Monomial factorizations via tensor decompositions

Mikael Sørensen[a], Lieven De Lathauwer[b], Nicholaos D. Sidiropoulos[a]

[a] *University of Virginia, Dept. of Electrical and Computer Engineering, Thornton Hall 351 McCormick Road, Charlottesville, VA 22904, USA, {ms8tz, nikos}@virginia.edu.*
[b] *Group Science, Engineering and Technology, KU Leuven - Kulak, E. Sabbelaan 53, 8500 Kortrijk, Belgium, and KU Leuven - STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, E.E. Dept. (ESAT), Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium. Lieven.DeLathauwer@kuleuven.be.*

Abstract

The Canonical Polyadic Decomposition (CPD), which decomposes a tensor into a sum of rank one terms, plays an important role in signal processing and machine learning. In this paper we extend the CPD framework to the more general case of monomial factorizations. This includes extensions of multilinear algebraic uniqueness conditions originally developed for the CPD. We obtain a deterministic condition for monomial factorizations that is both necessary and sufficient but which may be difficult to check in practice. We derive a deterministic relaxation that admits a constructive interpretation and is also easier to verify. Computationally, we reduce the monomial factorization problem into a CPD problem, which can be solved via a matrix EigenValue Decomposition (EVD). Under the given conditions, the discussed EVD-based algorithms are guaranteed to return the exact monomial factorization. Finally, we make a connection between monomial factorizations and the coupled block term decomposition, which allows us to translate monomial structures into low-rank structures.

Keywords: tensor, canonical polyadic decomposition, block term decomposition, coupled decomposition, monomial, uniqueness, eigenvalue decomposition.
2010 MSC: 15A15, 15A23

1. Introduction

Tensors have found many applications in signal processing and machine learning; see [1, 2] and references therein. The most well-known tensor decomposition is the Canonical Polyadic Decomposition (CPD), i.e., the decomposition into a minimal number of rank-one terms [3, 4]:

$$\mathbf{X} = (\mathbf{A} \odot \mathbf{B})\,\mathbf{S}^T \in \mathbb{C}^{IJ \times K}, \tag{1}$$

where '$\odot$' denotes the Khatri-Rao (columnwise Kronecker) product, $\mathbf{A} \in \mathbb{C}^{I \times R}$, $\mathbf{B} \in \mathbb{C}^{J \times R}$ and $\mathbf{S} \in \mathbb{C}^{K \times R}$. Note that the columns of $\mathbf{A} \odot \mathbf{B}$ correspond to vectorized rank-one matrices, explaining why (1) is referred to as a decomposition into rank-one terms. (A formal definition of the CPD will be provided in Section 1.2.) In signal processing, the CPD is related to the ESPRIT [5, 6] and ACMA [7] methods while in machine learning, it is related to the naive Bayes model [8, 9, 10, 11]. In [12, 13] we extended the CPD framework to coupled CPD and we have shown the usefulness of the latter decomposition

in sensor array processing [14], wireless communication [15] and in multidimensional harmonic retrieval [16, 17]. In this paper we will further extend the CPD framework to more general monomial structures. (A monomial is a product of variables, possibly with repetitions.) More precisely, we consider bilinear factorizations of the form

$$\mathbf{X} = \mathbf{A}\mathbf{S}^T \in \mathbb{C}^{I \times K}, \tag{2}$$

in which the columns of $\mathbf{A} \in \mathbb{C}^{I \times R}$ (or similarly the columns of $\mathbf{S} \in \mathbb{C}^{K \times R}$) satisfy monomial relations of the form

$$a_{p_1,r} \cdots a_{p_L,r} - a_{s_1,r} \cdots a_{s_L,r} = 0, \tag{3}$$

where $a_{m,r}$ denotes the m-th entry of the r-th column of $\mathbf{A}$. A bilinear factorization (2) exhibiting monomial structure of the form (3) will be referred to as a monomial factorization. (In Sections 3 and 4 it will become clear that (1) is a special case of (2).) To make things more tangible, let us consider a concrete example. In signal processing, the separation of digital communication signals is probably one of the earliest examples involving monomial structures. For instance, blind separation of M-PSK signals in which the entries of $\mathbf{S}$ in (2) take the form

$$s_{kr} = e^{\sqrt{-1}u_{kr}} \quad \text{with} \quad u_{kr} \in \big\{0, 2\pi/M, \ldots, 2\pi(M-1)/M\big\} \tag{4}$$

has been considered (e.g., [18, 19]). From (4) it is clear that $s_{k_1r}^M = s_{k_2r}^M$ for all $k_1, k_2 \in \{1, \ldots, K\}$. In other words, for every pair (k_1, k_2), with $k_1 < k_2$, we can exploit $C_K^2 = \frac{(K-1)K}{2}$ monomial relations of the form $s_{k_1r}^M - s_{k_2r}^M = 0$. In this paper we will explain how to translate this type of problems into a tensor decomposition problem. Another example, which will be discussed in Section 4.4, is the Binary Matrix Factorization (BMF):

$$\mathbf{X} = \mathbf{A}\mathbf{S}^T \in \mathbb{C}^{I \times K}, \tag{5}$$

where $\mathbf{A} \in \{0,1\}^{I \times R}$ is a binary matrix. BMFs of the form (5) play a role in binary latent variable modeling (e.g., [20, 21, 22]).

Monomial factorizations have the interesting property that they provide a framework that allows us to generalize the CPD model. As an example, the presented monomial factorization framework enables us to extend the CPD model (1) to the case of binary weighted rank-one terms (this will be made clear in Section 3.3):

$$\mathbf{X} = (\mathbf{D} * (\mathbf{A} \odot \mathbf{B}))\mathbf{S}^T \in \mathbb{C}^{IJ \times K}, \tag{6}$$

where '$*$' denotes the Hadamard (element-wise) product and $\mathbf{D} \in \{0,1\}^{IJ \times R}$. Binary weighted rank-one terms are of interest in clustering applications involving tensor structures (e.g., [23, 24]).

The paper is organised as follows. In the rest of the introduction we will first present the notation used throughout the paper, followed by a brief review of the CPD and the Block Term Decomposition (BTD) [25]. As our first contribution, we will in Section 2 present a new link between monomial factorizations of the form (2) and the coupled BTD [12, 13]. This connection enables us to translate the monomial constraint (3) into a low-rank constraint, which in turn allows us to treat the monomial factorization of a matrix as a tensor decomposition problem. Next, in Section 3 we will present

identifiability conditions. It will be explained that the presented identifiability conditions are extensions of well-known CPD uniqueness conditions developed in [26, 27, 28, 29] to the monomial case. We also explain that the monomial factorization framework can be used to generalize the CPD model (1) to the binary weighted CPD model (6). As our third contribution, we will in Section 4 extend the algebraic algorithm for CPD in [27, 30] to monomial factorizations. We also demonstrate how the presented algebraic algorithm can be adapted and used for the computation of a BMF of the form (5).

1.1. Notation

Vectors, matrices and tensors are denoted by lower case boldface, upper case boldface and upper case calligraphic letters, respectively. The r-th column, conjugate, transpose, conjugate-transpose, determinant, permanent, inverse, right-inverse, range and kernel of a matrix $\mathbf{A}$ are denoted by $\mathbf{a}_r$, $\mathbf{A}^*$, $\mathbf{A}^T$, $\mathbf{A}^H$, $|\mathbf{A}|$, $\overset{+}{|}\mathbf{A}\overset{+}{|}$, $\mathbf{A}^{-1}$, $\mathbf{A}^\dagger$, $\text{range}\,(\mathbf{A})$, $\ker(\mathbf{A})$, respectively. The dimension of a subspace S is denoted by $\dim(S)$. The symbols $\otimes$ and $\odot$ denote the Kronecker and Khatri-Rao product, defined as

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

in which $(\mathbf{A})_{mn} = a_{mn}$. The outer product of, say, three vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is denoted by $\mathbf{a}\circ\mathbf{b}\circ\mathbf{c}$, such that $(\mathbf{a}\circ\mathbf{b}\circ\mathbf{c})_{ijk} = a_i b_j c_k$. The number of non-zero entries of a vector $\mathbf{x}$ is denoted by $\omega(\mathbf{x})$ in the tensor decomposition literature, dating back to the work of Kruskal [31]. Let $\text{diag}(\mathbf{a}) \in \mathbb{C}^{J\times J}$ denote a diagonal matrix that holds a column vector $\mathbf{a}\in\mathbb{C}^{J\times 1}$ or a row vector $\mathbf{a}\in\mathbb{C}^{1\times J}$ on its diagonal. In some cases a diagonal matrix is holding row k of $\mathbf{A}\in\mathbb{C}^{I\times J}$ on its diagonal. This will be denoted by $D_k(\mathbf{A}) \in \mathbb{C}^{J\times J}$. Furthermore, let $\text{vec}(\mathbf{A})$ denote the vector obtained by stacking the columns of $\mathbf{A}\in\mathbb{C}^{I\times J}$ into a column vector $\text{vec}(\mathbf{A}) = [\mathbf{a}_1^T,\dots,\mathbf{a}_J^T]^T \in \mathbb{C}^{IJ}$. Let $\mathbf{e}_n^{(N)} \in \mathbb{C}^N$ denote the unit vector with unit entry at position n and zeros elsewhere. The all-ones vector is denoted by $\mathbf{1}_R = [1,\dots,1]^T \in \mathbb{C}^R$. 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 $\mathbf{A}$. 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 $\mathcal{C}_k(\mathbf{A}) \in \mathbb{C}^{C_I^k\times C_R^k}$. It is the matrix containing the determinants of all $k\times k$ submatrices of $\mathbf{A}$, arranged with the submatrix index sets in lexicographic order. See [28, 30, 32, 33] and references therein for a discussion.

1.2. Canonical Polyadic Decomposition (CPD)

Consider the 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 $\mathbf{s}\in\mathbb{C}^K$ such that $x_{ijk} = a_i b_j s_k$. A Polyadic Decomposition (PD) is a decomposition of $\mathcal{X}$ into a sum of rank-1 terms [3, 4]:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}^{(r)}\circ\mathbf{s}_r = \sum_{r=1}^{R} \mathbf{a}_r\circ\mathbf{b}_r\circ\mathbf{s}_r, \tag{7}$$

where $\mathbf{E}^{(r)} = \mathbf{a}_r\mathbf{b}_r^T = \mathbf{a}_r\circ\mathbf{b}_r \in \mathbb{C}^{I\times J}$ is a rank-one matrix. 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 (7) is called the Canonical PD (CPD) of $\mathcal{X}$.

1.2.1. Matrix representation

Consider the horizontal matrix slice $\mathbf{X}^{(i\cdot\cdot)} \in \mathbb{C}^{J\times K}$ of $\mathcal{X}$, defined by $(\mathbf{X}^{(i\cdot\cdot)})_{jk} = x_{ijk} = \sum_{r=1}^{R} a_{ir}b_{jr}s_{kr}$. The tensor $\mathcal{X}$ can be interpreted as a collection of matrix slices $\mathbf{X}^{(1\cdot\cdot)}, \ldots, \mathbf{X}^{(I\cdot\cdot)}$, each admitting the factorization $\mathbf{X}^{(i\cdot\cdot)} = \sum_{r=1}^{R} a_{ir}\mathbf{b}_r\mathbf{s}_r^T = \mathbf{B}D_i\left(\mathbf{A}\right)\mathbf{S}^T$. Stacking yields (1):

$$\mathbf{X} = \left[\mathbf{X}^{(1\cdot\cdot)T}, \ldots, \mathbf{X}^{(I\cdot\cdot)T}\right]^T = \left(\mathbf{A} \odot \mathbf{B}\right)\mathbf{S}^T. \tag{8}$$

1.2.2. Uniqueness conditions for CPD

The rank-1 tensors in (7) 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.

For cases where $\mathbf{S}$ in (8) has full column rank, the following necessary and sufficient uniqueness condition stated in Theorem 1.1 was obtained in [26] and later reformulated in terms of compound matrices in [28]. It makes use of the matrix

$$\mathbf{G}_{\text{CPD}}^{(2)} = \mathcal{C}_2\left(\mathbf{A}\right) \odot \mathcal{C}_2\left(\mathbf{B}\right) \in \mathbb{C}^{C_I^2 C_J^2 \times C_R^2} \tag{9}$$

and the vector

$$\mathbf{f}^{(2)}(\mathbf{d}) = [d_1d_2, d_1d_3, \ldots, d_{R-1}d_R]^T \in \mathbb{C}^{C_R^2}, \tag{10}$$

which consists of all distinct products of entries $d_r \cdot d_s$ with $r < s$ from the vector $\mathbf{d} = [d_1, \ldots, d_R]^T \in \mathbb{C}^R$.

Theorem 1.1. *Consider the PD of $\mathcal{X} \in \mathbb{C}^{I\times J\times K}$ in (7). Assume that $\boldsymbol{S}$ has full column rank. The rank of $\mathcal{X}$ is R and the CPD of $\mathcal{X}$ is unique if and only if the following implication holds*

$$\boldsymbol{G}_{CPD}^{(2)} \cdot \boldsymbol{f}^{(2)}(\boldsymbol{d}) = \boldsymbol{0} \Rightarrow \omega(\boldsymbol{d}) \leq 1, \tag{11}$$

for all structured vectors $\boldsymbol{f}^{(2)}(\boldsymbol{d})$ of the form (10).

In practice, condition (11) can be hard to check. However, as observed in [26, 27, 28], if $\mathbf{G}_{\text{CPD}}^{(2)}$ in (11) has full column rank, then $\mathbf{f}^{(2)}(\mathbf{d}) = \mathbf{0}$ and the condition is automatically satisfied. This fact leads to the following more easy to check uniqueness condition, which is only sufficient.

Theorem 1.2. *Consider the PD of $\mathcal{X} \in \mathbb{C}^{I\times J\times K}$ in (7). If*

$$\begin{cases} \boldsymbol{S} \text{ has full column rank,} \\ \boldsymbol{G}_{CPD}^{(2)} \text{ has full column rank,} \end{cases} \tag{12}$$

then the rank of $\mathcal{X}$ is R and the CPD of $\mathcal{X}$ is unique.

Furthermore, if condition (12) is satisfied, then the CPD of $\mathcal{X}$ can be computed via a matrix EVD [27, 30]. In short, the "CPD" of $\mathcal{X}$ can be converted into a "basic CPD" of an $(R \times R \times R)$ tensor $\mathcal{Q}$ of rank R, even in cases where $\max(I, J) < R$ [27, 30]. The latter CPD can be computed by means of a standard EVD (e.g., [3, 34]). In Section 4 we briefly discuss how to construct the tensor $\mathcal{Q}$ from $\mathcal{X}$ and how to retrieve the CPD factor matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{S}$ of $\mathcal{X}$ from the CPD of $\mathcal{Q}$.

More details about the CPD can be found in [3, 31, 27, 28, 29, 26, 30, 35, 2] and references therein.

1.3. Block Term Decomposition (BTD) and coupled BTD

The multilinear rank-$(P, P, 1)$ term decomposition of a tensor is an extension of the CPD (7), where each term in the decomposition now consists of the outer product of a vector and a matrix that is low-rank [25]. More formally, $\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{s}_r$ in (7) is replaced by $\mathbf{E}_r \circ \mathbf{s}_r$:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}_r \circ \mathbf{s}_r, \tag{13}$$

where $\mathbf{E}_r \in \mathbb{C}^{I \times J}$ is a rank-P matrix with $\min(I, J) > P$. Note that if $P = 1$, then (13) indeed reduces to (7) with $\mathbf{E}_r = \mathbf{a}_r \mathbf{b}_r^T = \mathbf{a}_r \circ \mathbf{b}_r$. We will consider the extension of (13) in which a set of tensors $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n \times J_n \times K}$, $n \in \{1, \ldots, N\}$ is decomposed into a sum of coupled multilinear rank-$(P, P, 1)$ terms [12], or coupled BTD for short:

$$\mathcal{X}^{(n)} = \sum_{r=1}^{R} \mathbf{E}_r^{(n)} \circ \mathbf{s}_r \ , \quad n \in \{1, \ldots, N\}, \tag{14}$$

where $\mathbf{E}_r^{(n)} \in \mathbb{C}^{I_n \times J_n}$ is a rank-P matrix with $\min(I_n, J_n) > P$ and $\mathbf{s}_r \in \mathbb{C}^K$. Note that $\{\mathbf{s}_r\}$ are shared between all $\mathcal{X}^{(n)}$, i.e., the third mode ensures the coupling. As in the CPD case, the rank of the coupled BTD is defined as the minimal number of coupled multilinear rank-$(P, P, 1)$ terms $\{\mathbf{E}_r^{(n)} \circ \mathbf{s}_r\}$ that yield $\mathcal{X}^{(1)}, \ldots, \mathcal{X}^{(N)}$. Since $\mathbf{E}_r^{(n)}$ is low-rank, we know that (14) can also be expressed in terms of a coupled PD:

$$\mathcal{X}^{(n)} = \sum_{r=1}^{R} \mathbf{E}_r^{(n)} \circ \mathbf{s}_r = \sum_{r=1}^{R} \mathbf{N}^{(n,r)} \mathbf{M}^{(n,r)T} \circ \mathbf{s}_r = \sum_{r=1}^{R} \sum_{p=1}^{P} \mathbf{n}_p^{(n,r)} \circ \mathbf{m}_p^{(n,r)} \circ \mathbf{s}_r, \tag{15}$$

where $\mathbf{E}_r^{(n)} = \mathbf{N}^{(n,r)} \mathbf{M}^{(n,r)T}$, in which $\mathbf{N}^{(n,r)} = [\mathbf{n}_1^{(n,r)}, \ldots, \mathbf{n}_P^{(n,r)}] \in \mathbb{C}^{I_n \times P}$ and $\mathbf{M}^{(n,r)} = [\mathbf{m}_1^{(n,r)}, \ldots, \mathbf{m}_P^{(n,r)}] \in \mathbb{C}^{J_n \times P}$ are rank-P matrices.

1.3.1. Matrix representation

Define

$$\mathbf{M}^{(n)} = \left[\mathbf{M}^{(n,1)}, \ldots, \mathbf{M}^{(n,R)}\right] \in \mathbb{C}^{I_n \times PR}, \tag{16}$$

$$\mathbf{N}^{(n)} = \left[\mathbf{N}^{(n,1)}, \ldots, \mathbf{N}^{(n,R)}\right] \in \mathbb{C}^{J_n \times PR}, \tag{17}$$

$$\mathbf{S}^{(\text{ext})} = \left[\mathbf{1}_P^T \otimes \mathbf{s}_1, \ldots, \mathbf{1}_P^T \otimes \mathbf{s}_R\right] \in \mathbb{C}^{K \times PR}. \tag{18}$$

Then the factorization (15) can also be expressed in terms of $\{\mathbf{M}^{(n)}, \mathbf{N}^{(n)}, \mathbf{S}^{(\text{ext})}\}$ as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(N)} \end{bmatrix} = \begin{bmatrix} \mathbf{M}^{(1)} \odot \mathbf{N}^{(1)} \\ \vdots \\ \mathbf{M}^{(N)} \odot \mathbf{N}^{(N)} \end{bmatrix} \mathbf{S}^{(\text{ext})T}. \tag{19}$$

1.3.2. Uniqueness condition for coupled BTD

The coupled BTD version of $\mathbf{G}_{\text{CPD}}^{(2)}$ in (9) is given by

$$\mathbf{G}_{\text{BTD}}^{(N,P+1)} = \begin{bmatrix} \mathcal{C}_{P+1}\left(\mathbf{M}^{(1)}\right) \odot \mathcal{C}_{P+1}\left(\mathbf{N}^{(1)}\right) \\ \vdots \\ \mathcal{C}_{P+1}\left(\mathbf{M}^{(N)}\right) \odot \mathcal{C}_{P+1}\left(\mathbf{N}^{(N)}\right) \end{bmatrix} \mathbf{P}_{\text{BTD}} \in \mathbb{C}^{N \times (C_{R+P}^{P+1} - R)}, \tag{20}$$

where $\mathbf{M}^{(n)} \in \mathbb{C}^{I_n \times PR}$ and $\mathbf{N}^{(n)} \in \mathbb{C}^{J_n \times PR}$ are given by (16) and (17), respectively, and $\mathbf{P}_{\text{BTD}} \in \mathbb{C}^{C_{PR}^{P+1} \times (C_{R+P}^{P+1} - R)}$ is the "compression" matrix that takes into account that each column vector $\mathbf{s}_r$ in (18) is repeated P times and that $|\mathbf{M}^{(n,r)}| = |\mathbf{N}^{(n,r)}| = 0$. The latter property implies that R columns of $\mathcal{C}_{P+1}(\mathbf{M}^{(n)}) \odot \mathcal{C}_{P+1}(\mathbf{N}^{(n)})$ are zero columns, which are eliminated by $\mathbf{P}_{\text{BTD}}$. Here we only state how $\mathbf{P}_{\text{BTD}}$ is constructed. The reasoning behind the construction $\mathbf{P}_{\text{BTD}}$ can be found in [12]. (More details can also be found in the proof of Theorem 2.1 in Appendix B.) The $C_{R+P}^{P+1} - R$ columns of $\mathbf{P}_{\text{BTD}}$ are indexed by the lexicographically ordered tuples in the set

$$\Gamma_c = \{(r_1, \ldots, r_{P+1}) \mid 1 \leq r_1 \leq \cdots \leq r_{P+1} \leq R\} \setminus \{(r, \ldots, r)\}_{r=1}^{R}.$$

Consider also the mapping $f_c : \{(r_1, \ldots, r_{P+1})\}^{P+1} \to \{1, 2, \ldots, C_{R+P}^{P+1} - R\}$ that returns the position of its argument in the set Γ_c. Similarly, the C_{PR}^{P+1} rows of $\mathbf{P}_{\text{BTD}}$ are indexed by the lexicographically ordered tuples in the set

$$\Gamma_r = \{(q_1, \ldots, q_{P+1}) \mid 1 \leq q_1 < \cdots < q_{P+1} \leq PR\}.$$

Likewise, we define the mapping $f_r : \{(q_1, \ldots, q_{P+1})\}^{P+1} \to \{1, 2, \ldots, C_{PR}^{P+1}\}$ that returns the position of its argument in the set Γ_r. The entries of $\mathbf{P}_{\text{BTD}}$ are now given by

$$(\mathbf{P}_{\text{BTD}})_{f_r(q_1,\ldots,q_{P+1}), f_c(r_1,\ldots,r_{P+1})} = \begin{cases} 1, & \text{if } \lceil \frac{q_1}{P} \rceil = r_1, \ldots, \lceil \frac{q_{P+1}}{P} \rceil = r_{P+1}, \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

Theorem 1.3. *Consider the coupled BTD of $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n \times J_n \times K}$, $n \in \{1, \ldots, N\}$ in (30). If*

$$\begin{cases} \mathbf{S} \text{ has full column rank,} \\ \boldsymbol{G}_{BTD}^{(N,P+1)} \text{ has full column rank,} \end{cases} \tag{22}$$

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

As in Theorem 1.2, if condition (22) in Theorem 1.3 is satisfied, then the coupled BTD of $\{\mathcal{X}^{(n)}\}$ can be computed via a matrix EVD [13]. More details will be provided in Section 4.1.1.

