-order overdetermined C-CPD with coupling in the first K modes.

Now we briefly explain how to compute the (K+1)th-order overdetermined C-CPD (24). When K = 2, (24) is equivalent to (16) and we can use the strategy explained in Subsection III. D for its computation. When K > 2, the tensor $\boldsymbol{\mathcal{W}}^{(m,n_1,\dots,n_K)}$ in (24) has order (K+1) > 3. We now follow [51], in which it is explained that, under conditions that are satisfied in our derivation, the CPD of a tensor of order higher than three can equivalently be expressed as a C-CPD of third-order tensors obtained by combining modes. This C-CPD can be computed by a GEVD. Therefore, any of the (K+1)th-order CPDs can be computed by GEVD, which yields us the factor matrices $\boldsymbol{B}^{(m,n_1)},\dots,\boldsymbol{B}^{(m,n_K)}$, for any choice of $(m,n_1,\dots,n_K)$. Then we work analogously with (17) and (18) to obtain all the factor matrices $\boldsymbol{B}^{(m,n)}$, m, n = 1,…,M. The details of this procedure are given in the supplementary materials.

As in the case K = 2, we may consider computing the C-CPD of all the tensors $\boldsymbol{\mathcal{W}}^{(m,n_1,\dots,n_K)}$ in (24) simultaneously via an optimization algorithm, while taking into account the coupling structure. After the C-CPD (24) has been computed, we use the strategy (19) to estimate the factor matrices $\tilde{\boldsymbol{A}}^{(m)}$, m = 1,…,M.

The following theorem summarizes our results concerning the uniqueness of DC-CPD for $K \in [2,M]$.

Theorem 4: Let $\boldsymbol{\mathcal{T}}^{(m,n_k)} = \left[\!\left[\boldsymbol{A}^{(m)}, \boldsymbol{A}^{(n_k)*}, \boldsymbol{C}^{(m,n_k)} \right]\!\right]_R$, m, n_k = 1,…,M, k = 1,...,K. We assume that $\boldsymbol{\Phi}^{(m)}_{n_1,\dots,n_K}$ in (22) and $\boldsymbol{C}^{(m,n_k)}$ have full column rank for all values $1 \le n_1 < \cdots < n_K \le M$ and $m = 1,\dots,M$, where $K \in [2,M]$ is a preset constant. Then we have:

(i) The DC-CPD consists of the sum of R double coupled rank-1 terms, i.e. the number of terms cannot be reduced.

(ii) The DC-CPD is unique.

(iii) The DC-CPD can be calculated algebraically.

The derivation of the algebraic DC-CPD algorithm in Subsection III.E provides a constructive proof of *Theorem 4*.

Analogous to the case K = 2, we can calculate the generic value of $R_{\max}$, for $K \in [2,M]$, by checking if the matrix $\boldsymbol{\Phi}^{(m)}_{n_1,\dots,n_K}$ has full column rank for a set of randomly generated factor matrices $\boldsymbol{A}^{(m)}$, m = 1,…,M. The results for K = 3 are given in Table I. We note that the uniqueness condition depends on the preset constant K. We have generally observed that increasing K relaxes the uniqueness condition of DC-CPD. For instance, Table I shows that for K = 3, a more relaxed generic uniqueness condition is obtained than for K = 2. We have also observed that the uniqueness condition of DC-CPD for K = 2 and K = 3 is more relaxed than that of CPD for K = 2 and K = 3, respectively. Instead of giving a formal proof, we limit ourselves to the following explanation. First, (22) says that $\boldsymbol{\Phi}^{(m)}_{n_1,\dots,n_K}$ has $\psi(\boldsymbol{a}^{(m)}_{r_1}\boldsymbol{a}^{(n_1)H}_{r_1},\dots,\boldsymbol{a}^{(m)}_{r_K}\boldsymbol{a}^{(n_K)H}_{r_K})$ as its columns, and (10) shows that $\psi(\boldsymbol{a}^{(m)}_{r_1}\boldsymbol{a}^{(n_1)H}_{r_1},\dots,\boldsymbol{a}^{(m)}_{r_K}\boldsymbol{a}^{(n_K)H}_{r_K})$ is obtained by concatenation of sub-vectors $\mathrm{vec}(\Phi_{(k_1,k_2)}(\boldsymbol{a}^{(m)}_{r_1}\boldsymbol{a}^{(n_1)H}_{r_1},\dots,\boldsymbol{a}^{(m)}_{r_K}\boldsymbol{a}^{(n_K)H}_{r_K}))$. The latter are obtained using the coupled rank-1 detection mapping defined in (9), $1 \le k_1 < k_2 \le K$. From (9) we have:

$$\mathrm{vec}(\Phi_{(k_1,k_2)}(\boldsymbol{a}^{(m)}_{r_1}\boldsymbol{a}^{(n_1)H}_{r_1},\dots,\boldsymbol{a}^{(m)}_{r_K}\boldsymbol{a}^{(n_K)H}_{r_K})) = \left[\left(\bigotimes_{k=1}^{K} \boldsymbol{a}^{(m)}_{r_k}\right) - \left(\bigotimes_{k\in[1,\dots k_1-1,k_2,k_1+1,\dots,k_2-1,k_1,k_2+1,\dots,K]} \boldsymbol{a}^{(m)}_{r_k}\right)\right] \otimes \left(\bigotimes_{l=1}^{K} \boldsymbol{a}^{(n_l)*}_{u_l}\right).$$

This equation shows that $\mathrm{vec}(\Phi_{(k_1,k_2)}(\boldsymbol{a}^{(m)}_{r_1}\boldsymbol{a}^{(n_1)H}_{r_1},\dots,\boldsymbol{a}^{(m)}_{r_K}\boldsymbol{a}^{(n_K)H}_{r_K}))$ is a sum of two vectors that have a Kronecker product structure. One can expect that in matrices with such Kronecker *product* structured columns, the likelihood of *linear* dependencies between columns will decrease as the number of factors 2K increases. Second, we note that when K increases, the number of rows of $\boldsymbol{\Phi}^{(m)}_{n_1,\dots,n_K}$ increases faster than the number of columns. (Note that the size of $\boldsymbol{\Phi}^{(m)}_{n_1,\dots,n_K}$ is $C_K^2 N^{2K} \times (R^K - R)$). Hence, one can again expect a more relaxed uniqueness condition for larger K. In particular, the most relaxed uniqueness condition can be expected for K = M, which implies that the coupled rank-1 detection mapping in (20) involves all the tensors in set $\Upsilon^{(m)}$. On the other hand, the size of $\Gamma(\boldsymbol{\mathcal{T}}^{(m,n_1)},\dots,\boldsymbol{\mathcal{T}}^{(m,n_K)})$ is

$C_K^2 N^{2K} \times R^K$, i.e., it depends exponentially on K such that increasing K increases the computational complexity significantly. For instance, for a DC-CPD with $N = 6$, $R = 23$, and $M = 3$, the matrix $\Gamma(\boldsymbol{\mathcal{T}}^{(m,n_1)},\ldots,\boldsymbol{\mathcal{T}}^{(m,n_K)})$ has size 1296×529 for $K = 2$, and has size 139968×12167 for $K = 3$. As such, if in a practical application uniqueness does not have to be ensured for a very high number of sources, we may limit ourselves to the minimal value $K = 2$ for efficiency of computation. Table III summarizes the algebraic algorithm in the general case.

