Bilinear factorizations subject to monomial equality constraints 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 bilinear factorizations subject to monomial equality constraints. This includes extensions of multilinear algebraic uniqueness conditions originally developed for the CPD. We obtain a deterministic uniqueness condition that admits a constructive interpretation. Computationally, we reduce the bilinear 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 bilinear factorization. Finally, we make a connection between bilinear factorizations subject to monomial equality constraints 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) in which a tensor $\mathcal{X} \in \mathbb{C}^{I\times J\times K}$ is decomposed into a sum of a minimal number of rank-one terms [3, 4]:

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

where $\mathbf{a}_r \in \mathbb{C}^I$, $\mathbf{b}_r \in \mathbb{C}^J$ and $\mathbf{s}_r \in \mathbb{C}^K$. The symbol '$\circ$' denotes the outer product, i.e., the (i, j, k)-th entry of $\mathcal{X}$ is equal to $x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} s_{kr}$ in which a_{ir} denotes i-th entry of $\mathbf{a}_r$ (similarly for $\mathbf{b}_r$ and $\mathbf{s}_r$). In this paper we will mainly consider a matrix unfolded version of $\mathcal{X}$ in which the entries x_{ijk} are stacked into a matrix $\mathbf{X} \in \mathbb{C}^{IJ\times K}$ with factorization

$$\mathbf{X} = (\mathbf{A} \odot \mathbf{B})\,\mathbf{S}^T, \tag{2}$$

where '$\odot$' denotes the Khatri–Rao (columnwise Kronecker) product, $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_R] \in \mathbb{C}^{I\times R}$, $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_R] \in \mathbb{C}^{J\times R}$ and $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_R] \in \mathbb{C}^{K\times R}$. A formal definition of the CPD and detailed explanation of matrix unfoldings of a tensor 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

$$\mathbf{X} = \mathbf{a}_1 \circ \mathbf{s}_1 + \cdots + \mathbf{a}_R \circ \mathbf{s}_R = \mathbf{a}_1\mathbf{s}_1^T + \cdots + \mathbf{a}_R\mathbf{s}_R^T = \mathbf{A}\mathbf{S}^T \in \mathbb{C}^{I\times K}, \tag{3}$$

in which the columns of $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_R] \in \mathbb{C}^{I\times R}$ (or similarly the columns of $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_R] \in \mathbb{C}^{K\times R}$) are subject to monomial equality constraints of the form

$$a_{p_1,r} \cdots a_{p_L,r} - a_{s_1,r} \cdots a_{s_L,r} = 0, \tag{4}$$

where $a_{m,r}$ denotes the m-th entry of the r-th column of $\mathbf{A}$. Since $a_{p_1,r} \cdots a_{p_L,r}$ and $a_{s_1,r} \cdots a_{s_L,r}$ are monomials of degree L, we sometimes say that the monomial equality constraint (4) is also of degree L. (In Sections 4 and 5 it will become clear that (2) is a special case of (3).)

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 (3) take the form

$$s_{kr} = e^{\sqrt{-1}u_{kr}} \quad \text{with} \quad u_{kr} \in \{0, 2\pi/M, \ldots, 2\pi(M-1)/M\} \tag{5}$$

has been considered (e.g., [18, 19]). From (5) 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 6.2, is the Binary Matrix Factorization (BMF):

$$\mathbf{X} = \mathbf{A}\mathbf{S}^T \in \mathbb{C}^{I\times K}, \tag{6}$$

where $\mathbf{A} \in \{0,1\}^{I\times R}$ is a binary matrix. BMFs of the form (6) play a role in binary latent variable modeling (e.g., [20, 21, 22]).

Bilinear factorizations subject to monomial equality constraints have the interesting property that they provide a framework that allows us to generalize the CPD model. As an example, the presented tensor decomposition framework for bilinear factorizations subject to monomial equality constraints enables us to extend the CPD model (2) to the case of binary weighted rank-one terms (this will be made clear in Section 6.1):

$$\mathbf{X} = (\mathbf{D} * (\mathbf{A} \odot \mathbf{B}))\mathbf{S}^T \in \mathbb{C}^{IJ\times K}, \tag{7}$$

where '$*$' denotes the Hadamard (element-wise) product and $\mathbf{D} \in \{0,1\}^{IJ \times R}$ is a binary matrix that is not fixed *a priori*. Binary weighted rank-one terms are of interest in clustering applications involving tensor structures (e.g., [23, 24]).

