Citation/Reference	Sorensen M.,Van Eeghem F., De Lathauwer L. (2017), **Blind multichannel deconvolution and convolutive extensions of canonical polyadic and block term decompositions** IEEE Transactions on Signal Processing , vol. 65, no. 15, Aug. 2017, 4132-4145.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://ieeexplore.ieee.org/document/7931673/
Journal homepage	http://ieeexplore.ieee.org/
Author contact	Mikael.Sorensen@kuleuven.be
Abstract	Tensor decompositions such as the canonical polyadic decomposition (CPD) or the block term decomposition (BTD) are basic tools for blind signal separation. Most of the literature concerns instantaneous mixtures/memoryless channels. In this paper, we focus on convolutive extensions. More precisely, we present a connection between convolutive CPD/BTD models and coupled but instantaneous CPD/BTD. We derive a new identifiability condition dedicated to convolutive low-rank factorization problems. We explain that under this condition, the convolutive extension of CPD/BTD can be computed by means of an algebraic method, guaranteeing perfect source separation in the noiseless case. In the inexact case, the algorithm can be used as a cheap initialization for an optimization-based method. We explain that, in contrast to the memoryless case, convolutive signal separation is in certain cases possible despite only two-way diversities (e.g., space × time).
IR	

(article begins on next page)

Blind multichannel deconvolution and convolutive extensions of canonical polyadic and block term decompositions

Mikael Sørensen, Frederik Van Eeghem, and Lieven De Lathauwer, *Fellow, IEEE*

***Abstract*—Tensor decompositions such as the Canonical Polyadic Decomposition (CPD) or the Block Term Decomposition (BTD) are basic tools for blind signal separation. Most of the literature concerns instantaneous mixtures / memoryless channels. In this paper we focus on convolutive extensions. More precisely, we present a connection between convolutive CPD/BTD models and coupled but instantaneous CPD/BTD. We derive a new identifiability condition dedicated to convolutive low-rank factorization problems. We explain that under this condition, the convolutive extension of CPD/BTD can be computed by means of an algebraic method, guaranteeing perfect source separation in the noiseless case. In the inexact case, the algorithm can be used as a cheap initialization for an optimization-based method. We explain that, in contrast to the memoryless case, convolutive signal separation is in certain cases possible despite only two-way diversities (e.g., space $\times$ time).**

***Index Terms*—tensor, coupled decomposition, canonical polyadic decomposition, block term decomposition, block-Toeplitz matrix, convolutive mixtures, source separation.**

I. Introduction

In convolutive signal separation it is assumed that the observed data matrix admits the factorization

$$\mathbf{X} = \mathbf{M}_{\mathrm{TB}}^{(1)}\mathbf{S}_{\mathrm{TB}}^{(1)T} + \cdots + \mathbf{M}_{\mathrm{TB}}^{(R)}\mathbf{S}_{\mathrm{TB}}^{(R)T} = \mathbf{M}_{\mathrm{TB}}\mathbf{S}_{\mathrm{TB}}^{T} \in \mathbb{C}^{I\times K}, \quad (1)$$

where $\mathbf{M}_{\mathrm{TB}} = [\mathbf{M}_{\mathrm{TB}}^{(1)}, \ldots, \mathbf{M}_{\mathrm{TB}}^{(R)}] \in \mathbb{C}^{I\times RL}$ is the mixing matrix in which $\mathbf{M}_{\mathrm{TB}}^{(r)} \in \mathbb{C}^{I\times L}$, $\mathbf{S}_{\mathrm{TB}} = [\mathbf{S}_{\mathrm{TB}}^{(1)}, \ldots, \mathbf{S}_{\mathrm{TB}}^{(R)}] \in \mathbb{C}^{K\times RL}$ is the signal matrix in which $\mathbf{S}_{\mathrm{TB}}^{(r)} \in \mathbb{C}^{K\times L}$ are Toeplitz matrices, and I and K denote the number of sensor outputs and data snapshots, respectively. The goal of convolutive signal separation is to recover the matrix $\mathbf{S}_{\mathrm{TB}}$, observing $\mathbf{X}$.

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

Research supported by: (1) Research Council KU Leuven: CoE EF/05/006 Optimization in Engineering (OPTEC), C1 project C16/15/059-nD, (2) F.W.O.: project G.0830.14N, G.0881.14N, (3) the Belgian Federal Science Policy Office: IUAP P7 (DYSCO II, Dynamical systems, control and optimization, 2012–2017), (4) EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no. 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

Early works in signal processing focused on the single-source case ($R = 1$), i.e., pure deconvolution without separation. Roughly speaking, it has been observed that by exploiting system structures, such as cyclostationarity, oversampling or widely separated receive antennas, the matrix $\mathbf{M}_{\mathrm{TB}}^{(1)}$ typically has full column rank. Under this condition, it suffices to capitalize on the Toeplitz structure of $\mathbf{S}_{\mathrm{TB}}^{(1)}$ for source recovery, see [23], [41], [45], [19], [21] and references therein for further details. In the convolutive signal separation case ($R \geq 2$), the Toeplitz structure of $\{\mathbf{S}_{\mathrm{TB}}^{(r)}\}$ in (1) is not enough to guarantee signal recovery. For this reason, several methods that impose additional assumptions on $\{\mathbf{S}_{\mathrm{TB}}^{(r)}\}$ have been proposed. Early methods assumed statistically independent sources (e.g., [3], [40], [39]). Later on also methods that impose deterministic constraints on $\{\mathbf{S}_{\mathrm{TB}}^{(r)}\}$, such as constant modulus or finite alphabet, were proposed (e.g., [19], [20], [43], [42]). More recently, tensor factorization approaches for the convolutive signal separation case ($R \geq 2$) have been suggested (e.g., [27], [7], [25]). In contrast to the previously mentioned methods, the tensor factorization approaches do not impose additional structural constraints on $\{\mathbf{S}_{\mathrm{TB}}^{(r)}\}$. Instead, low-rank assumptions on $\{\mathbf{M}_{\mathrm{TB}}^{(r)}\}$ are made. Note that this indeed yields multi-way extensions of the classical input-output relation (1) for Multiple-Input Multiple-Output (MIMO) linear systems of order L. However, the methods in [27], [7], [25] do not take the Toeplitz structure of $\{\mathbf{S}_{\mathrm{TB}}^{(r)}\}$ into account.

As our first contribution, we will develop an algebraic framework that takes both the Toeplitz structure of $\mathbf{S}_{\mathrm{TB}}$ and the assumed low-rank structure of $\mathbf{M}_{\mathrm{TB}}$ into account. In other words, we will extend the standard instantaneous ($L = 1$) source separation framework in which $\mathbf{M}_{\mathrm{TB}}$ is subject to a low-rank structure (e.g., [28]) to the convolutive case ($L \geq 2$). Interestingly, it will be explained that convolutive low-rank factorization problems lead to *coupled* tensor decomposition problems. This leads to a dedicated identifiability condition and an algebraic method that guarantees perfect separation in the noiseless case. Another way to see our contribution, is that we propose convolutive extensions of the Canonical Polyadic Decomposition (CPD) [14] and the Block Term Decomposition (BTD) [5]. These will first be translated into coupled versions of the basic, instanta-

neous decompositions [31], [36]. Convolutive CPD/BTD models have many applications in, for instance, wireless communication [27], [4], [7], [25] and neuroimaging [22]. We emphasize that coupled CPD/BTD is an important emerging concept for signal processing. We have already shown its relevance for multidimensional harmonic retrieval [34], [35] and sensor array processing [33]. It is also useful in multi-modal signal and data processing (e.g., [29], [1], [17], [11]). Part of this work appeared in the conference paper [30].

As an additional contribution, we will explain that, under certain conditions, the memory of the convolutive channel can act as a system diversity. This means that the triple diversity needed for the ordinary instantaneous ($L = 1$) low-rank source separation (e.g., space $\times$ time $\times$ oversampling in wireless communication [27], [4], [7], [24], [25], [26]) is not always a requirement in the convolutive case ($L \geq 2$). As an illustration, we show that for communication systems in which the channel enjoys "hidden" low-rank structures due to multipath propagation, convolutive signal separation is possible, despite only two-way diversities (space $\times$ time). We stress that this is not possible in the memoryless case ($L = 1$).

The paper is organized as follows. The rest of the introduction will present the notation. Section II reviews the low-rank tensor and block-Toeplitz factorizations used in the paper. In section III a link between convolutive low-rank factorizations and coupled tensor decompositions is established. It will also be explained that the source matrix $\mathbf{S}_{\text{TB}}$ in (1) can in many cases be recovered via Simultaneous matrix Diagonalization (SD) techniques. In Section IV we first demonstrate that the proposed framework leads to a dedicated uniqueness condition for the convolutive tensor decomposition problems considered in [27], [4], [7], [24], [25], [26], [22]. Second, we also demonstrate that convolutive channels can act as system diversities. Numerical experiments are reported in Section V. Section VI concludes the paper.

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

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

in which $(\mathbf{A})_{mn} = a_{mn}$. The outer product of N vectors $\mathbf{a}^{(n)} \in \mathbb{C}^{I_n}$ is denoted by $\mathbf{a}^{(1)} \otimes \mathbf{a}^{(2)} \otimes \cdots \otimes \mathbf{a}^{(N)} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$, such that

$$\left(\mathbf{a}^{(1)} \otimes \mathbf{a}^{(2)} \otimes \cdots \otimes \mathbf{a}^{(N)}\right)_{i_1,i_2,\dots,i_N} = a^{(1)}_{i_1} a^{(2)}_{i_2} \cdots a^{(N)}_{i_N}.$$

The transpose, conjugate, conjugate-tranpose, Moore-Penrose pseudo-inverse, column space and null space of a matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^\dagger$, $\text{range}(\mathbf{A})$ and $\ker(\mathbf{A})$, respectively. The k-rank of a matrix $\mathbf{A}$ is denoted by $k_{\mathbf{A}}$. It is equal to the largest integer $k_{\mathbf{A}}$ such that every subset of $k_{\mathbf{A}}$ columns of $\mathbf{A}$ is linearly independent.

The identity matrix is denoted by $\mathbf{I}_R \in \mathbb{C}^{R\times R}$ and the all-ones vector is denoted by $\mathbf{1}_R = [1, \dots, 1]^T \in \mathbb{C}^R$. The rth column of $\mathbf{I}_R$ is denoted by $\mathbf{e}_r^{(R)}$. Further, $\mathbf{0}_{M\times N} \in \mathbb{C}^{M\times N}$ and $\mathbf{0}_M \in \mathbb{C}^M$ denote the all-zero matrix and all-zero vector, respectively. $D_k(\mathbf{A}) \in \mathbb{C}^{J\times J}$ denotes the diagonal matrix holding row k of $\mathbf{A} \in \mathbb{C}^{I\times J}$ on its diagonal. Given $\mathcal{X} \in \mathbb{C}^{I_1\times I_2\times\cdots\times I_N}$, then $\text{Vec}(\mathcal{X}) \in \mathbb{C}^{\prod_{n=1}^N I_n}$ denotes the column vector

$$\text{Vec}(\mathcal{X}) = [x_{1,\dots,1,1}, x_{1,\dots,1,2}, \dots, x_{I_1,\dots,I_{N-1},I_N}]^T.$$

The reverse operation is $\text{Unvec}(\text{Vec}(\mathcal{X})) = \mathcal{X}$.

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

The binomial coefficient is denoted by $C_m^k = \frac{m!}{k!(m-k)!}$. The k-th compound matrix of $\mathbf{A} \in \mathbb{C}^{I\times R}$ is denoted by $C_k(\mathbf{A}) \in \mathbb{C}^{C_I^k \times C_R^k}$. It is the matrix containing the determinants of all $k \times k$ submatrices of $\mathbf{A}$, arranged with the submatrix index sets in lexicographic order [15].

II. Algebraic Prerequisites

A. Canonical Polyadic Decomposition (CPD)

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

$$\mathbb{C}^{I\times J\times K} \ni \mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r \,. \tag{2}$$

The rank of a tensor $\mathcal{X}$ is equal to the minimal number of rank-1 tensors that yield $\mathcal{X}$ in a linear combination. Assume that the rank of $\mathcal{X}$ is R, then (2) is called the CPD of $\mathcal{X}$. Let us stack the vectors $\{\mathbf{a}_r\}$, $\{\mathbf{b}_r\}$ and $\{\mathbf{c}_r\}$ into the matrices $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_R]$ and $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_R]$. The matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ will be referred to as the factor matrices of the PD or CPD of $\mathcal{X}$.

1) Matrix representation: Let $\mathbf{X}^{(i\cdot\cdot)} \in \mathbb{C}^{J\times K}$ denote the matrix such that $(\mathbf{X}^{(i\cdot\cdot)})_{jk} = x_{ijk}$, then $\mathbf{X}^{(i\cdot\cdot)} = \mathbf{B}D_i(\mathbf{A})\mathbf{C}^T$ and we obtain the following matrix representation of (2):

$$\mathbb{C}^{IJ\times K} \ni \mathbf{X} := \left[\mathbf{X}^{(1\cdot\cdot)T}, \dots, \mathbf{X}^{(I\cdot\cdot)T}\right]^T = (\mathbf{A}\odot\mathbf{B})\mathbf{C}^T. \tag{3}$$

2) Uniqueness: The rank-1 tensors in (2) can be arbitrarily permuted and the vectors within the same rank-1 tensor can be arbitrarily scaled provided the overall rank-1 term remains the same. We say that the CPD is unique when it is only subject to these trivial indeterminacies. In this paper at least one factor matrix has full column rank. For this case the following easy-to-check sufficient uniqueness condition has been obtained.

Theorem II.1. *Consider the PD of $\mathcal{X} \in \mathbb{C}^{I\times J\times K}$ in* (2)*. If*

$$\begin{cases} \boldsymbol{C} \text{ has full column rank,} \\ C_2(\boldsymbol{A}) \odot C_2(\boldsymbol{B}) \text{ has full column rank,} \end{cases} \tag{4}$$

then the rank of $\mathcal{X}$ is R and the CPD of $\mathcal{X}$ is unique [16], [8], [9], [10]. Generically[1]*, condition* (4) *is satisfied if $C_R^2 \leq C_I^2 C_J^2$ and $R \leq K$ [8], [37].*

An algorithm for computing the CPD of $\mathcal{X}$ under the condition (4) has been developed in [8]. In short, the CPD problem can be reduced to a square Simultaneous matrix Diagonalization (SD) problem, even in cases where $R > \max(I, J)$. The SD problem amounts to finding the CPD of a (partially symmetric) $(R\times R\times R)$ tensor $\mathcal{M}$ of rank R that has at least two factor matrices of full column rank. Hence, in the exact case, it can be solved by means of a standard matrix GEVD [18].