Besides the algebraic DC-CPD algorithm, we can also compute a DC-CPD via optimization based algorithms. For instance, the framework of structured data fusion (SDF) [64] is well-suited for the implementation of a nonlinear LS based DC-CPD algorithm. We can also derive an ALS based optimization for DC-CPD in an analogous manner to the CPD case. The derivation is given in the supplementary materials.

TABLE III
DC-CPD: ALGEBRAIC ALGORITHM, $2 \le K \le M$

Input: $\{\boldsymbol{\mathcal{T}}^{(m,n)} \in \mathbb{C}^{N\times N\times R} | 1 \le m,n \le M\}$ admitting DC-CPD (7).

1: Group $\boldsymbol{\mathcal{T}}^{(m,n)}$ into M sets $\Upsilon^{(m)}$ by (8), and select $K \in [2, M]$.

2: For each selection of K tensors in $\Upsilon^{(m)}$ perform the operation (10) defined through coupled rank-1 detection mapping (9), for all the frontal slices of these tensors by (20)–(22), to construct a matrix $\Gamma(\boldsymbol{\mathcal{T}}^{(m,n_1)},\ldots,\boldsymbol{\mathcal{T}}^{(m,n_K)})$.

3: Calculate the basis vectors $\boldsymbol{w}_r^{(m,n_1,\ldots,n_K)}$, $r = 1,\ldots,R$, of the null space of $\Gamma(\boldsymbol{\mathcal{T}}^{(m,n_1)},\ldots,\boldsymbol{\mathcal{T}}^{(m,n_K)})$.

4: For all the K-combinations of tensors in $\Upsilon^{(m)}$ and $m = 1,\ldots,M$, convert all the basis vectors $\boldsymbol{w}_r^{(m,n_1,\ldots,n_K)}$ into an overdetermined C-CPD, $r = 1,\ldots,R$, according to (23), (24).

5: Compute the nonsingular matrices $\boldsymbol{B}^{(m,n)}$ by solving the problem (24), and then we have $\tilde{\boldsymbol{C}}^{(m,n)} = \boldsymbol{B}^{(m,n)-T}$, $\tilde{\boldsymbol{A}}^{(m)} \odot \tilde{\boldsymbol{A}}^{(n)*} = \boldsymbol{T}_3^{(m,n)} \boldsymbol{B}^{(m,n)}$.

6: Compute matrix $\boldsymbol{G}_r$ by (19). Then the rth column of the factor matrices $\tilde{\boldsymbol{A}}^{(m)}$ can be computed as the left dominant singular vector of $\boldsymbol{G}_r$, $m = 1,\ldots,M$, $r = 1,\ldots,R$.

Output: Estimates of the factor matrices $\tilde{\boldsymbol{A}}^{(m)}$ and $\tilde{\boldsymbol{C}}^{(m,n)}$, $1 \le m,n \le M$.

Remark 3: By definition, DC-CPD consists of a set of CPDs with coupled factors. One can obviously compute each CPD and consider these individual CPD solutions as the solution to DC-CPD. It is then natural to ask what benefits DC-CPD can have over CPD. We briefly mention the following.

First, DC-CPD is able to provide better robustness to noise and model errors than CPD as it exploits more information and the way it is coupled. As in J-BSS, the coupling comes from the similarity of the sources across different datasets, DC-CPD is expected to perform well in finding components that are consistently present in multi-set signals. Second, DC-CPD components from distinct datasets are automatically aligned. In other words, DC-CPD avoids the permutation alignment / parameter pairing in a post-processing step, which can otherwise be difficult and time-consuming. Third, DC-CPD can be unique when none of its constituting CPDs is unique.

IV. EXPERIMENTS

In this section we present experiment results to demonstrate the performance of DC-CPD, in comparison with standard CPD. In experiment A, we discuss high-accuracy computation of exact decompositions. In experiment B, we compare DC-CPD and CPD in underdetermined J-BSS of synthetic noisy multi-set signals. In experiment C, we apply DC-CPD in underdetermined wideband array signal direction-of-arrival (DOA) estimation.

We use the following abbreviations:
- DC-CPD-ALG: algebraic DC-CPD algorithm.
- DC-CPD-ALS: DC-CPD via ALS.
- DC-CPD-SDF: DC-CPD via SDF.

DC-CPD-SDF is implemented using the 'sdf_nls.m' function in Tensorlab [65]. In the implementation of DC-CPD-ALG, we use the GNJD algorithm [39] to solve (16), (24). In the implementation of DC-CPD-ALS and DC-CPD-SDF, we initialize either randomly or by the results of DC-CPD-ALG, depending on the application. The initialization details will be given in each of the following examples. We use the 'cpd.m' function with default setting in Tensorlab for CPD. For CPD, we use either the auto-covariance tensor or the cross-covariance tensor in (4), which is denoted by CPD-A and CPD-C, respectively. In experiments B and C, the data tensors $\boldsymbol{\mathcal{T}}^{(m,n)}$ in (3) are approximated by their finite-sample version:

$$\left(\tilde{\boldsymbol{\mathcal{T}}}^{(m,n)}\right)_{(:,:,k)} = L^{-1} \cdot \left[\sum_{l=1}^{L} \left(\boldsymbol{X}^{(m,k)}\right)_{(:,l)} \left(\boldsymbol{X}^{(n,k)}\right)_{(:,l)}^{H}\right], \quad (25)$$

where $\boldsymbol{X}^{(m,k)} \in \mathbb{C}^{N\times L}$ denotes the kth temporal frame of $\boldsymbol{x}^{(m)}(t)$ in (3) with frame length L, overlapping with adjacent frames with overlap ratio $\alpha \in [0,1]$.

In experiment C, we use the benchmark codes 'stft.m' and 'istft.m' from [67] to perform the short-time Fourier transform (STFT) and the inverse STFT (ISTFT). All the experiments are performed on a computer with following configuration, CPU: Intel Core i7-4930MX 3.0GHz; Memory: 32GB; System: 64bit Windows 7; MATLAB R2013b.