We mention that using tools from algebraic geometry, generic identifiability conditions for certain bilinear factorizations subject to monomial equality constraints can be obtained (e.g., [25, 26, 27]). For instance, when the entries of $\mathbf{A}$ correspond to rational functions, in which the variables can be considered to be randomly drawn from absolutely continuous distributions, a generic uniqueness condition was obtained in [26]. The bilinear factorization (3) is also related to the so-called X-rank decomposition in which the columns of $\mathbf{A}$ belong to a variety, which again can be used to obtain generic uniqueness conditions (e.g., [27]). However, the entries cannot always be assumed to be drawn from an absolutely continuous distribution. For example, in separation of digital communication signals the entries of $\mathbf{A}$ may be restricted to a finite alphabet, i.e., $a_{i,r}$ must be a root of the polynomial $\prod_{m=1}^{M}(x - \beta_m) = 0$, where $\beta_m \in \mathbb{C}$. In this paper we limit the discussion to the binary case where $M = 2$, $\beta_1 = 0$ and $\beta_2 = 1$, corresponding to the BMF (6). Another example is the CPD of an incomplete tensor $\mathcal{X}$ in which entries are unobserved. We have shown in [28] that if the observed data pattern is structured, then variants of the bilinear factorization approach discussed in Sections 4 and 5 can be used to obtain identifiability conditions. In this paper we discuss the binary weighted variant (7) of CPD, which is another example where the entries of $\mathbf{A}$ are not randomly drawn from an absolutely continuous distribution but fits within our framework.

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 well-known CPD. Section 2 reviews the lesser known Block Term Decomposition (BTD) [29] and coupled BTD [12, 13]. As our first contribution, we will in Section 3 present a new link between bilinear factorizations subject to monomial equality constraints of the form (3) and the coupled BTD. This connection enables us to translate the monomial constraint (4) into a low-rank constraint, which in turn allows us to treat the matrix factorization (3) as a tensor decomposition problem. Next, in Section 4 we will present identifiability conditions. It will be explained that the presented identifiability is an extension of a well-known CPD uniqueness condition developed in [30, 31, 32, 33] to the monomial case. As our third contribution, we will in Section 5 extend the algebraic algorithm for CPD in [31, 34] to bilinear factorizations subject to monomial equality constraints. In Section 6 we explain that the tensor decomposition framework for bilinear factorizations subject to monomial equality constraints can be used to generalize the CPD model (2) to the binary weighted CPD model (7). We also demonstrate how the presented algebraic algorithm can be adapted and used for the computation of a BMF of the form (6).

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, rank, range and kernel of a matrix $\mathbf{A}$ are denoted by $\mathbf{a}_r$, $\mathbf{A}^*$, $\mathbf{A}^T$, $\mathbf{A}^H$, $|\mathbf{A}|$, $^{+}_{|}\mathbf{A}^{+}_{|}$, $\mathbf{A}^{-1}$, $\mathbf{A}^{\dagger}$, $\mathrm{rank}(\mathbf{A})$, $\mathrm{range}(\mathbf{A})$, $\ker(\mathbf{A})$, 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

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

in which $(\mathbf{A})_{mn} = a_{mn}$. The outer product of, 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 (the Hamming *weight*) of a vector $\mathbf{x}$ is denoted by $\omega(\mathbf{x})$ in the tensor decomposition literature, dating back to the work of Kruskal [35]. 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, i.e., $\text{vec}(\mathbf{A}) = [\mathbf{a}_1^T,\ldots,\mathbf{a}_J^T]^T \in \mathbb{C}^{IJ}$. 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 $\mathbf{1}_R = [1,\ldots,1]^T\in\mathbb{C}^R$. 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 $C_m^k = \binom{m}{k} = \frac{m!}{k!(m-k)!}$. 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 [32, 34, 36, 37] and references therein for a discussion. Finally, we let $\text{Sym}^L(\mathbb{C}^R)$ denote the vector space of all symmetric L-th order tensors defined on $\mathbb{C}^R$. The associated set of vectorized ("flattened") versions of the symmetric tensors in $\text{Sym}^L(\mathbb{C}^R)$ will be denoted by $\pi_S^{(L)}$, i.e., a symmetric tensor $\mathcal{X}\in\mathbb{C}^{R\times\cdots\times R}$ in $\text{Sym}^L(\mathbb{C}^R)$ is associated with a vector $\mathbf{x}\in\mathbb{C}^{R^L}$ in $\pi_S^{(L)}$.