The *objective of this paper* is to extend the CPD results discussed in this section to the case of monomial factorizations. More precisely, in Section 2 we explain that the monomial factorization can be interpreted as a coupled BTD. Next, in Section 3 we extend the uniqueness conditions stated in Theorems 1.1 and 1.2 to the case of bilinear models with factor matrices satisfying monomial relations. Finally, in Section 4 we extend the algebraic algorithm associated with Theorems 1.2 and 1.3 to the case of monomial factorizations.

2. Link between monomial factorization and coupled BTD

In Section 2.1 we explain how to represent the monomial structure (3) as a low-rank constraint on a particular matrix. Using this low-rank matrix, we will in Section 2.2 translate the monomial factorization (2) into the coupled BTD of the form (14) reviewed in Section 1.3.

2.1. Representation of monomial structure via low-rank matrix

We say that $\mathbf{X}$ admits an R-term monomial factorization enjoying N monomial relations of the form (3) if every column of $\mathbf{A}$ in (2) satisfies the monomial relations

$$\prod_{l=1}^{L} a_{lr}^{(+,n)} - \prod_{l=1}^{L} a_{lr}^{(-,n)} = 0, \quad r \in \{1,\dots,R\},\ n \in \{1,\dots,N\}, \tag{23}$$

where $\mathbf{a}_r^{(+,n)} \in \mathbb{C}^L$ and $\mathbf{a}_r^{(-,n)} \in \mathbb{C}^L$ are given by

$$\begin{cases} \mathbf{a}_r^{(+,n)} = [a_{1r}^{(+,n)},\dots,a_{Lr}^{(+,n)}]^T = \left[a_{p_{1,n},r},\dots,a_{p_{L,n},r}\right]^T, \\ \mathbf{a}_r^{(-,n)} = [a_{1r}^{(-,n)},\dots,a_{Lr}^{(-,n)}]^T = \left[a_{s_{1,n},r},\dots,a_{s_{L,n},r}\right]^T, \end{cases} \tag{24}$$

in which $a_{lr}^{(+,n)} = a_{p_{l,n},r}$ corresponds to the $p_{l,n}$-th entry of the r-th column of $\mathbf{A}$ (similarly for $a_{lr}^{(-,n)}$). Define the linearly structured matrix $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) \in \mathbb{C}^{L\times L}$:[1]

$$\mathbf{A}_L\left(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}\right) := \begin{bmatrix} a_{1r}^{(+,n)} & 0 & \cdots & 0 & (-1)^L \cdot a_{1r}^{(-,n)} \\ a_{2r}^{(-,n)} & a_{2r}^{(+,n)} & \ddots & & 0 \\ 0 & a_{3r}^{(-,n)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{Lr}^{(-,n)} & a_{Lr}^{(+,n)} \end{bmatrix} \tag{25}$$
$$= \operatorname{diag}(\mathbf{a}_r^{(+,n)}) + (\operatorname{diag}(\mathbf{a}_r^{(-,n)})\mathbf{P}) * (\mathbf{1}_L\mathbf{1}_L^T - \mathbf{e}_1^{(L)}\mathbf{e}_L^{(L)T} + (-1)^L\mathbf{e}_1^{(L)}\mathbf{e}_L^{(L)T}),$$

where $\mathbf{P}$ is the "cyclic" permutation matrix given by

$$\mathbf{P} = \sum_{l=1}^{L-1} \mathbf{e}_{l-1}^{(L)}\mathbf{e}_l^{(L)T} + \mathbf{e}_1^{(L)}\mathbf{e}_L^{(L)T} \in \mathbb{C}^{L\times L}. \tag{26}$$

[1] The matrix $\mathbf{1}_L\mathbf{1}_L^T - \mathbf{e}_1^{(L)}\mathbf{e}_L^{(L)T} + (-1)^L\mathbf{e}_1^{(L)}\mathbf{e}_L^{(L)T}$ takes the sign $(-1)^L$ in (25) into account.

From the cofactor expansion of $|\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})|$ along the first row, the connection between (23) and (25) becomes clear:

$$|\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})| = a_{1r}^{(+,n)} \cdot \begin{vmatrix} a_{2r}^{(+,n)} & 0 & \cdots & 0 \\ a_{3r}^{(-,n)} & \ddots & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ & & a_{Lr}^{(-,n)} & a_{Lr}^{(+,n)} \end{vmatrix} \tag{27}$$

$$+ (-1)^L a_{1r}^{(-,n)} (-1)^{L+1} \begin{vmatrix} a_{2r}^{(-,n)} & a_{2r}^{(+,n)} & & \\ 0 & a_{3r}^{(-,n)} & \ddots & \\ \vdots & \ddots & \ddots & a_{L-1r}^{(+,n)} \\ 0 & \cdots & 0 & a_{Lr}^{(-,n)} \end{vmatrix} = \prod_{l=1}^{L} a_{lr}^{(+,n)} - \prod_{l=1}^{L} a_{lr}^{(-,n)} = 0\,,$$

where we exploited that the two involved $(L-1) \times (L-1)$ minors in (27) are triangular. The determinant property (27) also explains that $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ is low-rank under the condition (23). In fact, since the minors in (27) do not vanish under condition (23), $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ will be a rank-$(L-1)$ matrix. The only possible exception is the trivial case where $\prod_{m=1}^{L} a_{p_m} = 0$ and $\prod_{n=1}^{L} a_{s_n} = 0$. In this section it will become clear that if $\prod_{m=1}^{L} a_{p_m} \neq 0$ or $\prod_{n=1}^{L} a_{s_n} \neq 0$, then the linear structure on $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ can be relaxed without affecting the identifiability. To the best of our knowledge, the representation of a monomial relation of the form (23) via the rank deficiency of the matrix in (25) is a novel contribution of this paper.

2.2. Monomial factorization via coupled BTD

Consider the bilinear factorization (2) in which the columns of $\mathbf{A}$ satisfy N monomial relations of the form (23). The mapping (25) together with the bilinear property of the monomial factorization enables us to transform (2) into a coupled BTD. In detail, for every monomial relation ($n \in \{1, \ldots, N\}$), we build a tensor $\mathcal{X}^{(n)} \in \mathbb{C}^{L \times L \times K}$ with matrix slices

$$\mathcal{X}^{(n)}(:,:,k) = \mathbf{A}_L(\mathbf{x}_k^{(+,n)}, \mathbf{x}_k^{(-,n)}) = \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) s_{kr},\ k \in \{1, \ldots, K\}, \tag{28}$$

in which $\mathbf{x}_k^{(+,n)} \in \mathbb{C}^L$ and $\mathbf{x}_k^{(-,n)} \in \mathbb{C}^L$ are constructed from the entries of the k-th column of $\mathbf{X}$ in accordance to the n-th monomial relation, so that (cf. Eq. (24)):

$$\begin{cases} \mathbf{x}_k^{(+,n)} = [x_{1k}^{(+,n)}, \ldots, x_{Lk}^{(+,n)}]^T = \left[x_{p_{1,n},k}, \ldots, x_{p_{L,n},k}\right]^T, \\ \mathbf{x}_k^{(-,n)} = [x_{1k}^{(-,n)}, \ldots, x_{Lk}^{(-,n)}]^T = \left[x_{s_{1,n},k}, \ldots, x_{s_{L,n},k}\right]^T. \end{cases} \tag{29}$$

Overall, this yields the coupled BTD:

$$\mathbb{C}^{L \times L \times K} \ni \mathcal{X}^{(n)} = \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) \circ \mathbf{s}_r, \quad n \in \{1, \ldots, N\}. \tag{30}$$

The key observation is that each equation in (28) corresponds to a BTD, and the collection of all equations in (28) is a coupled BTD.

Matrix representation. Let $\mathbf{X}^{(\cdot\cdot k,n)} \in \mathbb{C}^{L\times L}$ denote the k-th matrix slice of $\mathcal{X}^{(n)}$, defined according to $(\mathbf{X}^{(\cdot\cdot k,n)})_{ij} = (\mathcal{X}^{(n)})_{ijk}$. We now obtain the following matrix representation of (30):

$$\mathbf{X}^{(n)} = \left[\mathrm{vec}(\mathbf{X}^{(\cdot\cdot 1,n)}), \ldots, \mathrm{vec}(\mathbf{X}^{(\cdot\cdot K,n)})\right] = \mathbf{A}^{(n)}\mathbf{S}^T \in \mathbb{C}^{L^2\times K}, \quad n \in \{1,\ldots,N\},$$

where $\mathbf{A}^{(n)} = \left[\mathrm{vec}(\mathbf{A}_L(\mathbf{a}_1^{(+,n)}, \mathbf{a}_1^{(-,n)})), \ldots, \mathrm{vec}(\mathbf{A}_L(\mathbf{a}_R^{(+,n)}, \mathbf{a}_R^{(-,n)}))\right] \in \mathbb{C}^{L^2\times R}$ and $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_R] \in \mathbb{C}^{K\times R}$. Stacking yields

$$\mathbf{X} = \left[\mathbf{X}^{(1)T}, \ldots, \mathbf{X}^{(N)T}\right]^T = \mathbf{A}_{\mathrm{tot}}\mathbf{S}^T \in \mathbb{C}^{N\cdot L^2\times K}, \tag{31}$$

where $\mathbf{A}_{\mathrm{tot}} \in \mathbb{C}^{N\cdot L^2\times R}$ is given by

$$\mathbf{A}_{\mathrm{tot}} = \begin{bmatrix} \mathbf{A}^{(1)} \\ \vdots \\ \mathbf{A}^{(N)} \end{bmatrix} = \begin{bmatrix} \mathrm{vec}(\mathbf{A}_L(\mathbf{a}_1^{(+,1)}, \mathbf{a}_1^{(-,1)})) & \ldots & \mathrm{vec}(\mathbf{A}_L(\mathbf{a}_R^{(+,1)}, \mathbf{a}_R^{(-,1)})) \\ \vdots & \ddots & \vdots \\ \mathrm{vec}(\mathbf{A}_L(\mathbf{a}_1^{(+,N)}, \mathbf{a}_1^{(-,N)})) & \ldots & \mathrm{vec}(\mathbf{A}_L(\mathbf{a}_R^{(+,N)}, \mathbf{a}_R^{(-,N)})) \end{bmatrix}. \tag{32}$$

Uniqueness condition. Since $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ defined by (25) is low-rank, (30) corresponds to a coupled BTD. In more detail, let the rank of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ be equal to $L_{r,n} < L$, then it admits the low-rank factorization

$$\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) = \mathbf{N}^{(n,r)}\mathbf{M}^{(n,r)T} = \left[\widetilde{\mathbf{N}}^{(n,r)}, \mathbf{0}_{L,L-1-L_{r,n}}\right]\left[\widetilde{\mathbf{M}}^{(n,r)}, \mathbf{0}_{L,L-1-L_{r,n}}\right]^T, \tag{33}$$

where $\widetilde{\mathbf{N}}^{(n,r)} \in \mathbb{C}^{L\times L_{r,n}}$ and $\widetilde{\mathbf{M}}^{(n,r)} \in \mathbb{C}^{L\times L_{r,n}}$ are rank-$L_{r,n}$ matrices and $\mathbf{0}_{m,n}$ denotes an $(m\times n)$ zero matrix. Note that if $\omega(\mathbf{a}_r^{(+,n)}) = L$ or $\omega(\mathbf{a}_r^{(-,n)}) = L$, then $L_{r,n} = L-1$, as explained in Section 2.1. However, if $\omega(\mathbf{a}_r^{(+,n)}) < L$ and $\omega(\mathbf{a}_r^{(-,n)}) < L$, then $L_{r,n} < L-1$. Consider also $\mathbf{G}_{\mathrm{BTD}}^{(N,L)}$ given by (20) in which $\mathbf{M}^{(n)}$ and $\mathbf{N}^{(n)}$ are built from the $\mathbf{M}^{(n,r)}$ and $\mathbf{N}^{(n,r)}$ matrices in (33). We can now conclude that if for all $r \in \{1,\ldots,R\}$ there exists an $n \in \{1,\ldots,N\}$ such that $L_{r,n} = L-1$ and $\mathbf{G}_{\mathrm{BTD}}^{(N,L)}$ has full column rank, then the monomial factorization problem can be solved via the coupled BTD (30). Theorem 2.1 below summarizes the result.

Theorem 2.1. *Consider the coupled BTD of $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n\times J_n\times K}$, $n \in \{1,\ldots,N\}$ in (30). If*

$$\begin{cases} \mathbf{S} \text{ has full column rank,} \\ \boldsymbol{G}_{BTD}^{(N,L)} \text{ has full column rank,} \end{cases} \tag{34}$$

then the coupled BTD rank of $\{\mathcal{X}^{(n)}\}$ is R, the coupled BTD of $\{\mathcal{X}^{(n)}\}$ is unique, $\boldsymbol{A}_{tot}$ has full column rank, the monomial factorization of $\boldsymbol{X}$ in (2) is unique, and $\boldsymbol{A}$ in (2) has full column rank.

Proof. The result is an immmediate consequence of Theorem 1.3. The interested reader can find a proof in Appendix B in which the connection between Theorem 2.1 and the subsequent Theorem 3.2 is made apparent.

Note that in Theorem 2.1 we state that if condition (34) is satisfied, then $\mathbf{A}$ in (2) has full column rank. This is an obvious consequence of the uniqueness property of the full column rank factor matrix $\mathbf{S}$. Note also that we have dropped the linear structure on $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ and instead used the low-rank factorization $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) = \mathbf{N}^{(n,r)}\mathbf{M}^{(n,r)T}$ in the coupled BTD of $\{\mathcal{X}^{(n)}\}$.

As a final remark, we mention that the transformation step from a monomial factorization (2) to an "unconstrained" coupled BTD (15) can be understood as a generalization of the Hankelisation step used in the ESPRIT method [5] for the special case where the columns of $\mathbf{A}$ have exponential structure.

3. Identifiability conditions for monomial factorizations

We will now take the linear structure on $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ into account. In Section 3.1 we explain that this will lead to generalizations of the uniqueness conditions stated in Theorems 1.1 and 1.2 for CPD to monomial factorizations. Next, in Section 3.2 we formulate the presented uniqueness conditions in terms of null spaces, which will lead to a condition that is easier to comprehend.

3.1. Necessary and sufficient uniqueness conditions

3.1.1. Mixed discriminants

In Theorem 3.1 we present a uniqueness condition for the monomial factorization of $\mathbf{X}$. It will be based on the properties of the mixed discriminant discussed in this section. However, let us first briefly explain how mixed discriminants appear in our problem. The overall idea is to find a condition that ensures that $\mathbf{S}^T$ has a unique right-inverse (up to intrinsic column scaling and permutation ambiguities), denoted by $\mathbf{W}$. If $\mathbf{W}$ is unique, then $\mathbf{X}\mathbf{w}_r = \mathbf{a}_r$ is also unique and $\omega(\mathbf{S}^T\mathbf{w}_r) = 1$ for all $r \in \{1, \ldots, R\}$. This means that if $\mathbf{d}_r = \mathbf{S}^T\mathbf{w}_r$, then $\mathbf{X}^{(n)}\mathbf{w}_r = \sum_{s=1}^{R} \operatorname{vec}(\mathbf{A}^{(n,s)})d_{sr}$ is a vectorized rank-$(L-1)$ matrix. The latter property can be used to derive a condition that ensures the uniqueness of $\mathbf{W}$. Let $\mathbf{H}^{(r)} \in \mathbb{C}^{L\times L}$ be of the form (25), corresponding to $\mathbf{A}^{(n,r)}$ in which the superscript 'n' is omitted. Then in the proof of Theorem 3.1 it will become clear that the derivation of a uniqueness condition for $\mathbf{W}$, and therefore also for the monomial factorization, boils down to finding a compact expression of the determinant of $\mathbf{H}^{(1)}d_1 + \ldots + \mathbf{H}^{(R)}d_R$, where $d_r \in \mathbb{C}$:

$$\begin{aligned}\left|\sum_{r=1}^{R}\mathbf{H}^{(r)}d_r\right| &= \sum_{r_1,\ldots,r_L=1}^{R}\mathcal{D}(\mathbf{H}^{(r_1)},\ldots,\mathbf{H}^{(r_L)})\cdot d_{r_1}\cdots d_{r_L}\\ &= \sum_{(l_1,\ldots,l_R)\in[L]_R}\mathcal{D}\Big(\underbrace{\mathbf{H}^{(1)},\ldots,\mathbf{H}^{(1)}}_{l_1 \text{ times}},\ldots,\underbrace{\mathbf{H}^{(R)},\ldots,\mathbf{H}^{(R)}}_{l_R \text{ times}}\Big)d_1^{l_1}\cdots d_R^{l_R}, \qquad (35)\end{aligned}$$

where $[L]_R$ denotes the set of all weak compositions of L in R terms, i.e.,

$$[L]_R = \{(l_1,\ldots,l_R) \mid l_1 + \cdots + l_R = L \text{ and } l_1,\ldots,l_R \geq 0\}. \qquad (36)$$

Note that the cardinality of $[L]_R$ is C_{R+L-1}^{L}. The coefficients $\{\mathcal{D}(\mathbf{H}^{(r_1)},\ldots,\mathbf{H}^{(r_L)})\}$ in (35) are known as mixed discriminants and are given by

$$\mathcal{D}(\mathbf{H}^{(r_1)},\ldots,\mathbf{H}^{(r_L)}) = \frac{\partial^L\left|\mathbf{H}^{(r_1)}d_{r_1} + \cdots + \mathbf{H}^{(r_L)}d_{r_L}\right|}{\partial d_{r_1}\cdots\partial d_{r_L}}. \qquad (37)$$

The mixed discriminant of an L-tuple of square matrices $\mathbf{H}^{(r_1)} \in \mathbb{C}^{L\times L}, \ldots, \mathbf{H}^{(r_L)} \in \mathbb{C}^{L\times L}$ can also be defined as

$$\mathcal{D}\left(\mathbf{H}^{(r_1)}, \ldots, \mathbf{H}^{(r_L)}\right) = \frac{1}{L!} \sum_{\sigma\in S_L} \operatorname{sgn}(\sigma) \left| \left[\mathbf{h}^{(r_1)}_{\sigma(1)}, \mathbf{h}^{(r_2)}_{\sigma(2)}, \ldots, \mathbf{h}^{(r_L)}_{\sigma(L)}\right] \right|, \tag{38}$$

where $\mathbf{h}^{(r_l)}_{\sigma(l)}$ denotes the $\sigma(l)$-th column of $\mathbf{H}^{(r_l)}$, S_L denotes the set of all permutations of $1, 2, \ldots, L$ and $\operatorname{sgn}(\sigma)$ denotes the sign of the permutation σ. It can be shown that the two definitions (37) and (38) of the mixed discriminant are equivalent [36].