A. Results for exact decompositions

The data tensors are directly generated by (4), where both the real and imaginary parts of each entry of the factor matrices $\boldsymbol{A}^{(m)} \in \mathbb{C}^{N\times R}$ and $\boldsymbol{C}^{(m,n)} \in \mathbb{C}^{R\times R}$ are randomly drawn from a standard normal distribution, and no noise is added. We fix the third dimension of all the data tensors to R and let the number of mixing matrices $M = 4$. The mean relative error of the estimates of all the loading matrices is defined as:

$$\varepsilon = M^{-1} \sum_{m=1}^{M} \min_{\boldsymbol{\Pi}^{(m)},\boldsymbol{D}^{(m)}} \left(\left\| \boldsymbol{A}^{(m)} - \boldsymbol{\Pi}^{(m)} \boldsymbol{D}^{(m)} \tilde{\boldsymbol{A}}^{(m)} \right\|_F^2 \Big/ \left\| \boldsymbol{A}^{(m)} \right\|_F^2 \right), \quad (26)$$

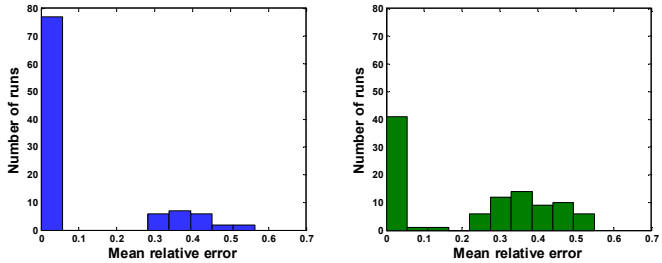
where $\tilde{\boldsymbol{A}}^{(m)}$ is the estimate of the true mixing matrix $\boldsymbol{A}^{(m)}$, $\boldsymbol{\Pi}^{(m)}$ is a permutation matrix and $\boldsymbol{D}^{(m)}$ is a diagonal matrix. For DC-CPD-ALS and DC-CPD-SDF, we try ten random initial values and select the one that gives the best fit after the first ten iterations to effectively initialize the algorithm. As there is no noise, we want to estimate the factor matrices with high precision. Therefore, we terminate the ALS iteration when $|\xi_{cur} - \xi_{prev}| / \xi_{prev} \le 10^{-12}$, where ξ_{cur} and ξ_{prev} denote the LS error in the current and previous iteration, respectively. For DC-CPD-SDF, we set the termination parameters 'TolFun' and 'TolX' in 'sdf_nls.m' to 10^{-12} and 10^{-6}, respectively. In addition, we set the maximal number of iterations to 1000 for

both DC-CPD-ALS and DC-CPD-SDF. The mean relative error and CPU time (denoted as '*t*') of DC-CPD-ALG, DC-CPD-ALS and DC-CPD-SDF for different values of *N* and *R* are reported in Table IV. Note that, under the listed conditions, the rank *R* is high and CPD has been proven to be generically not unique [18]. The average number of iterations of DC-CPD-ALS and DC-CPD-SDF (denoted as '*it*'), and the order of the coupled rank-1 detection mapping used in DC-CPD-ALG (denoted as '*K*') are also reported in Table IV. We select *K* as the minimal number for which DC-CPD-ALG finds the correct solution. All the results in Table IV are averaged over 100 independent runs.

TABLE IV
PERFORMANCE OF DC-CPD-ALG, DC-CPD-ALS, DC-CPD-SDF IN EXACT DECOMPOSITIONS, '*N*' IS THE NUMBER OF OBSERVATIONS, '*R*' IS THE NUMBER OF SOURCES, 'ε' IS THE MEAN RELATIVE ERROR, '*t*' IS THE MEAN CPU TIME, '*K*' IS THE ORDER OF THE COUPLED RANK-1 DETECTION MAPPING IN DC-CPD-ALG, AND '*it*' IS THE AVERAGE NUMBER OF ITERATIONS FOR DC-CPD-ALS AND DC-CPD-SDF

	DC-CPD-ALG	DC-CPD-ALS	DC-CPD-SDF
$N=3$ $R=5$	$\varepsilon = 2.14\times10^{-11}$ $t = 0.39$ sec. $K = 2$	$\varepsilon = 3.83\times10^{-10}$ $t = 9.48$ sec. $it = 693$	$\varepsilon = 2.07\times10^{-7}$ $t = 12.47$ sec. $it = 174$
$N=3$ $R=6$	$\varepsilon = 1.98\times10^{-12}$ $t = 1.50$ sec. $K = 3$	$\varepsilon = 0.106$ $t = 13.55$ sec. $it = 1000$	$\varepsilon = 0.230$ $t = 73.52$ sec. $it = 929$
$N=4$ $R=10$	$\varepsilon = 2.36\times10^{-12}$ $t = 3.09$ sec. $K = 2$	$\varepsilon = 0.079$ $t = 16.73$ sec. $it = 1000$	$\varepsilon = 0.0013$ $t = 71.36$ sec. $it = 793$
$N=4$ $R=11$	$\varepsilon = 2.04\times10^{-11}$ $t = 40.21$ sec. $K = 3$	$\varepsilon = 0.3622$ $t = 19.82$ sec. $it = 1000$	$\varepsilon = 0.4547$ $t = 95.22$ sec. $it = 1000$

We observe that in all the considered cases, DC-CPD-ALG has found the exact solution. DC-CPD-ALS and DC-CPD-SDF, however, did not always find the correct solution in the last three cases, even in the absence of noise. This can be seen from the histograms of the mean relative errors of DC-CPD-ALS and DC-CPD-SDF (By way of example, the histogram for setting $N = 3$, $R = 6$ is shown in Fig. 2). Further, Table IV shows that DC-CPD-ALG provides more efficient computation than DC-CPD-ALS and DC-CPD-SDF in exact decompositions. In practice, noise is usually present, and the performance of DC-CPD-ALG will deteriorate, but can still be used as a low-cost initialization for DC-CPD-ALS and DC-CPD-SDF, which directly fit the data tensor in the LS sense. This will be shown later.



(a) DC-CPD-ALS (b) DC-CPD-SDF

Fig. 2. Histogram of mean relative error for DC-CPD-ALS and DC-CPD-SDF in experiment A, $N = 3$, $R = 6$.

B. *Underdetermined J-BSS of multi-set signals*

We generate the sources in (2) from auxiliary variables $\boldsymbol{s}'_r(t)$:

$$\boldsymbol{s}_r(t) = \boldsymbol{Q}_r \boldsymbol{s}'_r(t), \tag{27}$$

where $\boldsymbol{s}_r(t) \triangleq [s_r^{(1)}(t),...,s_r^{(M)}(t)]^T \in \mathbb{C}^M$, $s_r^{(m)}(t)$ is the *r*th source at time instance *t* in the *m*th dataset and $\boldsymbol{Q}_r \in \mathbb{C}^{M\times M}$ is a full rank matrix used to introduce inter-set dependence between the corresponding source signals in different datasets. Both the real and imaginary part of each entry of $\boldsymbol{Q}_r$ are drawn from a standard normal distribution. The underlying generating vector $\boldsymbol{s}'_r(t) \triangleq [s_r'^{(1)}(t),...,s_r'^{(M)}(t)]^T$ consists of complex binary phase shift keying (BPSK) signals that are amplitude modulated across time slots. The mixtures are constructed as:

$$\boldsymbol{x}_m(t) = \sigma_s \frac{\boldsymbol{A}^{(m)}\boldsymbol{s}^{(m)}(t)}{\| \boldsymbol{A}^{(m)}\boldsymbol{s}^{(m)}(t) \|_F} + \sigma_n \frac{\boldsymbol{n}^{(m)}(t)}{\| \boldsymbol{n}^{(m)}(t) \|_F}, \; m = 1,...,M, \tag{28}$$