1.2. Canonical Polyadic Decomposition (CPD)

Consider a 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]:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}^{(r)}\circ\mathbf{s}_r = \sum_{r=1}^{R} \mathbf{a}_r\circ\mathbf{b}_r\circ\mathbf{s}_r, \tag{8}$$

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-1 matrix. The rank of the tensor $\mathcal{X}$ is equal to the minimal number of rank-1 tensors that yield $\mathcal{X}$ in a linear combination. When the rank of $\mathcal{X}$ is R, then (8) is called the Canonical PD (CPD) of $\mathcal{X}$.

1.2.1. Matrix representation

Consider the k-th frontal matrix slice $\mathbf{X}^{(\cdot\cdot k)}\in\mathbb{C}^{I\times J}$ of $\mathcal{X}$, defined by $(\mathbf{X}^{(\cdot\cdot k)})_{ij} = x_{ijk} = \sum_{r=1}^R a_{ir}b_{jr}s_{kr}$. The tensor $\mathcal{X}$ can be interpreted as a collection of matrix slices $\{\mathbf{X}^{(\cdot\cdot k)}\}$, each admitting a decomposition in rank-one terms

$$\mathbf{X}^{(\cdot\cdot k)} = \sum_{r=1}^{R}\mathbf{E}^{(r)}s_{kr} = \sum_{r=1}^{R}\mathbf{a}_r\mathbf{b}_r^T s_{kr}, \quad k\in\{1,\ldots,K\}. \tag{9}$$

Note that

$$\mathrm{vec}(\mathbf{X}^{(\cdot\cdot k)T}) = \sum_{r=1}^{R} (\mathbf{a}_r \otimes \mathbf{b}_r) s_{kr} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \ldots, \mathbf{a}_R \otimes \mathbf{b}_R]\mathbf{S}^T \mathbf{e}_k^{(K)}, \quad k \in \{1, \ldots, K\},$$

where $\mathbf{S}^T \mathbf{e}_k^{(K)}$ denotes the k-th column of $\mathbf{S}^T$. Stacking yields (2):

$$\mathbf{X} = \left[\mathrm{vec}(\mathbf{X}^{(\cdot\cdot 1)T}), \ldots, \mathrm{vec}(\mathbf{X}^{(\cdot\cdot K)T})\right] = [\mathbf{a}_1 \otimes \mathbf{b}_1, \ldots, \mathbf{a}_R \otimes \mathbf{b}_R]\mathbf{S}^T = (\mathbf{A} \odot \mathbf{B})\, \mathbf{S}^T. \quad (10)$$

The matrices $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_R] \in \mathbb{C}^{I\times R}$, $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_R] \in \mathbb{C}^{J\times R}$ and $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_R] \in \mathbb{C}^{K\times R}$ will sometimes be referred to as the factor matrices of the PD or CPD of $\mathcal{X}$.

1.2.2. Connection to bilinear factorizations subject to monomial equality constraints

Consider the CPD of $\mathcal{X}$ given by (8) in which $\mathbf{E}^{(r)} = \mathbf{a}_r \mathbf{b}_r^T$ is associated with the r-th column of $\mathbf{A} \odot \mathbf{B}$. The structure of $\mathbf{E}^{(r)}$ implies that any 2-by-2 submatrix of $\mathbf{E}^{(r)}$ is either a rank-0 or 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 bilinear factorization subject to monomial equality constraints involving $N = C_I^2 C_J^2$ monomial relations of degree $L = 2$ of the form

$$e^{(r)}_{i_1 j_1} e^{(r)}_{i_2 j_2} - e^{(r)}_{i_1 j_2} e^{(r)}_{i_2 j_1} = 0, \quad 1 \le i_1 < i_2 \le I,\ 1 \le j_1 < j_2 \le J. \quad (11)$$

Conversely, if the columns of $\mathbf{E}^{(r)}$ satisfy the monomial relations (11), then it admits the rank-1 factorization $\mathbf{E}^{(r)} = \mathbf{a}_r \mathbf{b}_r^T$. Two well-known CPD uniqueness conditions that rely on property (11) will be stated next.