From (38) it is clear that the mixed discriminant can be understood as an extension of the determinant. Indeed, if $\mathbf{H} := \mathbf{H}^{(r_1)} = \cdots = \mathbf{H}^{(r_L)}$, then (38) reduces to the determinant

$$\mathcal{D}\left(\underbrace{\mathbf{H}, \ldots, \mathbf{H}}_{L \text{ times}}\right) = \sum_{\sigma\in S_L} \operatorname{sgn}(\sigma) \prod_{l=1}^{L} \mathbf{h}_{l,\sigma(l)} = |\mathbf{H}|\,. \tag{39}$$

The mixed discriminant can also be understood as an extension of the permanent. More precisely, let $\mathbf{D}^{(1)} \in \mathbb{C}^{L\times L}, \ldots, \mathbf{D}^{(L)} \in \mathbb{C}^{L\times L}$ be diagonal matrices, then from (38) we obtain (a scaled version of) the permanent

$$\mathcal{D}\left(\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(L)}\right) = \frac{1}{L!} \sum_{\sigma\in S_L} \prod_{l=1}^{L} d^{(\sigma(l))}_{l,l} = \frac{1}{L!} \overset{+}{|}\mathbf{B}\overset{+}{|}, \tag{40}$$

where $\mathbf{B} \in \mathbb{C}^{L\times L}$ is given by $(\mathbf{B})_{il} = d^{(l)}_{ii}$. Furthermore, let $\mathbf{D}^{(1)} \in \mathbb{C}^{L\times L}, \ldots, \mathbf{D}^{(L)} \in \mathbb{C}^{L\times L}$ be diagonal matrices, then

$$\mathcal{D}\left(\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(L)}\right) = \mathcal{D}\left(\mathbf{D}^{(\sigma(1))}, \ldots, \mathbf{D}^{(\sigma(L))}\right) = \frac{1}{L!} \overset{+}{|}\mathbf{B}\overset{+}{|}, \quad \forall \sigma \in S_L, \tag{41}$$

which follows from the column permutation invariance property of the permanent, i.e., $\overset{+}{|}\mathbf{B}\overset{+}{|} = \overset{+}{|}\mathbf{B}\mathbf{\Pi}\overset{+}{|}$ for any permutation matrix $\mathbf{\Pi} \in \mathbb{C}^{L\times L}$. Note that the permanent can be seen as a signless version of the determinant (i.e., $\overset{+}{|}\mathbf{H}\overset{+}{|}$ is equal to (39) when $\operatorname{sgn}(\sigma)$ is dropped). This directly explains the permutation invariance property of the permanent. The three properties (39)–(41) of the mixed discriminant will be used in the derivation of Theorem 3.1. A further discussion of the mixed discriminant and its properties can be found in [36, 37]. A discussion of the properties of the permanent can be found in [32, 33].

3.1.2. Uniqueness condition in terms of dimension of column space

As mentioned earlier, in the proof of Theorem 3.1 we will make use of the expansion of $\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r\right|$ in terms of the scalars $d_1, \ldots, d_R$. The key ingredient in the derivation of condition (53) in Theorem 3.1 is the following identity

$$\begin{aligned}\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r\right| &= \sum_{\sigma\in S_L} \operatorname{sgn}(\sigma) \prod_{l=1}^{L} \left(\sum_{r=1}^{R} d_r \cdot (\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})_{l\sigma(l)}\right) \\ &= \prod_{l=1}^{L} \left(\sum_{r=1}^{R} d_r a_{lr}^{(+,n)}\right) - \prod_{l=1}^{L} \left(\sum_{r=1}^{R} d_r a_{lr}^{(-,n)}\right), \end{aligned} \tag{42}$$

where S_L denotes the set of all permutations of $1, 2, \ldots, L$, and $\mathrm{sgn}(\sigma)$ denotes the sign of the permutation σ. Note also that (42) directly follows from the patterned structure of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$. (See also equations (25) and (27).) Define

$$\mathbf{A}^{(+,r)} = \begin{bmatrix} \mathbf{a}_r^{(+,1)T} \\ \vdots \\ \mathbf{a}_r^{(+,N)T} \end{bmatrix} \in \mathbb{C}^{N\times L} \text{ and } \mathbf{A}^{(-,r)} = \begin{bmatrix} \mathbf{a}_r^{(-,1)T} \\ \vdots \\ \mathbf{a}_r^{(-,N)T} \end{bmatrix} \in \mathbb{C}^{N\times L}. \tag{43}$$

Due to properties (39)–(41), the expansion of the product-of-sums in (42) with respect to $d_1, \ldots, d_R$ yields the homogeneous polynomial (see the proof of Theorem 3.1 for more details):

$$\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r\right| = \prod_{l=1}^{L}\left(\sum_{r=1}^{R} d_r a_{lr}^{(+,n)}\right) - \prod_{l=1}^{L}\left(\sum_{r=1}^{R} d_r a_{lr}^{(-,n)}\right) \tag{44}$$

$$= \sum_{(l_1,\ldots,l_R)\in[L]_R} \Bigg[\mathcal{D}\Big(\underbrace{D_n(\mathbf{A}^{(+,1)}),\ldots,D_n(\mathbf{A}^{(+,1)})}_{l_1 \text{ times}},\ldots,\underbrace{D_n(\mathbf{A}^{(+,R)}),\ldots,D_n(\mathbf{A}^{(+,R)})}_{l_R \text{ times}}\Big)$$

$$- \mathcal{D}\Big(\underbrace{D_n(\mathbf{A}^{(-,1)}),\ldots,D_n(\mathbf{A}^{(-,1)})}_{l_1 \text{ times}},\ldots,\underbrace{D_n(\mathbf{A}^{(-,R)}),\ldots,D_n(\mathbf{A}^{(-,R)})}_{l_R \text{ times}}\Big)\Bigg] d_1^{l_1}\cdots d_R^{l_R}.$$

(Compare with (35).) In terms of the matrices and vectors defined below a compact expression of (44) will be introduced. For every weak composition of L in R terms (i.e., $l_1 + \cdots + l_R = L$ subject to $l_r \geq 0$) we define the square $(L \times L)$ matrices

$$\mathbf{A}_{(l_1,\ldots,l_R)}^{(+,n)} = \left[\mathbf{1}_{l_1}^T \otimes \mathbf{a}_1^{(+,n)}, \ldots, \mathbf{1}_{l_R}^T \otimes \mathbf{a}_R^{(+,n)}\right] \in \mathbb{C}^{L\times L}, \tag{45}$$

$$\mathbf{A}_{(l_1,\ldots,l_R)}^{(-,n)} = \left[\mathbf{1}_{l_1}^T \otimes \mathbf{a}_1^{(-,n)}, \ldots, \mathbf{1}_{l_R}^T \otimes \mathbf{a}_R^{(-,n)}\right] \in \mathbb{C}^{L\times L}. \tag{46}$$

From the matrices in (45) and (46), we also build the row vectors $\mathbf{g}_+^{(n,L)} \in \mathbb{C}^{1\times(C_{R+L-1}^L - R)}$ and $\mathbf{g}_-^{(n,L)} \in \mathbb{C}^{1\times(C_{R+L-1}^L - R)}$ whose entries are indexed by an R-tuple $(l_1, l_2, \ldots, l_R)$ with $0 \leq l_r \leq L-1$ and ordered lexicographically:

$$\mathbf{g}_+^{(n,L)} = \left[\overset{+}{|}\mathbf{A}_{(L-1,1,0,0,\ldots,0)}^{(+,n)}\overset{+}{|}, \overset{+}{|}\mathbf{A}_{(L-1,0,1,0,\ldots,0)}^{(+,n)}\overset{+}{|}, \ldots, \overset{+}{|}\mathbf{A}_{(0,\ldots,0,1,L-1)}^{(+,n)}\overset{+}{|}\right], \tag{47}$$

$$\mathbf{g}_-^{(n,L)} = \left[\overset{+}{|}\mathbf{A}_{(L-1,1,0,0,\ldots,0)}^{(-,n)}\overset{+}{|}, \overset{+}{|}\mathbf{A}_{(L-1,0,1,0,\ldots,0)}^{(-,n)}\overset{+}{|}, \ldots, \overset{+}{|}\mathbf{A}_{(0,\ldots,0,1,L-1)}^{(-,n)}\overset{+}{|}\right]. \tag{48}$$

Based on (47) and (48) we in turn build the row vector

$$\mathbf{g}_{\mathrm{MF}}^{(n,L)} = \left(\mathbf{g}_+^{(n,L)} - \mathbf{g}_-^{(n,L)}\right)\mathbf{D}_{\mathbf{W}}^{(L)} \in \mathbb{C}^{1\times(C_{R+L-1}^L - R)}, \tag{49}$$

in which the subscript 'MF' stands for Monomial Factorization and the diagonal weight matrix $\mathbf{D}_{\mathbf{W}}^{(L)} \in \mathbb{C}^{(C_{R+L-1}^L - R)\times(C_{R+L-1}^L - R)}$ is given by

$$\mathbf{D}_{\mathbf{W}}^{(L)} = \mathrm{diag}\left(w_{(L-1,1,0,0,\ldots,0)}^{(L)}, w_{(L-1,0,1,0,\ldots,0)}^{(L)}, \ldots, w_{(0,\ldots,0,1,L-1)}^{(L)}\right), \tag{50}$$

where the scalar $w^{(L)}_{(l_1,l_2,\ldots,l_R)} = \frac{1}{l_1!l_2!\cdots l_R!}$ takes into account that, due to the column permutation invariance property of the permanent, $\overset{+}{|}\mathbf{A}^{(+,n)}_{(l_1,l_2,\ldots,l_R)}\overset{+}{|}$ and $\overset{+}{|}\mathbf{A}^{(-,n)}_{(l_1,l_2,\ldots,l_R)}\overset{+}{|}$ appear $\frac{L!}{l_1!l_2!\cdots l_R!}$ times in the expansion of $\left|\sum_{r=1}^{R}\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})d_r\right|$ and that each permanent is scaled by the factor $\frac{1}{L!}$ (see (40)), as will be made clear in the proof of Theorem 3.1. Overall, we build

$$\mathbf{G}^{(N,L)}_{\mathrm{MF}} = \left[\mathbf{g}^{(1,L)T}_{\mathrm{MF}}, \mathbf{g}^{(2,L)T}_{\mathrm{MF}}, \ldots, \mathbf{g}^{(N,L)T}_{\mathrm{MF}}\right]^T \in \mathbb{C}^{N\times(C^L_{R+L-1}-R)}. \tag{51}$$

It can be verified that (51) is an extension of (9) to the monomial case, i.e., if $\mathbf{X}$ satisfies the CPD factorization (8) with full column rank $\mathbf{S}$, then $\mathbf{G}^{(N,L)}_{\mathrm{MF}}$ reduces to $\mathbf{G}^{(2)}_{\mathrm{CPD}}$. Note that in the former case there are two superscripts. Namely, 'N' and 'L' that indicate the number of monomial constraints / equations and the degree of the involved monomials, respectively. In the CPD case we have $N = C_I^2C_J^2$ and $L = 2$. Finally, we will also make use of the vector

$$\mathbf{f}^{(L)}(\mathbf{d}) = [d_1^{L-1}d_2, d_1^{L-1}d_3, \ldots, d_{R-1}d_R^{L-1}]^T \in \mathbb{C}^{(C^L_{R+L-1}-R)}. \tag{52}$$

Comparing (10) with (52), it is clear that the latter is also an extension of the former. More precisely, $\mathbf{f}^{(L)}(\mathbf{d})$ consists of all C^L_{R+L-1} distinct entries of $\mathbf{d}\otimes\cdots\otimes\mathbf{d}$ minus the R entries $d_1^L,\ldots,d_R^L$.

Theorem 3.1. *Consider the monomial factorization of $\boldsymbol{X}$ in (2) with N monomial relations of the form (3). Assume that $\boldsymbol{S}$ has full column rank. The monomial factorization of $\boldsymbol{X}$ is unique, $\boldsymbol{A}$ in (2) has full column rank and $\boldsymbol{A}_{tot}$ in (31) has full column rank if and only if the following implication holds*

$$\boldsymbol{G}^{(N,L)}_{MF}\cdot\boldsymbol{f}^{(L)}(\boldsymbol{d}) = \boldsymbol{0} \Rightarrow \omega(\boldsymbol{d}) \leq 1, \tag{53}$$

for all structured vectors $\boldsymbol{f}^{(L)}(\boldsymbol{d})$ of the form (52).

Proof. See Appendix A.

As (11) in the CPD case, condition (53) can be hard to check. Analogous, we observe that if $\mathbf{G}^{(N,L)}_{\mathrm{MF}}$ in (53) has full column rank, then $\mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}$ and the condition is automatically satisfied. This leads to the following more easy to check sufficient uniqueness condition.

Theorem 3.2. *Consider the monomial factorization of $\boldsymbol{X}$ in (2) with N monomial relations of the form (3). If*

$$\begin{cases} \boldsymbol{S} \text{ has full column rank,} \\ \boldsymbol{G}^{(N,L)}_{MF} \text{ has full column rank,} \end{cases} \tag{54}$$

then the monomial factorization of $\boldsymbol{X}$ is unique, $\boldsymbol{A}$ in (2) has full column rank and $\boldsymbol{A}_{tot}$ in (31) has full column rank.

3.2. Uniqueness condition in terms of dimension of null space

Theorem 3.3 below provides an alternative formulation of Theorem 3.1, which may be more easy to comprehend.

Theorem 3.3 makes use of a matrix $\mathbf{\Psi}^{(N,L)} \in \mathbb{C}^{N \times R^L}$, defined as

$$\mathbf{\Psi}^{(N,L)} = \begin{bmatrix} \boldsymbol{\psi}^{(1,L)} \\ \vdots \\ \boldsymbol{\psi}^{(N,L)} \end{bmatrix} = \begin{bmatrix} (\widetilde{\mathbf{a}}_1^{(+,1)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(+,1)})^T \\ \vdots \\ (\widetilde{\mathbf{a}}_1^{(+,N)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(+,N)})^T \end{bmatrix} - \begin{bmatrix} (\widetilde{\mathbf{a}}_1^{(-,1)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(-,1)})^T \\ \vdots \\ (\widetilde{\mathbf{a}}_1^{(-,N)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(-,N)})^T \end{bmatrix}, \quad (55)$$

where $\widetilde{\mathbf{a}}_l^{(+,n)} = [a_{l1}^{(+,n)}, \ldots, a_{lR}^{(+,n)}]^T \in \mathbb{C}^R$ and $\widetilde{\mathbf{a}}_l^{(-,n)} = [a_{l1}^{(-,n)}, \ldots, a_{lR}^{(-,n)}]^T \in \mathbb{C}^R$. Note that

$$\begin{aligned} \left| \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r \right| &= \left(\widetilde{\mathbf{a}}_1^{(+,n)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(+,n)} - \widetilde{\mathbf{a}}_1^{(-,n)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(-,n)} \right)^T (\mathbf{d} \otimes \cdots \otimes \mathbf{d}) \\ &= \boldsymbol{\psi}^{(n,L)}(\mathbf{d} \otimes \cdots \otimes \mathbf{d}), \end{aligned} \quad (56)$$

where $\mathbf{d} = [d_1 \ldots, d_R]^T \in \mathbb{C}^R$. Theorem 3.3 will also make use of the subspace

$$\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}, \quad (57)$$

where $\pi_S^{(L)}$ denotes the subspace of vectorized R^L symmetric tensors. The link between Theorem 3.1 and Theorem 3.3 follows from the observation that condition (53) can also be expressed as

$$\mathbf{f}^{(L)}(\mathbf{d}) \in \ker(\mathbf{G}_{\mathrm{MF}}^{(N,L)}) \Rightarrow \omega(\mathbf{d}) \leq 1. \quad (58)$$

Since

$$\mathbf{G}_{\mathrm{MF}}^{(N,L)} \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0} \Leftrightarrow \mathbf{\Psi}^{(N,L)}(\mathbf{d} \otimes \cdots \otimes \mathbf{d}) = \mathbf{0}, \quad (59)$$

as explained in Appendix C, condition (58) can in turn be expressed as

$$\mathbf{d} \otimes \cdots \otimes \mathbf{d} \in \ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)} \Rightarrow \omega(\mathbf{d}) \leq 1. \quad (60)$$

Note that this also means that the dimension of the subspace $\ker(\mathbf{G}_{\mathrm{MF}}^{(N,L)})$ is equal to the dimension of the subspace $\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$. Theorems 3.3 and 3.4 below are reformulations of Theorems 3.1 and 3.2 in terms of $\ker\left(\mathbf{\Psi}^{(N,L)}\right) \cap \pi_S^{(L)}$.

Theorem 3.3. *Consider the monomial factorization of* $\boldsymbol{X}$ *in* (2) *with* N *monomial relations of the form* (3)*. Assume that* $\boldsymbol{S}$ *has full column rank. The monomial factorization of* $\boldsymbol{X}$ *is unique,* $\boldsymbol{A}$ *in* (2) *has full column rank and* $\boldsymbol{A}_{tot}$ *in* (31) *has full column rank if and only if the following implication holds*

$$\boldsymbol{d} \otimes \cdots \otimes \boldsymbol{d} \in ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)} \Rightarrow \omega(\boldsymbol{d}) \leq 1, \quad (61)$$

where $\mathbf{\Psi}^{(N,L)}$ *is given by* (55)*.*

Theorem 3.4. *Consider the monomial factorization of* $\boldsymbol{X}$ *in* (2) *with* N *monomial relations of the form* (3)*. If*

$$\begin{cases} \boldsymbol{S} \text{ has full column rank,} \\ ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)} \text{ is an } R\text{-dimensional subspace,} \end{cases} \tag{62}$$

then the monomial factorization of $\boldsymbol{X}$ *is unique,* $\boldsymbol{A}$ *in* (2) *has full column rank and* $\boldsymbol{A}_{tot}$ *in* (31) *has full column rank.*

3.3. Application: Extension of CPD to (0,1)-binary weighted CPD

3.3.1. CPD

Consider the CPD of $\mathcal{X}$ given by (7) in which $\mathbf{E}^{(r)} = \mathbf{a}_r \mathbf{b}_r^T$ is associated with the r-th column of $\mathbf{A} \odot \mathbf{B}$. Recall that any 2-by-2 submatrix of $\mathbf{E}^{(r)}$ is a rank-1 matrix, i.e., $\begin{vmatrix} e^{(r)}_{i_1 j_1} & e^{(r)}_{i_1 j_2} \\ e^{(r)}_{i_2 j_1} & e^{(r)}_{i_2 j_2} \end{vmatrix} = e^{(r)}_{i_1 j_1} e^{(r)}_{i_2 j_2} - e^{(r)}_{i_1 j_2} e^{(r)}_{i_2 j_1} = 0$. Since there are C_I^2 ways of selecting two rows of $\mathbf{E}^{(r)}$ and C_J^2 ways of selecting two columns of $\mathbf{E}^{(r)}$, it is clear that the CPD can be interpreted as a monomial factorization involving $N = C_I^2 C_J^2$ monomial relations of the form (23) with $L = 2$, $\mathbf{a}_r^{(+,n)} = [e^{(r)}_{i_1 j_1}, e^{(r)}_{i_2 j_2}]^T$and $\mathbf{a}_r^{(-,n)} = [e^{(r)}_{i_1 j_2}, e^{(r)}_{i_2 j_1}]^T$, where the superscript $'n'$ is associated with the tuple (i_1, i_2, j_1, j_2).

3.3.2. Binary weighted CPD

A nice property of monomial factorizations is that they allow us to extend the CPD model (1) to binary weighted CPD (6) in which $\mathbf{E}^{(r)}$ in (7) now takes the form $\mathbf{E}^{(r)} = \mathbf{D}^{(r)} * (\mathbf{a}_r \mathbf{b}_r^T)$, where $\mathbf{D}^{(r)} \in \{0,1\}^{I \times J}$ is a binary "connectivity" matrix. This means that the tensor representation (7) extends to

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}^{(r)} \circ \mathbf{s}_r = \sum_{r=1}^{R} (\mathbf{D}^{(r)} * (\mathbf{a}_r \mathbf{b}_r^T)) \circ \mathbf{s}_r \tag{63}$$

for binary weighted CPD. From (8) and (63) it is clear that (6) is a matrix representation of the binary weighted CPD of $\mathcal{X}$ in which $\mathbf{D} = [\text{vec}(\mathbf{D}^{(1)T}), \ldots, \text{vec}(\mathbf{D}^{(R)T})] \in \mathbb{C}^{IJ \times R}$. Since $\mathbf{E}^{(r)}$ is not necessarily a low-rank matrix, the CPD modeling approach cannot be used for binary weighted CPD. However, it can be verified that any 2-by-2 submatrix of $\mathbf{E}^{(r)} = \mathbf{D}^{(r)} * (\mathbf{a}_r \mathbf{b}_r^T)$ must satisfy the monomial relation $e^{(r)}_{i_1 j_1} e^{(r)}_{i_2 j_2} e^{(r)}_{i_1 j_2} e^{(r)}_{i_2 j_1} \cdot (e^{(r)}_{i_1 j_1} e^{(r)}_{i_2 j_2} - e^{(r)}_{i_1 j_2} e^{(r)}_{i_2 j_1}) = 0$. We can now conclude that the binary weighted CPD of a tensor can be interpreted as a monomial factorization involving $N = C_I^2 C_J^2$ monomial relations of the form (23) with $L = 6$ and

$$\begin{cases} \mathbf{a}_r^{(+,n)} = [e^{(r)}_{i_1 j_1}, e^{(r)}_{i_2 j_2} e^{(r)}_{i_1 j_2}, e^{(r)}_{i_2 j_1}, e^{(r)}_{i_1 j_1}, e^{(r)}_{i_2 j_2}]^T, \\ \mathbf{a}_r^{(-,n)} = [e^{(r)}_{i_1 j_1}, e^{(r)}_{i_2 j_2}, e^{(r)}_{i_1 j_2}, e^{(r)}_{i_2 j_1}, e^{(r)}_{i_1 j_2}, e^{(r)}_{i_2 j_1}]^T, \end{cases} \tag{64}$$

in which the superscript $'n'$ is associated with the tuple (i_1, i_2, j_1, j_2).