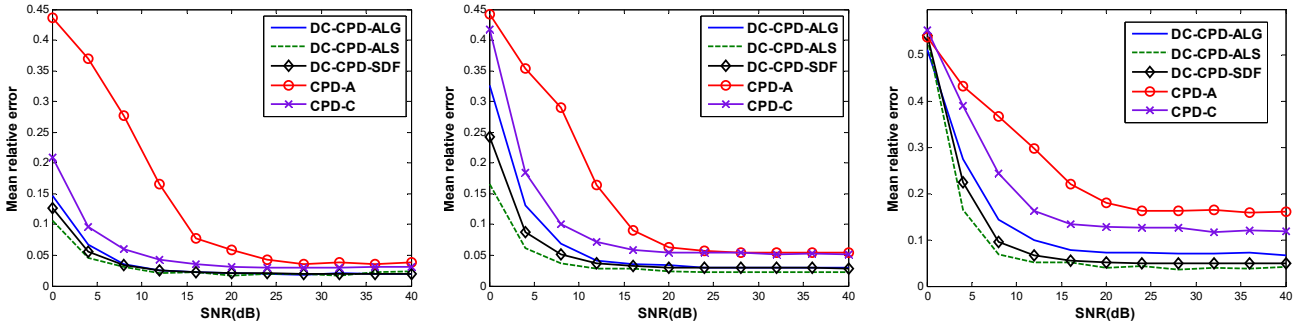
where the real and imaginary part of each entry of $\boldsymbol{A}^{(m)}$ are drawn from a standard normal distribution. Vector $\boldsymbol{n}^{(m)}(t)$ denotes the white Gaussian noise term added in the *m*th dataset, and σ_s, σ_n denote the signal level and noise level, respectively. The signal-to-noise ratio (SNR) is defined as:

$$\text{SNR} = 20\log_{10}(\sigma_s / \sigma_n). \tag{29}$$

DC-CPD-ALS and DC-CPD-SDF are initialized with the results of DC-CPD-ALG. The ALS iteration in DC-CPD-ALS is terminated when $|\xi_{cur} - \xi_{prev}| / \xi_{prev} \le 10^{-6}$. The termination parameters 'TolFun' and 'TolX' of the 'sdf_nls.m' function used in DC-CPD-SDF are set to 10^{-6} and 10^{-3}, respectively. The order of the coupled rank-1 detection mapping used in DC-CPD-ALG is fixed to its minimum value $K = 2$. We also include CPD-A and CPD-C in the comparison.

We fix the number of snapshots $T = 40000$, the framelength $L = 1000$, and the overlap ratio $\alpha = 0.5$, which are used in (25), and let SNR vary from 0dB to 40dB. The mean relative error ε versus SNR is plotted in Fig. 3 to demonstrate the performance of the compared algorithms in the following three settings: (a) a slightly underdetermined case $N = 3$, $R = 4$, $M = 3$, (b) a medium underdetermined case $N = 4$, $R = 6$, $M = 4$, and (c) a highly underdetermined case $N = 4$, $R = 8$, $M = 4$. We perform 200 Monte Carlo runs to calculate each point in the curves.

We can clearly see from Fig. 3 that all three DC-CPD algorithms outperform both CPD-A and CPD-C, because they use the information in both the auto-covariance and cross-covariance tensors, as well as the way the tensors are coupled. The CPD based BSS methods, on the other hand, use only auto-covariance or cross-covariance tensors. In addition, we observe that DC-CPD-ALS and DC-CPD-SDF yield more accurate results than DC-CPD-ALG. This shows that DC-CPD-ALS and DC-CPD-SDF, which directly optimize LS based cost function are more accurate than DC-CPD-ALG in the presence of noise and errors caused by finite sample effects. Overall, the difference is not very large. As expected, it is more significant when the problem is more challenging.



(a) $N = 3$, $R = 4$, $M = 3$ (b) $N = 4$, $R = 6$, $M = 4$ (c) $N = 4$, $R = 8$, $M = 4$

Fig. 3. Mean relative error of DC-CPD-ALG, DC-CPD-ALS, DC-CPD-SDF, CPD-A, and CPD-C vs. SNR in experiment B. The plots illustrate the performance in (a) a slightly underdetermined setting, (b) a medium underdetermined setting, and (c) a highly underdetermined setting.

C. Underdetermined wideband DOA estimation

Wideband DOA estimation is a challenging problem in array processing as the conventional methods based on linear instantaneous mixing model only apply to narrowband signals. The conventional way to handle wideband signals is to transform them into multiple narrowband signals in different frequency bins, use narrowband techniques for each bin separately, and integrate these results (see [68] and references therein). In this example, we will illustrate how J-BSS can be used to fuse the signals in adjacent frequency bins to improve the performance of wideband DOA estimation. In particular, we are interested in the challenging underdetermined problem.

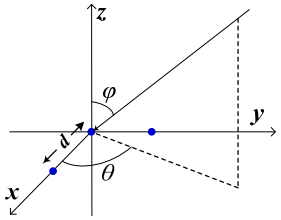


Fig. 4. Array configuration and angle definition in experiment C.

A small scaled L-shaped array of three antennas is used (see Fig. 4). The narrowband component of the wideband array signal in the fth frequency bin can be expressed as:

$$\boldsymbol{y}^{(f)}(t) = \sum_r \boldsymbol{a}_r^{(f)} s_r^{(f)}(t) + \boldsymbol{n}^{(f)}(t), \tag{30}$$

where $s_r^{(f)}(t)$ is the narrowband component of the rth source at the fth frequency bin and $\boldsymbol{n}^{(f)}(t)$ is the corresponding noise component. Vector $\boldsymbol{a}_r^{(f)}$ denotes the steering vector of $s_r^{(f)}(t)$, which is formulated as:

$$\boldsymbol{a}_r^{(f)} \triangleq \exp\left(-\mathrm{i}2\pi c^{-1} f\right) \cdot \boldsymbol{K}\boldsymbol{p}_{(\theta_r,\varphi_r)}, \tag{31}$$

where $\boldsymbol{p}_{(\theta,\varphi)} \triangleq [\cos\theta\sin\varphi, \sin\theta\sin\varphi, \cos\varphi]^T$ is the direction vector associated with azimuth θ and elevation φ (see Fig. 4 for the definition of the angles), and θ_r and φ_r are the azimuth and elevation of $s_r^{(f)}(t)$, respectively. Matrix $\boldsymbol{K} \in \mathbb{R}^{3\times 3}$ describes the spatial configuration of the L-shaped array, taken as:

$$\boldsymbol{K} = [(0,d,0)^T, (0,0,0)^T, (d,0,0)^T]^T. \tag{32}$$

In (32), each row of $\boldsymbol{K}$ holds the Cartesian coordinates of the corresponding antenna, d denotes the interspacing between two adjacent antennas, and c denotes the wave propagation speed.