1.2.3. Uniqueness conditions for CPD

The rank-1 tensors in (8) 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 (10) has full column rank, the following necessary and sufficient uniqueness condition stated in Theorem 1.1 was obtained in [30] and later reformulated in terms of compound matrices in [32]. The derivations in [30, 32] are based on Kruskal's permutation lemma [35]. Theorem 1.1 makes use of the matrix

$$\mathbf{G}^{(2)}_{\mathrm{CPD}} = \mathcal{C}_2(\mathbf{A}) \odot \mathcal{C}_2(\mathbf{B}) \in \mathbb{C}^{C_I^2 C_J^2 \times C_R^2} \quad (12)$$

and the vector

$$\mathbf{f}^{(2)}(\mathbf{d}) = [d_1 d_2, d_1 d_3, \ldots, d_{R-1} d_R]^T \in \mathbb{C}^{C_R^2}, \quad (13)$$

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. *[30, Condition B and Eq. (16)], [32, Theorem 1.11] Consider an R-term PD of $\mathcal{X} \in \mathbb{C}^{I\times J\times K}$ in* (8)*. 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*

$$\boldsymbol{G}^{(2)}_{CPD} \cdot \boldsymbol{f}^{(2)}(\boldsymbol{d}) = \boldsymbol{0} \Rightarrow \omega(\boldsymbol{d}) \leq 1, \tag{14}$$

for all structured vectors $\boldsymbol{f}^{(2)}(\boldsymbol{d})$ *of the form* (13)*.*

In practice, condition (14) can be hard to check. However, as observed in [30, 31, 32], if $\mathbf{G}^{(2)}_{\mathrm{CPD}}$ in (14) 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. *[30, Condition B and Eq. (17)], [32, Theorem 1.12], [31, Remark 1, p. 652] Consider an R-term PD of* $\mathcal{X} \in \mathbb{C}^{I\times J\times K}$ *in* (8)*. If*

$$\begin{cases} \boldsymbol{S} \text{ has full column rank,} \\ \boldsymbol{G}^{(2)}_{CPD} \text{ has full column rank,} \end{cases} \tag{15}$$

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

Furthermore, if condition (15) is satisfied, then the CPD of $\mathcal{X}$ can be computed via a matrix EVD [31, 34]. 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 $\max(I, J) < R$ [31, 34]. The latter CPD can be computed by means of a standard EVD (e.g., [3, 38]). In Section 5 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, 35, 31, 32, 33, 30, 34, 39, 2] and references therein.

2. Review of Block Term Decomposition (BTD) and coupled BTD

2.1. Block Term Decomposition (BTD)

The multilinear rank-$(P, P, 1)$ term decomposition of a tensor is an extension of the CPD (8), where each term in the decomposition now consists of the outer product of a vector and a matrix that is low-rank [29]. More formally, $\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{s}_r$ in (8) is replaced by $\mathbf{E}_r \circ \mathbf{s}_r$:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}_r \circ \mathbf{s}_r, \tag{16}$$

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 (16) indeed reduces to (8) with $\mathbf{E}_r = \mathbf{a}_r\mathbf{b}_r^T = \mathbf{a}_r \circ \mathbf{b}_r$.

Connection to polyadic decomposition. Since $\mathbf{E}_r$ is low-rank, we know that (16) can also be expressed in terms of a PD:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}_r \circ \mathbf{s}_r = \sum_{r=1}^{R} \mathbf{M}^{(r)}\mathbf{N}^{(r)T} \circ \mathbf{s}_r = \sum_{r=1}^{R}\sum_{p=1}^{P} \mathbf{m}^{(r)}_p \circ \mathbf{n}^{(r)}_p \circ \mathbf{s}_r, \tag{17}$$

where $\mathbf{E}_r = \mathbf{M}^{(r)}\mathbf{N}^{(r)T}$, in which $\mathbf{M}^{(r)} = [\mathbf{m}^{(r)}_1, \ldots, \mathbf{m}^{(r)}_P] \in \mathbb{C}^{I\times P}$ is a rank-P matrix and $\mathbf{N}^{(r)} = [\mathbf{n}^{(r)}_1, \ldots, \mathbf{n}^{(r)}_P] \in \mathbb{C}^{J\times P}$ is a rank-P matrix.