B. Coupled CPD

An extension of the CPD of a tensor $\mathcal{X}$ is the coupled CPD of a set of tensors $\mathcal{X}^{(1)}, \ldots, \mathcal{X}^{(N)}$ in which the rank-1 terms associated with the different tensors are connected. More specifically, we consider the situation where a collection of tensors $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n\times J_n\times K}$, $n \in \{1, \ldots, N\}$, admits the coupled PD

$$\mathcal{X}^{(n)} = \sum_{r=1}^{R} \mathbf{a}_r^{(n)} \otimes \mathbf{b}_r^{(n)} \otimes \mathbf{c}_r, \quad n \in \{1, \ldots, N\}, \tag{5}$$

with factor matrices $\mathbf{A}^{(n)} = [\mathbf{a}_1^{(n)}, \ldots, \mathbf{a}_R^{(n)}]$, $\mathbf{B}^{(n)} = [\mathbf{b}_1^{(n)}, \ldots, \mathbf{b}_R^{(n)}]$ and common $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_R]$. We define the coupled rank of $\{\mathcal{X}^{(n)}\}$ as the minimal number of coupled rank-1 tensors $\mathbf{a}_r^{(n)} \otimes \mathbf{b}_r^{(n)} \otimes \mathbf{c}_r$ that yield $\{\mathcal{X}^{(n)}\}$ in a linear combination. Assume that the coupled rank of $\{\mathcal{X}^{(n)}\}$ is R, then (5) will be called the coupled CPD of $\{\mathcal{X}^{(n)}\}$.

1) Matrix representation: The coupled PD or CPD of $\{\mathcal{X}^{(n)}\}$ given by (5) has the following matrix representation

$$\mathbf{X} = \left[\mathbf{X}^{(1)T}, \ldots, \mathbf{X}^{(N)T}\right]^T = \mathbf{F}\mathbf{C}^T \in \mathbb{C}^{(\sum_{n=1}^{N} I_n J_n)\times K}, \tag{6}$$

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

2) Uniqueness: As in the CPD case, the coupled rank-1 tensors in (5) can be arbitrarily permuted and the vectors within the same coupled rank-1 tensor can be arbitrarily scaled provided the overall coupled rank-1 term remains the same. Likewise, we say that the coupled CPD is unique when it is only subject to these trivial indeterminacies. Uniqueness conditions for the coupled CPD were derived in [31]. Theorem II.2 extends Theorem II.1 to coupled CPD. It makes use of the matrix

$$\mathbf{Z}^{(N)} = \begin{bmatrix} C_2\left(\mathbf{A}^{(1)}\right) \odot C_2\left(\mathbf{B}^{(1)}\right) \\ \vdots \\ C_2\left(\mathbf{A}^{(N)}\right) \odot C_2\left(\mathbf{B}^{(N)}\right) \end{bmatrix} \in \mathbb{C}^{\left(\sum_{n=1}^{N} C_{I_n}^2 C_{J_n}^2\right)\times C_R^2}. \tag{7}$$

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

Theorem II.2. *Consider the coupled PD of $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n\times J_n\times K}$, $n \in \{1, \ldots, N\}$ in* (5)*. If*

$$\begin{cases} \mathbf{C} \text{ in (6) has full column rank,} \\ \boldsymbol{Z}^{(N)} \text{ in (7) has full column rank,} \end{cases} \tag{8}$$

then the coupled rank of $\{\mathcal{X}^{(n)}\}$ is R and the coupled CPD of $\{\mathcal{X}^{(n)}\}$ is unique [31]. Generically, condition (8) *is satisfied if $C_R^2 \leq \sum_{n=1}^{N} C_{I_n}^2 C_{J_n}^2$ and $R \leq K$.*

An adaptation of the SD method to coupled CPD under the condition (8) has been described in [36]. Briefly, the coupled CPD (5) is reduced to a CPD of a (partially symmetric) $(R \times R \times R)$ tensor $\mathcal{M}$ of rank R that has at least two factor matrices of full column rank.

C. (Coupled) Block-Term Decomposition (BTD)

Another extension of the CPD (2) is the multilinear rank-$(P, P, 1)$ term decomposition, where each term in the decomposition now consists of the outer product of a vector and matrix that is low-rank (instead of rank-1). Formally, we consider the decomposition of the tensor(s) $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n\times J_n\times K}$, $n \in \{1, \ldots, N\}$, into (coupled) multilinear rank-$(P, P, 1)$ terms, or (coupled) BTD for short:

$$\mathcal{X}^{(n)} = \sum_{r=1}^{R} \sum_{p=1}^{P} \mathbf{a}_p^{(r,n)} \otimes \mathbf{b}_p^{(r,n)} \otimes \mathbf{c}^{(r)} = \sum_{r=1}^{R} \left(\mathbf{A}^{(r,n)} \mathbf{B}^{(r,n)T}\right) \otimes \mathbf{c}^{(r)}, \tag{9}$$

with factor matrices $\mathbf{A}^{(r,n)} = [\mathbf{a}_1^{(r,n)}, \ldots, \mathbf{a}_P^{(r,n)}] \in \mathbb{C}^{I_n\times P}$, $\mathbf{B}^{(r,n)} = [\mathbf{b}_1^{(r,n)}, \ldots, \mathbf{b}_P^{(r,n)}] \in \mathbb{C}^{J_n\times P}$ and $\mathbf{C} = [\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(R)}] \in \mathbb{C}^{K\times R}$. The cases where $N = 1$ correspond to ordinary BTD [5], [6] while the cases where $N > 1$ correspond to coupled BTD [31], [36]. Note also that if $P = 1$, then (9) reduces to coupled CPD ($N > 1$) or even ordinary CPD ($N = 1$). The (coupled) BTD rank of $\{\mathcal{X}^{(n)}\}$ is equal to the minimal number of (coupled) multilinear rank-$(P, P, 1)$ tensors that yield $\{\mathcal{X}^{(n)}\}$ in a linear combination.

1) Matrix representation: The (coupled) BTD of (9) has the following matrix representation

$$\mathbf{X} = \mathbf{F}\mathbf{C}^T, \tag{10}$$

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

2) Uniqueness: The (coupled) multilinear rank-$(P, P, 1)$ tensors in (9) can be arbitrarily permuted and the matrices/vectors within the same term can be arbitrarily transformed/scaled provided that the overall (coupled) multilinear rank-$(P, P, 1)$ tensors remain the same. Formally, we can allow $\frac{1}{\alpha}[(\mathbf{A}^{(r,n)}\mathbf{T}^{(r,n)})(\mathbf{B}^{(r,n)}(\mathbf{T}^{(r,n)})^{-T})^T] \otimes (\alpha\mathbf{c}^{(r)})$ for arbitrary nonzero α and nonsingular $\mathbf{T}^{(r,n)}$. We say that the (coupled) BTD is unique when it is only subject to these trivial indeterminacies.

Theorem II.3 below extends Theorem II.2 to coupled BTD. The extension of (7) makes use of the column-selection matrix $\mathbf{P}_{\mathrm{BTD}} \in \mathbb{C}^{C_{RP}^{P+1}\times(C_{R+P}^{P+1}-R)}$ that takes into account that each column vector $\mathbf{c}^{(r)}$ in (9) is repeated P times. Here we only state how $\mathbf{P}_{\mathrm{BTD}}$ can be constructed, see [36] for further explanation. The $C_{R+P}^{P+1} - R$ columns of

$\mathbf{P}_{\text{BTD}}$ are indexed by the lexicographically ordered tuples in the set

$$\Gamma_c = \{(j_1, \ldots, j_{P+1}) \mid 1 \le j_1 \le \cdots \le j_{P+1} \le R\} \setminus \{(j, \ldots, j)\}_{j=1}^{R}.$$

Consider also the mapping $f_c : \mathbb{N}^{P+1} \to \{1, 2, \ldots, C_{R+P}^{P+1} - R\}$ that returns the position of its argument in the set Γ_c. Similarly, the C_{RP}^{P+1} rows of $\mathbf{P}_{\text{BTD}}$ are indexed by the lexicographically ordered tuples in the set

$$\Gamma_r = \{(i_1, \ldots, i_{P+1}) \mid 1 \le i_1 < \cdots < i_{P+1} \le RP\}.$$

Likewise, we define the mapping $f_r : \mathbb{N}^{P+1} \to \{1, 2, \ldots, C_{RP}^{P+1}\}$ that returns the position of its argument in the set Γ_r. The entries of $\mathbf{P}_{\text{BTD}}$ are given by

$$(\mathbf{P}_{\text{BTD}})_{f_r(i_1,\ldots,i_{P+1}),f_c(j_1,\ldots,j_{P+1})} = \begin{cases} 1, & \text{if } \lceil \frac{i_1}{P} \rceil = j_1, \ldots, \lceil \frac{i_{P+1}}{P} \rceil = j_{P+1}, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The extension of $\mathbf{Z}^{(N)}$ in (7) to the BTD case, denoted by $\mathbf{Z}_P^{(N)} \in \mathbb{C}^{\left(\sum_{n=1}^{N} C_{I_n}^{P+1} C_{J_n}^{P+1}\right) \times (C_{R+P}^{P+1} - R)}$, is defined as

$$\mathbf{Z}_P^{(N)} = \begin{bmatrix} C_{P+1}\left(\mathbf{A}^{(1)}\right) \odot C_{P+1}\left(\mathbf{B}^{(1)}\right) \\ \vdots \\ C_{P+1}\left(\mathbf{A}^{(N)}\right) \odot C_{P+1}\left(\mathbf{B}^{(N)}\right) \end{bmatrix} \mathbf{P}_{\text{BTD}}, \tag{12}$$

where $\mathbf{A}^{(n)} = [\mathbf{A}^{(1,n)} \ \ldots \ \mathbf{A}^{(R,n)}] \in \mathbb{C}^{I_n \times RP}$ and $\mathbf{B}^{(n)} = [\mathbf{B}^{(1,n)} \ \ldots \ \mathbf{B}^{(R,n)}] \in \mathbb{C}^{J_n \times RP}$.

Theorem II.3. *Consider the (coupled) BTD of $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n \times J_n \times K}$, $n \in \{1, \ldots, N\}$ in (9). If*

$$\begin{cases} \mathbf{C} \text{ in (10) has full column rank,} \\ \mathbf{Z}_P^{(N)} \text{ in (12) has full column rank,} \end{cases}$$

then the (coupled) BTD rank of $\{\mathcal{X}^{(n)}\}$ is R and the (coupled) BTD of $\{\mathcal{X}^{(n)}\}$ is unique [36].

An extension of the SD method to the (coupled) BTD has been developed in [36].

D. Toeplitz-block and block-Toeplitz matrices

In this paper we consider Toeplitz-block matrices of the form (cf. (1)):

$$\mathbf{S}_{\text{TB}} = [\mathbf{S}_{\text{TB}}^{(1)}, \ldots, \mathbf{S}_{\text{TB}}^{(R)}] \in \mathbb{C}^{K \times RL},$$

where $\mathbf{S}_{\text{TB}}^{(r)} \in \mathbb{C}^{K \times L}$ is the Toeplitz matrix

$$\mathbf{S}_{\text{TB}}^{(r)} = \left[\mathbf{s}_1^{(r)}, \ldots, \mathbf{s}_L^{(r)}\right] = \begin{bmatrix} s_1^{(r)} & s_0^{(r)} & \cdots & s_{1-L}^{(r)} \\ s_2^{(r)} & s_1^{(r)} & \cdots & s_{2-L}^{(r)} \\ \vdots & \vdots & \ddots & \vdots \\ s_K^{(r)} & s_{K-1}^{(r)} & \cdots & s_{K-L+1}^{(r)} \end{bmatrix}, \tag{13}$$

in which $\mathbf{s}_l^{(r)}$ denotes the lth column of $\mathbf{S}_{\text{TB}}^{(r)}$. We will also work with block-Toeplitz matrices, which are of the form

$$\mathbf{S}_{\text{BT}} = [\mathbf{S}_{\text{BT}}^{(1)}, \ldots, \mathbf{S}_{\text{BT}}^{(L)}] = \begin{bmatrix} \mathbf{p}_L & \cdots & \mathbf{p}_1 \\ \vdots & \ddots & \vdots \\ \mathbf{p}_K & & \mathbf{p}_L \\ \vdots & \ddots & \vdots \\ \mathbf{p}_{K+L-1} & \cdots & \mathbf{p}_K \end{bmatrix} \in \mathbb{C}^{K \times LR}, \tag{14}$$

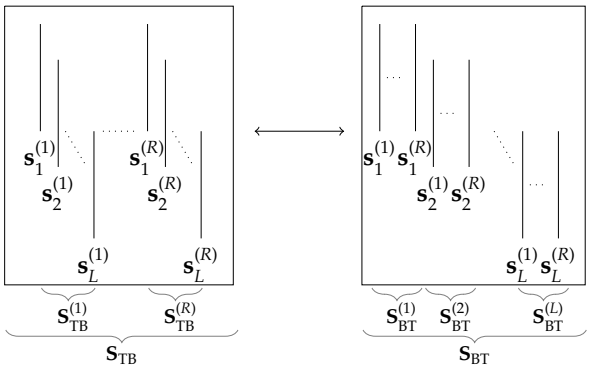


Fig. 1. Relation between the Toeplitz-block matrix $\mathbf{S}_{\text{TB}}$ and the block-Toeplitz matrix $\mathbf{S}_{\text{BT}}$.

where $\mathbf{S}_{\text{BT}}^{(l)} \in \mathbb{C}^{K \times R}$ and where $\mathbf{p}_k \in \mathbb{C}^{1 \times R}$ are row vectors. We have $\mathbf{S}_{\text{BT}}^{(l)} = [\mathbf{s}_l^{(1)}, \ldots, \mathbf{s}_l^{(R)}]$ or equivalently $\mathbf{p}_k = [s_k^{(1)}, \ldots, s_k^{(R)}]$, so that $\mathbf{S}_{\text{TB}}$ and $\mathbf{S}_{\text{BT}}$ are just column permuted versions of each other, i.e., there exists a column permutation matrix $\mathbf{\Pi}_{\mathbf{S}} \in \mathbb{C}^{LR \times LR}$ such that $\mathbf{S}_{\text{TB}}\mathbf{\Pi}_{\mathbf{S}} = \mathbf{S}_{\text{BT}}$. The relation between $\mathbf{S}_{\text{TB}}$ and $\mathbf{S}_{\text{BT}}$ is depicted in Figure 1.