Let us end this section with a concrete example that demonstrates that the connectivity pattern of $\mathbf{D}$ makes the uniqueness properties of the binary weighted CPD of $\mathcal{X}$

in (6) different from the uniqueness properties the CPD of $\mathcal{X}$ in (7). Let $I = J = 5$, $K = R = 3$, and

$$\mathbf{A} = \begin{bmatrix} 2 & 8 & 4 \\ 7 & 3 & 4 \\ 2 & 10 & 3 \\ 4 & 8 & 5 \\ 6 & 5 & 5 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 8 & 5 & 6 \\ 8 & 3 & 6 \\ 7 & 10 & 2 \\ 4 & 9 & 3 \\ 8 & 6 & 5 \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} 6 & 8 & 10 \\ 8 & 2 & 8 \\ 6 & 1 & 4 \end{bmatrix},$$

$$\mathbf{D}^{(1)} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix}, \mathbf{D}^{(2)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \mathbf{D}^{(3)} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

It can be verified that condition (54) in Theorem 3.2 is satisfied, implying the uniqueness of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{S}$. It can also be verified that the ranks of $\mathbf{E}^{(1)} = \mathbf{D}^{(1)} * (\mathbf{a}_1\mathbf{b}_1^T)$, $\mathbf{E}^{(2)} = \mathbf{D}^{(2)} * (\mathbf{a}_2\mathbf{b}_2^T)$ and $\mathbf{E}^{(3)} = \mathbf{D}^{(3)} * (\mathbf{a}_3\mathbf{b}_3^T)$ are 4, 3 and 5, respectively, i.e., the connectivity pattern implies that the binary weighted CPD factor matrices are not even required to be low rank. If they were low rank, BTD would have applied; this exemplifies that binary weighted CPD can guarantee uniqueness in cases where other tensor decompositions fail. Observe also that in contrast to the ordinary CPD, the local low-rank properties of the binary weighted CPD factor matrices can vary. For example, $[\mathbf{e}_1^{(I)}\ \mathbf{e}_2^{(I)}]^T\mathbf{E}^{(1)}[\mathbf{e}_1^{(J)}\ \mathbf{e}_2^{(J)}] = \left[\begin{smallmatrix} 16 & 16 \\ 56 & 0 \end{smallmatrix}\right]$ is a rank-2 matrix while $[\mathbf{e}_{I-1}^{(I)}\ \mathbf{e}_I^{(I)}]^T\mathbf{E}^{(1)}[\mathbf{e}_{J-I}^{(J)}\ \mathbf{e}_J^{(J)}] = \left[\begin{smallmatrix} 16 & 32 \\ 24 & 48 \end{smallmatrix}\right]$ is a rank-1 matrix.

The connectivity pattern of $\mathbf{D}$ also affects the identifiability of the CPD factor matrices $\mathbf{A}$ and $\mathbf{B}$. To see this, consider again the above example, but now we set $\mathbf{a}_1 = \mathbf{a}_2$ and $\mathbf{b}_1 = \mathbf{b}_2$. The binary weighted CPD of $\mathcal{X}$ in (63) is still unique, despite the rank-deficient matrix $\mathbf{A} \odot \mathbf{B}$. In contrast, the CPD of $\mathcal{X}$ (i.e., $\mathbf{D} = \mathbf{1}_{IJ}\mathbf{1}_R^T$) is not unique.

4. Algorithms for monomial factorizations

In this section we will present algebraic algorithms tailored for the monomial factorization of $\mathbf{X}$. In Section 4.1 we first derive an algorithm that exploits the link between monomial factorizations and coupled BTD, as discussed in Section 2.2. Next, in Section 4.2 we derive an algorithm that also exploits the linear structure on $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$, which is based on the column space formulation in Section 3.1.2. Finally, in Section 4.3 we present an alternative algorithm that relies on the null space formulation in Section 3.2.

4.1. Condition and algorithm based on the coupled BTD formulation

In this section we review the algorithm for coupled BTD developed in [13], which is an extension of the algebraic algorithm for CPD developed in [27, 30]. This leads to a constructive interpretation of Theorem 2.1 for monomial factorizations. First in Section 4.2 we will explain how to incorporate the linear structure of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ into the algorithm, leading to a constructive interpretation of Theorem 3.2.

4.1.1. Exploiting low-rank structure

Let $\mathbf{X}^{(n,k)}$ denote the k-th matrix slice of $\mathcal{X}^{(n)}$ in (30) so that from (15) we obtain

$$\mathbf{X}^{(n,k)} = \mathcal{X}^{(n)}(:,:,k) = \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) s_{kr} = \sum_{r=1}^{R} \mathbf{N}^{(n,r)} \mathbf{M}^{(n,r)T} s_{kr}. \tag{65}$$

Assume that condition (22) in Theorem 2.1 is satisfied. This implies that we can w.l.o.g. assume that $K = R$, i.e., $\mathbf{S}$ is square nonsingular.[2] The "monomial" structure of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ in (65) will first be exploited in Section 4.2. For the moment we will only use the low-rank factorization $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) = \mathbf{N}^{(n,r)} \mathbf{M}^{(n,r)T}$. The overall idea behind the algebraic algorithm for coupled BTD is to look for a matrix $\mathbf{W} \in \mathbb{C}^{R \times R}$ that separates the R BTD terms $\{\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) \circ \mathbf{s}_r\}$ in (30), i.e.,

$$[\mathbf{d}_1, \ldots, \mathbf{d}_R] = \mathbf{S}^T \mathbf{W} \quad \text{and} \quad \mathbf{W} = \mathbf{S}^{-T} \mathbf{\Pi} \mathbf{\Delta}, \tag{66}$$

where $\mathbf{\Pi}$ is a permutation matrix and $\mathbf{\Delta}$ is a nonsingular diagonal matrix. Note that the separation property of $\mathbf{d}_r$ implies that $\omega(\mathbf{d}_r) = 1$, say entry $d_{sr} = (\mathbf{d}_r)_s$ corresponds to the nonzero entry of $\mathbf{d}_r$. This in turn implies from (65) that $\mathbf{X}\mathbf{w}_r = \mathbf{a}_s d_{sr} \Leftrightarrow \sum_{t=1}^{R} \mathbf{X}^{(n,t)} w_{tr} = \mathbf{A}_L(\mathbf{a}_s^{(+,n)}, \mathbf{a}_s^{(-,n)}) d_{sr}$, $n \in \{1, \ldots, N\}$. The column permutation ambiguity of the separation problem implies that we can w.l.o.g. drop the subscripts of $\mathbf{d}_r$ and $\mathbf{w}_r$, i.e., we now write $\mathbf{d}$ and $\mathbf{w}$. By capitalizing on the low-rank property of $\mathbf{A}_L(\mathbf{a}_s^{(+,n)}, \mathbf{a}_s^{(-,n)})$, $\mathbf{w}$ must satisfy the relations $\left|\sum_{r=1}^{R} \mathbf{X}^{(n,r)} w_r\right| = \left|\mathbf{A}_L(\mathbf{a}_s^{(+,n)}, \mathbf{a}_s^{(-,n)}) d_s\right| = 0$, $n \in \{1, \ldots, N\}$, where d_s corresponds to the nonzero entry of $\mathbf{d}$. However, since $\mathbf{w}$ is unknown, we obtain the homogeneous polynomial of degree L:

$$\left|\sum_{r=1}^{R} \mathbf{X}^{(n,r)} w_r\right| = \sum_{r_1, \ldots, r_L = 1}^{R} w_{r_1} \cdots w_{r_L} \mathcal{D}(\mathbf{X}^{(n,r_1)}, \ldots, \mathbf{X}^{(n,r_L)}). \tag{67}$$

Observe that in (67), the mixed discriminants $\{\mathcal{D}(\mathbf{X}^{(n,r_1)}, \ldots, \mathbf{X}^{(n,r_L)})\}$ given by (37) are known while the coefficients in the vector $\mathbf{w} = [w_1, \ldots, w_R]^T$, corresponding to a column of $\mathbf{W}$ in (66), are unobserved. We will now explain how to find the columns of $\mathbf{W}$, each with separation property $\left|\sum_{r=1}^{R} \mathbf{X}^{(n,r)} w_r\right| = \left|\mathbf{A}_L(\mathbf{a}_s^{(+,n)}, \mathbf{a}_s^{(-,n)}) d_s\right| = 0$, $n \in \{1, \ldots, N\}$, via the observable mixed discriminants $\{\mathcal{D}(\mathbf{X}^{(n,r_1)}, \ldots, \mathbf{X}^{(n,r_L)})\}$. Let $\mathbf{r}_L(\mathcal{X}^{(n)}) \in \mathbb{C}^{1 \times R^L}$ be the row vector consisting of all possible mixed discriminants involving L matrix slices of $\mathcal{X}^{(n)}$, i.e.,

$$\mathbf{r}_L(\mathcal{X}^{(n)}) = \left[\psi^{(n)}_{(L,0,0,\ldots,0)}, \psi^{(n)}_{(L-1,1,0,\ldots,0)}, \ldots, \psi^{(n)}_{(0,0\ldots,0,L)}\right], \quad n \in \{1, \ldots, N\}, \tag{68}$$

in which the entries $\{\psi^{(n)}_{(l_1,\ldots,l_R)}\}$ are indexed by the R-tuple $(l_1, \ldots, l_R)$ with $0 \leq l_k \leq L$ and property $l_1 + \cdots + l_R = L$, that is

$$\psi^{(n)}_{(l_1,\ldots,l_R)} = \mathcal{D}\Big(\underbrace{\mathbf{X}^{(n,1)}, \ldots, \mathbf{X}^{(n,1)}}_{l_1 \text{ times}}, \ldots, \underbrace{\mathbf{X}^{(n,R)}, \ldots, \mathbf{X}^{(n,R)}}_{l_R \text{ times}}\Big).$$

[2] Theorem 2.1 tells us that if $\mathbf{G}^{(N,L)}_{\text{BTD}}$ and $\mathbf{S}$ have full column rank, then $\mathbf{A}$ also has full column rank. Consequently, the $(I \times K)$ matrix $\mathbf{X} = \mathbf{A}\mathbf{S}^T$ can always be reduced to an $(I \times R)$ matrix.

The entries of the tuple $(l_1, \ldots, l_R)$ indicate how many times the different matrix slices of $\mathcal{X}^{(n)}$ are involved in the construction of the mixed discriminant. It can be verified that (see [30, 13] for details):

$$\mathbf{r}_L(\mathcal{X}^{(n)}) = (\mathcal{C}_L(\mathbf{M}^{(n)}) \odot \mathcal{C}_L(\mathbf{N}^{(n)}))\mathcal{R}_L(\mathbf{S}^{(\text{ext})})^T, \quad n \in \{1, \ldots, N\}, \tag{69}$$

where the matrix $\mathcal{R}_L(\mathbf{S}^{(\text{ext})}) \in \mathbb{C}^{R^L \times C^L_{R(L-1)}}$ is given by

$$\mathcal{R}_L(\mathbf{S}^{(\text{ext})}) = L! \left[\text{Sym}(\mathbf{s}_1^{(\text{ext})} \otimes \cdots \otimes \mathbf{s}_L^{(\text{ext})}), \ldots, \text{Sym}(\mathbf{s}_{R(L-1)-L+1}^{(\text{ext})} \otimes \cdots \otimes \mathbf{s}_{R(L-1)}^{(\text{ext})})\right], \tag{70}$$

in which the columns of $\mathcal{R}_L(\mathbf{S}^{(\text{ext})})$ consist of the symmetric part of the rank-one tensors $\{\mathbf{s}_{r_1}^{(\text{ext})} \otimes \cdots \otimes \mathbf{s}_{r_L}^{(\text{ext})}\}$, i.e., $\text{Sym}(\mathbf{s}_{r_1}^{(\text{ext})} \otimes \cdots \otimes \mathbf{s}_{r_L}^{(\text{ext})}) = \frac{1}{L!}\sum_{\sigma \in S_L} \mathbf{s}_{r_{\sigma(1)}}^{(\text{ext})} \otimes \cdots \otimes \mathbf{s}_{r_{\sigma(L)}}^{(\text{ext})}$. Stacking the row-vectors (69) yields

$$\mathbf{R}^{(N,L)} = \begin{bmatrix} \mathbf{r}_L(\mathcal{X}^{(1)}) \\ \vdots \\ \mathbf{r}_L(\mathcal{X}^{(N)}) \end{bmatrix} = \begin{bmatrix} \mathcal{C}_L(\mathbf{M}^{(1)}) \odot \mathcal{C}_L(\mathbf{N}^{(1)}) \\ \vdots \\ \mathcal{C}_L(\mathbf{M}^{(N)}) \odot \mathcal{C}_L(\mathbf{N}^{(N)}) \end{bmatrix} \mathcal{R}_L(\mathbf{S}^{(\text{ext})})^T \in \mathbb{C}^{N \times R^L}. \tag{71}$$

Let

$$\mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})}) \in \mathbb{C}^{R^L \times (C^L_{R+L-1} - R)} \tag{72}$$

be the matrix formed by the distinct columns of $\mathcal{R}_L(\mathbf{S}^{(\text{ext})})$, corresponding to the columns $\text{Sym}(\mathbf{s}_{r_1} \otimes \cdots \otimes \mathbf{s}_{r_L})$, $1 \leq r_1 \leq \cdots \leq r_L \leq R$, $r_1 \neq r_L$. A more detailed discussion about the structure of $\mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})$ can be found in Appendix E. Similar to the construction of $\mathbf{G}_{\text{BTD}}^{(N,L)}$, see (20), this matrix takes into account that column $\mathbf{s}_l$ is repeated $L-1$ times in $\mathbf{S}^{(\text{ext})}$ and that the R columns $\text{Sym}(\mathbf{s}_r \otimes \cdots \otimes \mathbf{s}_r)$, $1 \leq r \leq R$, are associated with all-zero columns in the Khatri–Rao products in (71). Overall, we obtain a compact version of (71) (see also (B.7)–(B.9)):

$$\mathbf{R}^{(N,L)} = \begin{bmatrix} \mathcal{C}_L(\mathbf{M}^{(1)}) \odot \mathcal{C}_L(\mathbf{N}^{(1)}) \\ \vdots \\ \mathcal{C}_L(\mathbf{M}^{(N)}) \odot \mathcal{C}_L(\mathbf{N}^{(N)}) \end{bmatrix} \mathbf{P}_{\text{BTD}} \cdot \mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})^T = \mathbf{G}_{\text{BTD}}^{(N,L)} \cdot \mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})^T, \tag{73}$$

where the "compression" matrix $\mathbf{P}_{\text{BTD}} \in \mathbb{C}^{C^L_{R(L-1)} \times (C^L_{R+L-1} - R)}$ is given by (21). It now turns out that if condition (34) in Theorem 2.1 is satisfied, then thanks to relation (73), the problem of finding the coupled BTD of $\{\mathcal{X}^{(n)}\}$ can be translated into an ordinary CPD problem, as will be explained next.

4.1.2. Reduction to CPD

We will now explain that if $\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}$ is an R-dimensional subspace (we will justify it below), then $\mathbf{W}$ can be obtained via a CPD. Observe that relation (73) implies that

$$\ker(\mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})^T) \cap \pi_S^{(L)} \subseteq \ker(\mathbf{G}_{\text{BTD}}^{(N,L)}\mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})^T) \cap \pi_S^{(L)} = \ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}.$$

Due to [12, Lemma 5.3] and the fact that since $\mathbf{S}$ is nonsingular, we have that

$$\begin{cases} \dim\big(\ker(\mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})^T) \cap \pi_S^{(L)}\big) = R, \\ \mathbf{w}_r \otimes \cdots \otimes \mathbf{w}_r \in \ker(\mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})^T) \cap \pi_S^{(L)}, \quad \forall r \in \{1, \ldots, R\}. \end{cases} \tag{74}$$

Thus,

$$\mathbf{R}^{(N,L)}(\mathbf{W} \odot \cdots \odot \mathbf{W}) = \mathbf{0}. \tag{75}$$

From (75) it is clear that if $\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}$ is an R-dimensional subspace, then $\mathbf{W}$ can be obtained from any basis for $\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}$ via a CPD. In more detail, since $\dim\big(\ker(\mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})^T) \cap \pi_S^{(L)}\big) = \dim\big(\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}\big) = R$, the vectors $\{\mathbf{w}_r \otimes \cdots \otimes \mathbf{w}_r\}_{r=1}^R$ constitute a basis for both of the involved subspaces, i.e.,

$$\text{range}\,(\mathbf{W} \odot \cdots \odot \mathbf{W}) = \ker(\mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})^T) \cap \pi_S^{(L)} = \ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}. \tag{76}$$

This implies that $\mathbf{W}$ can be uniquely determined from $\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}$. In other words, if $\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}$ is an R-dimensional subspace, then $\mathbf{S} = \mathbf{W}^{-T}$ and $\mathbf{A} = \mathbf{X}\mathbf{W}$ are unique (up to intrinsic column scaling and permutation ambiguities). We will now argue that if condition (22) in Theorem 2.1 is satisfied, then $\dim\big(\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}\big) = R$. Relation (73) and the full column rank assumption on $\mathbf{G}_{\text{BTD}}^{(N,L)}$ imply that

$$\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)} = \ker(\mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})^T) \cap \pi_S^{(L)}, \tag{77}$$

i.e., we also have that $\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)} \subseteq \ker(\mathcal{R}_L^{(\text{dist})}(\mathbf{S}^{(\text{ext})})^T) \cap \pi_S^{(L)}$. From property (74) and relation (77) it is now clear that if condition (22) in Theorem 2.1 is satisfied, then the key assumption that $\dim\big(\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}\big) = R$ holds true.

An important practical implication of (76) is that $\mathbf{W} = \mathbf{S}^{-T}$ can be obtained via a CPD. In detail, let the columns of $\mathbf{Q}$ constitute a basis for $\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}$ and assume that the dimension of this subspace is minimal (i.e., R), then there exists a nonsingular change-of-basis matrix $\mathbf{F} \in \mathbb{C}^{R \times R}$ such that

$$\mathbf{Q} = (\mathbf{W} \odot \cdots \odot \mathbf{W})\mathbf{F}^T. \tag{78}$$

It is clear that relation (78) corresponds to a matrix representation of the CPD of a tensor $\mathcal{Q} \in \mathbb{C}^{R \times \cdots \times R}$ of order $L+1$:

$$\mathcal{Q} = \sum_{r=1}^{R} \mathbf{w}_r \circ \cdots \circ \mathbf{w}_r \circ \mathbf{f}_r\,. \tag{79}$$

Since all the factor matrices of the CPD in (79) have full column rank, we know that $\mathbf{W}$ can be recovered from (79) via an EVD (e.g., [3, 34]). Theorem 4.1 below summarizes the above uniqueness result for monomial factorizations, which is based on a constructive interpretation of Theorem 2.1.

Theorem 4.1. *Consider the coupled BTD of* $\mathcal{X}^{(n)} \in \mathbb{C}^{L \times L \times K}$, $n \in \{1, \ldots, N\}$ *in* (30). *If*

$$\begin{cases} \boldsymbol{S} \text{ has full column rank,} \\ ker(\boldsymbol{R}^{(N,L)}) \cap \pi_S^{(L)} \text{ is an } R\text{-dimensional subspace,} \end{cases} \tag{80}$$

then the coupled BTD rank of $\{\mathcal{X}^{(n)}\}$ *is* R*, the coupled BTD of* $\{\mathcal{X}^{(n)}\}$ *is unique,* $\boldsymbol{A}_{tot}$ *has full column rank, the monomial factorization of* $\boldsymbol{X}$ *in* (2) *is unique,* $\boldsymbol{A}$ *in* (2) *has full column rank, and the coupled BTD of* $\{\mathcal{X}^{(n)}\}$ *can be reduced to the CPD of* $\mathcal{Q}$ *in* (79)*. The latter decomposition can be computed algebraically via an EVD.*

The constructive interpretation of Theorem 2.1 is summarized as Algorithm 1.

Algorithm 1 Summary of constructive interpretation of Theorem 2.1.

1. Build $\mathbf{R}^{(N,L)}$ in (71).
2. Obtain matrix $\mathbf{Q}$ whose columns form a basis for $\ker(\mathbf{R}^{(N,L)}) \cap \pi_S^{(L)}$.
3. Build tensor $\mathcal{Q}$ in (79) from $\mathbf{Q}$.
4. Obtain $\mathbf{W}$ from CPD of $\mathcal{Q}$.
5. Compute $\mathbf{A} = \mathbf{X}\mathbf{W}$ and $\mathbf{S} = \mathbf{W}^{-T}$.