Matrix representation. Similar to (9), the tensor $\mathcal{X}$ given by (16) can be interpreted as a collection of matrix slices $\{\mathbf{X}^{(\cdot\cdot k)}\}$, each of which can be written as

$$\mathbf{X}^{(\cdot\cdot k)} = \sum_{r=1}^{R} \mathbf{E}_r s_{kr} = \sum_{r=1}^{R} \left(\sum_{p=1}^{P} \mathbf{m}_p^{(r)} \mathbf{n}_p^{(r)T} \right) s_{kr}, \quad k \in \{1, \dots, K\}. \tag{18}$$

Note that $\mathrm{vec}(\mathbf{X}^{(\cdot\cdot k)T}) = \sum_{r=1}^{R} \sum_{p=1}^{P} (\mathbf{m}_p^{(r)} \otimes \mathbf{n}_p^{(r)}) s_{kr}$, $k \in \{1, \dots, K\}$. Define

$$\mathbf{M} = \left[\mathbf{M}^{(1)}, \dots, \mathbf{M}^{(R)}\right] \in \mathbb{C}^{I \times PR}, \tag{19}$$

$$\mathbf{N} = \left[\mathbf{N}^{(1)}, \dots, \mathbf{N}^{(R)}\right] \in \mathbb{C}^{J \times PR}, \tag{20}$$

$$\mathbf{S}^{(\mathrm{ext})} = \left[\mathbf{1}_P^T \otimes \mathbf{s}_1, \dots, \mathbf{1}_P^T \otimes \mathbf{s}_R\right] \in \mathbb{C}^{K \times PR}. \tag{21}$$

Then, according to (10), the decomposition (17) can also be expressed in terms of the matrices $\mathbf{M}$, $\mathbf{N}$ and $\mathbf{S}^{(\mathrm{ext})}$ as follows:

$$\mathbf{X} = [\mathrm{vec}(\mathbf{X}^{(\cdot\cdot 1)T}), \dots, \mathrm{vec}(\mathbf{X}^{(\cdot\cdot K)T})] = (\mathbf{M} \odot \mathbf{N})\mathbf{S}^{(\mathrm{ext})T}. \tag{22}$$

By exploiting the structure of $\mathbf{S}^{(\mathrm{ext})}$, relation (22) can also be written more compactly as

$$\mathbf{X} = [\mathrm{vec}(\mathbf{E}_1), \dots, \mathrm{vec}(\mathbf{E}_R)]\mathbf{S}^T, \tag{23}$$

where we recall that $\mathbf{E}_r = \mathbf{M}^{(r)}\mathbf{N}^{(r)T}$.

2.2. Extension to coupled BTD

In this paper we will consider the extension of (16) in which a set of tensors $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n \times J_n \times K}$, $n \in \{1, \dots, N\}$ is decomposed into a sum of coupled BTDs [12]:

$$\mathcal{X}^{(n)} = \sum_{r=1}^{R} \mathbf{E}_r^{(n)} \circ \mathbf{s}_r, \quad n \in \{1, \dots, N\}, \tag{24}$$

where

$$\mathbf{E}_r^{(n)} := \mathbf{M}^{(n,r)}\mathbf{N}^{(n,r)T} = [\widetilde{\mathbf{M}}^{(n,r)}, \mathbf{0}_{I_n, P-P_{r,n}}][\widetilde{\mathbf{N}}^{(n,r)}, \mathbf{0}_{J_n, P-P_{r,n}}]^T \in \mathbb{C}^{I_n \times J_n} \tag{25}$$

is a matrix with $\mathrm{rank}(\mathbf{E}_r^{(n)}) = P_{r,n} \leq P$ and $\min(I_n, J_n) > P$, $\widetilde{\mathbf{M}}^{(n,r)} \in \mathbb{C}^{I_n \times P_{r,n}}$ and $\widetilde{\mathbf{N}}^{(n,r)} \in \mathbb{C}^{J_n \times P_{r,n}}$ are rank-$P_{r,n}$ matrices, and $\mathbf{0}_{m,n}$ denotes an $(m \times n)$ zero matrix. More precisely, we consider the coupled decomposition (24) subject to

$$\max_{1 \leq n \leq N} \mathrm{rank}(\mathbf{E}_r^{(n)}) = P \text{ and } \mathbf{s}_r \neq \mathbf{0}, \quad \forall r \in \{1, \dots, R\}. \tag{26}$$