The block-Toeplitz matrix $\mathbf{S}_{\text{BT}}$ can also be expressed as

$$\mathbf{S}_{\text{BT}} = \sum_{k=1}^{K+L-1} \mathbf{E}_k \otimes \mathbf{p}_k \in \mathbb{C}^{K \times LR}, \tag{15}$$

where $\mathbf{E}_k \in \mathbb{C}^{K \times L}$ are Toeplitz basis matrices of the form

$$(\mathbf{E}_k)_{ij} = \begin{cases} 1, & i = j + k - L, \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

E. Block-Toeplitz factorization

The column permutation relation between $\mathbf{S}_{\text{TB}}$ and $\mathbf{S}_{\text{BT}}$ implies that the Toeplitz-block factorization in (1) is equivalent to the block-Toeplitz factorization

$$\mathbf{X} = \mathbf{M}_{\text{BT}}^{(1)}\mathbf{S}_{\text{BT}}^{(1)T} + \cdots + \mathbf{M}_{\text{BT}}^{(L)}\mathbf{S}_{\text{BT}}^{(L)T} = \mathbf{M}_{\text{BT}}\mathbf{S}_{\text{BT}}^{T} \in \mathbb{C}^{I \times K}, \tag{17}$$

where $\mathbf{M}_{\text{BT}} = [\mathbf{M}_{\text{BT}}^{(1)}, \ldots, \mathbf{M}_{\text{BT}}^{(L)}] \in \mathbb{C}^{I \times RL}$ in which $\mathbf{M}_{\text{BT}}^{(l)} \in \mathbb{C}^{I \times R}$, and where $\mathbf{S}_{\text{BT}} = [\mathbf{S}_{\text{BT}}^{(1)}, \ldots, \mathbf{S}_{\text{BT}}^{(L)}] \in \mathbb{C}^{K \times LR}$ is given by (14). Indeed, we have

$$\mathbf{X} = \mathbf{M}_{\text{TB}}\mathbf{S}_{\text{TB}}^{T} = \mathbf{M}_{\text{TB}}\mathbf{\Pi}_{\mathbf{S}} \cdot \mathbf{\Pi}_{\mathbf{S}}^{T}\mathbf{S}_{\text{TB}}^{T} = \mathbf{M}_{\text{BT}}\mathbf{S}_{\text{BT}}^{T},$$

where $\mathbf{M}_{\text{BT}} = \mathbf{M}_{\text{TB}}\mathbf{\Pi}_{\mathbf{S}}$ and $\mathbf{S}_{\text{BT}} = \mathbf{S}_{\text{TB}}\mathbf{\Pi}_{\mathbf{S}}$.

1) Essential uniqueness of block-Toeplitz factorization: Let $\widehat{\mathbf{M}}_{\text{BT}} = [\widehat{\mathbf{M}}_{\text{BT}}^{(1)}, \ldots, \widehat{\mathbf{M}}_{\text{BT}}^{(L)}]$ and the block-Toeplitz matrix $\widehat{\mathbf{S}}_{\text{BT}} = [\widehat{\mathbf{S}}_{\text{BT}}^{(1)}, \ldots, \widehat{\mathbf{S}}_{\text{BT}}^{(L)}]$ yield an alternative block-Toeplitz decomposition $\mathbf{X} = \mathbf{M}_{\text{BT}}\mathbf{S}_{\text{BT}}^{T} = \widehat{\mathbf{M}}_{\text{BT}}\widehat{\mathbf{S}}_{\text{BT}}^{T}$. In words, $\widehat{\mathbf{M}}_{\text{BT}}$ corresponds to an alternative mixing matrix. (Similarly, $\widehat{\mathbf{S}}_{\text{BT}}$ denotes an alternative block-Toeplitz structured signal matrix.) Since $\mathbf{M}_{\text{BT}}$ is unconstrained, there always exists an alternative decomposition. We call the factorization in (17) essentially unique if any alternative

$(\widehat{\mathbf{M}}_{\mathrm{BT}}, \widehat{\mathbf{S}}_{\mathrm{BT}})$ is related to $(\mathbf{M}_{\mathrm{BT}}, \mathbf{S}_{\mathrm{BT}})$ via a nonsingular matrix $\mathbf{G} \in \mathbb{C}^{R\times R}$ as follows

$$\mathbf{M}_{\mathrm{BT}}^{(l)} = \widehat{\mathbf{M}}_{\mathrm{BT}}^{(l)}\mathbf{G}^T \text{ and } \mathbf{S}_{\mathrm{BT}}^{(l)} = \widehat{\mathbf{S}}_{\mathrm{BT}}^{(l)}\mathbf{G}^{-1} \text{ for every } l \in \{1,\dots,L\}.$$

Since $\mathbf{S}_{\mathrm{BT}}$ is only assumed to be block-Toeplitz structured, $\mathbf{G}$ is an intrinsic indeterminacy. Several conditions for essential uniqueness can be found in the literature. In this paper we will make use of Lemma II.4 below, which can for instance be found in [19], [20] (although formulated in a slightly different way).

Lemma II.4. *Consider the block-Toeplitz factorization of $\boldsymbol{X} \in \mathbb{C}^{I\times K}$ in* (17)*. Define*

$$\boldsymbol{H}_{\boldsymbol{S}_{BT}} = \left[\overline{\boldsymbol{S}}_{BT}, \underline{\boldsymbol{S}}_{BT}^{(L)}\right] = \left[\overline{\boldsymbol{S}}_{BT}^{(1)}, \overline{\boldsymbol{S}}_{BT}^{(2)}, \dots, \overline{\boldsymbol{S}}_{BT}^{(L)}, \underline{\boldsymbol{S}}_{BT}^{(L)}\right] \in \mathbb{C}^{(K-1)\times(L+1)R}, \tag{18}$$

where $\boldsymbol{S}_{BT} = [\boldsymbol{S}_{BT}^{(1)}, \dots, \boldsymbol{S}_{BT}^{(L)}] \in \mathbb{C}^{K\times LR}$ is block-Toeplitz of the form (14)*. If*

$$\begin{cases} \boldsymbol{M}_{BT} \text{ in (17) has full column rank,} \\ \boldsymbol{H}_{\boldsymbol{S}_{BT}} \text{ in (18) has full column rank,} \end{cases} \tag{19}$$

then the block-Toeplitz factorization of $\boldsymbol{X}$ is essentially unique.

In Subsection III-A we explain that Lemma II.4 can be used to compute the block-Toeplitz factorization.

2) Temporal smoothing: If the block-Toeplitz matrix $\mathbf{S}_{\mathrm{BT}}$ in (17) is sufficiently tall, then it is possible to apply temporal smoothing in a preprocessing step (e.g., [43], [20]). The goal of temporal smoothing is to extend Lemma II.4 to cases where $\mathbf{M}_{\mathrm{BT}}$ in (17) does not have full column rank. In short, by replacing $\mathbf{X}$, $\mathbf{M}_{\mathrm{BT}}$ and $\mathbf{S}_{\mathrm{BT}}$ in (17) with $\mathbf{X}_{\mathrm{BT,sm}}^{(m)}$, $\mathbf{M}_{\mathrm{BT,sm}}^{(m)}$ and $\mathbf{S}_{\mathrm{BT,sm}}^{(m)}$ defined below, Lemma II.4 can still guarantee essential uniqueness of the block-Toeplitz factorization of $\mathbf{X}$ in (17), despite rank-deficient $\mathbf{M}_{\mathrm{BT}}$.

Observe from (14) that

$$\mathbf{S}_{\mathrm{BT}}^T = \begin{bmatrix} \mathbf{p}_L^T & \mathbf{p}_{L+1}^T & \cdots & \mathbf{p}_{K+L-2}^T & \mathbf{p}_{K+L-1}^T \\ \mathbf{p}_{L-1}^T & \mathbf{p}_L^T & \cdots & \mathbf{p}_{K+L-3}^T & \mathbf{p}_{K+L-2}^T \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{p}_1^T & \mathbf{p}_2^T & \cdots & \mathbf{p}_{K-1}^T & \mathbf{p}_K^T \end{bmatrix}.$$

Let $\mathbf{x}_k$ denote the kth column of $\mathbf{X}$ in (17). It is clear that

$$\mathbf{x}_k = \mathbf{M}_{\mathrm{BT}} \begin{bmatrix} \mathbf{p}_{k+L-1}^T \\ \vdots \\ \mathbf{p}_k^T \end{bmatrix}.$$

Stacking yields

$$\mathbb{C}^{Im\times(K-m+1)} \ni \mathbf{X}_{\mathrm{BT,sm}}^{(m)} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{K-m+1} \\ \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_{K-m+2} \\ \vdots & \iddots & \iddots & \vdots \\ \mathbf{x}_m & \mathbf{x}_{m+1} & \cdots & \mathbf{x}_K \end{bmatrix} = \mathbf{M}_{\mathrm{BT,sm}}^{(m)} \mathbf{S}_{\mathrm{BT,sm}}^{(m)T},$$

in which $\mathbf{M}_{\mathrm{BT,sm}}^{(m)} \in \mathbb{C}^{Im\times(L+m-1)R}$ and $\mathbf{S}_{\mathrm{BT,sm}}^{(m)} \in \mathbb{C}^{(K-m+1)\times(L+m-1)R}$ are given by

$$\mathbf{M}_{\mathrm{BT,sm}}^{(m)} = \begin{bmatrix} & & & \iddots & \iddots & & \iddots \\ & & \mathbf{M}_{\mathrm{BT}}^{(1)} & \mathbf{M}_{\mathrm{BT}}^{(2)} & \cdots & \cdots & \mathbf{M}_{\mathrm{BT}}^{(L)} \\ & \mathbf{M}_{\mathrm{BT}}^{(1)} & \mathbf{M}_{\mathrm{BT}}^{(2)} & \cdots & \cdots & \mathbf{M}_{\mathrm{BT}}^{(L)} & \\ \mathbf{M}_{\mathrm{BT}}^{(1)} & \mathbf{M}_{\mathrm{BT}}^{(2)} & \cdots & \cdots & \mathbf{M}_{\mathrm{BT}}^{(L)} & & \end{bmatrix}, \tag{20}$$

$$\mathbf{S}_{\mathrm{BT,sm}}^{(m)} = \begin{bmatrix} \mathbf{p}_{m+L-1} & \mathbf{p}_{m+L-2} & \cdots & \mathbf{p}_1 \\ \mathbf{p}_{m+L} & \mathbf{p}_{m+L-1} & \cdots & \mathbf{p}_2 \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{p}_{K+L-1} & \mathbf{p}_{K+L-2} & \cdots & \mathbf{p}_{K-m+1} \end{bmatrix}. \tag{21}$$

Finally, the partitioning $\mathbf{S}_{\mathrm{BT,sm}}^{(m)} = [\mathbf{S}_{\mathrm{BT,sm}}^{(1,m)}, \dots, \mathbf{S}_{\mathrm{BT,sm}}^{(L+m-1,m)}]$ with $\mathbf{S}_{\mathrm{BT,sm}}^{(l,m)} \in \mathbb{C}^{(K-m+1)\times R}$, leads to the temporally smoothed version of $\mathbf{H}_{\mathbf{S}_{\mathrm{BT}}}$ in (18):

$$\mathbf{H}_{\mathbf{S}_{\mathrm{BT,sm}}^{(m)}} = [\overline{\mathbf{S}}_{\mathrm{BT,sm}}^{(m)}\ ,\ \underline{\mathbf{S}}_{\mathrm{BT,sm}}^{(L+m-1,m)}] \in \mathbb{C}^{(K-m+1)\times(L+m)R}. \tag{22}$$

Hence, if the matrices $\mathbf{M}_{\mathrm{BT,sm}}^{(m)}$ and $\mathbf{H}_{\mathbf{S}_{\mathrm{BT,sm}}^{(m)}}$ have full column rank, Lemma II.4 guarantees the essential uniqueness of the block-Toeplitz factorization of $\mathbf{X}_{\mathrm{BT,sm}}^{(m)}$. Since $\mathbf{S}_{\mathrm{BT,sm}}^{(m)}$ is essentially unique, $\mathbf{S}_{\mathrm{BT}}$ is essentially unique. Furthermore, since $\mathbf{H}_{\mathbf{S}_{\mathrm{BT,sm}}^{(m)}}$ has full column rank, $\overline{\mathbf{S}}_{\mathrm{BT,sm}}^{(m)}$ has full column rank, which obviously means that $\mathbf{S}_{\mathrm{BT,sm}}^{(m)}$ has full column rank. Observe also that the submatrix consisting of the last LR columns of $\mathbf{S}_{\mathrm{BT,sm}}^{(m)}$ is equal to the submatrix consisting of the first $K-m+1$ rows of $\mathbf{S}_{\mathrm{BT}}$. The last two facts imply that $\mathbf{S}_{\mathrm{BT}}$ also has full column rank. Consequently, $\mathbf{M}_{\mathrm{BT}} = \mathbf{X}(\mathbf{S}_{\mathrm{BT}}^T)^\dagger$ is also essentially unique.

Note that the full column rank assumptions on $\mathbf{M}_{\mathrm{BT,sm}}^{(m)}$ and $\mathbf{H}_{\mathbf{S}_{\mathrm{BT,sm}}^{(m)}}$ require that $I > R$ and that the smoothing parameter m is chosen in the interval $\lceil\frac{(L-1)R}{I-R}\rceil \le m \le \lfloor\frac{K+1-(L-1)R}{R+1}\rfloor$.

III. Convolutive extensions of tensor decompositions

In Section III-A we explain that convolutive low-rank factorizations are connected to coupled decompositions. Next, in Sections III-B and III-C, we show that the link between convolutive low-rank factorizations and coupled decompositions leads to improved identifiability results and algebraic algorithms.

A. Deconvolution leads to coupled decompositions

Recall that in the block-Toeplitz factorization (17), $\mathbf{S}_{\mathrm{BT}} \in \mathbb{C}^{K\times LR}$ can only be found up to the intrinsic indeterminacy $\mathbf{S}_{\mathrm{BT}} = \widehat{\mathbf{S}}_{\mathrm{BT}}(\mathbf{I}_L \otimes \mathbf{G})$, where $\mathbf{G} \in \mathbb{C}^{R\times R}$ is a nonsingular matrix and $\widehat{\mathbf{S}}_{\mathrm{BT}} \in \mathbb{C}^{K\times LR}$ is a block-Toeplitz matrix such that $\mathrm{range}(\mathbf{S}_{\mathrm{BT}}) = \mathrm{range}\left(\widehat{\mathbf{S}}_{\mathrm{BT}}\right)$. Expression (17) can therefore also be written as

$$\mathbf{X} = \mathbf{M}_{\mathrm{BT}}(\mathbf{I}_L \otimes \mathbf{G}^T)\widehat{\mathbf{S}}_{\mathrm{BT}}^T. \tag{23}$$

A two-step procedure for finding $\mathbf{M}_{\text{BT}}$, $\mathbf{G}$ and $\widehat{\mathbf{S}}_{\text{BT}}$ from $\mathbf{X}$ will now be outlined. First, by capitalizing on the block-Toeplitz structure of $\mathbf{S}_{\text{BT}}$, $\widehat{\mathbf{S}}_{\text{BT}}$ in (23) will be determined using Lemma II.4. Next, by capitalizing on the low-rank structure of $\mathbf{M}_{\text{BT}}$, the remaining unknowns $\mathbf{M}_{\text{BT}}$ and $\mathbf{G}$ in (23) will be obtained via coupled tensor decompositions. The block-Toeplitz factorization step can be understood as a deconvolution that reduces the original convolutive source separation problem into an instantaneous blind source separation problem. The coupled low-rank factorization can in turn be interpreted as a subsequent unmixing step that solves the latter problem.