We convert the source signals into nine narrowband components by STFT with frame length $F = 16$, mix the narrowband components according to (30), and back-transform via ISTFT to generate the wideband mixture in the time domain. The relative bandwidth of each narrowband bin is 0.1, and thus the signal in each frequency bin is a narrowband signal. White Gaussian noise is then added in the time domain.

To estimate the DOAs, we convert the wideband mixture to the time-frequency domain via STFT with frame length $F = 16$, so that there are nine frequency bins. We select the signals in the fourth and fifth frequency bin as the multi-set data. The reason for this selection is that we have observed that only between two neighboring frequency bins the inter-set dependence required for J-BSS is sufficiently strong, while the selected two frequency bins account for approximately 50% of the total amount of energy of the observed mixture.

Denoting the selected multi-set data as $\boldsymbol{x}^{(m)}(t)$, $m = 1, 2$, we have:

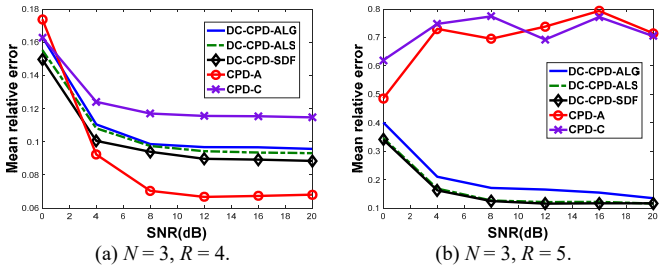
$$\begin{aligned} \boldsymbol{x}^{(1)}(t) &= \boldsymbol{y}^{(4)}(t) = \textstyle\sum_r \boldsymbol{a}_r^{(4)} \boldsymbol{s}_r^{(4)T} + \boldsymbol{n}^{(4)}(t), \\ \boldsymbol{x}^{(2)}(t) &= \boldsymbol{y}^{(5)}(t) = \textstyle\sum_r \boldsymbol{a}_r^{(5)} \boldsymbol{s}_r^{(5)T} + \boldsymbol{n}^{(5)}(t). \end{aligned} \tag{33}$$

We calculate by (25), for a chosen framelength L and overlapping ratio α, the second-order covariance tensors $\tilde{\boldsymbol{\mathcal{T}}}^{(m,n)}$ with $1 \le m, n \le 2$ as the data tensors for DC-CPD. We also include CPD-A and CPD-C in the comparison. We initialize DC-CPD-ALS and DC-CPD-SDF by the results of DC-CPD-ALG. The ALS iteration in DC-CPD-ALS is terminated when $|\varepsilon_{cur} - \varepsilon_{prev}|/\varepsilon_{prev} \le 10^{-12}$. The termination parameters 'TolFun' and 'TolX' in the 'sdf_nls.m' function used in DC-CPD-SDF are set to 10^{-8} and 10^{-4}, respectively. The order of the coupled rank-1 detection mapping used in DC-CPD-ALG is fixed to its minimum value $K = 2$. After DC-CPD or CPD has been performed to estimate $\boldsymbol{a}_r^{(4)}, \boldsymbol{a}_r^{(5)}$, the DOAs are estimated by solving (31). Note that the accuracy of DOA estimation depends on how well the steering vectors $\boldsymbol{a}_r^{(4)}, \boldsymbol{a}_r^{(5)}$ are identified. We here use the mean relative estimation error, as defined in (26), to evaluate the performance of the compared algorithms.

We consider the following two underdetermined cases: 1) a slightly underdetermined setting: four sources with $(\theta_1, \varphi_1) = (22°, 11°)$, $(\theta_2, \varphi_2) = (67°, 33°)$, $(\theta_3, \varphi_3) = (112°, 56°)$, and

$(\theta_4,\varphi_4)=(157°,78°)$; 2) a highly underdetermined setting: five sources with $(\theta_1,\varphi_1)=(18°,9°)$, $(\theta_2,\varphi_2)=(54°,27°)$, $(\theta_3,\varphi_3)=(90°,45°)$, $(\theta_4,\varphi_4)=(126°,63°)$, and $(\theta_5,\varphi_5)=(162°,81°)$.

The SNR is defined in (29) with σ_s and σ_n denoting the signal and noise levels in the time domain, respectively. The number of snapshots is denoted as T. In the experiment, we fix T = 204800 and let SNR vary from 0dB to 20dB. The frame-length is L = 6400, and the overlapping ratio is α = 0.5. The mean relative error of DC-CPD-ALG, DC-CPD-ALS, DC-CPD-SDF, CPD-A, and CPD-C in the above two cases is plotted in Fig. 5.



(a) $N=3$, $R=4$. (b) $N=3$, $R=5$.

Fig. 5. Mean relative error of DC-CPD-ALG, DC-CPD-ALS, DC-CPD-SDF, CPD-A, and CPD-C in wideband DOA estimation. The plots illustrate the performance in (a) a slightly underdetermined setting, and (b) a highly underdetermined setting.

In Fig. 5 (a) we observe that all the algorithms provide quite accurate results in the slightly underdetermined setting. DC-CPD-ALG and DC-CPD-ALS have almost identical accuracy in this case, while DC-CPD-SDF is slightly better. Comparing DC-CPD and CPD, we note that all the DC-CPD algorithms outperform CPD-C for all SNR levels. On the other hand, CPD-A performs better than DC-CPD for high SNR. This is mainly because the noise components across adjacent frequency bins are strongly correlated in this example, and thus the cross-covariance tensors are rather noisy. This also explains the relatively poor performance of CPD-C.

In Fig. 5 (b) we observe that CPD-A and CPD-C do not provide the correct result in the highly underdetermined setting. Note that the uniqueness condition of CPD is not satisfied. On the other hand, DC-CPD-ALG, DC-CPD-ALS and DC-CPD-SDF still perform quite well. DC-CPD-ALS and DC-CPD-SDF perform slightly better than DC-CPD-ALG. Put the other way, the optimization based algorithms could not improve the algebraic algorithm much.

V. Conclusion

In this study, we have considered the problem of DC-CPD for J-BSS of multi-set signals. We have shown how, via the use of second-order statistics, a DC-CPD problem can be obtained from the multi-set signals in J-BSS. We have proposed an algebraic DC-CPD algorithm, via the use of the coupled rank-1 detection mapping. Deterministic and generic uniqueness conditions of DC-CPD have been studied, and shown to be more relaxed than their CPD equivalents. The proposed algorithm is deterministic and returns the exact solution in the noiseless case, as long as the uniqueness condition is satisfied. When noise is present, the proposed algorithm can serve to provide a low-cost initialization for optimization based algorithms. Experiments are performed to compare DC-CPD and CPD, with particular emphasis on underdetermined J-BSS problems. The results have shown the superiority of using DC-CPD over CPD for J-BSS applications, with respect to uniqueness and accuracy.