An important observation is that condition (26) does not prevent that $\mathrm{rank}(\mathbf{E}_r^{(n)}) < P$ for some pairs (r, n). Note also that the vectors $\mathbf{s}_1, \dots, \mathbf{s}_R \in \mathbb{C}^K$ in (24) are shared between all $\mathcal{X}^{(n)}$, i.e., the third mode induces the *coupling.* It is the coupling of $\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(N)}$ via $\{\mathbf{s}_r\}$ that makes the coupled BTD useful for studying bilinear matrix factorizations subject to monomial equality constraints, as will be explained in Section 3. As in the CPD case, the rank of the coupled BTD is defined as the minimal number of coupled terms $\{\mathbf{E}_r^{(n)} \circ \mathbf{s}_r\}$ with property (26) that yield $\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(N)}$.

Matrix representation. Similar to (19) and (20), we define

$$\mathbf{M}^{(n)} = \left[\mathbf{M}^{(n,1)}, \dots, \mathbf{M}^{(n,R)}\right] \in \mathbb{C}^{I_n \times PR}, \tag{27}$$

$$\mathbf{N}^{(n)} = \left[\mathbf{N}^{(n,1)}, \dots, \mathbf{N}^{(n,R)}\right] \in \mathbb{C}^{J_n \times PR}. \tag{28}$$

We know from (22) that the matrix representation $\mathbf{X}^{(n)}$ of $\mathcal{X}^{(n)}$ admits the factorization

$$\mathbf{X}^{(n)} = (\mathbf{M}^{(n)} \odot \mathbf{N}^{(n)})\mathbf{S}^{(\text{ext})T}, \quad n \in \{1, \dots, N\}. \tag{29}$$

Similar to (23), relation (29) can be written more compactly as

$$\mathbf{X}^{(n)} = [\text{vec}(\mathbf{E}_1^{(n)}), \dots, \text{vec}(\mathbf{E}_R^{(n)})]\mathbf{S}^T, \quad n \in \{1, \dots, N\}, \tag{30}$$

where $\mathbf{E}_r^{(n)} = \mathbf{M}^{(n,r)}\mathbf{N}^{(n,r)T}$.

2.3. Uniqueness condition for (coupled) BTD

The coupled BTD version of $\mathbf{G}_{\text{CPD}}^{(2)}$ in (12) is given by

$$\mathbf{G}_{\text{BTD}}^{(N,P+1)} = \begin{bmatrix} \mathcal{C}_{P+1}(\mathbf{M}^{(1)}) \odot \mathcal{C}_{P+1}(\mathbf{N}^{(1)}) \\ \vdots \\ \mathcal{C}_{P+1}(\mathbf{M}^{(N)}) \odot \mathcal{C}_{P+1}(\mathbf{N}^{(N)}) \end{bmatrix} \mathbf{P}_{\text{BTD}} \in \mathbb{C}^{N \times (C_{R+P}^{P+1} - R)}, \tag{31}$$

where the row-vectors $\mathcal{C}_{P+1}(\mathbf{M}^{(n)}) \in \mathbb{C}^{1 \times C_{PR}^{P+1}}$ and $\mathcal{C}_{P+1}(\mathbf{N}^{(n)}) \in \mathbb{C}^{1 \times C_{PR}^{P+1}}$ take into account that the matrices $\mathbf{M}^{(n,r)}$ and $\mathbf{N}^{(n,r)}$ in (27)–(28) can be rank-P matrices. The stacking of the matrices $\{\mathcal{C}_{P+1}(\mathbf{M}^{(n)}) \odot \mathcal{C}_{P+1}(\mathbf{N}^{(n)})\}$ is a consequence of the coupling between $\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(N)}$ via the shared factor $\mathbf{S}$. The matrix $\mathbf{P}_{\text{BTD}} \in \mathbb{C}^{C_{PR}^{P+1} \times (C_{R+P}^{P+1} - R)}$ is the "compression" matrix that takes into account that each column vector $\mathbf{s}_r$ in (21) is repeated P times. The reasoning behind the construction $\mathbf{P}_{\text{BTD}}$ can be found in [13, p. 1032]. The $C_{R+P}^{P+1} - R$ columns of $\mathbf{P}_{\text{BTD}}$ are indexed by the lexicographically ordered tuples in the set