Step 1: Find S_{BT} up to intrinsic block-Toeplitz factorization indeterminacy: Consider an alternative factorization of $\mathbf{X}$ in (17), denoted by $\mathbf{X} = \widehat{\mathbf{M}}_{\text{BT}}\widehat{\mathbf{S}}_{\text{BT}}^T$ with $\widehat{\mathbf{M}}_{\text{BT}} \in \mathbb{C}^{I\times L\widehat{R}}$, $\widehat{\mathbf{S}}_{\text{BT}} \in \mathbb{C}^{K\times L\widehat{R}}$, and $\widehat{R} \leq R$. Recall from (18) that $\mathbf{H}_{\mathbf{S}_{\text{BT}}} = [\overline{\mathbf{S}}_{\text{BT}}, \underline{\mathbf{S}}_{\text{BT}}^{(L)}]$. Consequently, if $\mathbf{H}_{\mathbf{S}_{\text{BT}}}$ has full column rank, then $\overline{\mathbf{S}}_{\text{BT}}$ (and $\mathbf{S}_{\text{BT}}$) must also have full column rank. Condition (19) therefore implies that $\mathbf{M}_{\text{BT}}$ and $\mathbf{S}_{\text{BT}}$ both have full column rank. Since $\mathbf{X} = \mathbf{M}_{\text{BT}}\mathbf{S}_{\text{BT}}^T = \widehat{\mathbf{M}}_{\text{BT}}\widehat{\mathbf{S}}_{\text{BT}}^T$ condition (19) also implies that $\widehat{\mathbf{M}}_{\text{BT}}$ and $\widehat{\mathbf{S}}_{\text{BT}}$ have full column rank, that $\widehat{R} = R$ and that $\text{range}\left(\mathbf{X}^T\right) = \text{range}\left(\mathbf{S}_{\text{BT}}\right) = \text{range}\left(\widehat{\mathbf{S}}_{\text{BT}}\right)$. The latter property implies that for some nonsingular matrix $\mathbf{N} \in \mathbb{C}^{LR\times LR}$ we have

$$\widehat{\mathbf{S}}_{\text{BT}} \cdot \mathbf{N} = \left(\sum_{k=1}^{K+L-1} \mathbf{E}_k \otimes \widehat{\mathbf{p}}_k\right) \cdot \mathbf{N} = \sum_{k=1}^{K+L-1} \mathbf{E}_k \otimes \mathbf{p}_k = \mathbf{S}_{\text{BT}}, \quad (24)$$

where $\mathbf{E}_k \in \mathbb{C}^{K\times L}$ is the Toeplitz basis matrix of the form (16) and $\widehat{\mathbf{S}}_{\text{BT}}$ is built from the row vectors $\widehat{\mathbf{p}}_1 \in \mathbb{C}^{1\times R}, \ldots, \widehat{\mathbf{p}}_{K+L-1} \in \mathbb{C}^{1\times R}$. If condition (19) in Lemma II.4 is satisfied, then any alternative $\widehat{\mathbf{S}}_{\text{BT}}$ must be equal to $\mathbf{S}_{\text{BT}}$ up to the intrinsic ambiguity $\widehat{\mathbf{S}}_{\text{BT}} = (\mathbf{I}_L \otimes \mathbf{G})\mathbf{S}_{\text{BT}}$. In other words, if condition (19) is satisfied, then $\mathbf{N}$ in (24) reduces to $\mathbf{N} = \mathbf{I}_L \otimes \mathbf{G}$.

In practice we can find $\mathbf{S}_{\text{BT}}$ up to the intrinsic indeterminacy as follows: first we compute the column space $\text{range}\left(\mathbf{X}^T\right) = \text{range}\left(\mathbf{S}_{\text{BT}}\right)$ and then find $\mathbf{S}_{\text{BT}}$ from it up to the intrinsic indeterminacies. In more detail, let the columns of $\mathbf{U} \in \mathbb{C}^{K\times LR}$ constitute a basis for $\text{range}\left(\mathbf{X}^T\right)$. By making use of relation (15), $\mathbf{S}_{\text{BT}}$ can be expressed in terms of $\{\mathbf{p}_k\}$. The latter variables can in turn be determined from the system of homogeneous linear equations

$$\mathbf{U}\mathbf{N} = \mathbf{S}_{\text{BT}} = \sum_{k=1}^{K+L-1} \mathbf{E}_k \otimes \mathbf{p}_k \Leftrightarrow \mathbf{U}\mathbf{N} - \sum_{k=1}^{K+L-1} \mathbf{E}_k \otimes \mathbf{p}_k = \mathbf{0}_{K\times LR}, \quad (25)$$

where $\mathbf{N}$ and $\{\mathbf{p}_k\}$ are the unknowns. Note that, since $\mathbf{U}$ contains an arbitrary basis for $\text{range}\left(\mathbf{S}_{\text{BT}}\right)$, the nonsingular matrix $\mathbf{N}$ is not necessarily block-diagonal. To summarize, if $\{\widehat{\mathbf{N}}, \widehat{\mathbf{p}}_1, \ldots, \widehat{\mathbf{p}}_{K+L-1}\}$ is a solution to (25) and if condition (19) in Lemma II.4 is satisfied, then $\widehat{\mathbf{S}}_{\text{BT}} = \sum_{k=1}^{K+L-1} \mathbf{E}_k \otimes \widehat{\mathbf{p}}_k$ is equal to $\mathbf{S}_{\text{BT}}$ up to the intrinsic block-Toeplitz factorization ambiguity, i.e., $\widehat{\mathbf{S}}_{\text{BT}} = \sum_{k=1}^{K+L-1} \mathbf{E}_k \otimes \widehat{\mathbf{p}}_k = \mathbf{S}_{\text{BT}}(\mathbf{I}_L \otimes \mathbf{G}^{-1})$ in which $\mathbf{G} \in \mathbb{C}^{R\times R}$ is nonsingular. Several computational variants of this procedure can be found in the literature (e.g., [19], [23]). More details can also be found in the technical report [32].

If $\mathbf{M}_{\text{BT}}$ does not have full column rank, then temporal smoothing can be used in a preprocessing step, as discussed in Subsection II-E. In short, if $\mathbf{M}_{\text{BT,sm}}^{(m)}$ and $\mathbf{H}_{\mathbf{S}_{\text{BT,sm}}^{(m)}}$ have full column rank, and the columns of $\mathbf{U}^{(m)} \in \mathbb{C}^{(K-m+1)\times(L+m-1)R}$ constitute a basis for $\text{range}\left(\mathbf{X}_{\text{BT,sm}}^{(m)T}\right)$, then $\mathbf{S}_{\text{BT}}$ can be determined up to the ambiguity matrix $\mathbf{G}$ via the temporally smoothed version of (25):

$$\mathbf{U}^{(m)}\mathbf{N}^{(m)} = \mathbf{S}_{\text{BT,sm}}^{(m)}, \quad (26)$$

where $\mathbf{N}^{(m)} \in \mathbb{C}^{(L+m-1)R\times(L+m-1)R}$ is a nonsingular matrix. Section IV will provide illustrative examples where temporal smoothing is used to overcome rank deficiency of $\mathbf{M}_{\text{BT}}$.

Step 2: Find M_{BT} and G via coupled factorization: Once $\widehat{\mathbf{S}}_{\text{BT}}$ has been determined, (23) can be reduced to

$$\mathbf{Y} = \mathbf{X}(\widehat{\mathbf{S}}_{\text{BT}}^T)^\dagger = \mathbf{M}_{\text{BT}}\mathbf{S}_{\text{BT}}^T(\widehat{\mathbf{S}}_{\text{BT}}^T)^\dagger = \mathbf{M}_{\text{BT}}(\mathbf{I}_L \otimes \mathbf{G}^T) \in \mathbb{C}^{I\times LR}. \quad (27)$$

The partitioning $\mathbf{Y} = [\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(L)}]$, in which $\mathbf{Y}^{(l)} \in \mathbb{C}^{I\times R}$, yields the following reshaped version of (27):

$$\mathbf{Y}_{\text{tot}} := \begin{bmatrix} \mathbf{Y}^{(1)} \\ \vdots \\ \mathbf{Y}^{(L)} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{\text{BT}}^{(1)} \\ \vdots \\ \mathbf{M}_{\text{BT}}^{(L)} \end{bmatrix} \mathbf{G}^T \in \mathbb{C}^{IL\times R}, \quad (28)$$

where 'tot' stands for total. Perhaps surprisingly, the deconvolution has resulted in a *coupled* decomposition. Several variants may be distinguished, depending on the specific low-rank structure in $\{\mathbf{M}_{\text{BT}}^{(l)}\}$. The next two subsections will consider basic CPD and BTD examples where $\mathbf{M}_{\text{BT}} = \mathbf{A} \odot \mathbf{B}$ and $\mathbf{M}_{\text{BT}} = [\text{Vec}(\mathbf{B}^{(1)}\mathbf{A}^{(1)T}) \ \ldots \ \text{Vec}(\mathbf{B}^{(R)}\mathbf{A}^{(R)T})]$, respectively. Other examples can be found in the technical report [32].

B. Convolutive extension of CPD

From (28) it is clear that the main difference between low-rank constrained instantaneous ($L = 1$) and convolutive ($L \geq 2$) signal separation problems is that the former leads to a single decomposition while the latter leads to a coupled decomposition. As an illustrative example, consider the basic case where the mixing matrix $\mathbf{M}_{\text{BT}}$ is subject to a Khatri–Rao low-rank constraint:

$$\mathbf{X} = \mathbf{M}_{\text{BT}}\mathbf{S}_{\text{BT}}^T = (\mathbf{A} \odot \mathbf{B})\mathbf{S}_{\text{BT}}^T, \quad (29)$$

where

$$\mathbf{A} = [\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(L)}] \in \mathbb{C}^{I\times LR}, \qquad \mathbf{A}^{(l)} \in \mathbb{C}^{I\times R},$$
$$\mathbf{B} = [\mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(L)}] \in \mathbb{C}^{J\times LR}, \qquad \mathbf{B}^{(l)} \in \mathbb{C}^{J\times R},$$
$$\mathbf{M}_{\text{BT}} = [\mathbf{M}_{\text{BT}}^{(1)}, \ldots, \mathbf{M}_{\text{BT}}^{(L)}] \in \mathbb{C}^{IJ\times LR}, \quad \mathbf{M}_{\text{BT}}^{(l)} = \mathbf{A}^{(l)} \odot \mathbf{B}^{(l)} \in \mathbb{C}^{IJ\times R}.$$

Comparing (3) with (29) it is clear that, in the instantaneous case, the factorization of $\mathbf{X}$ may just be seen as a CPD. We now explain in detail that the factorization of

$\mathbf{X}$ is unique under relaxed conditions in the convolutive case.

Let us assume that condition (19) in Lemma II.4 is satisfied, so that we perform the deconvolution as in Section III-A. Substitution of $\mathbf{M}_{\mathrm{BT}}^{(l)} = \mathbf{A}^{(l)} \odot \mathbf{B}^{(l)}$ in (28) yields:

$$\mathbf{Y}_{\mathrm{tot}} = \begin{bmatrix} \mathbf{Y}^{(1)} \\ \vdots \\ \mathbf{Y}^{(L)} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{\mathrm{BT}}^{(1)} \\ \vdots \\ \mathbf{M}_{\mathrm{BT}}^{(L)} \end{bmatrix} \mathbf{G}^T = \begin{bmatrix} \mathbf{A}^{(1)} \odot \mathbf{B}^{(1)} \\ \vdots \\ \mathbf{A}^{(L)} \odot \mathbf{B}^{(L)} \end{bmatrix} \mathbf{G}^T \in \mathbb{C}^{IJL \times R}. \tag{30}$$

From Theorem II.2 we know that $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{G}$ can be recovered from $\mathbf{Y}_{\mathrm{tot}}$ if the matrix

$$\mathbf{Z}^{(L)} = \begin{bmatrix} C_2\left(\mathbf{A}^{(1)}\right) \odot C_2\left(\mathbf{B}^{(1)}\right) \\ \vdots \\ C_2\left(\mathbf{A}^{(L)}\right) \odot C_2\left(\mathbf{B}^{(L)}\right) \end{bmatrix} \in \mathbb{C}^{L \cdot C_I^2 C_J^2 \times C_R^2} \tag{31}$$

has full column rank. (Recall also that $\mathbf{Z}^{(L)}$ generically has full column rank if $L \cdot C_I^2 C_J^2 \geq C_R^2$.) Proposition III.1 below summarizes the uniqueness condition obtained by combining Lemma II.4 and Theorem II.2. Lemma II.4 ensures that the block-Toeplitz factorization (29) can be reduced to the coupled CPD (30). Theorem II.2 in turn ensures that the latter is unique, implying uniqueness of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{S}_{\mathrm{BT}}$.

Proposition III.1. *Consider the decomposition of $\boldsymbol{X} \in \mathbb{C}^{IJ \times K}$ in* (29)*, with $\boldsymbol{M}_{BT}^{(l)} = \boldsymbol{A}^{(l)} \odot \boldsymbol{B}^{(l)}$, $l \in \{1, \ldots, L\}$, and $\boldsymbol{S}_{BT}$ block-Toeplitz. If*

$$\begin{cases} \boldsymbol{H}_{\boldsymbol{S}_{BT}} \text{ in (18) has full column rank,} \\ \boldsymbol{M}_{BT} \text{ in (29) has full column rank,} \\ \boldsymbol{Z}^{(L)} \text{ in (31) has full column rank,} \end{cases} \tag{32}$$

then the decomposition of $\boldsymbol{X}$ is unique.

Note that under condition (32) the factorization of $\mathbf{X}$ can in the exact case be reduced to a matrix GEVD. More precisely, after solving the system of homogeneous linear equations (25), the problem reduces to the computation of the coupled CPD (30), which can be done by means of a GEVD.

Let us now illustrate the relaxation of the uniqueness condition.