4.2. Condition and algorithm based on column space formulation

Note that in the preceding discussion, the linear structure of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ was not used. In this section we explain how it can be exploited. From (28) we know that $\mathbf{X}^{(l,n)} = \mathbf{A}_L(\mathbf{x}_l^{(+,n)}, \mathbf{x}_l^{(-,n)}) = \text{diag}(\mathbf{x}_l^{(+,n)}) + (\text{diag}(\mathbf{x}_l^{(-,n)})\mathbf{P}) * (\mathbf{1}_L\mathbf{1}_L^T - \mathbf{e}_L^{(L)}\mathbf{e}_L^{(L)T} + (-1)^L\mathbf{e}_L^{(L)}\mathbf{e}_L^{(L)T})$, where $\mathbf{x}_l^{(+,n)} \in \mathbb{C}^L$ and $\mathbf{x}_l^{(-,n)} \in \mathbb{C}^L$ are of the form (29) and $\mathbf{P} \in \mathbb{C}^{L\times L}$ is the "cyclic" permutation matrix given by (26). Observe that if $\mathbf{X}^{(l,n)} = \text{diag}(\mathbf{x}_l^{(+,n)}) + (\text{diag}(\mathbf{x}_l^{(-,n)})\mathbf{P}) * (\mathbf{1}_L\mathbf{1}_L^T - \mathbf{e}_L^{(L)}\mathbf{e}_L^{(L)T} + (-1)^L\mathbf{e}_L^{(L)}\mathbf{e}_L^{(L)T})$, where $l \in \{1, \ldots, L\}$, then from definition (38), property (40) and relation (42) we obtain

$$\mathcal{D}\left(\mathbf{X}^{(1,n)}, \ldots, \mathbf{X}^{(L,n)}\right) = \frac{1}{L!}\sum_{\sigma\in S_L} \text{sgn}(\sigma) \left| \left[\mathbf{x}_{\sigma(1)}^{(1,n)}, \mathbf{x}_{\sigma(2)}^{(2,n)}, \ldots, \mathbf{x}_{\sigma(L)}^{(L,n)}\right] \underbrace{\mathbf{\Pi}_{\sigma^{-1}}\mathbf{\Pi}_\sigma}_{=\mathbf{I}_L} \right| = \tag{81}$$

$$\frac{1}{L!}\sum_{\sigma\in S_L} \underbrace{\text{sgn}(\sigma)\,|\mathbf{\Pi}_\sigma|}_{=1} \left|[\mathbf{x}_1^{(\sigma(1),n)}, \mathbf{x}_2^{(\sigma(2),n)}, \ldots, \mathbf{x}_L^{(\sigma(L),n)}]\right| =$$

$$\frac{1}{L!}\sum_{\sigma\in S_L}\prod_{l=1}^{L} x_{l\sigma(l)}^{(+,n)} - \prod_{l=1}^{L} x_{l\sigma(l)}^{(-,n)} = \frac{1}{L!}\Big|[\mathbf{x}_1^{(+,n)}, \ldots, \mathbf{x}_L^{(+,n)}]\Big|^+ - \frac{1}{L!}\Big|[\mathbf{x}_1^{(-,n)}, \ldots, \mathbf{x}_L^{(-,n)}]\Big|^+,$$

where $\mathbf{\Pi}_\sigma$ denotes the permutation matrix associated with σ (similarly for $\mathbf{\Pi}_{\sigma^{-1}}$). Hence, in the context of monomial factorizations, relation (81) implies that the mixed discriminants $\{\psi_{(l_1,\ldots,l_R)}^{(n)}\}$ in (69) simplify to

$$\psi_{(l_1,\ldots,l_R)}^{(n)} = \mathcal{D}\Big(\underbrace{\mathbf{X}^{(n,1)}, \ldots, \mathbf{X}^{(n,1)}}_{l_1 \text{ times}}, \ldots, \underbrace{\mathbf{X}^{(n,R)}, \ldots, \mathbf{X}^{(n,R)}}_{l_R \text{ times}}\Big) = \tag{82}$$

$$\frac{1}{L!}\Big|\Big[\mathbf{1}_{l_1}^T \otimes \mathbf{x}_1^{(+,n)}, \ldots, \mathbf{1}_{l_R}^T \otimes \mathbf{x}_R^{(+,n)}\Big]\Big|^+ - \frac{1}{L!}\Big|\Big[\mathbf{1}_{l_1}^T \otimes \mathbf{x}_1^{(-,n)}, \ldots, \mathbf{1}_{l_R}^T \otimes \mathbf{x}_R^{(-,n)}\Big]\Big|^+.$$

It can be verified that $\mathbf{r}_L(\mathcal{X}^{(n)})$ in (69) has C^L_{R+L-1} distinct entries; actually the rows of $\mathbf{R}^{(N,L)}$ represent vectorized symmetric tensors. Indeed, the permutation invariance property of the permanent implies that $\mathbf{R}^{(N,L)}$ only has at most C^L_{R+L-1} distinct columns. In detail, for every weak composition of L in R terms (i.e., $l_1 + \cdots + l_R = L$ subject to $l_r \geq 0$) we define the square matrices

$$\mathbf{X}^{(+,n)}_{(l_1,\ldots,l_R)} = \left[\mathbf{1}^T_{l_1} \otimes \mathbf{x}^{(+,n)}_1, \ldots, \mathbf{1}^T_{l_R} \otimes \mathbf{x}^{(+,n)}_R\right] \in \mathbb{C}^{L\times L}, \tag{83}$$

$$\mathbf{X}^{(-,n)}_{(l_1,\ldots,l_R)} = \left[\mathbf{1}^T_{l_1} \otimes \mathbf{x}^{(-,n)}_1, \ldots, \mathbf{1}^T_{l_R} \otimes \mathbf{x}^{(-,n)}_R\right] \in \mathbb{C}^{L\times L}. \tag{84}$$

Note that (83) and (84) correspond to (45) and (46), but built form the entries of the observed data matrix $\mathbf{X}$. Based on (83) and (84), we also build the row vectors $\mathbf{r}^{(n,L)}_+ \in \mathbb{C}^{1\times C^L_{R+L-1}}$ and $\mathbf{r}^{(n,L)}_- \in \mathbb{C}^{1\times C^L_{R+L-1}}$ whose entries are indexed by an R-tuple $(l_1, l_2, \ldots, l_R)$ with $0 \leq l_r \leq L$ and ordered lexicographically:

$$\mathbf{r}^{(n,L)}_+ = \left[\overset{+}{|}\mathbf{X}^{(+,n)}_{(L,0,0,\ldots,0)}\overset{+}{|}, \overset{+}{|}\mathbf{X}^{(+,n)}_{(L-1,1,0,\ldots,0)}\overset{+}{|}, \ldots, \overset{+}{|}\mathbf{X}^{(+,n)}_{(0,\ldots,0,0,L)}\overset{+}{|}\right], \tag{85}$$

$$\mathbf{r}^{(n,L)}_- = \left[\overset{+}{|}\mathbf{X}^{(-,n)}_{(L,0,0,\ldots,0)}\overset{+}{|}, \overset{+}{|}\mathbf{X}^{(-,n)}_{(L-1,1,0,\ldots,0)}\overset{+}{|}, \ldots, \overset{+}{|}\mathbf{X}^{(-,n)}_{(0,\ldots,0,0,L)}\overset{+}{|}\right]. \tag{86}$$

Here, (85) and (86) correspond to (47) and (48), but based on entries of the observed data matrix $\mathbf{X}$. The only difference is that the R tuples $(L, 0, \ldots, 0)$, $\ldots$, $(0, \ldots, 0, L)$ are also included. From (69) and (82) it is clear that when the linear structure of $\mathbf{A}_L(\mathbf{a}^{(+,n)}_r, \mathbf{a}^{(-,n)}_r)$ is taken into account, then $\mathbf{r}^{(n,L)}_+ - \mathbf{r}^{(n,L)}_-$ consists of the distinct entries of $\mathbf{r}_L(\mathcal{X}^{(n)})$. In other words, $\mathbf{r}^{(n,L)}_+ - \mathbf{r}^{(n,L)}_-$ imposes the linear structure of $\mathbf{A}_L(\mathbf{a}^{(+,n)}_r, \mathbf{a}^{(-,n)}_r)$ onto $\mathbf{r}_L(\mathcal{X}^{(n)})$. Hence, by exploiting the monomial structure, the computation of $\mathbf{R}^{(N,L)}$ given by (73) reduces to

$$\mathbf{R}^{(N,L)}_{\mathrm{MF}} = \left(\begin{bmatrix} \mathbf{r}^{(1,L)}_+ \\ \vdots \\ \mathbf{r}^{(N,L)}_+ \end{bmatrix} - \begin{bmatrix} \mathbf{r}^{(1,L)}_- \\ \vdots \\ \mathbf{r}^{(N,L)}_- \end{bmatrix}\right) \widetilde{\mathbf{D}}^{(L)}_{\mathbf{W}} \in \mathbb{C}^{N\times C^L_{R+L-1}}, \tag{87}$$

where

$$\widetilde{\mathbf{D}}^{(L)}_{\mathbf{W}} = \mathrm{diag}\left(w^{(L)}_{(L,0,0,\ldots,0)}, w^{(L)}_{(L-1,1,0,\ldots,0)}, \ldots, w^{(L)}_{(0,\ldots,0,L)}\right) \in \mathbb{C}^{C^L_{R+L-1}\times C^L_{R+L-1}}, \tag{88}$$

with diagonal entires $w^{(L)}_{(l_1,l_2,\ldots,l_R)} = \frac{1}{l_1!l_2!\cdots l_R!}$. Observe that (88) is a diagonal column weight matrix of the form (50) in which the R scalars $w^{(L)}_{(L,0,\ldots,0)}, \ldots, w^{(L)}_{(0,\ldots,0,L)}$ are now also included. Due to relation (87), (75) reduces to

$$\mathbf{R}^{(N,L)}_{\mathrm{MF}} \cdot [\mathbf{f}^{(L)}_{\mathrm{MF}}(\mathbf{w}_1), \ldots, \mathbf{f}^{(L)}_{\mathrm{MF}}(\mathbf{w}_R)] = \mathbf{0}, \tag{89}$$

where $\mathbf{f}^{(L)}_{\mathrm{MF}}(\mathbf{w}) \in \mathbb{C}^{C^L_{R+L-1}}$ consists of the distinct entries of $\mathbf{w} \otimes \cdots \otimes \mathbf{w}$, i.e., $\mathbf{f}^{(L)}_{\mathrm{MF}}(\mathbf{w}) = [w^L_1, w^{L-1}_1 w_2, \ldots, w^L_R]^T$. Hence, if $\ker(\mathbf{R}^{(N,L)}_{\mathrm{MF}})$ is an R-dimensional subspace, then $\mathbf{W}$

can be obtained from (89). We will now explain that if condition (54) in Theorem 3.2 is satisfied, then $\ker(\mathbf{R}_{\mathrm{MF}}^{(N,L)})$ is indeed R-dimensional. Using relation (81), we show in Appendix F that

$$\mathbf{R}_{\mathrm{MF}}^{(N,L)} = \mathbf{G}_{\mathrm{MF}}^{(N,L)} \cdot \mathcal{R}_L^{(\mathrm{dist})}(\mathbf{S}^{(\mathrm{ext})})^T. \tag{90}$$

Note that (90) can be seen as an extension of the identity (73) from CPD/BTD to monomial factorizations. From (90) it is clear that if $\mathbf{G}_{\mathrm{MF}}^{(N,L)}$ has full column rank, then $\ker(\mathbf{R}_{\mathrm{MF}}^{(N,L)}) = \ker(\mathcal{R}_L^{(\mathrm{dist})}(\mathbf{S}^{(\mathrm{ext})})^T)$. We can now proceed as in Section 4.1.2. Theorem 4.2 below summarizes the above uniqueness result for monomial factorizations, which is based on a constructive interpretation of Theorem 3.2.

Theorem 4.2. *Consider the monomial factorization of $\boldsymbol{X}$ in (2). If*

$$\begin{cases} \boldsymbol{S} \text{ has full column rank,} \\ ker(\boldsymbol{R}_{MF}^{(N,L)}) \text{ is an } R\text{-dimensional subspace,} \end{cases} \tag{91}$$

then $\boldsymbol{A}$ and $\boldsymbol{S}$ in (2) are unique, $\boldsymbol{A}$ has full column rank, and the monomial factorization of $\boldsymbol{X}$ can be reduced to the CPD of $\mathcal{Q}$ in (79). The latter decomposition can be computed algebraically via an EVD.

Note that the only difference between Theorems 4.1 and 4.2 is that in the latter case $\mathcal{Q}$ is obtained from $\mathbf{R}_{\mathrm{MF}}^{(N,L)}$ given by (90), in which the linear structure of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ is taken into account. Hence, by replacing $\mathbf{R}^{(N,L)}$ with $\mathbf{R}_{\mathrm{MF}}^{(N,L)}$ in Step 1 Algorithm 1, the algorithm also summarizes the constructive interpretation of Theorem 4.2.

4.3. Condition and algorithm based on the null space formulation

In this section an alternative algebraic algorithm for the monomial factorization problem will be outlined that is based on the null space formulation discussed in Section 3.2. We also explain that this null space approach is equivalent to the approach discussed in Section 4.2. Assume that condition (62) in Theorem 3.4 is satisfied. W.l.o.g. we can assume that $\mathbf{S}$ is square nonsingular ($K = R$). Note that from (56) there exist R linearly independent vectors $\mathbf{w}_1, \ldots, \mathbf{w}_R$, each with property $\omega(\mathbf{S}^T\mathbf{w}_r) = 1$, such that for every $n \in \{1, \ldots, N\}$ we have that

$$\left| \sum_{q=1}^{R} \mathbf{A}_L(\mathbf{a}_q^{(+,n)}, \mathbf{a}_q^{(-,n)}) d_{qr} \right| = \left(\prod_{l=1}^{L} a_{ls}^{(+,n)} - \prod_{l=1}^{L} a_{ls}^{(-,n)} \right) d_{sr}^L = 0, \tag{92}$$

where d_{sr} denotes the s-th entry of $\mathbf{d}_r = \mathbf{S}^T\mathbf{w}_r$. W.l.o.g. we assume that $\mathbf{d}_r = \mathbf{e}_r^{(R)} \cdot \alpha_r$, where $\alpha_r \in \mathbb{C}$. From $\mathbf{X}\mathbf{w}_r = \mathbf{A}\mathbf{S}^T\mathbf{w}_r = \mathbf{A}\mathbf{d}_r = \mathbf{a}_r \cdot \alpha_r$ we conclude that

$$a_{lr}^{(+,n)} = \mathbf{e}_{p_{l,n}}^{(I)T}\mathbf{X}\mathbf{w}_r \quad \text{and} \quad a_{lr}^{(-,n)} = \mathbf{e}_{s_{l,n}}^{(I)T}\mathbf{X}\mathbf{w}_r. \tag{93}$$

Plugging (93) into (23) yields

$$a_{1r}^{(+,n)} \cdots a_{Lr}^{(+,n)} - a_{1r}^{(-,n)} \cdots a_{Lr}^{(-,n)} = 0 \Leftrightarrow \mathbf{p}_L^{(n)T} \cdot \big(\underbrace{\mathbf{w}_r \otimes \cdots \otimes \mathbf{w}_r}_{L \text{ times}} \big) = 0, \tag{94}$$

where $\mathbf{p}_L^{(n)} := ((\mathbf{X}^T\mathbf{e}_{p_{1,n}}^{(I)}) \otimes \cdots \otimes (\mathbf{X}^T\mathbf{e}_{p_{L,n}}^{(I)})) - ((\mathbf{X}^T\mathbf{e}_{s_{1,n}}^{(I)}) \otimes \cdots \otimes (\mathbf{X}^T\mathbf{e}_{s_{L,n}}^{(I)})) \in \mathbb{C}^{R^L}$. Stacking yields

$$\mathbf{P}^{(N,L)} \cdot (\mathbf{w}_r \otimes \cdots \otimes \mathbf{w}_r) = \mathbf{0}\,, \tag{95}$$

where $\mathbf{P}^{(N,L)} = [\mathbf{p}_L^{(1)}, \ldots, \mathbf{p}_L^{(N)}]^T \in \mathbb{C}^{N \times R^L}$. Since $\mathbf{P}^{(N,L)}(\mathbf{W} \odot \cdots \odot \mathbf{W}) = \mathbf{0}$, we know that there exist at least R linearly independent vectors $\{\mathbf{w}_r \otimes \cdots \otimes \mathbf{w}_r\}$, each with property $\mathbf{w}_r \otimes \cdots \otimes \mathbf{w}_r \in \ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}$, and each built from a column of $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_R] = \mathbf{S}^{-T}\mathbf{\Pi}\mathbf{\Delta}$. Hence, if the dimension of $\ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}$ is R (this will be justified below) and the columns of $\mathbf{Q}$ form a basis for $\ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}$, then there exists a nonsingular change-of-basis matrix $\mathbf{F} \in \mathbb{C}^{R \times R}$ such that

$$\mathbf{Q} = (\mathbf{W} \odot \cdots \odot \mathbf{W})\mathbf{F}^T. \tag{96}$$

It is clear that (96) corresponds to a matrix representation of a CPD of the form (79). This again implies the uniqueness of $\mathbf{A}$ and $\mathbf{S}$, as already explained in Section 4.1.2. We will now argue that if condition (62) in Theorem 3.4 is satisfied, then the subspace $\ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}$ is indeed R-dimensional. In Appendix D we explain that $\ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}$ is R-dimensional if and only if $\ker(\mathbf{R}_{\text{MF}}^{(N,L)})$ is R-dimensional. From the identity (90) we know that if $\mathbf{G}_{\text{MF}}^{(N,L)}$ has full column rank, then $\ker(\mathbf{R}_{\text{MF}}^{(N,L)})$ is R-dimensional, which demonstrates our claim.

Theorem 4.3 below summarizes the above uniqueness result for monomial factorizations, which is based on a constructive interpretation of Theorem 3.4.

Theorem 4.3. *Consider the monomial factorization of $\boldsymbol{X}$ in* (2) *with N monomial relations of the form* (3)*. If*

$$\begin{cases} \boldsymbol{S} \text{ has full column rank,} \\ ker(\boldsymbol{P}^{(N,L)}) \cap \pi_S^{(L)} \text{ is an } R\text{-dimensional subspace,} \end{cases} \tag{97}$$

then $\boldsymbol{A}$ and $\boldsymbol{S}$ in (2) *are unique, $\boldsymbol{A}$ has full column rank, and the monomial factorization of $\boldsymbol{X}$ can be reduced to the CPD of $\mathcal{Q}$ in* (79)*. The latter decomposition can be computed algebraically via an EVD.*

We stress that, although Theorems 4.2 and 4.3 were derived in seemingly different ways, they are actually equivalent. Again, by replacing $\mathbf{R}^{(N,L)}$ with $\mathbf{P}^{(N,L)}$ in Step 1 in Algorithm 1, the algorithm also summarizes the constructive interpretation of Theorem 3.4. As a final remark, we mention that the discussed null space approach can be understood as an extension of the ACMA method [7] for blind separation of constant modulus signals to monomial factorizations.

4.4. Application: Binary Matrix Factorization (BMF)

Consider the BMF (5). It is clear that not every BMF is unique (up to the intrinsic column permutation ambiguity). For example, if $\mathbf{X} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right][\mathbf{s}_1, \mathbf{s}_2]^T$, then $\mathbf{X} = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right][\mathbf{s}_1, \mathbf{s}_2 - \mathbf{s}_1]^T$ yields an alternative BMF. Below we explain how the presented framework can be used to obtain identifiability conditions and algebraic algorithms for BMFs of the form (5).