Appendix I
Proofs

Theorem 1:

Assuming that $\boldsymbol{X}^{(1)},...,\boldsymbol{X}^{(K)}$ are nonzero matrices, we have, by definition, that $\text{vec}(\Phi_{(k_1,k_2)}(\boldsymbol{X}^{(1)},...,\boldsymbol{X}^{(K)}))$ in (10) is an all-zero vector if and only if $\boldsymbol{X}^{(k_1)}$ and $\boldsymbol{X}^{(k_2)}$ are both rank-1 and have identical column spaces. Then *Theorem 1* holds as $\psi(\boldsymbol{X}^{(1)},...,\boldsymbol{X}^{(K)})$ is defined as the concatenation of $\text{vec}(\Phi_{(k_1,k_2)}(\boldsymbol{X}^{(1)},...,\boldsymbol{X}^{(K)}))$ for all possible values of (k_1,k_2). ■

Theorem 3:

We prove the following partial statements:

(i) $\ker(\Gamma(\boldsymbol{\mathcal{T}}^{(m,n_1)},...,\boldsymbol{\mathcal{T}}^{(m,n_K)}))=\ker((\boldsymbol{P}^{(m)}_{n_1,...,n_K})^T)$.

(ii) $\dim(\Gamma(\boldsymbol{\mathcal{T}}^{(m,n_1)},...,\boldsymbol{\mathcal{T}}^{(m,n_K)}))=R$.

(iii) The basis vectors in $\ker(\boldsymbol{\Gamma}^{(m,n_1,...,n_K)})$, denoted as $\boldsymbol{w}_r^{(m,n_1,...,n_k)}\in\mathbb{C}^{R^K}$, $r=1,...,R$, can be written as a linear combination of vectors $\boldsymbol{b}_u^{(m,n_1)}\otimes\cdots\otimes\boldsymbol{b}_u^{(m,n_K)}$, $u=1,\ldots,R$, where $[\boldsymbol{b}_1^{(m,n_k)},...,\boldsymbol{b}_R^{(m,n_k)}]=\boldsymbol{B}^{(m,n_k)}\triangleq(\boldsymbol{C}^{(m,n_k)}\boldsymbol{D}^{(m,n_k)}\boldsymbol{\Pi})^{-T}\in\mathbb{C}^{R\times R}$, $\boldsymbol{D}^{(m,n_k)}$ is a diagonal matrix, and $\boldsymbol{\Pi}$ is a permutation matrix that is common to all the matrices $\boldsymbol{B}^{(m,n_k)}$, $k=1,\ldots,K$.

Note that $m,n_1,...,n_K$ are fixed in the above statements.

First, as $\boldsymbol{\Phi}^{(m)}_{n_1,...,nK}$ is assumed to have full column rank, and $\Gamma(\boldsymbol{\mathcal{T}}^{(m,n_1)},...,\boldsymbol{\mathcal{T}}^{(m,n_K)})=\boldsymbol{\Phi}^{(m)}_{n_1,...,n_K}(\boldsymbol{P}^{(m)}_{n_1,...,n_K})^T$, we directly have $\ker(\Gamma(\boldsymbol{\mathcal{T}}^{(m,n_1)},...,\boldsymbol{\mathcal{T}}^{(m,n_K)}))=\ker((\boldsymbol{P}^{(m)}_{n_1,...,n_K})^T)$.

Second, as $\boldsymbol{C}^{(m,n_1)},...,\boldsymbol{C}^{(m,n_K)}$ are assumed to have full column rank, the Kronecker product $\boldsymbol{C}^{(m,n_1)}\otimes\cdots\otimes\boldsymbol{C}^{(m,n_K)}$ also has full column rank. The matrix $\boldsymbol{P}^{(m)}_{n_1,...,n_K}\in\mathbb{C}^{R^K\times(R^K-R)}$ has full column rank as well, since its columns constitute a subset of the columns of $\boldsymbol{C}^{(m,n_1)}\otimes\cdots\otimes\boldsymbol{C}^{(m,n_K)}$. By the rank nullity theorem we have:

$$\dim[\ker((\boldsymbol{P}^{(m)}_{n_1,...,n_K})^T)]=R^K-\text{rank}((\boldsymbol{P}^{(m)}_{n_1,...,n_K})^T)=R.$$

Third, we prove that $\boldsymbol{b}_r^{(m,n_1)}\otimes\cdots\otimes\boldsymbol{b}_r^{(m,n_K)}$ is in $\ker((\boldsymbol{P}^{(m)}_{n_1,...,nK})^T)$, $r=1,...,R$. This is equivalent to proving the following for all the values $1\le r_1,...,r_K\le R$:

$$\begin{aligned}&\left(\boldsymbol{c}_{r_1}^{(m,n_1)}\otimes\cdots\otimes\boldsymbol{c}_{r_K}^{(m,n_K)}\right)^T\left(\boldsymbol{b}_r^{(m,n_1)}\otimes\cdots\otimes\boldsymbol{b}_r^{(m,n_K)}\right)\\&=\left(\boldsymbol{c}_{r_1}^{(m,n_1)T}\boldsymbol{b}_r^{(m,n_1)}\right)\otimes\cdots\otimes\left(\boldsymbol{c}_{r_K}^{(m,n_K)T}\boldsymbol{b}_r^{(m,n_K)}\right)=0.\end{aligned}\tag{34}$$

Recall that $\boldsymbol{b}_r^{(m,n_k)}$ is the rth column of $(\tilde{\boldsymbol{C}}^{(m,n_k)})^{-T}$, where $\tilde{\boldsymbol{C}}^{(m,n_k)}=\boldsymbol{C}^{(m,n_k)}\boldsymbol{D}^{(m,n_k)}\boldsymbol{\Pi}$. We assume the matrix $\boldsymbol{\Pi}$ permutes the columns of $\boldsymbol{C}^{(m,n_k)}$ in such a way that $(\boldsymbol{C}^{(m,n_k)}\boldsymbol{\Pi})_{:,r}=\boldsymbol{c}_u^{(m,n_1)}$, where $r=\pi(u)$, $1\le u\le R$, and $(\pi(1),...,\pi(R))$ is a permutation of $(1,...,R)$. We denote $\tilde{\boldsymbol{c}}_r^{(m,n_k)}$ as the rth column of $\tilde{\boldsymbol{C}}^{(m,n_k)}$. By definition we have $\tilde{\boldsymbol{c}}_r^{(m,n_k)}=d_r^{(m,n_k)}\boldsymbol{c}_u^{(m,n_k)}$, where $d_r^{(m,n_k)}$ is the rth entry on the diagonal of $\boldsymbol{D}^{(m,n_k)}$.

Therefore, we have $\boldsymbol{c}_{r_k}^{(m,n_k)T}\boldsymbol{b}_r^{(m,n_k)}=0$ if $r=\pi(u)$ and $u\ne r_k$.