First, observe that, by taking the convolutive structure in (29) into account, we have gone from the $(C_I^2 C_J^2 \times C_{LR}^2)$ matrix $C_2(\mathbf{A}) \odot C_2(\mathbf{B})$ to the taller $(L \cdot C_I^2 C_J^2 \times C_R^2)$ matrix $\mathbf{Z}^{(L)}$. Obviously the latter matrix has full column rank under relaxed conditions. (For instance, $\mathbf{Z}^{(L)}$ can have full column rank even if only $\mathbf{M}_{\mathrm{BT}}^{(1)} = \mathbf{A}^{(1)} \odot \mathbf{B}^{(1)}$ is Khatri–Rao structured, i.e., the remaining $\mathbf{M}_{\mathrm{BT}}^{(2)}, \ldots, \mathbf{M}_{\mathrm{BT}}^{(L)}$ are allowed to be unstructured.) By way of example, Proposition III.1 does not prevent $k_{\mathbf{A}} = 1$ and/or $k_{\mathbf{B}} = 1$. This is in contrast to ordinary CPD where $k_{\mathbf{A}} \geq 2$ and $k_{\mathbf{B}} \geq 2$ are necessary for uniqueness (e.g., [38]). Let us consider $\mathbf{X} \in \mathbb{C}^{IJ \times K}$ in (29) with $I = 3$, $J = 3$, $K = 10$, $L = 2$ and $R = 3$. If apart from the constraints $\mathbf{a}_1^{(1)} = \mathbf{a}_2^{(1)}$ and $\mathbf{b}_1^{(2)} = \mathbf{b}_2^{(2)}$ the parameters are randomly drawn, Proposition III.1 guarantees the generic uniqueness of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{S}_{\mathrm{BT}}$, despite $k_{\mathbf{A}} = k_{\mathbf{B}} = 1$.

Second, we show that the temporally smoothed version of Proposition III.1 can handle rank deficient $\mathbf{M}_{\mathrm{BT}}$. Again, this is in contrast to ordinary CPD where full column rank of $\mathbf{M}_{\mathrm{BT}}$ is necessary for uniqueness (e.g., [38]). Let us consider $\mathbf{X} \in \mathbb{C}^{IJ \times K}$ in (29) with $I = 3$, $J = 3$, $K = 10$, $L = 2$ and $R = 2$. If apart from the constraint $\mathbf{m}_{\mathrm{BT},1} = \mathbf{a}_1^{(1)} \otimes \mathbf{b}_1^{(1)} = \mathbf{a}_2^{(2)} \otimes \mathbf{b}_2^{(2)} = \mathbf{m}_{\mathrm{BT},4}$ the parameters are randomly drawn, the smoothed version of Proposition III.1 with $m = 2$ guarantees the generic uniqueness of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{S}_{\mathrm{BT}}$, despite the fact that $\mathbf{M}_{\mathrm{BT}}$ is rank deficient.

C. Convolutive extension of BTD

A straightforward variant of the convolutive CPD model (29) is the convolutive extension of BTD:

$$\mathbf{X} = \mathbf{M}_{\mathrm{BT}} \mathbf{S}_{\mathrm{BT}}^T = [\mathrm{Vec}(\mathbf{B}^{(1)} \mathbf{A}^{(1)T}), \ldots, \mathrm{Vec}(\mathbf{B}^{(R)} \mathbf{A}^{(R)T})] \mathbf{S}_{\mathrm{BT}}^T. \tag{33}$$

Replacing the $\mathbf{M}_{\mathrm{BT}} = \mathbf{A} \odot \mathbf{B}$ matrix in (29) by the $\mathbf{M}_{\mathrm{BT}}$ matrix in (33) yields the coupled BTD variant of (30):

$$\mathbf{Y}_{\mathrm{tot}} = \begin{bmatrix} \mathrm{Vec}(\mathbf{B}^{(1,1)} \mathbf{A}^{(1,1)T}), \ldots, \mathrm{Vec}(\mathbf{B}^{(R,L)} \mathbf{A}^{(R,L)T}) \\ \vdots \\ \mathrm{Vec}(\mathbf{B}^{(1,1)} \mathbf{A}^{(1,1)T}), \ldots, \mathrm{Vec}(\mathbf{B}^{(R,L)} \mathbf{A}^{(R,L)T}) \end{bmatrix} \mathbf{G}^T, \tag{34}$$

where $\mathbf{A}^{(r,l)} \in \mathbb{C}^{I \times R}$ is the lth block matrix of $\mathbf{A}^{(r)} = [\mathbf{A}^{(r,1)}, \ldots, \mathbf{A}^{(r,L)}] \in \mathbb{C}^{I \times LR}$ and $\mathbf{B}^{(r,l)} \in \mathbb{C}^{J \times R}$ is the lth block matrix of $\mathbf{B}^{(r)} = [\mathbf{B}^{(r,1)}, \ldots, \mathbf{B}^{(r,L)}] \in \mathbb{C}^{J \times LR}$. Comparing (10) with (34), it is clear that the decomposition of $\mathbf{Y}_{\mathrm{tot}}$ corresponds to the coupled BTD discussed in Section II-C.

IV. Applications in array processing

In order to demonstrate that the introduced convolutive tensor decomposition framework is suitable for signal processing, we first review in Section IV-A a data model used for convolutive source separation in wireless communication [27], [4], [7], [25]. Next, in Section IV-B, we explain that the presented convolutive tensor decomposition framework provides an algebraic algorithm and identifiability conditions for this particular application. Finally, in Section IV-C, we consider two data modeling variants that show that structure of the channel mixing matrix can act as a system diversity. Remarkably, this property enables convolutive signal separation in some cases where only two-way diversities are used.

A. Data model

Let us demonstrate that our approach allows us to improve existing results for convolutive signal separation. We consider the problem of convolutive signal separation in synchronised DS-CDMA systems in which the convolution takes place in the far field and the transmitted signals arrive via a few dominant paths. More precisely, we consider a communication system in which

the receiver is equipped with I collocated antennas and the baseband signal received by the ith antenna at the jth oversampling period of the kth symbol period is

$$x_{(i-1)J+j,k} = \sum_{r=1}^{R} \sum_{p=1}^{P} a_i^{(p,r)} \sum_{l=1}^{L} b_{j+(l-1)J}^{(p,r)} s_{k-l+1}^{(r)}, \tag{35}$$

where R, J, P and L denote the number of users, the oversampling factor, the number of dominant paths and the channel length (i.e., interference occurs over L symbols), respectively, $a_i^{(p,r)} \in \mathbb{C}$ denotes the antenna response associated with the ith antenna and the pth path of the rth user, $b_{j+(l-1)J}^{(p,r)} \in \mathbb{C}$ is a sample of the channel response associated with the pth path of the rth user, and $s_{k-l+1}^{(r)}$ is a symbol transmitted by the rth user. See [27], [4], [7], [25] and references therein for further details. We note in passing that convolutive source separation problems of the form (35) also occur in neuroimaging [22].[2] Hence, the results presented in this section are not necessarily limited to telecommunication applications.

Let us stack the coefficients $\{a_i^{(p,r)}\}$ into the antenna response vectors $\mathbf{a}^{(p,r)} \in \mathbb{C}^I$ as follows $(\mathbf{a}^{(p,r)})_i = a_i^{(p,r)}$. Build the channel matrices $\mathbf{B}^{(p,r)} \in \mathbb{C}^{J \times L}$ according to

$$\mathbf{B}^{(p,r)} = \begin{bmatrix} b_1^{(p,r)} & b_{1+J}^{(p,r)} & \cdots & b_{1+(L-1)J}^{(p,r)} \\ b_2^{(p,r)} & b_{2+J}^{(p,r)} & \cdots & b_{2+(L-1)J}^{(p,r)} \\ \vdots & \vdots & \cdots & \vdots \\ b_J^{(p,r)} & b_{J+J}^{(p,r)} & \cdots & b_{J+(L-1)J}^{(p,r)} \end{bmatrix}.$$

Construct also the Toeplitz matrices $\mathbf{S}_{\mathrm{TB}}^{(r)} \in \mathbb{C}^{K \times L}$ according to (13). Define $\mathbf{X}^{(i\cdot)} \in \mathbb{C}^{J \times K}$ according to $(\mathbf{X}^{(i\cdot)})_{jk} = x_{(i-1)J+j,k}$, then $\mathbf{X}^{(i\cdot)} = \sum_{r=1}^{R} \sum_{p=1}^{P} a_i^{(p,r)} \mathbf{B}^{(p,r)} \mathbf{S}_{\mathrm{TB}}^{(r)T}$. Stacking yields the observation matrix $\mathbf{X} \in \mathbb{C}^{IJ \times K}$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1\cdot)} \\ \vdots \\ \mathbf{X}^{(I\cdot)} \end{bmatrix} = \mathbf{M}_{\mathrm{TB}}^{(1)} \mathbf{S}_{\mathrm{TB}}^{(1)T} + \cdots + \mathbf{M}_{\mathrm{TB}}^{(R)} \mathbf{S}_{\mathrm{TB}}^{(R)T} = \mathbf{M}_{\mathrm{TB}} \mathbf{S}_{\mathrm{TB}}^{T}, \tag{36}$$

where

$$\mathbf{M}_{\mathrm{TB}}^{(r)} = \sum_{p=1}^{P} \mathbf{a}^{(p,r)} \otimes \mathbf{B}^{(p,r)} \in \mathbb{C}^{IJ \times L}.$$

Equivalently, we can write

$$\mathbf{X} = \mathbf{M}_{\mathrm{TB}} \mathbf{S}_{\mathrm{TB}}^{T}, = \mathbf{M}_{\mathrm{TB}} \mathbf{\Pi}_{\mathbf{S}} \cdot \mathbf{\Pi}_{\mathbf{S}}^{T} \mathbf{S}_{\mathrm{TB}}^{T} = \mathbf{M}_{\mathrm{BT}} \mathbf{S}_{\mathrm{BT}}^{T}, \tag{37}$$

with

$$\mathbf{M}_{\mathrm{BT}}^{(l)} = \Big[\sum_{p=1}^{P} \mathbf{a}^{(p,1)} \otimes \mathbf{b}_l^{(p,1)}, \ldots, \sum_{p=1}^{P} \mathbf{a}^{(p,R)} \otimes \mathbf{b}_l^{(p,R)} \Big] \in \mathbb{C}^{IJ \times R}. \tag{38}$$

We conclude from (37) and (38) that the convolutive source separation problems in [27], [4], [7], [25], [22] are instances of the general problem addressed in this paper.

[2]The convolutive 'ConvCP' model in [22] corresponds to (35) with $P = 1$.

B. Convolution leads to improved identifiability

Assume that condition (19) in Lemma II.4 is satisfied, then we know from Section III that the block-Toeplitz factorization of $\mathbf{X}$ in (37) can be reduced to the coupled low-rank factorization (cf. (28)):

$$\mathbf{Y}_{\mathrm{tot}} = \begin{bmatrix} \mathbf{Y}^{(1)} \\ \vdots \\ \mathbf{Y}^{(L)} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{\mathrm{BT}}^{(1)} \\ \vdots \\ \mathbf{M}_{\mathrm{BT}}^{(L)} \end{bmatrix} \mathbf{G}^T \in \mathbb{C}^{IJL \times R}, \tag{39}$$

where $\mathbf{M}_{\mathrm{BT}}^{(l)}$ is given by (38) and where $\mathbf{G} \in \mathbb{C}^{R \times R}$ is nonsingular. Note that (39) corresponds to a matrix representation of a coupled BTD of the form (9) with $N = L$, $\mathbf{C} = \mathbf{G}$, and $\mathbf{A}^{(l,r)} = \mathbf{A}^{(r)} := [\mathbf{a}^{(1,r)}, \ldots, \mathbf{a}^{(P,r)}] \in \mathbb{C}^{I \times P}$, $\forall l \in \{1, \ldots, L\}$. Consequently, the combination of relation (39) and Theorem II.3 already yields an identifiability condition. However, a more relaxed condition can be obtained by additionally taking the constraint $\mathbf{A}^{(l,r)} = \mathbf{A}^{(r)}$, $\forall l \in \{1, \ldots, L\}$ into account. In short, by permuting the rows of $\mathbf{Y}_{\mathrm{tot}}$, the factorization in (39) reduces to a single BTD. In more detail, from (38) and (39) it is clear that the ith block row $\mathbf{Y}^{(i,l)} \in \mathbb{C}^{J \times R}$ of $\mathbf{Y}^{(l)} = [\mathbf{Y}^{(1,l)T}, \ldots, \mathbf{Y}^{(I,l)T}]^T \in \mathbb{C}^{IJ \times R}$ admits the factorization

$$\mathbf{Y}^{(i,l)} = \Big[\sum_{p=1}^{P} a_i^{(p,1)} \mathbf{b}_l^{(p,1)}, \ldots, \sum_{p=1}^{P} a_i^{(p,R)} \mathbf{b}_l^{(p,R)} \Big] \mathbf{G}^T.$$

It can now be verified that the following row-permuted version of $\mathbf{Y}_{\mathrm{tot}}$ admits the factorization

$$\begin{aligned} \check{\mathbf{Y}}_{\mathrm{tot}} &= \begin{bmatrix} \check{\mathbf{Y}}^{(1)} \\ \vdots \\ \check{\mathbf{Y}}^{(I)} \end{bmatrix} = \Big[\sum_{p=1}^{P} \mathbf{a}^{(p,1)} \otimes \check{\mathbf{b}}^{(p,1)}, \ldots, \sum_{p=1}^{P} \mathbf{a}^{(p,R)} \otimes \check{\mathbf{b}}^{(p,R)} \Big] \\ &= \Big[\mathrm{Vec}\Big(\check{\mathbf{B}}^{(1)} \mathbf{A}^{(1)T} \Big), \ldots, \mathrm{Vec}\Big(\check{\mathbf{B}}^{(R)} \mathbf{A}^{(R)T} \Big) \Big] \mathbf{G}^T \in \mathbb{C}^{IJL \times R}, \end{aligned} \tag{40}$$

where $\check{\mathbf{Y}}^{(i)} = [\mathbf{Y}^{(i,1)T}, \ldots, \mathbf{Y}^{(i,L)T}]^T \in \mathbb{C}^{JL \times R}$, $\check{\mathbf{b}}^{(p,r)} = \mathrm{Vec}(\mathbf{B}^{(p,r)}) \in \mathbb{C}^{JL}$, $\check{\mathbf{B}}^{(r)} = [\check{\mathbf{b}}^{(1,r)}, \ldots, \check{\mathbf{b}}^{(P,r)}] \in \mathbb{C}^{JL \times P}$ and $\mathbf{A}^{(r)} = [\mathbf{a}^{(1,r)}, \ldots, \mathbf{a}^{(P,r)}] \in \mathbb{C}^{I \times P}$. Comparing (10) with (40), it is clear that the decomposition of $\check{\mathbf{Y}}_{\mathrm{tot}}$ corresponds to a BTD of a tensor for which a uniqueness condition has already been stated in Theorem II.3. Proposition IV.1 below summarizes the obtained uniqueness condition for convolutive source separation problems with input-output relation of the form (35). It makes use of the matrix

$$\mathbf{Z}_P = \Big(C_{P+1}(\mathbf{A}) \odot C_{P+1}\Big(\check{\mathbf{B}}\Big) \Big) \mathbf{P}_{\mathrm{BTD}} \in \mathbb{C}^{C_I^{P+1} C_{JL}^{P+1} \times (C_{R+P}^{P+1} - R)}, \tag{41}$$

where $\mathbf{A} = [\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(R)}] \in \mathbb{C}^{I \times PR}$ and $\check{\mathbf{B}} = [\check{\mathbf{B}}^{(1)}, \ldots, \check{\mathbf{B}}^{(R)}] \in \mathbb{C}^{JL \times PR}$, and where the column selection matrix $\mathbf{P}_{\mathrm{BTD}} \in \mathbb{C}^{C_{RP}^{P+1} \times (C_{R+P}^{P+1} - R)}$ is defined according to (11).