4.4.1. Nonhomogeneous formulation

Let us first explain how Theorem 4.3 can be adapted to solve the BMF problem using a variant of the ACMA method. The property $a_{i,r} \in \{0,1\}$ can be expressed via the nonhomogeneous (polynomial) relation

$$a_{i,r}(a_{i,r}-1) = a_{i,r}a_{i,r} - a_{i,r} = 0. \tag{98}$$

Following the procedure in Section 4.3, we obtain

$$a_{i,r}(a_{i,r}-1) = a_{i,r}a_{i,r} - a_{i,r} = 0 \Leftrightarrow \left[\mathbf{p}_1^{(i)}, \mathbf{p}_2^{(i)}\right] \begin{bmatrix} \mathbf{w}_r \\ \mathbf{w}_r \otimes \mathbf{w}_r \end{bmatrix} = 0, \quad 1 \leq i \leq I, \tag{99}$$

where $\mathbf{p}_1^{(i)} = \mathbf{e}_i^{(I)T}\mathbf{X}$ and $\mathbf{p}_2^{(i)} = (\mathbf{e}_i^{(I)T}\mathbf{X}) \otimes (\mathbf{e}_i^{(I)T}\mathbf{X})$. Stacking yields

$$\mathbf{P}_{\text{BMF}}^{(I)} \cdot \begin{bmatrix} \mathbf{w}_r \\ \mathbf{w}_r \otimes \mathbf{w}_r \end{bmatrix} = \mathbf{0}, \quad \text{where } \mathbf{P}_{\text{BMF}}^{(I)} = \begin{bmatrix} \mathbf{p}_1^{(1)} & \mathbf{p}_2^{(1)} \\ \vdots & \vdots \\ \mathbf{p}_1^{(I)} & \mathbf{p}_2^{(I)} \end{bmatrix} \in \mathbb{C}^{I \times (R+R^2)}. \tag{100}$$

We know that there exist at least R linearly independent vectors $\{\left[\begin{smallmatrix} \mathbf{w}_r \\ \mathbf{w}_r \otimes \mathbf{w}_r \end{smallmatrix}\right]\}$ in the subspace $\ker(\mathbf{P}_{\text{BMF}}^{(I)}) \cap (\mathbb{C}^R \times \pi_S^{(2)})$. Consequently, if the dimension of the subspace $\ker(\mathbf{P}_{\text{BMF}}^{(I)}) \cap (\mathbb{C}^R \times \pi_S^{(2)})$ is minimal (i.e., R), then the problem of finding $\{\mathbf{w}_r\}$ from (100) can be reduced to a CPD problem. In more detail, let the columns of $\mathbf{M} \in \mathbb{C}^{(R+R^2)\times R}$ denote a basis for the subspace $\ker(\mathbf{P}_{\text{BMF}}^{(I)}) \cap (\mathbb{C}^R \times \pi_S^{(2)})$. Partition $\mathbf{M}$ as follows $\mathbf{M} = \left[\begin{smallmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{smallmatrix}\right]$ in which $\mathbf{M}_1 \in \mathbb{C}^{R\times R}$ and $\mathbf{M}_2 \in \mathbb{C}^{R^2\times R}$. Then there exists a nonsingular matrix $\mathbf{F} \in \mathbb{C}^{R\times R}$ such that we obtain the coupled factorization $\mathbf{W}\mathbf{F}^T = \mathbf{M}_1$ and $(\mathbf{W} \odot \mathbf{W})\mathbf{F}^T = \mathbf{M}_2$. The latter factorization corresponds to a matrix representation of a CPD $\mathcal{M}_2 = \sum_{r=1}^R \mathbf{w}_r \circ \mathbf{w}_r \circ \mathbf{f}_r \in \mathbb{C}^{R\times R\times R}$ that can be computed via an EVD. This also implies the uniqueness of $\mathbf{A} = \mathbf{X}\mathbf{W}$. (By a similar reasoning Theorem 4.2 and its associated algorithm can be adapted to nonhomogeneous relations, such as (99).)

4.4.2. Homogeneous formulation

Even though Theorems 4.2 and 4.3 and their associated algorithms can be adapted to the nonhomogeneous case, the basic coupled BTD approach discussed in Section 4.1 cannot be used to solve the BMF problem via (99). Observe that $\mathbf{X} = \mathbf{A}\mathbf{S}^T = \mathbf{A}\mathbf{D} \cdot \mathbf{D}^{-1}\mathbf{S}^T$ for any nonsingular diagonal matrix $\mathbf{D} \in \mathbb{C}^{R\times R}$. This property implies that in the context of BMF, the binary constraint $\mathbf{A} \in \{0,1\}^{I\times R}$ can be relaxed to a binary zero-constant constraint $\mathbf{A} \in \{0,\alpha\}^{I\times R}$, where $\alpha \in \mathbb{C}$. Thus, as an alternative to (99), we propose the following homogeneous (monomial) formulation of the binary constraint of $a_{ir} \in \{0,1\}$:

$$a_{i_1,r}a_{i_1,r}a_{i_2,r} - a_{i_1,r}a_{i_2,r}a_{i_2,r} = 0, \quad 1 \leq i_1 < i_2 \leq I. \tag{101}$$

From (101) it is clear that the BMF (5) can also be interpreted as a basic monomial factorization of the form (2) with $N = C_I^2$, $L = 3$, $\mathbf{a}_r^{(+,n)} = [a_{i_1,r}, a_{i_1,r}, a_{i_2,r}]^T$ and $\mathbf{a}_r^{(-,n)} = [a_{i_1,r}, a_{i_2,r}, a_{i_2,r}]^T$ and where the superscript $'n'$ is associated with the pair (i_1, i_2). Compared to (99), an interesting property of the homogeneous formulation

(101) is that it can lead to relaxed identifiability conditions. As an example, consider the case where $\mathbf{S} \in \mathbb{C}^{K\times 6}$ has full column rank with $R = 6$ and

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}^T \in \mathbb{C}^{13\times 6}$$

with $I = 13$. It can be verified that condition (97) in Theorem 4.3 is satisfied when $\mathbf{P}^{(N,L)}$ is built from the C_I^2 equations associated with (101). In contrast, the nonhomogeneous approach based on (98) together with Theorem 4.3 cannot guarantee uniqueness for this BMF.

4.4.3. Numerical experiments

Let us end the section with an illustrative example. Consider (5) with $K = 50$, $R = 4$ and varying I. The goal is to estimate $\mathbf{S}$ from $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where $\mathbf{N}$ is an unstructured perturbation matrix. In each trial of the Monte Carlo experiment, the entries of $\mathbf{S}$ and $\mathbf{N}$ are randomly drawn from a Gaussian distribution with zero mean and unit variance while the entries of $\mathbf{A}$ are randomly drawn from a Bernoulli distribution in which the random variable takes the values 0 and 1 with equal probability. The following Signal-to-Noise Ratio (SNR) measure will be used: $\text{SNR} = 10\log(\|\mathbf{X}\|_F^2/\|\mathbf{N}\|_F^2)$. As a performance measure we use the distance between $\mathbf{S}$ and its estimate, $\widehat{\mathbf{S}}$. The distance is measured according to the criterion $\text{P}(\mathbf{S}) = \min_{\mathbf{\Pi\Lambda}} \|\mathbf{S} - \widehat{\mathbf{S}}\mathbf{\Pi\Lambda}\|_F / \|\mathbf{S}\|_F$, where $\mathbf{\Pi}$ and $\mathbf{\Lambda}$ denote a permutation matrix and a diagonal matrix, respectively. We compare the two algebraic methods associated with Theorem 4.3 when the nonhomogeneous (99) and homogeneous (101) approaches are used. The mean $\text{P}(\mathbf{S})$ over 100 Monte Carlo runs in which $I = 20$, $I = 30$ and $I = 40$ are shown in Figure 1. We observe that for the case $I = 20$, the homogeneous method performed better than the nonhomogeneous method. For the case $I = 40$, the homogeneous and nonhomogeneous methods performed about the same.

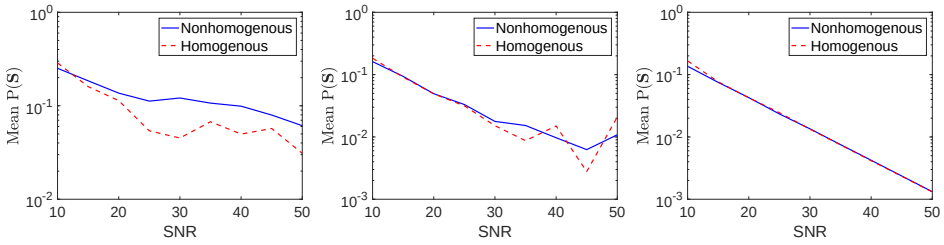


Figure 1: Mean $\text{P}(\mathbf{S})$ values over 100 Monte Carlo runs when (left) $I = 20$, (middle) $I = 30$ and (right) $I = 40$.

5. Conclusion

In this paper we studied bilinear factorizations of which one of the factor matrices is subject to monomial constraints and the other has full column rank. We have explained that such monomial factorizations generalize the CPD of higher-order tensors

in which at least one factor matrix has full column rank, and we have extended the framework accordingly. More precisely, we first presented a link between coupled BTD and monomial factorizations that allowed us to relax the monomial constraint into a low-rank constraint. By taking into account the specific linear structure that captures the monomial constraint, we have even reduced the monomial factorization to a CPD. We demonstrated that BMFs can be interpreted as monomial factorizations. This led to an algorithm and relatively easy-to-check uniqueness conditions for BMF. Finally, the framework of monomial factorizations enabled us to extend the CPD model to the decomposition of a tensor into a sum of binary weighted rank-one terms. This allows one to handle weighting structures that change over different slices.

Acknowledgements

The research of Mikael Sørensen and Nicholaos D. Sidiropulos were supported by NSF ECCS-1608961, NSF ECCS-1807660, and IIS-1704074.

The research of Lieven De Lathauwer was supported by: (1) Research Council KU Leuven: C1 project C16/15/059-nD, (2) F.W.O.: the Fund for Scientific Research–Flanders (Belgium), EOS Project no 30468160 (SeLMA), (3) 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.

Appendix A. Proof of Theorem 3.1

PROOF. **Sufficiency of condition** (53)**.** Let the pair $(\widehat{\mathbf{A}}_{\text{tot}}, \widehat{\mathbf{S}})$ be an alternative decomposition of (31) with $\widehat{R} \leq R$ terms so that

$$\mathbf{A}_{\text{tot}}\mathbf{S}^T = \widehat{\mathbf{A}}_{\text{tot}}\widehat{\mathbf{S}}^T. \tag{A.1}$$

We first establish uniqueness of $\mathbf{S}$, i.e., we provide a condition that ensures that $\mathbf{S} = \widehat{\mathbf{S}}\mathbf{\Pi}\Delta_{\mathbf{S}}$, where $\mathbf{\Pi}$ is an $(R \times R)$ column permutation matrix and $\Delta_{\mathbf{S}}$ is an $(R \times R)$ nonsingular diagonal matrix. In short, Kruskal's permutation lemma [31][3] ensures the uniqueness of $\mathbf{S}$ if $\omega(\mathbf{S}^T\mathbf{z}) \leq \omega(\widehat{\mathbf{S}}^T\mathbf{z})$ for every vector $\mathbf{z} \in \mathbb{C}^K$ such that $\omega(\widehat{\mathbf{S}}^T\mathbf{z}) \leq 1$. We will explain that condition (53) implies that $\mathbf{A}_{\text{tot}}$ has full column rank. For the moment we just assume that it is a fact. This fact together with the assumption that $\mathbf{S}$ has full column rank implies that $\widehat{\mathbf{S}}$ must also have full column rank (recall that $\widehat{R} \leq R \leq K$) and that $\widehat{R} = R$. Denote $\mathbf{d} = \mathbf{S}^T\mathbf{z}$ and $\widehat{\mathbf{d}} = \widehat{\mathbf{S}}^T\mathbf{z}$. Kruskal's permutation lemma now guarantees uniqueness of $\mathbf{S}$ if $\omega(\mathbf{d}) \leq \omega(\widehat{\mathbf{d}})$ for every $\omega(\widehat{\mathbf{d}}) \leq R - r_{\widehat{\mathbf{S}}} + 1 = 1$, where $r_{\widehat{\mathbf{S}}}$

[3] Kruskal's permutation lemma is not limited to the case where $\mathbf{S}$ has full column rank; see for instance [28]. In general, consider two matrices $\mathbf{S} \in \mathbb{C}^{K\times R}$ and $\widehat{\mathbf{S}} \in \mathbb{C}^{K\times \widehat{R}}$ with no zero columns and $\widehat{R} \leq R$. Let $r_{\widehat{\mathbf{S}}}$ denote the rank of $\widehat{\mathbf{S}}$. Kruskal's permutation lemma states that if for every $\mathbf{z} \in \mathbb{C}^K$, we have that $\omega(\widehat{\mathbf{S}}^T\mathbf{z}) \leq R - r_{\widehat{\mathbf{S}}} + 1 \Rightarrow \omega(\mathbf{S}^T\mathbf{z}) \leq \omega(\widehat{\mathbf{S}}^T\mathbf{z})$, then $\widehat{R} = R$ and $\mathbf{S} = \widehat{\mathbf{S}}\mathbf{\Pi}\Delta_{\mathbf{S}}$, where $\mathbf{\Pi}$ is an $(R \times R)$ column permutation matrix and $\Delta_{\mathbf{S}}$ is an $(R \times R)$ nonsingular diagonal matrix.

denotes the rank of $\widehat{\mathbf{S}}$. Thus, we only have to verify that this condition holds for the two cases $\omega(\widehat{\mathbf{d}}) = 0$ and $\omega(\widehat{\mathbf{d}}) = 1$.

Case $\omega(\widehat{\mathbf{d}}) = 0$. Let us first consider the case $\omega(\widehat{\mathbf{d}}) = 0 \Leftrightarrow \widehat{\mathbf{S}}^T\mathbf{z} = \mathbf{0}$. As mentioned before, it will be explained that condition (53) implies that $\mathbf{A}_{\text{tot}}$ has full column rank. Hence, we know from (A.1) that $\mathbf{A}_{\text{tot}}\mathbf{S}^T\mathbf{z} = \widehat{\mathbf{A}}_{\text{tot}}\widehat{\mathbf{S}}^T\mathbf{z} = \mathbf{0} \Leftrightarrow \mathbf{S}^T\mathbf{z} = \mathbf{0}$, where we took into account that $\widehat{\mathbf{d}} = \widehat{\mathbf{S}}^T\mathbf{z} = \mathbf{0}$. In other words, we must have that $\mathbf{d} = \mathbf{S}^T\mathbf{z} = \mathbf{0}$ for all $\mathbf{z} \in \mathbb{C}^K$ such that $\omega(\widehat{\mathbf{d}}) = 0$. We conclude that the inequality condition $0 = \omega(\mathbf{S}^T\mathbf{z}) \leq \omega(\widehat{\mathbf{S}}^T\mathbf{z}) = 0$ in Kruskal's permutation lemma is satisfied.

Case $\omega(\widehat{\mathbf{d}}) = 1$. Consider again a vector $\mathbf{z} \in \mathbb{C}^K$ so that from (A.1) we obtain

$$\mathbf{A}_{\text{tot}}\mathbf{S}^T\mathbf{z} = \widehat{\mathbf{A}}_{\text{tot}}\widehat{\mathbf{S}}^T\mathbf{z}, \tag{A.2}$$

where $\mathbf{A}_{\text{tot}}$ and $\widehat{\mathbf{A}}_{\text{tot}}$ are of the form (32). Recall that $\mathbf{d} = \mathbf{S}^T\mathbf{z}$ and $\widehat{\mathbf{d}} = \widehat{\mathbf{S}}^T\mathbf{z}$. We assume that the vector $\mathbf{z} \in \mathbb{C}^K$ is chosen so that $\omega(\widehat{\mathbf{d}}) = \omega(\widehat{\mathbf{C}}^T\mathbf{z}) = 1$. Reshaping the vectors $\mathbf{A}^{(n)}\mathbf{S}^T\mathbf{z} = \mathbf{A}^{(n)}\mathbf{d}$ and $\widehat{\mathbf{A}}^{(n)}\widehat{\mathbf{S}}^T\mathbf{z} = \widehat{\mathbf{A}}^{(n)}\widehat{\mathbf{d}}$ in (A.2) into $(L \times L)$ matrices yields

$$\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r = \sum_{r=1}^{R} \mathbf{A}_L(\widehat{\mathbf{a}}_r^{(+,n)}, \widehat{\mathbf{a}}_r^{(-,n)})\widehat{d}_r\,, \quad n \in \{1, \ldots, N\}. \tag{A.3}$$

Since $\omega(\widehat{\mathbf{d}}) = 1$ and $\mathbf{A}_L(\widehat{\mathbf{a}}_r^{(+,n)}, \widehat{\mathbf{a}}_r^{(-,n)})$ is a rank-$(L-1)$ matrix, we know that

$$\left|\sum_{r=1}^{R} \mathbf{A}_L(\widehat{\mathbf{a}}_r^{(+,n)}, \widehat{\mathbf{a}}_r^{(-,n)})\widehat{d}_r\right| = \left|\mathbf{A}_L(\widehat{\mathbf{a}}_r^{(+,n)}, \widehat{\mathbf{a}}_r^{(-,n)})\right| \widehat{d}_r^L = 0, \quad n \in \{1, \ldots, N\}.$$

Consequently, the determinant of the left-hand side of (A.3) must vanish as well:

$$\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r\right| = 0, \quad n \in \{1, \ldots, N\}. \tag{A.4}$$

Expansion of (A.4) in terms of $d_1, \ldots, d_R$ now yields

$$\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r\right| = \prod_{l=1}^{L}\left(\sum_{r=1}^{R} a_{lr}^{(+,n)}d_r\right) - \prod_{l=1}^{L}\left(\sum_{r=1}^{R} a_{lr}^{(-,n)}d_r\right) \tag{A.5}$$

$$= \left|\sum_{r=1}^{R} D_n(\mathbf{A}^{(+,r)})d_r\right| - \left|\sum_{r=1}^{R} D_n(\mathbf{A}^{(-,r)})d_r\right| \tag{A.6}$$

$$= \sum_{(l_1,\ldots,l_R)\in[L]_R} \left[\mathcal{D}\left(\underbrace{D_n(\mathbf{A}^{(+,1)}), \ldots, D_n(\mathbf{A}^{(+,1)})}_{l_1 \text{ times}}, \ldots, \underbrace{D_n(\mathbf{A}^{(+,R)}), \ldots, D_n(\mathbf{A}^{(+,R)})}_{l_R \text{ times}}\right)\right. \tag{A.7}$$

$$\left. - \mathcal{D}\left(\underbrace{D_n(\mathbf{A}^{(-,1)}), \ldots, D_n(\mathbf{A}^{(-,1)})}_{l_1 \text{ times}}, \ldots, \underbrace{D_n(\mathbf{A}^{(-,R)}), \ldots, D_n(\mathbf{A}^{(-,R)})}_{l_R \text{ times}}\right)\right] d_1^{l_1} \cdots d_R^{l_R}$$

$$= \sum_{(l_1,\ldots,l_R)\in[L]_R} \frac{1}{l_1! \cdots l_R!}\left(\left|\mathbf{A}^{(+,n)}_{(l_1,\ldots,l_R)}\right|^{+} - \left|\mathbf{A}^{(-,n)}_{(l_1,\ldots,l_R)}\right|^{+}\right) d_1^{l_1} \cdots d_R^{l_R} = 0, \tag{A.8}$$

where (A.6) follows from the definition (43), (A.7) follows from the definition of the mixed discriminant (see equations (35) and (37)), and (A.8) follows from properties (40)–(41). Due to property (39), we also know that if

$$(l_1, \ldots, l_R) \in \Omega := \{(L, 0, 0 \ldots, 0), (0, L, 0 \ldots, 0), \ldots, (0, 0 \ldots, 0, L)\},$$

then $\overset{+}{|}\mathbf{A}^{(+,n)}_{(l_1,\ldots,l_R)}\overset{+}{|} - \overset{+}{|}\mathbf{A}^{(-,n)}_{(l_1,\ldots,l_R)}\overset{+}{|} = \prod_{l=1}^{L} a_{lr}^{(+,n)} - \prod_{l=1}^{L} a_{lr}^{(-,n)} = 0$. Consequently, (A.8) simplifies to

$$\mathbf{g}^{(n,L)}_{\mathrm{MF}} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = 0, \quad n \in \{1, \ldots, N\}, \tag{A.9}$$

where $\mathbf{g}^{(n,L)}_{\mathrm{MF}}$ and $\mathbf{f}^{(L)}(\mathbf{d})$ are given by (49) and (52), respectively. Stacking yields

$$\mathbf{G}^{(N,L)}_{\mathrm{MF}} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}\,, \tag{A.10}$$

where $\mathbf{G}^{(N,L)}_{\mathrm{MF}}$ is given by (51). Condition (53) now implies that the inequality condition $\omega(\mathbf{d}) = \omega(\mathbf{S}^T\mathbf{z}) \leq \omega(\widehat{\mathbf{S}}^T\mathbf{z}) = \omega(\widehat{\mathbf{d}}) = 1$ in Kruskal's permutation lemma is satisfied. We conclude that condition (53) is sufficient for the uniqueness of $\mathbf{S}$. This fact together with the full column rank assumption of $\mathbf{S}$ implies the uniqueness of $\mathbf{A}$ and therefore also the uniqueness of the monomial factorization of $\mathbf{X}$.

Condition (53) implies that $\mathbf{A}_{\mathbf{tot}}$ has full column rank. Let us now explain that condition (53) implies that $\mathbf{A}_{\mathrm{tot}}$ has full column rank. From relations (A.4) and (A.10) it is clear that there cannot exist a vector $\mathbf{d} \in \mathbb{C}^R$ with properties $\omega(\mathbf{d}) \geq 2$ and $\mathbf{A}_{\mathrm{tot}}\mathbf{d} = \mathbf{0}$ (i.e., $\mathbf{A}_{\mathrm{tot}}$ does not have full column rank), since it would imply that $\mathbf{G}^{(N,L)}_{\mathrm{MF}} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}$ and thereby contradict condition (53). In more details, the identity $\mathbf{G}^{(N,L)}_{\mathrm{MF}} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}$ means that all the determinants of the form $\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r\right|$ vanish. Since the determinant $\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r\right|$ built from $[\mathbf{a}_{\mathrm{tot},1}d_1, \ldots, \mathbf{a}_{\mathrm{tot},R}d_R]$ vanishes for every $\mathbf{d}$ with property $\mathbf{A}_{\mathrm{tot}}\mathbf{d} = \mathbf{0}$,[4] there always exists a vector $\mathbf{d}$ with properties $\omega(\mathbf{d}) \geq 2$, $\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r = \mathbf{0}$ and $\mathbf{G}^{(N,L)}_{\mathrm{MF}} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = 0$ when $\mathbf{A}_{\mathrm{tot}}$ does not have full column rank. This explains that if condition (53) is satisfied, then $\mathbf{A}_{\mathrm{tot}}$ must have full column rank.