Since at least two of the indices $r_1,...,r_K$ are distinct according to the definition of $\boldsymbol{P}^{(m)}_{m_1,...,n_K}$, at least one term in $\boldsymbol{c}^{(m,n_1)T}_{r_1}\boldsymbol{b}^{(m,n_1)}_{r},...,\boldsymbol{c}^{(m,n_K)T}_{r_K}\boldsymbol{b}^{(m,n_K)}_{r}$ is equal to zero, which proves (34). Moreover, as $\boldsymbol{b}^{(m,n_k)}_{1},...,\boldsymbol{b}^{(m,n_k)}_{R}$ are linearly independent, the set $\{\boldsymbol{b}^{(m,n_1)}_{r}\otimes\cdots\otimes\boldsymbol{b}^{(m,n_K)}_{r}, r=1,...,R\}$ contains R linearly independent vectors, and thus is a maximal set of linearly independent vectors of $\ker(\Gamma(\boldsymbol{\mathcal{T}}^{(m,n_1)},...,\boldsymbol{\mathcal{T}}^{(m,n_K)}))$. Therefore, the basis vectors $\boldsymbol{w}^{(m,n_1,...,n_k)}_{r}$ of $\ker(\boldsymbol{\Gamma}^{(m,n_1,...,n_K)})$ can be expressed as a linear combination of $\boldsymbol{b}^{(m,n_1)}_{r}\otimes\cdots\otimes\boldsymbol{b}^{(m,n_K)}_{r}$, $r=1,...,R$. ■

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Dr. Ignat Domanov and Dr. Mikael Sorensen for their insightful comments on the algebraic DC-CPD algorithm.

REFERENCES

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

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

[3] M. Mørup, "Applications of tensor (multiway array) factorizations and decompositions in data mining," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 1, no. 1, pp. 24–40, Jan. 2011.

[4] F.-Y. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikanen, T. Ristaniemi, "Tensor decomposition of EEG signals: a brief review," *J. Neurosci*. *Meth*., vol. 248, pp. 59–69, Jun. 2015.

[5] A. Cichocki, D. Mandic, A.-H. Phan, C. Caiafa, G.-X. Zhou, Q.-B. Zhao, L. De Lathauwer, "Tensor decompositions for signal processing applications. From two-way to multiway component analysis," *IEEE Signal Process. Mag.,* vol. 32, no. 2, pp. 145–163, Mar. 2015.

[6] P. Comon, L. De Lathauwer, "Algebraic identification of under-determined mixtures," in *Handbook of blind source separation, independent component analysis and applications*. P. Comon, and C. Jutten, ed., U.K.: Academic Press, 2010, pp. 325–365.

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

[8] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, C. Faloutsos, "Tensor decomposition for signal processing and machine learning," arXiv: 1607.01668v1. 2016.

[9] R. Harshman, M. Lundy, "PARAFAC: Parallel factor analysis," *Comput. Stat. Data Anal.,* vol. 18, no. 1, pp. 39–72, Aug. 1994.

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

[11] L. De Lathauwer, J. Castaing, "Blind identification of under-determined mixtures by simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp.1096–1105, Mar. 2008.

[12] I. Domanov, L. De Lathauwer, "Canonical polyadic decomposition of third-order tensors: reduction to generalized eigenvalue decomposition," *SIAM J. Matrix Anal. Appl*., vol. 35, no. 2, pp. 636–660, May 2014.

[13] I. Domanov, L. De Lathauwer, "Canonical polyadic decomposition of third-order tensors: relaxed uniqueness conditions and algebraic algorithm," *Linear Algebra Appl*., vol. 513, pp. 342–375, Jan. 2017.

[14] L. Sorber, M. Van Barel, L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: canonical polyadic decomposition, decomposition in rank-(L_r, L_r, 1) terms, and a new generalization," *SIAM J. Matrix Anal. Appl*., vol. 23, no. 2, pp. 695–720, Apr. 2013.

[15] P. Paatero, "The multilinear engine—a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model," *J. Comput. Graph. Stat.,* vol. 8, no. 4, pp. 854–888, Dec. 1999.

[16] A.-H Phan, P. Tichavsky, A. Cichocki, "Low complexity damped Gauss-Newton algorithms for CANDECOMP/PARAFAC," *SIAM J. Matrix Anal. Appl*., vol. 34, no. 1, pp. 126–147, Jan. 2013.

[17] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics," *Linear Algebra Appl*., vol. 18, no. 2, pp. 95–138, Dec. 1977.

[18] L. Chiantini, G. Ottaviani, "On generic identifiability of 3-tensors of small rank," *SIAM J. Matrix Anal. Appl*., vol. 33, no. 3, pp. 1018–1037, Jul. 2012.

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

[20] I. Domanov, L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part II: uniqueness of the overall decomposition," *SIAM J. Matrix Anal. Appl*., vol. 34, no. 3, pp. 876–903, Jul. 2013.

[21] I. Domanov, L. De Lathauwer, "Generic uniqueness conditions for the canonical polyadic decomposition and INDSCAL," *SIAM J. Matrix Anal. Appl*., vol. 36, no. 4, pp. 1567–1589, Nov. 2015.

[22] M. Sorensen, L. De Lathauwer, "New uniqueness conditions for the canonical polyadic decomposition of third-order tensors," *SIAM J. Matrix Anal. Appl*., vol. 36, no. 4, pp. 1381–1403, Oct. 2015.

[23] A. Cichocki, R. Zdunek, A.-H. Phan, S. Amari, "Multi-way array (tensor) factorizations and decompositions," in *Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation*. Singapore: John Wiley & Sons, 2009, pp. 337–432.

[24] G.-X. Zhou, A. Cichocki, "Canonical polyadic decomposition based on a single mode blind source separation," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 523–526, Aug. 2012.

[25] M. De Vos, D. Nion, S. Van Huffel, L. De Lathauwer, "A combination of parallel factor and independent component analysis," *Signal Process*., vol. 92, no. 12, pp. 2990–2999, Dec. 2012.

[26] M. Sorensen, L. De Lathauwer, "Blind signal separation via tensor decomposition with Vandermonde factor: Canonical polyadic decomposition," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5507–5519, Nov. 2012.

[27] L. De Lathauwer, J. Castaing, "Tensor-based techniques for the blind separation of DS-CDMA signals," *Signal Process.*, vol. 87, no. 2, pp. 322–336, Feb. 2007.

[28] M. Sorensen, L. De Lathauwer, P. Comon, S. Icart, L. Deneire, "Canonical polyadic decomposition with a columnwise orthonormal factor matrix," *SIAM J. Matrix Anal. Appl*., vol. 33, no. 4, pp. 1190–1213, Oct. 2012.

[29] J. Sui, T. Adalı, Q.-B. Yu, V. Calhoun, "A Review of multivariate methods for multimodal fusion of brain imaging data," *J. Neurosci. Meth*., vol. 204, pp. 68–81, Feb. 2012.

[30] T. Kim, H. T. Attias, S.-Y. Lee, T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. Audio, Speech, Lang. Process*., vol. 15, no. 1, pp. 70–79, Jan. 2007.

[31] J.-H. Lee, T.-W. Lee, F. Jolesz, S.-S. Yoo, "Independent vector analysis (IVA): multivariate approach for fMRI group study," *NeuroImage*, vol. 40, no. 1, pp. 86–109, Mar. 2008.

[32] M. Anderson, T. Adalı, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: algorithms and performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1672–1683, Apr. 2012.