Proposition IV.1. *Consider the decomposition of $\boldsymbol{X} \in \mathbb{C}^{IJ \times K}$ in* (37)*, with $\boldsymbol{M}_{BT}^{(l)} = [Vec(\boldsymbol{B}^{(l,1)} \boldsymbol{A}^{(1)T}), \ldots, Vec(\boldsymbol{B}^{(l,R)} \boldsymbol{A}^{(R)T})]$, $l \in$*

$\{1,\ldots,L\}$, *and* $\boldsymbol{S}_{BT}$ *block-Toeplitz. If*

$$\begin{cases} \boldsymbol{H}_{\mathbf{S}_{BT}} \text{ in (18) has full column rank,} \\ \boldsymbol{M}_{BT} \text{ in (37) has full column rank,} \\ \boldsymbol{Z}_P \text{ in (41) has full column rank,} \end{cases} \tag{42}$$

then the decomposition of $\boldsymbol{X}$ *is unique.*

Note that under condition (42), the factorization of $\mathbf{X}$ can be computed algebraically via a two-step procedure as in Section III. Namely, we first compute the block-Toeplitz factorization of $\mathbf{X}$ followed by the computation of the BTD expressed by (40), using for instance the algebraic algorithms in [36].

We will now show that Proposition IV.1 yields uniqueness conditions that are more relaxed than the results reported in the literature. Since most of the existing results (e.g., [27], [24], [7], [26]) are limited to cases where $P = 1$, we will postpone the discussion of cases where $P \geq 2$ until Section IV-C.

The first uniqueness condition for matrix factorizations of the form (36) with $P = 1$ was obtained in [27]. In short, a link between the GEVD and the PARALIND model [2] was derived that generically guarantees uniqueness if the following dimensionality constraints are satisfied (see [27] for a deterministic uniqueness condition):

$$P = 1, \quad \min(J, K) \geq R \cdot L \quad \text{and} \quad I \geq 2\,. \tag{43}$$

More relaxed conditions have since been derived in [24], [7], [26]. The most relaxed uniqueness condition was obtained in [26]. Roughly speaking, a link between joint block-diagonalization and BTD was established that requires the following dimensionality constraints to be satisfied (see [26] for a deterministic uniqueness condition):

$$P = 1, \quad K \geq R \cdot L, \quad I \geq L \quad \text{and} \quad C_I^2 C_J^2 \geq C_R^2 \cdot L^2\,. \tag{44}$$

Note that both conditions (43) and (44) require oversampling by a factor $J \geq 2$ and assume single-path propagation ($P = 1$). In cases where $P = 1$ and $J \geq 2$, we expect that condition (42) in Proposition IV.1 is satisfied if $K \geq R(L+1)$, $I \geq L$, and $C_I^2 C_{JL}^2 \geq C_R^2$. (These are the dimensionality constraints under which $\mathbf{H}_{\mathbf{S}_{BT}}$, $\mathbf{M}_{BT}$ and $\mathbf{Z}_P$ are expected to have full column rank generically.) Note that the latter inequality imposes less restrictive conditions on I and J than the last inequality condition in (44). On the other hand, the first inequality imposes a slightly more restrictive condition on K. However, since K corresponds to the number of symbol periods we typically have that $K >> RL$, making the constraint $K \geq R(L+1)$ less significant. Overall, we expect that Proposition IV.1 yields relaxed uniqueness conditions in practice. As it has not been proved that these matrices generically have full column rank under the mentioned conditions, we resort to a tool with which generic properties can be checked for specific parameter choices. This tool is the following lemma.

Lemma IV.2. *Let* $f : \mathbb{C}^n \to \mathbb{C}$ *be an analytic function. If there exists an element* $\boldsymbol{x} \in \mathbb{C}^n$ *such that* $f(\boldsymbol{x}) \neq 0$*, then the set* $\{\boldsymbol{x} \mid f(\boldsymbol{x}) = 0\}$ *is of Lebesgue measure zero (e.g., [13]).*

Recall that an $I \times R$ matrix has full column rank R if it has a non-vanishing $R \times R$ minor. A minor is an analytic function (namely, it is a polynomial in several variables). Lemma IV.2 implies that in order to obtain a generic bound on R, it suffices to numerically check condition (42) for a random block-Toeplitz matrix $\mathbf{S}_{BT}$ and a mixing matrix $\mathbf{M}_{BT}$ of the form (38) with $P = 1$ and built from random $\{\mathbf{a}^{(1,r)}\}$ and $\{\mathbf{B}^{(1,r)}\}$.

Table I indicates the maximum value of R for which conditions (42), (43) and (44) generically guarantee uniqueness for cases with $J = 4$, $K > R(L+1)$, $L = 2$ or $L = 3$, and I varying. In the cases listed in Table I, condition (42) in Proposition IV.1 indeed relaxes the bound on R significantly. See [30] for more comparisons.

Another benefit of the presented framework is that it also provides an algebraic solution to the convolutive signal separation problem. Recall that the proposed algebraic method consists of a two-step procedure, namely a deconvolution and a tensor decomposition step. The complexity of the deconvolution step for this problem is of the order $O\left(4(IJL)^2 + 22(K+L-1)^3\right)$ when the noise subspace method discussed in [32] is implemented via the R-SVD method [12]. The computation of the subsequent CPD has a complexity of the order $O(I^2J^2L^2R^2)$ when a simultaneous matrix diagonalization method is used [36]. In comparison, the iterative alternating least squares method for this problem has a complexity of the order $O\left(IJK \cdot ((K+L-1)^2 + L^2 + 1)R^2\right)$ per iteration [7].

C. Convolution acts as a system diversity

In this section we discuss two scenarios in which convolutive source separation is possible, despite only two-way diversities (space × time). Instead of three-way diversities (space × oversampling × time), we assume that the propagation channel has "hidden" low-rank structure.

1) Coherent channels: Consider again a system with input-output relation of the form (35) but now without oversampling (i.e., $J = 1$). This implies that the submatrices in $\mathbf{M}_{BT} = [\mathbf{M}_{BT}^{(1)}, \ldots, \mathbf{M}_{BT}^{(L)}]$ reduce to $\mathbf{M}_{BT}^{(l)} = [\sum_{p=1}^{P} \mathbf{a}^{(p,1)} b_l^{(p,1)}, \ldots, \sum_{p=1}^{P} \mathbf{a}^{(p,R)} b_l^{(p,R)}]$. The presence of coherent channels ($P > 1$) is now necessary for uniqueness. Namely, if $P = 1$, then $\mathbf{X} = \sum_{l=1}^{L} \mathbf{M}_{BT}^{(l)} \mathbf{S}_{BT}^{(l)T} = \sum_{r=1}^{R} \mathbf{M}_{TB}^{(r)} \mathbf{S}_{TB}^{(r)T} = \sum_{r=1}^{R} \mathbf{a}^{(1,r)} \mathbf{b}^{(1,r)T} \mathbf{S}_{TB}^{(r)T}$, so that the recovery of $\mathbf{S}_{TB}^{(r)}$ is impossible.[3] Furthermore, in order to obtain a

[3]Note that $\mathbf{Sb} = \mathbf{T_b p}_{tot}$, where $\mathbf{p}_{tot} = [s_{1-L}, \ldots, s_K]^T \in \mathbb{C}^{K+L-1}$ is the total generator vector of the Toeplitz matrix $\mathbf{S} = \begin{bmatrix} s_1 & \cdots & s_{1-L} \\ \vdots & \ddots & \vdots \\ s_K & \cdots & s_{K-L+1} \end{bmatrix} \in \mathbb{C}^{K \times L}$ and $\mathbf{T_b} = \begin{bmatrix} b_L & \cdots & b_1 & & \\ & b_L & \cdots & b_1 & \\ & & \ddots & & \ddots \end{bmatrix} \in \mathbb{C}^{K \times (K+L-1)}$ is the banded Toeplitz matrix built from the vector $\mathbf{b} = [b_1, \ldots, b_L]^T \in \mathbb{C}^L$. It is now clear that any vector $\mathbf{p}_{tot} + \hat{\mathbf{p}}_{tot}$ with $\hat{\mathbf{p}}_{tot} \in \ker(\mathbf{T_b})$ yields an alternative factorization of $\mathbf{Sb}$, i.e., $\mathbf{Sb} = \mathbf{T_b p}_{tot} = \mathbf{T_b}(\mathbf{p}_{tot} + \hat{\mathbf{p}}_{tot}) = (\mathbf{S} + \hat{\mathbf{S}})\mathbf{b}$ where $\hat{\mathbf{S}}$ is the Toeplitz matrix built from $\hat{\mathbf{p}}_{tot}$.

L	2									3								
I	2	3	4	5	6	7	8	9	10	2	3	4	5	6	7	8	9	10
Condition (42)	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**	**20**	**2**	**4**	**5**	**6**	**8**	**9**	**10**	**12**	**13**
Condition (43), [27]	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1
Condition (44), [26]	2	3	4	6	7	8	9	10	12	1	1	1	1	5	5	6	7	8

TABLE I
MAXIMUM VALUE OF R IN (36) WITH $J = 4$ AND $K \geq R(L+1)$ FOR WHICH CONDITIONS (42), (44) AND (43) ARE SATISFIED.

uniqueness condition and an algorithm based on Proposition IV.1, the inequality $P < L$ must be satisfied as well. The reason is that if $P \geq L$ and $J = 1$, then $\check{\mathbf{M}}_{\text{tot}}$ in (40) does not have any low-rank structure, implying that $\mathbf{Z}_P$ is not even defined. However, the condition $P < L$ implies that $\mathbf{M}_{\text{BT}}$ will be rank deficient. Indeed, if $P < L$ and $J = 1$, then the columns of $\mathbf{M}_{\text{TB}}^{(r)} = [\sum_{p=1}^{P} \mathbf{a}^{(p,r)} b_1^{(p,r)}, \ldots, \sum_{p=1}^{P} \mathbf{a}^{(p,r)} b_L^{(p,r)}]$ are linearly dependent, implying that $\mathbf{M}_{\text{BT}} = \mathbf{M}_{\text{TB}} \boldsymbol{\Pi}_{\mathbf{S}}$ cannot have full column rank. Fortunately, temporal smoothing ($m > 1$) can in several cases be used to overcome this rank deficiency. Let $\mathbf{X}_{\text{BT,sm}}^{(m)} = \mathbf{M}_{\text{BT,sm}}^{(m)} \mathbf{S}_{\text{BT,sm}}^{(m)T}$ denote the temporally smoothed version of $\mathbf{X}$, in which $\mathbf{M}_{\text{BT,sm}}^{(m)}$ is of the form (20) with $\mathbf{M}_{\text{BT}}^{(l)} = [\sum_{p=1}^{P} \mathbf{a}^{(p,1)} b_l^{(p,1)}, \ldots, \sum_{p=1}^{P} \mathbf{a}^{(p,R)} b_l^{(p,R)}]$ and in which $\mathbf{S}_{\text{BT,sm}}^{(m)}$ is given by (21). Proposition IV.1 now states that if there exists an integer m, such that

$$\begin{cases} \mathbf{M}_{\text{BT,sm}}^{(m)} \text{ in (20) has full column rank,} \\ \mathbf{H}_{\mathbf{S}_{\text{BT,sm}}^{(m)}} \text{ in (22) has full column rank,} \\ \mathbf{Z}_P \text{ in (41) has full column rank,} \end{cases} \tag{45}$$

then the decomposition of $\mathbf{X}$ is unique. Furthermore, we know that the decomposition of $\mathbf{X}$ can be found by computing a block-Toeplitz factorization followed by computing a BTD. Note that in the exact case, the latter decomposition can be computed via GEVD, as briefly mentioned in Subsection II-C. To summarize, if $1 < P < L$ and $m > 1$, then condition (45) states that convolutive ($L \geq 2$) signal separation is possible despite only two-way diversity ($I \times K$), i.e., no temporal oversampling is needed ($J = 1$). The reason is that the third mode results from deconvolution. This is in contrast to the instantaneous case ($L = 1$) where three-way diversity ($I \times J \times K$) is necessary, even in the case of coherent channels ($P \geq 2$). Indeed, if $J = 1$ and $L = 1$, then $\mathbf{X} = \mathbf{M}_{\text{BT}} \mathbf{S}_{\text{BT}}^T = \mathbf{M}_{\text{BT}} \mathbf{N} \cdot \mathbf{N}^{-1} \mathbf{S}_{\text{BT}}^T$ for any nonsingular matrix $\mathbf{N} \in \mathbb{C}^{R \times R}$.

Table II lists the maximum value of R for which condition (45) is generically satisfied with $J = 1$, $K = 80$, and varying L, P, I. The listed values of m in Table II were obtained by varying m over all possible values $\{\lceil \frac{(L-1)R}{I-R} \rceil, \ldots, \lfloor \frac{K+1-(L-1)R}{R+1} \rfloor\}$ and retaining the best result.