Necessity of condition (53). Assume now on the contrary that condition (53) is not satisfied, i.e., there exists a vector $\mathbf{d} \in \mathbb{C}^R$ with properties $\omega(\mathbf{d}) \geq 2$ and $\mathbf{G}^{(N,L)}_{\mathrm{MF}} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}$. Recall that the identity $\mathbf{G}^{(N,L)}_{\mathrm{MF}} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}$ means that all the determinants of the form $|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r|$ vanish (cf. Eq. (A.4)) and consequently we know that for every $n \in \{1, \ldots, N\}$, there exist vectors $\widehat{\mathbf{a}}^{(+,n)} \in \mathbb{C}^L$ and $\widehat{\mathbf{a}}^{(-,n)} \in \mathbb{C}^L$ such that the vector $\mathrm{vec}\big(\mathbf{A}_L(\widehat{\mathbf{a}}^{(+,n)}, \widehat{\mathbf{a}}^{(-,n)})\big) = \sum_{r=1}^{R} \mathrm{vec}(\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}))d_r = \mathbf{A}^{(n)}\mathbf{d}$ has the property

$$|\mathbf{A}_L(\widehat{\mathbf{a}}^{(+,n)}, \widehat{\mathbf{a}}^{(-,n)})| = \left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r\right| = 0 \text{ for every } n \in \{1, \ldots, N\}.$$

[4] Note that $\mathbf{A}_{\mathrm{tot}}\mathbf{d} = 0$ means that $\sum_{r=1}^{R} \mathrm{vec}(\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r) = \mathbf{0}$, which in turn means that $\sum_{r=1}^{R} \mathbf{a}_r^{(+,n)}d_r = \mathbf{0}$ and $\sum_{r=1}^{R} \mathbf{a}_r^{(-,n)}d_r = \mathbf{0}$. It is now clear that $\mathbf{A}_{\mathrm{tot}}\mathbf{d} = 0$ implies that $\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r\right| = 0$.

The position of the elements of $\widehat{\mathbf{a}}^{(+,n)}$ and $\widehat{\mathbf{a}}^{(-,n)}$ in $\widehat{\mathbf{a}} = \mathbf{A}\mathbf{d}$ correspond to the position of the elements of $\mathbf{a}_r^{(+,n)}$ and $\mathbf{a}_r^{(-,n)}$ in $\mathbf{a}_r$. In other words, $\widehat{\mathbf{a}} = \mathbf{A}\mathbf{d} = \mathbf{a}_1 d_1 + \cdots + \mathbf{a}_R d_R$ is a vector that satisfies the monomial relation (23) for every $n \in \{1, \ldots, N\}$. Since $\omega(\mathbf{d}) \geq 2$, $\widehat{\mathbf{a}}$ can only be proportional to a column of $\mathbf{A}$ if the columns of $\mathbf{A}$ are linearly dependent. The latter would imply nonuniqueness of the monomial factorization of $\mathbf{X}$. Assume now that $\widehat{\mathbf{a}}$ is not proportional to any column of $\mathbf{A}$. Since $\omega(\mathbf{d}) \geq 2$ we can w.l.o.g. assume that $d_1 \neq 0$ so that $\mathbf{a}_1 = \frac{1}{d_1}\widehat{\mathbf{a}} - \frac{d_2}{d_1}\mathbf{a}_2 - \cdots - \frac{d_R}{d_1}\mathbf{a}_R$ and consequently the latter factorization in

$$\mathbf{X} = \mathbf{A}\mathbf{S}^T = [\widehat{\mathbf{a}}, \mathbf{a}_2, \ldots, \mathbf{a}_R] \cdot \left(\begin{bmatrix} \frac{1}{d_1} & 0 & \cdots & 0 \\ -\frac{d_2}{d_1} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ -\frac{d_R}{d_1} & 0 & \cdots & 1 \end{bmatrix} \mathbf{S}^T \right) \tag{A.11}$$

yields an alternative monomial factorization of $\mathbf{X}$. We conclude that condition (53) is also necessary for the uniqueness of $\mathbf{A}$ and $\mathbf{S}$.

Condition (53) implies that A has full column rank. We have just established that condition (53) implies that $\mathbf{S}$ is unique, which in turn implies that $\mathbf{A}$ has full column rank. Indeed, if $\mathbf{A}$ does not have full column rank, then $\dim(\ker(\mathbf{A})) \geq 1$ and $\mathbf{X} = \mathbf{A}\mathbf{S}^T = \mathbf{A}(\mathbf{S}^T + \mathbf{n}\mathbf{c}^T)$ for any $\mathbf{n} \in \ker(\mathbf{A})$ and $\mathbf{c} \in \mathbb{C}^K$. This means that there exists an alternative $R-1$ term monomial factorization of $\mathbf{X}$. For example, since $\dim(\ker(\mathbf{A})) \geq 1$ we know that there exists an $\mathbf{n} \in \ker(\mathbf{A})$ with $n_i \neq 0$. Set $\mathbf{c} = -\frac{1}{n_i}\mathbf{s}_i$, then $\mathbf{X} = \mathbf{A}\mathbf{S}^T = \mathbf{A}(\mathbf{S}^T + \mathbf{n}\mathbf{c}^T) = [\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_R] \cdot [\mathbf{s}_1, \ldots, \mathbf{s}_{i-1}, \mathbf{s}_{i+1}, \ldots, \mathbf{s}_R]^T$, which contradicts the uniqueness property of $\mathbf{S}$.

Appendix B. Proof of Theorem 2.1

Proof. The proof is analogous to the proof of Theorem 3.2. In short, the difference is that we will now make use of the low-rank factorization $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) = \mathbf{N}^{(r,n)}\mathbf{M}^{(r,n)T}$, where $\mathbf{N}^{(r,n)} \in \mathbf{C}^{L \times (L-1)}$ and $\mathbf{M}^{(r,n)} \in \mathbf{C}^{L \times (L-1)}$ are rank-$(L-1)$ matrices. The expansion of the determinant in (A.4) can now be written as

$$\left| \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r \right| = \left| \sum_{r=1}^{R} \mathbf{N}^{(r,n)}\mathbf{M}^{(r,n)T} d_r \right| = 0, \quad n \in \{1, \ldots, N\}, \tag{B.1}$$

or more compactly as

$$\left| \mathbf{N}^{(n)} \mathrm{Diag}\left(\mathbf{d}^{(\mathrm{ext})}\right) \mathbf{M}^{(n)T} \right| = 0, \quad n \in \{1, \ldots, N\}, \tag{B.2}$$

where $\mathbf{M}^{(n)}$ and $\mathbf{N}^{(n)}$ are of the form (16) and (17), respectively, and $\mathrm{Diag}\left(\mathbf{d}^{(\mathrm{ext})}\right)$ is a diagonal matrix that holds the column vector

$$\mathbf{d}^{(\mathrm{ext})} = \left[d_1 \mathbf{1}_{L-1}^T, \ldots, d_R \mathbf{1}_{L-1}^T\right]^T \in \mathbb{C}^{(L-1)R} \tag{B.3}$$

on its diagonal. Relation (B.2) can be expressed in terms of compound matrices as follows

$$C_L\left(\mathbf{N}^{(n)}\right) C_L\left(\mathrm{Diag}\left(\mathbf{d}^{(\mathrm{ext})}\right)\right) C_L\left(\mathbf{M}^{(n)T}\right) = 0, \quad n \in \{1,\ldots,N\}. \tag{B.4}$$

Consider the set

$$S = \{(i_1, i_2, \ldots, i_L) \mid 1 \leq i_1 < i_2 < \cdots < i_L \leq R(L-1)\}, \tag{B.5}$$

in which the L-tuples are ordered lexicographically and indexed by $S(1), \ldots, S(C^L_{R(L-1)})$. Let

$$\mathbf{d}^{(\mathrm{ext},L)} = \left[d^{(\mathrm{ext})}_{S(1)}, d^{(\mathrm{ext})}_{S(2)}, \ldots, d^{(\mathrm{ext})}_{S(C^L_{R(L-1)})}\right]^T \in \mathbb{C}^{C^L_{R(L-1)}}, \tag{B.6}$$

where $d^{(\mathrm{ext})}_{S(j)} = d^{(\mathrm{ext})}_{i_1} d^{(\mathrm{ext})}_{i_2} \cdots d^{(\mathrm{ext})}_{i_L}$ with $S(j) = (i_1, i_2, \ldots, i_L)$. Relation (B.4) can now also be expressed as

$$\left(C_L\left(\mathbf{M}^{(n)}\right) \odot C_L\left(\mathbf{N}^{(n)}\right)\right) \mathbf{d}^{(\mathrm{ext},L)} = 0, \quad n \in \{1,\ldots,N\}. \tag{B.7}$$

We will now take into account that coefficient d_l is repeated $L-1$ times in $\mathbf{d}^{(\mathrm{ext})}$. Observe that

$$d^{(\mathrm{ext})}_{(i_1,i_2,\ldots,i_L)} = d^{(\mathrm{ext})}_{i_1} d^{(\mathrm{ext})}_{i_2} \cdots d^{(\mathrm{ext})}_{i_L} = d_{\lceil \frac{i_1}{L-1} \rceil} d_{\lceil \frac{i_2}{L-1} \rceil} \cdots d_{\lceil \frac{i_L}{L-1} \rceil} = d_{j_1} d_{j_2} \cdots d_{j_L}, \tag{B.8}$$

where $j_1 = \lceil \frac{i_1}{L-1} \rceil$, $j_2 = \lceil \frac{i_2}{L-1} \rceil$, ..., $j_L = \lceil \frac{i_L}{L-1} \rceil$. By also taking into account that $|\mathbf{M}^{(n,r)}| = |\mathbf{N}^{(n,r)}| = 0$ (which are associated with the term d^L_r), it is now clear that the matrix $\mathbf{P}_{\mathrm{BTD}} \in \mathbb{C}^{C^L_{R(L-1)} \times (C^L_{R+L-1} - R)}$, defined by (21), can be used to obtain the following compact version of (B.7):

$$\left(C_L\left(\mathbf{M}^{(n)}\right) \odot C_L\left(\mathbf{N}^{(n)}\right)\right) \mathbf{P}_{\mathrm{BTD}} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = 0, \quad n \in \{1,\ldots,N\}, \tag{B.9}$$

where $\mathbf{f}^{(L)}(\mathbf{d})$ is given by (52). Stacking yields

$$\mathbf{G}^{(N,L)}_{\mathrm{BTD}} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}, \tag{B.10}$$

where $\mathbf{G}^{(N,L)}_{\mathrm{BTD}}$ is of the form (20). By substituting $\mathbf{G}^{(N,L)}_{\mathrm{MF}}$ in (A.10) with $\mathbf{G}^{(N,L)}_{\mathrm{BTD}}$ we can now proceed as in the proof of Theorem 3.2.

Appendix C. Connection between Theorem 3.1 and Theorem 3.3

Let us now explain relation (59). Consider the right-hand side of (59). Observe that if the symmetry constraint on $\mathbf{d} \otimes \cdots \otimes \mathbf{d}$ in $\mathbf{\Psi}^{(N,L)}(\mathbf{d} \otimes \cdots \otimes \mathbf{d})$ is imposed, then

$$\boldsymbol{\psi}^{(n,L)}(\mathbf{d} \otimes \cdots \otimes \mathbf{d}) = \prod_{l=1}^{L}\left(\sum_{r=1}^{R} a^{(+,n)}_{lr} d_r\right) - \prod_{l=1}^{L}\left(\sum_{r=1}^{R} a^{(-,n)}_{lr} d_r\right) = \tag{C.1}$$

$$\sum_{(l_1,\ldots,l_R)\in[L]_R}\left[\sum_{\sigma\in S_{(l_1,\ldots,l_R)}} \prod_{l=1}^{L}\left(a^{(+,n)}_{l\sigma(l)} - a^{(-,n)}_{l\sigma(l)}\right)\right] d_1^{l_1} \cdots d_R^{l_R},$$

where $S_{(l_1,\ldots,l_R)} = \Big\{$set of all distinct permutations of the L-element tuple $(\underbrace{1,\ldots,1}_{l_1 \text{ times}},\ldots,\underbrace{R,\ldots,R}_{l_R \text{ times}})$ with property $l_1+\cdots+l_R = L\Big\}$. Note that for a given sequence $(l_1,\ldots,l_R)$ there are $\frac{L!}{l_1!\cdots l_R!}$ such permutations. In order to see that (C.1) corresponds to the n-th row of $\mathbf{G}_{\mathrm{MF}}^{(N,L)}$, observe that

$$\begin{aligned}\overset{+}{|}\mathbf{A}^{(+,n)}_{(l_1,\ldots,l_R)}\overset{+}{|} - \overset{+}{|}\mathbf{A}^{(-,n)}_{(l_1,\ldots,l_R)}\overset{+}{|} &= \sum_{\sigma\in S_L}\prod_{l=1}^{L}\left(\mathbf{A}^{(+,n)}_{(l_1,\ldots,l_R)} - \mathbf{A}^{(-,n)}_{(l_1,\ldots,l_R)}\right)_{l\sigma(l)} \\ &= l_1!l_2!\cdots l_R!\left[\sum_{\sigma\in S_{(l_1,\ldots,l_R)}}\left(\prod_{l=1}^{L} a^{(+,n)}_{l\sigma(l)} - a^{(-,n)}_{l\sigma(l)}\right)\right], \quad \text{(C.2)}\end{aligned}$$

where $\mathbf{A}^{(+,n)}_{(l_1,\ldots,l_R)}$ and $\mathbf{A}^{(-,n)}_{(l_1,\ldots,l_R)}$ are given by (45) and (46), respectively, and the column permutation invariance property of the permanents is taken into account. Combining (C.1) and (C.2) yields

$$\begin{aligned}&\prod_{l=1}^{L}\left(\sum_{r=1}^{R} a^{(+,n)}_{lr}d_r\right) - \prod_{l=1}^{L}\left(\sum_{r=1}^{R} a^{(-,n)}_{lr}d_r\right) = \\ &\sum_{(l_1,\ldots,l_R)\in[L]_R}\frac{1}{l_1!l_2!\cdots l_R!}\left(\overset{+}{|}\mathbf{A}^{(+,n)}_{(l_1,\ldots,l_R)}\overset{+}{|} - \overset{+}{|}\mathbf{A}^{(-,n)}_{(l_1,\ldots,l_R)}\overset{+}{|}\right)d_1^{l_1}\cdots d_R^{l_R}. \quad \text{(C.3)}\end{aligned}$$

Comparing (C.3) with (49), it is clear that the right-hand side of (59) indeed corresponds to its the left-hand side.

Appendix D. Connection between Theorems 4.2 and 4.3

Based on the connection between Theorems 3.2 and 3.4, we will now explain that condition (91) in Theorem 4.2 is equivalent to condition (97) in Theorem 4.3. Observe that a relation of the form (94) can be expressed as (compare with (C.1)):

$$\mathbf{p}_L^{(n)T}\cdot(\mathbf{w}\otimes\cdots\otimes\mathbf{w}) = \prod_{l=1}^{L}\left(\sum_{r=1}^{R} x^{(n)}_{p_l r}w_r\right) - \prod_{l=1}^{L}\left(\sum_{r=1}^{R} x^{(n)}_{s_l r}w_r\right). \quad \text{(D.1)}$$

By taking this symmetry property of $\mathbf{w}\otimes\cdots\otimes\mathbf{w}$ into account, relation (D.1) can be expressed more compactly as (compare with (C.3)):

$$\mathbf{p}_L^{(n)T}\cdot(\mathbf{w}\otimes\cdots\otimes\mathbf{w}) = \sum_{(l_1,\ldots,l_R)\in[L]_R}\frac{1}{l_1!\cdots l_R!}(\overset{+}{|}\mathbf{X}^{(+,n)}_{(l_1,\ldots,l_R)}\overset{+}{|} - \overset{+}{|}\mathbf{X}^{(-,n)}_{(l_1,\ldots,l_R)}\overset{+}{|})w_1^{l_1}\cdots w_R^{l_R}, \quad \text{(D.2)}$$

where $\mathbf{X}^{(+,n)}_{(l_1,\ldots,l_R)} \in \mathbb{C}^{L\times L}$ and $\mathbf{X}^{(-,n)}_{(l_1,\ldots,l_R)} \in \mathbb{C}^{L\times L}$ are given by (83) and (84), respectively. Stacking yields $\mathbf{R}_{\mathrm{MF}}^{(N,L)}$ given by (87). Consequently, $\mathbf{W}$ can also be obtained from $\ker(\mathbf{R}_{\mathrm{MF}}^{(N,L)})$, which is equivalent to obtaining it from $\ker(\mathbf{P}^{(N,L)})\cap\pi_S^{(L)}$.

Appendix E. Structure of $\mathcal{R}_L^{(\mathrm{dist})}(\mathbf{S}^{(\mathrm{ext})})$

Recall from (70) and (72) that

$$\mathcal{R}_L^{(\mathrm{dist})}(\mathbf{S}^{(\mathrm{ext})}) = L!\left[\mathrm{Sym}(\mathbf{s}_1 \otimes \mathbf{s}_1 \otimes \cdots \otimes \mathbf{s}_2), \ldots, \mathrm{Sym}(\mathbf{s}_{R-1} \otimes \mathbf{s}_{R-1} \otimes \cdots \otimes \mathbf{s}_R)\right],$$

where $\mathrm{Sym}(\mathbf{s}_{r_1} \otimes \mathbf{s}_{r_2} \otimes \cdots \otimes \mathbf{s}_{r_L}) = \frac{1}{L!}\sum_{\sigma \in S_L} \mathbf{s}_{r_{\sigma(1)}} \otimes \mathbf{s}_{r_{\sigma(2)}} \otimes \cdots \otimes \mathbf{s}_{r_{\sigma(L)}} \in \mathbb{C}^{R^L}$. In other words, the matrix $\mathcal{R}_L^{(\mathrm{dist})}(\mathbf{S}^{(\mathrm{ext})})$ is formed by the columns $\mathrm{Sym}(\mathbf{s}_{r_1} \otimes \mathbf{s}_{r_2} \otimes \cdots \otimes \mathbf{s}_{r_L})$, $1 \le r_1 \le \cdots \le r_L \le R$, $r_1 \ne r_L$. The vector $\mathrm{Sym}(\mathbf{s}_{r_1} \otimes \mathbf{s}_{r_2} \otimes \cdots \otimes \mathbf{s}_{r_L})$ in turn consists of the symmetric part of the rank-one tensor $\mathbf{s}_{r_1} \otimes \cdots \otimes \mathbf{s}_{r_L}$. Let the entries of $\mathrm{Sym}(\mathbf{s}_{r_1} \otimes \mathbf{s}_{r_2} \otimes \cdots \otimes \mathbf{s}_{r_L})$ be indexed by the L-tuple $(k_1, \ldots, k_L)$. Observe that

$$L! \cdot \mathrm{Sym}(\mathbf{s}_{r_1} \otimes \mathbf{s}_{r_2} \otimes \cdots \otimes \mathbf{s}_{r_L}) = \begin{bmatrix} \overset{+}{|}[\mathbf{s}_{r_1}^{(1,1,\ldots,1)}, \mathbf{s}_{r_2}^{(1,1,\ldots,1)}, \ldots, \mathbf{s}_{r_L}^{(1,1,\ldots,1)}]\overset{+}{|} \\ \overset{+}{|}[\mathbf{s}_{r_1}^{(1,1,\ldots,2)}, \mathbf{s}_{r_2}^{(1,1,\ldots,2)}, \ldots, \mathbf{s}_{r_L}^{(1,1,\ldots,2)}]\overset{+}{|} \\ \vdots \\ \overset{+}{|}[\mathbf{s}_{r_1}^{(R,R,\ldots,R)}, \mathbf{s}_{r_2}^{(R,R,\ldots,R)}, \ldots, \mathbf{s}_{r_L}^{(R,R,\ldots,R)}]\overset{+}{|} \end{bmatrix} \in \mathbb{C}^{R^L}, \tag{E.1}$$

where $\mathbf{s}_{r_l}^{(k_1,\ldots,k_L)} = [s_{k_1,r_l}, \ldots, s_{k_L,r_l}]^T \in \mathbb{C}^L$. Thus,

$$\overset{+}{|}[\mathbf{s}_{r_1}^{(k_1,\ldots,k_L)}, \mathbf{s}_{r_2}^{(k_1,\ldots,k_L)}, \ldots, \mathbf{s}_{r_L}^{(k_1,\ldots,k_L)}]\overset{+}{|} = \sum_{\sigma \in S_L} \prod_{l=1}^{L} s_{k_l,r_{\sigma(l)}} = \sum_{\sigma \in S_L} \prod_{l=1}^{L} s_{k_{\sigma(l)},r_l}. \tag{E.2}$$

We assume that lexicographical ordering is used, which means that if $\mathbf{e}_{(k_1,k_2,\ldots,k_L)}^{(R^L)} \in \mathbb{C}^{R^L}$ denotes the unit vector with unit entry at position $(k_1, k_2, \ldots, k_L)$, then

$$L! \cdot \mathrm{Sym}(\mathbf{s}_{r_1} \otimes \mathbf{s}_{r_2} \otimes \cdots \otimes \mathbf{s}_{r_L})^T \mathbf{e}_{(k_1,\ldots,k_L)}^{(R^L)} = \overset{+}{|}[\mathbf{s}_{r_1}^{(k_1,\ldots,k_L)}, \mathbf{s}_{r_2}^{(k_1,\ldots,k_L)}, \ldots, \mathbf{s}_{r_L}^{(k_1,\ldots,k_L)}]\overset{+}{|}. \tag{E.3}$$

Relation (E.3) will be used in Appendix F.