[33] Y.-O. Li, T. Adalı, W. Wang, V. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3918–3929, Oct. 2009.

[34] N. M. Correa, T. Eichele, T. Adalı, Y.-O. Li, V. D. Calhoun, "Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI," *NeuroImage*, vol. 50, no. 4, pp. 1438–1445, May 2010.

[35] N. M. Correa, T. Adalı, V. D. Calhoun, "Canonical correlation analysis for data fusion and group analysis: Examining applications

of medical imaging data," *IEEE Signal Process. Mag.,* vol. 27, no. 4, pp. 39–50, Jul. 2010.

[36] J. Chatel-Goldman, M. Congedo, R. Phlypo, "Joint BSS as a natural analysis framework for EEG-hyperscanning," in *Proc. ICASSP'2013*, Vancouver, Canada, May 26–30, 2013.

[37] X.-L. Li, T. Adalı, M. Anderson, "Joint blind source separation by generalized joint diagonalization of cumulant matrices," *Signal Process.*, vol. 91, no. 10, pp. 2314–2322, Oct. 2011.

[38] M. Congedo, R. Phlypo. J. Chatel-Goldman, "Orthogonal and non-orthogonal joint blind source separation in the least-squares sense," in *Proc. EUSIPCO'2012*, Bucharest, Romania, Aug. 27–31, 2012.

[39] X.-F. Gong, X.-L. Wang, Q.-H. Lin, "Generalized non-orthogonal joint diagonalization with LU decomposition and successive rotations," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1322–1334, Mar. 2015.

[40] D. Lahat, T. Adalı, C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[41] R. Harshman, M. Lundy. (1984b), "Data preprocessing and the extended PARAFAC model," in H. Law, C. Snyder Jr, J. Hattie, R. McDonald, ed., *Research methods for multimode data analysis*, Praeger, New York, 1984, pp. 216–284.

[42] C. Beckmann, S. Smith, "Tensorial extensions of independent component analysis for multisubject fMRI analysis," *NeuroImage*, vol. 25, no. 1, pp. 294–311, Mar. 2005.

[43] Y. Guo, G. Pagnoni, "A unified framework for group independent component analysis for multi-subject fMRI data," *NeuroImage*, vol. 42, no. 3, pp. 1078–1093, Sep. 2008.

[44] A. Groves, C. Beckmann, S. Smith, M. Woolrich, "Linked independent component analysis for multimodal data fusion," *NeuroImage*, vol. 54, no. 3, pp. 2198–2217, Feb. 2011.

[45] E. Acar, R. Bro, A. K. Smilde, "Data fusion in metabolomics using coupled matrix and tensor factorizations," *Proc. IEEE*, vol. 103, no. 9, pp. 1602–1620, Sep. 2015.

[46] M. Sorensen, L. De Lathauwer, "Multiple invariance ESPRIT for nonuniform linear arrays: a coupled canonical polyadic decomposition approach," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3693–3704, Jul. 2016.

[47] Y.-N. Yu, A. Petropulu, "PARAFAC-based blind estimation of possibly underdetermined convolutive MIMO systems," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 111–124, Jan. 2008.

[48] H. Becker, P. Comon, L. Albera, "Tensor-based preprocessing of combined EEG/MEG data," in *Proc. EUSIPCO'2012*, Bucharest, Romania, Aug. 27–31, 2012, pp. 275–279.

[49] X.-F. Gong, Y. -N. Hao, Q. -H. Lin, "Joint canonical polyadic decomposition of two tensors with one shared loading matrix" in *Proc. MLSP'2013*, Southampton, U. K., Sep. 22–25, 2013.

[50] T. Yokota, A. Cichocki, Y. Yamashita, "Linked PARAFAC / CP tensor decomposition and its fast implementation for multi-block tensor analysis," in *Proc. ICONIP'2012*, Doha, Qatar, Nov. 12–15, 2012, pp. 84–92.

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

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

[53] M. Sorensen, L. De Lathauwer, "Multidimensional harmonic retrieval via coupled canonical polyadic decompositions—Part I: model and identifiability," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 517–527, Jan. 2017.

[54] M. Sorensen, L. De Lathauwer, "Multidimensional harmonic retrieval via coupled canonical polyadic decompositions—Part II: algorithm and multirate sampling," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 528–539, Jan. 2017.

[55] M. Sorensen, L. De Lathauwer, "Coupled tensor decompositions in array processing," ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep. 13–241, 2014.

[56] R. Farias, J. Cohen, P. Comon, "Exploring multimodal data fusion through joint decompositions with flexible couplings," *IEEE Trans. Signal Process.,* vol. 64, no. 18, pp. 4830–4844, Sept. 2016.

[57] B. Rivet, M. Duda, A. Guérin-Dugué, C. Jutten, P. Comon, "Multimodal approach to estimate the ocular movements during EEG recordings: a coupled tensor factorization method," in *Proc. EMBC'2015*, Milan, Italy, Aug. 25–29, 2015.

[58] G.-X. Zhou, Q.-B. Zhao, Y. Zhang, T. Adalı, S.-L. Xie, A. Cichocki, "Linked component analysis from matrices to high order tensors: applications to biomedical data," *Proc. IEEE*, vol. 104, no. 2, pp. 310–331, Feb. 2016.

[59] M. Sorensen, L. De Lathauwer, "Double coupled canonical polyadic decomposition with applications," ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep. 13–145, 2013.

[60] O. Debals, L. De Lathauwer, "Stochastic and deterministic tensorization for blind signal separation," in *Proc. LVA/ICA'2015*, Liberec, Czech Republic, Aug. 25–28, 2015, pp. 3–13.

[61] J.-F Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors," *in Proc. ICASSP'1991*, Toronto, Canada, Apr. 14-17, 1991, pp. 3109-3112.

[62] R. Sands, F. Young, "Component models for three-way data: an alternating least squares algorithm with optimal scaling features," *Psychometrika,* vol. 45, No.1, pp. 39–67, Mar. 1980.

[63] L. Wang, H. Ding, F. -L. Yin, "A region-growing permutation alignment approach in frequency-domain blind source separation of speech mixtures," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 3, pp. 549–557, Mar. 2011.

[64] L. Sorber, M. Van Barel, L. De Lathauwer, "Structured data fusion," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 586–600, Jun. 2015.

[65] N. Vervliet. O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, *Tensorlab 3.0*, Mar. 2016. [Online], Available: http://www.tensorlab.net/.

[66] R. Gunning, H. Rossi, "Holomorphic functions" in *Analytic functions of several complex variables.* U.S.: Prentice-Hall, Inc, Englewood Cliffs, N.J., 1965, pp. 1–63.

[67] E. Vincent, Evaluation and audio source separation software [Online], Available: http://www.irisa.fr/metiss/members/evincent/softwareA.

[68] Q. Shen, W. Liu, W. Cui, S. -L. Wu, Y. -M. Zhang, M. Amin, "Low-complexity direction-of-arrival estimation based on wideband co-prime arrays," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 9, pp. 1445–1456, Sep. 2015.