2) Widely separated antenna arrays: As a second illustration, we consider a scenario in which N widely separated antenna arrays are used. Assume that the output of the ith sensor of the nth receive antenna array at the kth data snapshot is of the form (35) with $J = 1$ and $P = 1$ (i.e., there is no oversampling and the angle spread is small), that is

$$x_{ik}^{(n)} = \sum_{r=1}^{R} a_i^{(n,r)} \sum_{l=1}^{L} b_l^{(n,r)} s_{k-l+1}^{(r)},$$

where $a_i^{(n,r)}$ is the response of the ith antenna of the nth antenna array to the rth signal, $b_l^{(n,r)}$ is a filter coefficient associated with the channel between the nth antenna array and the rth user and $s_{k-l+1}^{(r)}$ is a symbol transmitted by the rth user. In other words, N versions of (35) with $J = 1$ and $P = 1$ are observed. Stacking yields

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(N)} \end{bmatrix} = \mathbf{M}_{\text{TB}} \mathbf{S}_{\text{TB}}^T = \mathbf{M}_{\text{TB}} \boldsymbol{\Pi}_{\mathbf{S}} \cdot \boldsymbol{\Pi}_{\mathbf{S}}^T \mathbf{S}_{\text{TB}}^T = \mathbf{M}_{\text{BT}} \mathbf{S}_{\text{BT}}^T, \tag{46}$$

where the submatrices of $\mathbf{M}_{\text{BT}} = [\mathbf{M}_{\text{BT}}^{(1)}, \ldots, \mathbf{M}_{\text{BT}}^{(L)}]$ are given by

$$\mathbf{M}_{\text{BT}}^{(l)} = \begin{bmatrix} \mathbf{M}_{\text{BT}}^{(1,l)} \\ \vdots \\ \mathbf{M}_{\text{BT}}^{(N,l)} \end{bmatrix} = [\mathbf{m}_l^{(1)}, \ldots, \mathbf{m}_l^{(R)}],$$

in which $\mathbf{M}_{\text{BT}}^{(n,l)} = [\mathbf{a}^{(n,1)} b_l^{(n,1)}, \ldots, \mathbf{a}^{(n,R)} b_l^{(n,R)}] \in \mathbb{C}^{I_n \times R}$. The structure implies that the nth block row of $\mathbf{M}_{\text{BT}} \in \mathbb{C}^{(\sum_{n=1}^{N} I_n) \times R}$ can be rearranged into a matrix that has Khatri–Rao structure:

$$\check{\mathbf{F}}^{(n)} := \begin{bmatrix} \mathbf{M}_{\text{BT}}^{(n,1)} \\ \vdots \\ \mathbf{M}_{\text{BT}}^{(n,L)} \end{bmatrix} = \mathbf{B}^{(n)} \odot \mathbf{A}^{(n)} \in \mathbb{C}^{LI_n \times R}, \tag{47}$$

where $\mathbf{B}^{(n)} = [\mathbf{b}^{(n,1)}, \ldots, \mathbf{b}^{(n,R)}] \in \mathbb{C}^{L \times R}$ and $\mathbf{A}^{(n)} = [\mathbf{a}^{(n,1)}, \ldots, \mathbf{a}^{(n,R)}] \in \mathbb{C}^{I_n \times R}$, $n \in \{1, \ldots, N\}$. Stacking yields

$$\check{\mathbf{F}} := \begin{bmatrix} \check{\mathbf{F}}^{(1)} \\ \vdots \\ \check{\mathbf{F}}^{(N)} \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{(1)} \odot \mathbf{A}^{(1)} \\ \vdots \\ \mathbf{B}^{(N)} \odot \mathbf{A}^{(N)} \end{bmatrix}. \tag{48}$$

Relation (48), in turn, yields the following variant of $\mathbf{Y}_{\text{tot}}$ in (30):

$$\check{\mathbf{Y}}_{\text{tot}} = \begin{bmatrix} \check{\mathbf{Y}}^{(1)} \\ \vdots \\ \check{\mathbf{Y}}^{(N)} \end{bmatrix} = \begin{bmatrix} \check{\mathbf{F}}^{(1)} \\ \vdots \\ \check{\mathbf{F}}^{(N)} \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{(1)} \odot \mathbf{A}^{(1)} \\ \vdots \\ \mathbf{B}^{(N)} \odot \mathbf{A}^{(N)} \end{bmatrix} \mathbf{G}^T, \tag{49}$$

with $\check{\mathbf{Y}}^{(n)}$ defined in analogy with $\check{\mathbf{F}}^{(n)}$ in (47), i.e., $\check{\mathbf{Y}}^{(n)} = [\mathbf{Y}^{(n,1)T}, \ldots, \mathbf{Y}^{(n,L)T}]^T$ in which $\mathbf{Y}^{(n,l)} \in \mathbb{C}^{I_n \times R}$ is the nth block row of $\mathbf{Y}^{(l)} = [\mathbf{Y}^{(1,l)T}, \ldots, \mathbf{Y}^{(N,l)T}]^T \in \mathbb{C}^{(\sum_{n=1}^{N} I_n) \times R}$.

L	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5
P	2	2	2	2	2	2	2	2	3	3	3	3	2	2	2	2	3	3	3	3	4	4	4	4
I	5	6	7	8	5	6	7	8	5	6	7	8	5	6	7	8	5	6	7	8	5	6	7	8
m	3	4	5	6	12	6	8	5	2	3	4	5	6	8	6	7	6	8	6	7	1	2	3	4
R	3	4	5	6	4	4	5	5	2	3	4	5	3	4	4	5	3	4	4	5	1	2	3	4

TABLE II
MAXIMUM VALUE OF R IN (36) FOR WHICH CONDITION (45) GENERICALLY GUARANTEES UNIQUENESS WITH $J = 1$ AND $K = 80$.

Condition (50) below is an adaptation of condition (32) in Proposition III.1 to matrix factorizations of the form (46):

$$\begin{cases} \mathbf{M}_{\mathrm{BT}} \text{ in (20) has full column rank,} \\ \mathbf{H}_{\mathbf{S}_{\mathrm{BT}}} \text{ in (22) has full column rank,} \\ \mathbf{Z}^{(N)} \text{ in (51) has full column rank,} \end{cases} \tag{50}$$

where $\mathbf{Z}^{(L)}$ in (32) has been replaced by the matrix

$$\mathbf{Z}^{(N)} = \begin{bmatrix} C_2(\mathbf{A}^{(1)}) \odot C_2(\mathbf{B}^{(1)}) \\ \vdots \\ C_2(\mathbf{A}^{(N)}) \odot C_2(\mathbf{B}^{(N)}) \end{bmatrix} \in \mathbb{C}^{(\sum_{n=1}^{N} C_{I_n}^2 C_L^2) \times C_R^2}, \tag{51}$$

which takes the coupled CPD low-rank structure of (49) into account. Similarly, the two-step block-Toeplitz factorization and coupled CPD procedure associated with Proposition III.1 can be adapted to this problem.

Since no oversampling is used, the factorization of $\mathbf{X}$ in (46) is not unique in the instantaneous case ($L = 1$). (Recall that if $L = 1$, then $\mathbf{X} = \mathbf{M}_{\mathrm{BT}}\mathbf{S}_{\mathrm{BT}}^T = \mathbf{M}_{\mathrm{BT}}\mathbf{N} \cdot \mathbf{N}^{-1}\mathbf{S}_{\mathrm{BT}}^T$ for any nonsingular matrix $\mathbf{N} \in \mathbb{C}^{R \times R}$.) The necessity of a convolutive channel ($L \geq 2$) can intuitively be understood as a requirement for the factorization in (49) to have coupled CPD structure. Indeed, if $L = 1$, then (49) reduces to the factorization $\check{\mathbf{Y}}_{\mathrm{tot}} = \left[D_1(\mathbf{B}^{(1)})\mathbf{A}^{(1)T}, \ldots, D_1(\mathbf{B}^{(N)})\mathbf{A}^{(N)T}\right]^T \mathbf{G}^T$, which is clearly not unique.

Since it is assumed that the angle spread is small, another necessary condition is that widely separated antenna arrays are used ($N \geq 2$), even if $L \geq 2$. In fact, if $N = 1$, then (46) reduces to $\mathbf{X} = \mathbf{X}^{(1)} = \sum_{r=1}^{R} \mathbf{a}^{(1,r)}\mathbf{b}^{(1,r)T}\mathbf{S}_{\mathrm{TB}}^{(r)T}$, making signal recovery impossible as explained in Subsection IV-C1. The necessity of $N \geq 2$ can also intuitively be understood as a condition for the channel matrix $\mathbf{M}_{\mathrm{BT}}$ to have full column rank, despite the use of temporal smoothing ($m \geq 1$).[4]

Table III lists the maximum value for R in (46) for which a smoothed ($m \geq 1$) version of condition (50) generically guarantees uniqueness with fixed $K = 80$, $N = 2$, and varying L, I_1 and I_2. The listed values of m in Table III were obtained by varying m over all possible values $\{\lceil \frac{(L-1)R}{I_1+I_2-R} \rceil, \ldots, \lfloor \frac{K+1-(L-1)R}{R+1} \rfloor\}$ and retaining the best result.

L	2	2	2	2	2	2	3	3	3	3	3	3
$I_1 = I_2$	3	4	5	6	7	8	3	4	5	6	7	8
m	2	2	2	2	2	3	10	6	5	4	4	4
R	4	5	6	8	9	11	5	6	7	8	9	10

TABLE III
MAXIMUM VALUE OF R IN (46) FOR WHICH A SMOOTHED ($m \geq 1$) VERSION OF CONDITION (50) GENERICALLY GUARANTEES UNIQUENESS WITH $K = 80$ AND $N = 2$.

V. NUMERICAL EXPERIMENTS

Consider the low-rank and Toeplitz structured factorization of $\mathbf{X}$ in (1). The goal is to estimate $\mathbf{S}_{\mathrm{TB}}$ from $\mathbf{T} = \mathbf{X} + \beta\mathbf{N}$, where $\mathbf{N}$ is an unstructured perturbation matrix and $\beta \in \mathbb{R}$ controls the noise level. The Signal-to-Noise Ratio (SNR) is measured as: SNR [dB] = $10\log(\|\mathbf{X}\|_F^2/\|\beta\mathbf{N}\|_F^2)$. The true $\{\mathbf{p}_k\}$ and estimated $\{\widehat{\mathbf{p}}_k\}$ source variables can be stored in the matrices (cf. Eq. (14)):

$$\mathbf{P} = \left[\mathbf{p}_1^T, \ldots, \mathbf{p}_{K+L-1}^T\right]^T, \quad \widehat{\mathbf{P}} = \left[\widehat{\mathbf{p}}_1^T, \ldots, \widehat{\mathbf{p}}_{K+L-1}^T\right]^T.$$

The performance is measured in terms of the Relative Error Norm (REN), defined as $\mathrm{REN}(\mathbf{P}) = 20\log_{10}(\|\mathbf{P} - \widehat{\mathbf{P}}\boldsymbol{\Delta}\boldsymbol{\Pi}\|_{\mathrm{F}}/\|\mathbf{P}\|_{\mathrm{F}})$, where the matrices $\boldsymbol{\Delta}$ and $\boldsymbol{\Pi}$ represent the optimal scaling and permutation matrix, respectively.

A. Case 1: Coherent channels

Consider a convolutive mixture of $R = 2$ sources with channel length $L = 4$, measured with $I = 6$ collocated antennas. The channel has large angle spread with $P = 2$ dominant paths. Note that due to the lack of a third diversity (e.g., oversampling), no alternative algorithms have been presented in the literature. The temporal smoothing factor is set to $m = 6$. The sources are QPSK signals of length $K = 80$. Figure 2(a) shows the mean $\mathrm{REN}(\mathbf{P})$ values over 100 trials as a function of SNR for the (i) algebraic approach presented in this paper, the optimization-based nonlinear least squares solver `sdf_nls.m` in Tensorlab [44] initialized using (ii) a single random initialization, (iii) the best out of five random initializations, or (iv) algebraic initialization

[4] It is easy to understand from an example that $N \geq 2$ is necessary for the temporally smoothed channel matrix $\mathbf{M}_{\mathrm{BT,sm}}^{(m)}$ to have full column rank. Consider the case where $N = 1$, $R = 2$ and $L = 2$, i.e., $\mathbf{M}_{\mathrm{BT}} = [\mathbf{M}_{\mathrm{BT}}^{(1)}, \mathbf{M}_{\mathrm{BT}}^{(2)}] = [\mathbf{a}^{(1,1)}b_1^{(1,1)}, \mathbf{a}^{(1,1)}b_2^{(1,1)}, \mathbf{a}^{(1,2)}b_1^{(1,2)}, \mathbf{a}^{(1,2)}b_2^{(1,2)}]$. There are two possible choices of m, namely $m = 1$ and $m = 2$. Clearly, $\mathbf{M}_{\mathrm{BT,sm}}^{(1)} = \mathbf{M}_{\mathrm{BT}}$ does not have full column rank. The matrix $\mathbf{M}_{\mathrm{BT,sm}}^{(2)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{a}^{(1,1)}b_1^{(1,1)} & \mathbf{a}^{(1,1)}b_2^{(1,1)} & \mathbf{a}^{(1,2)}b_1^{(1,2)} & \mathbf{a}^{(1,2)}b_2^{(1,2)} \\ \mathbf{a}^{(1,1)}b_1^{(1,1)} & \mathbf{a}^{(1,1)}b_2^{(1,1)} & \mathbf{a}^{(1,2)}b_1^{(1,2)} & \mathbf{a}^{(1,2)}b_2^{(1,2)} & \mathbf{0} & \mathbf{0} \end{bmatrix}$ does not have full column rank either, e.g., the columns $\mathbf{m}_{\mathrm{BT,sm},1}^{(2)}$, $\mathbf{m}_{\mathrm{BT,sm},3}^{(2)}$ and $\mathbf{m}_{\mathrm{BT,sm},5}^{(2)}$ are linearly dependent. The reasoning can be generalized to arbitrary L and R.

using the presented method. The methods are referred to as 'Algebraic', 'Tensorlab 1 random init', 'Tensorlab 5 random init' and 'Tensorlab algebraic init', respectively. The figure shows that the algebraic algorithm provides a reasonable estimate in the presence of noise. Especially at higher SNR levels, the optimization-based method initialized with the algebraic method performs better than the randomly initialized approach since it starts close to the true solution. We also observe that the 'Tensorlab 5 random init' and 'Tensorlab algebraic init' methods perform about the same. However, due to the cheap initalization provided by the presented algebraic method, the latter approach was about five times faster (corresponding to the number of random initializations used by 'Tensorlab 5 random init') than the former approach.

Figure 2(b) shows the mean REN $(\mathbf{P})$ values over 100 trials as a function of SNR for the proposed algebraic approach in which the signal length (K) is varying. As expected, the performance improves as the number of observations increases.