Appendix F. Derivation of identity (90)

Since $\mathbf{x}_k^{(+,n)} = \sum_{r=1}^{R} \mathbf{a}_r^{(+,n)} s_{kr} \in \mathbb{C}^L$, the entries of the left-hand side of (90) are given by

$$\begin{aligned} \overset{+}{|}[\mathbf{x}_{k_1}^{(+,n)}, \ldots, \mathbf{x}_{k_L}^{(+,n)}]\overset{+}{|} &= \sum_{\sigma \in S_L} \prod_{l=1}^{L} \left(\sum_{r=1}^{R} a_{l,r}^{(+,n)} s_{k_{\sigma(l)},r} \right) \\ &= \sum_{1 \le r_1,\ldots,r_L \le R} \left(\prod_{l=1}^{L} a_{l,r_l}^{(+,n)} \right) \left(\sum_{\sigma \in S_L} \prod_{l=1}^{L} s_{k_{\sigma(l)},r_l} \right) \\ &= \sum_{1 \le r_1,\ldots,r_L \le R} \left(\prod_{l=1}^{L} a_{l,r_l}^{(+,n)} \right) \overset{+}{|}[\mathbf{s}_{r_1}^{(k_1,\ldots,k_L)}, \ldots, \mathbf{s}_{r_L}^{(k_1,\ldots,k_L)}]\overset{+}{|}. \end{aligned} \tag{F.1}$$

Using (F.1) and the column permutation invariance property of the permanent we will now derive the identity (90). We first exploit the fact that

$$\overset{+}{|}[\mathbf{s}_{r_{\sigma(1)}}^{(k_1,\ldots,k_L)},\ldots,\mathbf{s}_{r_{\sigma(L)}}^{(k_1,\ldots,k_L)}]\overset{+}{|} = \overset{+}{|}[\mathbf{s}_{r_1}^{(k_1,\ldots,k_L)},\ldots,\mathbf{s}_{r_L}^{(k_1,\ldots,k_L)}]\overset{+}{|}, \quad \forall \sigma \in S_L. \tag{F.2}$$

Let $\Xi_{(l_1,\ldots,l_R)}$ denote the set of all distinct permutations of $\Big(\underbrace{1,\ldots,1}_{l_1},\ldots,\underbrace{R,\ldots,R}_{l_R}\Big)$ subject to $l_1+l_2+\cdots+l_R = L$ and $l_r \geq 0$. Note that the set $\Xi_{(l_1,\ldots,l_R)}$ contains $\frac{L!}{l_1!l_2!\cdots l_R!}$ elements. Note also that due to the permutation invariance property of the permanent, for any L-tuple $(r_1,\ldots,r_L) \in \Xi_{(l_1,\ldots,l_R)}$, we have that

$$\overset{+}{|}[\mathbf{s}_{r_1}^{(k_1,\ldots,k_L)},\ldots,\mathbf{s}_{r_L}^{(k_1,\ldots,k_L)}]\overset{+}{|} = \overset{+}{|}[\mathbf{1}_{l_1}^T \otimes \mathbf{s}_1^{(k_1,\ldots,k_L)},\ldots,\mathbf{1}_{l_R}^T \otimes \mathbf{s}_R^{(k_1,\ldots,k_L)}]\overset{+}{|}. \tag{F.3}$$

Using (F.2) and (F.3), relation (F.1) can also be written as

$$\sum_{1\leq r_1,\ldots,r_L\leq R}\left(\prod_{l=1}^{L} a_{l,r_l}^{(+,n)}\right)\overset{+}{|}[\mathbf{s}_{r_1}^{(k_1,\ldots,k_L)},\ldots,\mathbf{s}_{r_L}^{(k_1,\ldots,k_L)}]\overset{+}{|} = \tag{F.4}$$

$$\sum_{(l_1,\ldots,l_R)\in[L]_R}\ \sum_{(r_1,\ldots,r_L)\in\Xi_{(l_1,\ldots,l_R)}}\prod_{l=1}^{L} a_{l,r_l}^{(+,n)}\overset{+}{|}[\mathbf{1}_{l_1}^T \otimes \mathbf{s}_1^{(k_1,\ldots,k_L)},\ldots,\mathbf{1}_{l_R}^T \otimes \mathbf{s}_R^{(k_1,\ldots,k_L)}]\overset{+}{|}.$$

We will now find a compact expression for the part involving the entries $\{a_{l,r_l}^{(+,n)}\}$ in (F.4). Note that each L-tuple $(r_1,\ldots,r_L) \in \Xi_{(l_1,\ldots,l_R)}$ is associated with a product

$$a_{1,r_1}^{(+,n)} a_{2,r_2}^{(+,n)} \cdots a_{L,r_L}^{(+,n)} \tag{F.5}$$

in (F.4). Since l_r of the elements $\{r_1,\ldots,r_L\}$ in (F.5) are equal to 'r', we can replace (F.5) by an average of all its permutation invariant versions, i.e.,

$$a_{1,r_1}^{(+,n)} a_{2,r_2}^{(+,n)} \cdots a_{L,r_L}^{(+,n)} = \frac{1}{l_1!l_2!\cdots l_R!}\sum_{\tau\in T_{(r_1,\ldots,r_L)}} a_{1,r_{\tau(1)}}^{(+,n)} a_{2,r_{\tau(2)}}^{(+,n)} \cdots a_{L,r_{\tau(L)}}^{(+,n)}, \tag{F.6}$$

where $T_{(r_1,\ldots,r_L)}$ denotes the set of all $l_1!l_2!\cdots l_R!$ permutations of $(r_1,\ldots,r_L)$ with property $a_{1,r_1}^{(+,n)} a_{2,r_2}^{(+,n)} \cdots a_{L,r_L}^{(+,n)} = a_{1,r_{\tau(1)}}^{(+,n)} a_{2,r_{\tau(2)}}^{(+,n)} \cdots a_{L,r_{\tau(L)}}^{(+,n)}$, $\forall \tau \in T_{(r_1,\ldots,r_L)}$. Note that

$$\sum_{(r_1,\ldots,r_L)\in\Xi_{(l_1,\ldots,l_R)}} \frac{1}{l_1!\cdots l_R!}\sum_{\tau\in T_{(r_1,\ldots,r_L)}}\prod_{l=1}^{L} a_{l,r_{\tau(l)}}^{(+,n)} = \frac{1}{l_1!\cdots l_R!}\sum_{\sigma\in S_L}\prod_{l=1}^{L} a_{l,r_{\sigma(l)}}^{(+,n)}. \tag{F.7}$$

Similar to (F.3), we also observe that the permutation invariance property of the permanent implies that for any L-tuple $(r_1,\ldots,r_L) \in \Xi_{(l_1,\ldots,l_R)}$, we have that

$$\sum_{\sigma\in S_L}\prod_{l=1}^{L} a_{l,r_{\sigma(l)}}^{(+,n)} = \overset{+}{|}[\mathbf{1}_{l_1}^T \otimes \mathbf{a}_1^{(+,n)},\ldots,\mathbf{1}_{l_R}^T \otimes \mathbf{a}_R^{(+,n)}]\overset{+}{|} = \overset{+}{|}\mathbf{A}_{(l_1,\ldots,l_R)}^{(+,n)}\overset{+}{|}. \tag{F.8}$$

By inserting (F.6) in (F.4) and making use of relations (E.3), (F.7) and (F.8), we obtain

$$\sum_{(l_1,\ldots,l_R)\in[L]_R}\sum_{(r_1,\ldots,r_L)\in\Xi_{(l_1,\ldots,l_R)}}\frac{1}{l_1!\cdots l_R!}\sum_{\tau\in T_{(r_1,\ldots,r_L)}}\prod_{l=1}^{L}a_{l,r_{\tau(l)}}^{(+,n)} \tag{F.9}$$

$$\cdot\,\Big|^{+}[\mathbf{1}_{l_1}^T\otimes\mathbf{s}_1^{(k_1,\ldots,k_L)},\ldots,\mathbf{1}_{l_R}^T\otimes\mathbf{s}_R^{(k_1,\ldots,k_L)}]\Big|^{+}=$$

$$\sum_{(l_1,\ldots,l_R)\in[L]_R}\frac{1}{l_1!\cdots l_R!}\sum_{\sigma\in S_L}\prod_{l=1}^{L}a_{l,r_{\sigma(l)}}^{(+,n)}\,\Big|^{+}[\mathbf{1}_{l_1}^T\otimes\mathbf{s}_1^{(k_1,\ldots,k_L)},\ldots,\mathbf{1}_{l_R}^T\otimes\mathbf{s}_R^{(k_1,\ldots,k_L)}]\Big|^{+}=$$

$$\sum_{(l_1,\ldots,l_R)\in[L]_R}\frac{1}{l_1!\cdots l_R!}\,\Big|^{+}\mathbf{A}_{(l_1,\ldots,l_R)}^{(+,n)}\Big|^{+}\cdot\,\Big|^{+}[\mathbf{1}_{l_1}^T\otimes\mathbf{s}_1^{(k_1,\ldots,k_L)},\ldots,\mathbf{1}_{l_R}^T\otimes\mathbf{s}_R^{(k_1,\ldots,k_L)}]\Big|^{+}=$$

$$\mathbf{g}_{+}^{(n,L)}\mathbf{D}_{\mathbf{W}}^{(L)}\mathcal{R}_L^{(\mathrm{dist})}(\mathbf{S}^{(\mathrm{ext})})^T\mathbf{e}_{(k_1,\ldots,k_L)}^{(R^L)}+$$

$$\sum_{(l_1,\ldots,l_R)\in\Omega}\,\Big|^{+}\mathbf{A}_{(l_1,\ldots,l_R)}^{(+,n)}\Big|^{+}\cdot\,\Big|^{+}[\mathbf{1}_{l_1}^T\otimes\mathbf{s}_1^{(k_1,\ldots,k_L)},\ldots,\mathbf{1}_{l_R}^T\otimes\mathbf{s}_R^{(k_1,\ldots,k_L)}]\Big|^{+},$$

where $\Omega:=\{(L,0,0\ldots,0),(0,L,0\ldots,0),\ldots,(0,0\ldots,0,L)\}$. To summarize, from (F.1) and (F.9) we obtain

$$\Big|^{+}[\mathbf{x}_{k_1}^{(+,n)},\ldots,\mathbf{x}_{k_L}^{(+,n)}]\Big|^{+}=\mathbf{g}_{+}^{(n,L)}\mathbf{D}_{\mathbf{W}}^{(L)}\mathcal{R}_L^{(\mathrm{dist})}(\mathbf{S}^{(\mathrm{ext})})^T\mathbf{e}_{(k_1,k_2,\ldots,k_L)}^{(R^L)}+$$

$$\sum_{(l_1,\ldots,l_R)\in\Omega}\,\Big|^{+}\mathbf{A}_{(l_1,\ldots,l_R)}^{(+,n)}\Big|^{+}\cdot\,\Big|^{+}[\mathbf{1}_{l_1}^T\otimes\mathbf{s}_1^{(k_1,k_2,\ldots,k_L)},\ldots,\mathbf{1}_{l_R}^T\otimes\mathbf{s}_R^{(k_1,k_2,\ldots,k_L)}]\Big|^{+}. \tag{F.10}$$

Using $\mathbf{x}_k^{(-,n)}=\sum_{r=1}^{R}\mathbf{a}_r^{(-,n)}s_{kr}\in\mathbb{C}^L$, then by a similar reasoning, we obtain

$$\Big|^{+}[\mathbf{x}_{k_1}^{(-,n)},\ldots,\mathbf{x}_{k_L}^{(-,n)}]\Big|^{+}=\mathbf{g}_{-}^{(n,L)}\mathbf{D}_{\mathbf{W}}^{(L)}\mathcal{R}_L^{(\mathrm{dist})}(\mathbf{S}^{(\mathrm{ext})})^T\mathbf{e}_{(k_1,k_2,\ldots,k_L)}^{(R^L)}+$$

$$\sum_{(l_1,\ldots,l_R)\in\Omega}\,\Big|^{+}\mathbf{A}_{(l_1,\ldots,l_R)}^{(-,n)}\Big|^{+}\cdot\,\Big|^{+}[\mathbf{1}_{l_1}^T\otimes\mathbf{s}_1^{(k_1,k_2,\ldots,k_L)},\ldots,\mathbf{1}_{l_R}^T\otimes\mathbf{s}_R^{(k_1,k_2,\ldots,k_L)}]\Big|^{+}. \tag{F.11}$$

Since $\Big|^{+}\mathbf{A}_{(l_1,\ldots,l_R)}^{(+,n)}\Big|^{+}-\Big|^{+}\mathbf{A}_{(l_1,\ldots,l_R)}^{(-,n)}\Big|^{+}=0$ for every $(l_1,\ldots,l_R)\in\Omega$, the combination of (F.10) and (F.11) yields

$$\Big|^{+}[\mathbf{x}_{k_1}^{(+,n)},\ldots,\mathbf{x}_{k_L}^{(+,n)}]\Big|^{+}-\Big|^{+}[\mathbf{x}_{k_1}^{(-,n)},\ldots,\mathbf{x}_{k_L}^{(-,n)}]\Big|^{+}=$$

$$(\mathbf{g}_{+}^{(n,L)}-\mathbf{g}_{-}^{(n,L)})\mathbf{D}_{\mathbf{W}}^{(L)}\mathcal{R}_L^{(\mathrm{dist})}(\mathbf{S}^{(\mathrm{ext})})^T\mathbf{e}_{(k_1,k_2,\ldots,k_L)}^{(R^L)}. \tag{F.12}$$

From (49) and (E.3) it is clear that (F.12) corresponds to $(k_1,k_2,\ldots,k_L)$-th column entry of the n-th row of $\mathbf{R}_{\mathrm{MF}}^{(N,L)}$. This proves our claim that $\mathbf{R}_{\mathrm{MF}}^{(N,L)}$ admits the factorization (90).

References

[1] A. Cichocki, D. Mandic, C. Caifa, A. Phan, G. Zhou, L. De Lathauwer, Tensor decompositions for signal processing applications: From two-way to multiway component analysis, IEEE Signal Processing Magazine 32 (2) (2015) 145–163.

[2] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, C. Faloutsos, Tensor Decomposition for Signal Processing and Machine Learning, IEEE Trans. Signal Processing 65 (13) (2017) 3551–3582.
[3] R. A. Harshman, Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis, UCLA Working Papers in Phonetics 16 (1970) 1–84.
[4] J. D. Carroll, J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart–Young" decomposition, Psychometrika 35 (3) (1970) 283–319.
[5] R. Roy, T. Kailath, Estimation of signal parameters via rotational invariance techniques, IEEE Trans. ASSP 32 (7) (1989) 984–995.
[6] N. D. Sidiropoulos, R. Bro, G. B. Giannakis, Parallel factor analysis in sensor array processing, IEEE Trans. Signal Processing 48 (8) (2000) 2377–2388.
[7] A.-J. van der Veen, A. Paulraj, An Analytical Constant Modulus Algorithm, IEEE Trans. Signal Process. 44 (5) (1996) 1136–1155.
[8] A. Shashua, T. Hazan, Non-negative tensor factorization with applications to statistics and computer vision, *Proceedings of the 22nd international conference on Machine learning (ICML'05)*, 792–799, Bonn, Germany, August 07-11, 2005.
[9] E. Allman, C. Matias, J. Rhodes, Identifiability of parameters in latent structure models with many observed variables, The Annals of Statistics 37 (6) (2009) 3099–3132.
[10] L.-H. Lim, P. Comon, Nonnegative approximations of nonnegative tensors, J. Chemometrics 23 (7–8) (2009) 432–441.
[11] N. Kargas, N. D. Sidiropoulos, X. Fu, Tensors, Learning, and 'Kolmogorov Extension' for Finite-alphabet Random Vectors, IEEE Trans. Signal Processing 66 (18) (2018) 4854–4868.
[12] M. Sørensen, L. De Lathauwer, Coupled Canonical Polyadic Decompositions and (Coupled) Decompositions in Multilinear rank-$(L_{r,n}, L_{r,n}, 1)$ terms — Part I: Uniqueness, SIAM J. Matrix Anal. Appl. 36 (2) (2015) 496–522.
[13] M. Sørensen, I. Domanov, L. De Lathauwer, Coupled Canonical Polyadic Decompositions and (Coupled) Decompositions in Multilinear rank-$(L_{r,n}, L_{r,n}, 1)$ terms — Part II: Algorithms, SIAM J. Matrix Anal. Appl. 36 (3) (2015) 1015–1045.
[14] M. Sørensen, I. Domanov, L. De Lathauwer, Coupled canonical polyadic decompositions and multiple shift-invariance in array processing, IEEE Trans. Signal Processing 66 (14) (2018) 3665–3680.
[15] M. Sørensen, F. Van Eeghem, L. De Lathauwer, Blind multichannel deconvolution and convolutive extensions of canonical polyadic and block term decompositions, IEEE Trans. Signal Processing 65 (15) (2017) 4132–4145.
[16] M. Sørensen, L. De Lathauwer, Multidimensional harmonic retrieval via coupled canonical polyadic decomposition — Part I: Model and identifiability, IEEE Trans. Signal Processing 65 (2) (2017) 517–527.
[17] M. Sørensen, L. De Lathauwer, Multidimensional harmonic retrieval via coupled canonical polyadic decomposition — Part II: Algorithm and multirate sampling, IEEE Trans. Signal Processing 65 (2) (2017) 528–539.
[18] A.-J. van der Veen, Analytical Method for Blind Binary Signal Separation, IEEE Trans. Signal Process. 45 (4) (1997) 1078–1082.
[19] O. Grellier, P. Comon, Analytical Blind Discrete Source Separation, in: Proc. EUSIPCO, 2000.
[20] Z. Ghahramani, T. L. Griffiths, Infinite latent feature models and the Indian buffet process, in: Advances in Neural Information Processing Systems 18, Cambridge, MA, USA, 2005.
[21] M. Slawski, M. Hein, P. Lutsik, Matrix factorization with binary components, in: Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13), Dec. 05-10, Lake Tahoe, Nevada, USA, 2013.
[22] A. Jaffe, R. Weiss, B. Nadler, S. Carmi, Y. Kluger, Learning binary latent variable models: A tensor eigenpair approach, in: Proc. of the 35th International Conference on Machine Learning, July 10-15, Stockholm Sweden, 2018.
[23] B. W. Bader, R. A. Harshman, T. G. Kolda, Temporal analysis of semantic graphs using ASALSAN, in: Proc. of the Seventh IEEE International Conference on Data Mining (ICDM 2007), Oct. 28-31, Omaha, Nebraska, USA, 2007.
[24] E. E. Papalexakis, N. D. Sidiropolous, R. Bro, From K-Means to Higher-Way Co-Clustering: Multilinear Decomposition With Sparse Latent Factors, IEEE Trans. Signal Process. 61 (2013) 493–506.
[25] L. De Lathauwer, Decomposition of a higher-order tensor in block terms — Part II: Definitions and uniqueness, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 1033–1066.
[26] T. Jiang, N. D. Sidiropoulos, Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear model with constant modulus constraints, IEEE Trans. Signal

Process. 52 (9) (2004) 2625–2636.
[27] L. De Lathauwer, A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization, SIAM J. Matrix Anal. Appl. 28 (3) (2006) 642–666.
[28] I. Domanov, L. De Lathauwer, On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part I: Basic results and uniqueness of one factor matrix, SIAM J. Matrix Anal. Appl. 34 (3) (2013) 855–875.
[29] I. Domanov, L. De Lathauwer, On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part II: Overall uniqueness, SIAM J. Matrix Anal. Appl. 34 (3) (2013) 876–903.
[30] I. Domanov, L. De Lathauwer, Canonical polyadic decomposition of third-order tensors: reduction to generalized eigenvalue decomposition, SIAM J. Matrix Anal. Appl. 35 (2) (2014) 636–660.
[31] J. B. Kruskal, Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics, Linear Algebra and its Applications 18 (1977) 95–138.
[32] M. Marcus, H. Minc, On the relation between the determinant and the permanent, Illinois J. Math. 5 (1961) 376–381.
[33] H. Minc, Permanents, Cambridge University Press, 1984.
[34] S. E. Leurgans, R. T. Ross, R. B. Abel, A decomposition of three-way arrays, SIAM J. Matrix Anal. Appl. 14 (1993) 1064–1083.
[35] I. Domanov, L. De Lathauwer, Canonical polyadic decomposition of third-order tensors: Relaxed uniqueness conditions and algebraic algorithm, Linear Algebra Appl. 513 (2017) 342–375.
[36] R. B. Bapat, Mixed discriminants of positive semidefinite matrices, Linear Algebra and its Applications 126 (1989) 107–124.
[37] R. B. Bapat, T. E. S. Raghavan, Nonnegative matrices and applications, Cambridge University Press, 1997.