B. Case 2: Widely separated antenna arrays

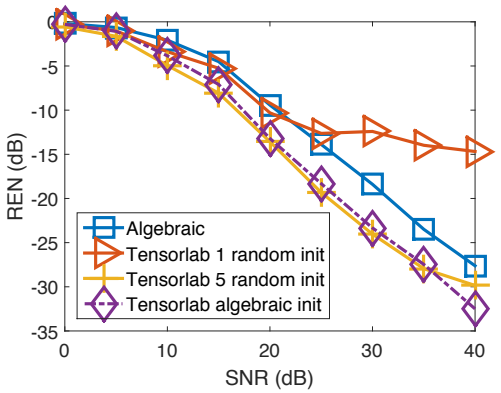
Consider a convolutive mixture of $R = 3$ sources with channel length $L = 2$, a single dominant path $P = 1$, measured in $N = 2$ widely separated antenna arrays consisting of $I = 5$ antennas each. Again, due to the lack of a third diversity, no alternative algorithms have been presented in the literature. The temporal smoothing factor is $m = 2$ and the sources are QPSK signals of length $K = 80$. Figure 3(a) shows the mean REN $(\mathbf{P})$ values over 100 trials as a function of SNR for the 'Algebraic', 'Tensorlab 1 random init', 'Tensorlab 5 random init' and 'Tensorlab algebraic init' methods. We can draw the same conclusions as in the previous Case 1.

Figure 3(b) shows the mean REN $(\mathbf{P})$ values over 100 trials as a function of SNR for the proposed algebraic approach in which the number of signals (R) is varying. At 20 SNR we observe that there is a difference of about 20 dB between $R = 2$ and $R = 6$ (i.e., 4 sources added).

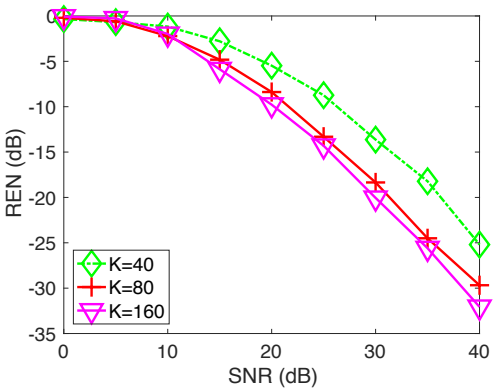
C. Case 3: Oversampled system

Consider a convolutive mixture of $R = 2$ sources with channel length $L = 2$, a single dominant path $P = 1$, measured by a receiver equipped with $I = 5$ collocated antennas. The temporal smoothing factor is $m = 6$ and the sources are QPSK signals of length $K = 80$. The observed data have been oversampled by a factor $J = 4$. Due to the oversampling the measured data enjoy a triple diversity (space × oversampling × time), which implies that tensor methods that ignore the convolutive nature of the channel (i.e., the block-Toeplitz structure of $\mathbf{S}_{\mathrm{BT}}$ is ignored) can be used to solve the signal separation problem. We compare the proposed algebraic convolutive tensor method with a randomly initialized BTD method in Tensorlab that ignores the block-Toeplitz structure of $\mathbf{S}_{\mathrm{BT}}$. In the latter method the deconvolution is performed independently from the separation in a subsequent step. Figures 4(a)–4(b) show the mean and median REN $(\mathbf{P})$ values over 100 trials as a function of SNR for the 'Algebraic' and 'Tensorlab 1 random init' methods. Comparing Figure 4(a) with Figure 4(b) it is clear that 'Tensorlab 1 random init' is sensitive w.r.t. its initialization.

Figure 4(c) shows the mean REN $(\mathbf{P})$ values over 100 trials as a function of SNR for the proposed algebraic approach in which $K = 20$ and the smoothing parameter (m) is varying. The performance drops as m increases. The reason is that in the deconvolution step, the structure of the row space of the $IJm \times (K - m + 1)$ observation matrix $\mathbf{X}^{(m)}_{\mathrm{BT,sm}}$ is exploited, which is affected by the choice of m. More precisely, increasing m will increase the number of unknown variables in the $(L+m-1)R \times (L+m-1)R$ matrix $\mathbf{N}^{(m)}$ in (26) while at the same time decreasing the number of rows in the $(K-m+1) \times (L+m-1)R$ matrix $\mathbf{U}^{(m)}$. This makes the linear system (26) less well-conditioned. For this particular example, it is mainly for $m = 5$ that a decrease of accuracy can be seen.



(a) Comparison between algebraic and optimization-based methods.

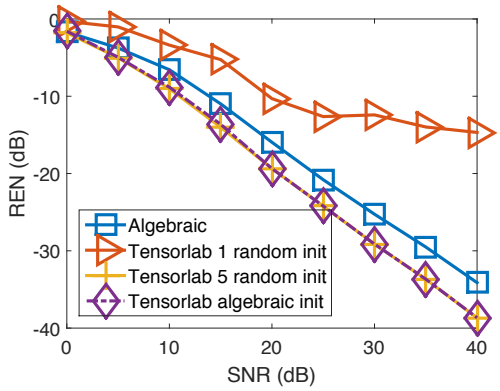


(b) Algebraic method and varying signal length.

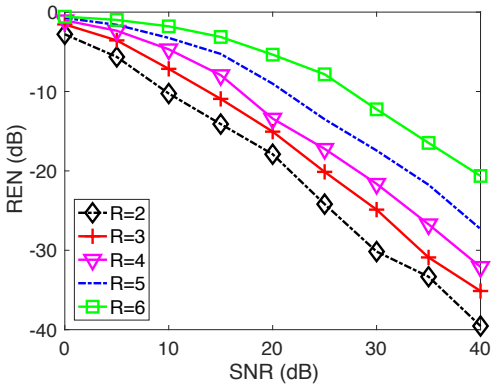
Fig. 2. Case 1: Coherent channels ($L = 4$ and $P = 2$).

VI. Conclusion

Low-rank tensor decompositions play an important role in blind source separation. This paper presented



(a) Comparison between algebraic and optimization-based methods.



(b) Algebraic method and varying number of signal sources.

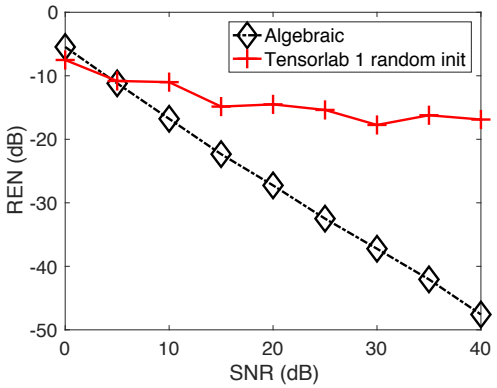
Fig. 3. Case 2: Widely separated antenna arrays ($L = 2$ and $N = 2$).

a generalization of the instantaneous low-rank source separation modeling framework to the convolutive case. In particular, we made a connection between convolutive extensions of CPD/BTD and the coupled CPD/BTD model. This led to a new deterministic uniqueness condition for convolutive low-rank source separation that is easy to check. By simultaneously exploiting the Toeplitz and low-rank structures of the problem, we explained that the convolutive channel can act as a system diversity that in some cases enables source separation despite only two-way diversities. The link between convolutive low-rank factorizations and coupled decompositions also led to an algebraic method that under mild conditions is guaranteed to find the decomposition in the exact case. In the presence of noise, it can be used to initialize an optimization-based method. Numerical experiments confirmed that the presented algorithm can be used as an efficient initialization method.

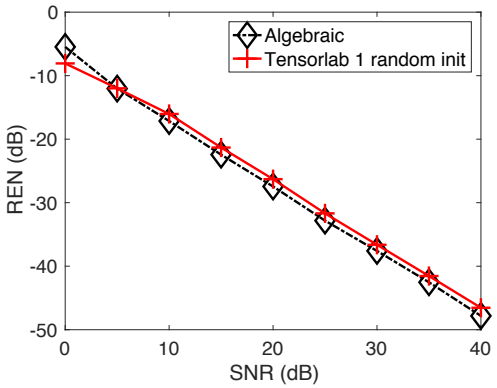
References

[1] E. Acar, R. Bro, and A. Smilde, "Data fusion in metabolics using coupled matrix tensor factorizations," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1602–1620, Sep. 2015.

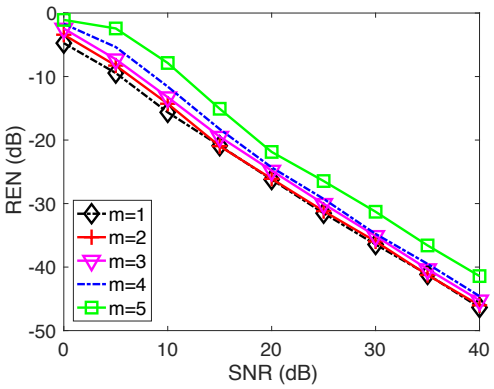
[2] R. Bro, R. A. Harshman, N. D. Sidiropoulos, and M. E. Lundy, "Modeling multi-way data with linearly dependent loadings," *J. Chemometrics*, vol. 23, pp. 324–340, Feb. 2009.



(a) Mean REN performance of algebraic convolutive tensor method and optimization-based BTD method.



(b) Median REN performance of algebraic convolutive tensor method and optimization-based BTD method.



(c) Algebraic method and varying smoothing parameter.

Fig. 4. Case 3: Oversampled antenna array system ($L = 2$ and $J = 4$).

[3] P. Comon, "MA identification using fourth order cumulants," *Signal Processing*, vol. 26, pp. 381–388, 1992.

[4] A. L. F. De Almeida, G. Favier, and J. C. M. Mota, "PARAFAC-based unified tensor modeling for wireless communication systems with application to blind multiuser equalization," *Signal Processing*, vol. 87, pp. 337–351, 2007.

[5] L. De Lathauwer, "Decomposition of a higher-order tensor in block terms — Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, pp. 1033–1066, 2008.

[6] ——, "Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1451–1474, 2011.

[7] L. De Lathauwer and A. de Baynast, "Blind deconvolution of DS-CDMA signals by means of decomposition in rank-(1,L,L) terms," *IEEE Trans. Signal Processing*, vol. 56, no. 4, pp. 1562–1571, Mar. 2008.

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

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

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

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

[12] G. H. Golub and C. F. Van Loan, *Matrix computations*, 4th ed. John Hopkins university press, 2013.

[13] R. C. Gunning and H. Rossi, *Analytic functions in several complex variables*. Prentice-Hall, 1965.

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

[15] R. A. Horn and C. Johnson, *Matrix analysis*, 2nd ed. Cambridge University Press, 2013.

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

[17] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

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

[19] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communication," *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2714–2723, Nov. 1995.

[20] ——, "Smart antennas in wireless systems: Uplink multiuser blind channel and sequence detection," *IEEE Trans. Communications*, vol. 45, no. 2, pp. 187–199, Feb. 1997.

[21] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Processing*, vol. 50, pp. 83–99, 1996.

[22] M. Mørup, L. K. Hansen, and K. H. Madsen, "Modeling latency and shape change in trial based neuroimaging data," in *Proc. Asilomar*, Nov. 6-9, Pacific Grove, CA, USA, 2011.

[23] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[24] D. Nion and L. De Lathauwer, "A tensor-based blind DS-CDMA receiver using simultaneous matrix diagonalization," in *Proc. of the VIII IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2007)*, Helsinki, Finland.

[25] ——, "A block component model-based blind DS-CDMA receiver," *IEEE Trans. Signal Processing*, vol. 56, no. 11, pp. 5567–5579, Nov. 2008.

[26] ——, "A link between the decomposition of a third-order tensor in rank$-(L, L, 1)$ terms and joint block diagonalization," in *Proceedings of the 3rd IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2009.

[27] N. D. Sidiropoulos and G. Z. Dimić, "Blind multiuser detection in W-CDMA systems with large delay spread," *IEEE Signal Proc. Letters*, vol. 8, no. 9, pp. 87–89, Sep. 2001.

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

[29] L. Sorber, M. Van Barel, and L. De Lathauwer, "Structured data fusion," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 586–600, Jun. 2015.

[30] M. Sørensen and L. De Lathauwer, "Tensor decompositions with block-Toeplitz structure and applications in signal processing," in *Proc. Asilomar*, Nov. 6-9, Pacific Grove, CA, USA, 2011.

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

[32] ——, "Tensor decompositions with a block-Toeplitz factor matrix," ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep. 11-216, 2015.

[33] ——, "Multiple invariance ESPRIT for nonuniform linear arrays: A coupled canonical polyadic decomposition approach," *IEEE Trans. Signal Processing*, vol. 64, no. 4, pp. 3693–3704, Jul. 2016.

[34] ——, "Multidimensional harmonic retrieval via coupled canonical polyadic decomposition — Part I: Model and identifiability," *IEEE Trans. Signal Processing*, vol. 65, no. 2, pp. 517–527, Jan. 2017.

[35] ——, "Multidimensional harmonic retrieval via coupled canonical polyadic decomposition — Part II: Algorithm and multirate sampling," *IEEE Trans. Signal Processing*, vol. 65, no. 2, pp. 528–539, Jan. 2017.

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

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

[38] A. Stegeman and N. D. Sidiropoulos, "On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition," *Linear Algebra and its Applications*, vol. 420, pp. 540–552, 2007.

[39] L. Tong, "Identification of multichannel MA parameters using higher-order statistics," *Signal Processing*, vol. 53, pp. 195–209, 1996.

[40] L. Tong, Y. Inouye, and R. Liu, "A finite-step global convergence algorithm for the parameter estimation of multichannel MA processes," *IEEE Trans. Signal Process.*, vol. 10, pp. 2547–2558, 1992.

[41] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind identification and equalization based on second-order statistics: A frequency domain approach," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 329–333, 1995.

[42] A.-J. van der Veen, "Algebraic methods for deterministic blind beamforming," *Proceedings of the IEEE*, vol. 10, pp. 1987–2008, 1998.

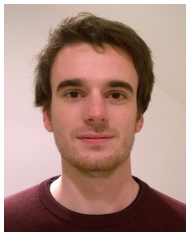
[43] A.-J. van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 173–190, Jan. 1997.

[44] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, *Tensorlab v3.0*, Available online, March 2016,. [Online]. Available: http://www.tensorlab.net/.

[45] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.



Mikael Sørensen received the Master's degree from Aalborg University, Denmark, and the Ph.D. degree from University of Nice, France, in 2006 and 2010, respectively, both in electrical engineering. From 2010 to 2017 he was a Postdoctoral Fellow with the KU Leuven, Belgium. Past research interests include applied linear algebra, tensor decompositions, and tensor-based array and signal processing. His current research activities are focused on tensor methods for machine learning and big data analytics.



Frederik Van Eeghem obtained his M.Sc. degree in Mathematical Engineering from KU Leuven, Belgium, in 2014. He is currently a Ph.D. candidate affiliated with the KU Leuven, Belgium. His research interests include blind system identification, independent component analysis and tensor decompositions.



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