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

Consider also the mapping $f_c : \{(r_1, \dots, r_{P+1})\} \to \{1, 2, \dots, C_{R+P}^{P+1} - R\}$ 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

$$\Gamma_r = \{(q_1, \dots, q_{P+1}) \mid 1 \le q_1 < \dots < q_{P+1} \le PR\}.$$

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

$$(\mathbf{P}_{\text{BTD}})_{f_r(q_1,\dots,q_{P+1}), f_c(r_1,\dots,r_{P+1})} = \begin{cases} 1, & \text{if } \lceil \frac{q_1}{P} \rceil = r_1, \dots, \lceil \frac{q_{P+1}}{P} \rceil = r_{P+1}, \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

It can be verified that when $N = 1$ and $P = 1$, then (31) reduces to (12), i.e., $\mathbf{G}_{\text{BTD}}^{(1,2)} = \mathbf{G}_{\text{CPD}}^{(2)}$. Theorem 2.1 below is an extension of Theorem 1.2 to the coupled BTD case.

Theorem 2.1. *[13, Algorithm 5 and identity (5.28) in Section 5.2.3] Consider an R-term coupled BTD of $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n \times J_n \times K}$, $n \in \{1,\ldots,N\}$ in* (24)*. If*

$$\begin{cases} \mathbf{S} \text{ has full column rank,} \\ \boldsymbol{G}_{BTD}^{(N,P+1)} \text{ has full column rank,} \end{cases} \tag{33}$$

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

We stress that $\operatorname{rank}(\mathbf{E}_r^{(n)}) < P$ is permitted, as long as condition (26) is satisfied. Fortunately, statement (i) in Lemma 2.2 below asserts that the full column rank assumption of $\mathbf{G}_{\text{BTD}}^{(N,P+1)}$ implies that condition (26) is satisfied. This fact will be useful in Section 3.

Lemma 2.2. *[13, Lemma S.3.1] Assume that the matrix $\boldsymbol{G}_{BTD}^{(N,P+1)} \in \mathbb{C}^{N\times(C_{R+P}^{P+1}-R)}$ given by* (31) *has full column rank. Then*

(i) $\max_{1\le n\le N} r(\boldsymbol{E}_r^{(n)}) = P$ *for all* $r \in \{1,\ldots,R\}$,

(ii) the matrix $\begin{bmatrix} vec(\boldsymbol{E}_1^{(1)}) & \cdots & vec(\boldsymbol{E}_R^{(1)}) \\ \vdots & \ddots & \vdots \\ vec(\boldsymbol{E}_1^{(N)}) & \cdots & vec(\boldsymbol{E}_R^{(N)}) \end{bmatrix}$ *has full column rank.*

As in Theorem 1.2, if condition (33) in Theorem 2.1 is satisfied, then the coupled BTD of $\{\mathcal{X}^{(n)}\}$ can be computed via a matrix EVD [13].

In this paper we extend the CPD/BTD results discussed in this section to the case of bilinear matrix factorizations subject to monomial equality constraints. More precisely, in Section 3 we explain that the bilinear matrix factorization subject to monomial equality constraints can be interpreted as a coupled BTD. Next, in Section 4 we extend the uniqueness conditions stated in Theorems 1.2 and 2.1 to the case of bilinear models with factor matrices satisfying monomial relations. Finally, in Section 5 we extend the algebraic algorithm associated with Theorems 1.2 and 2.1 to the case of bilinear matrix factorizations subject to monomial equality constraints.

3. Link between bilinear factorizations subject to monomial equality constraints and coupled BTD

In Section 3.1 we explain how to represent the monomial structure (4) as a low-rank constraint on a particular matrix. Using this low-rank matrix, we will in Section 3.2 translate the bilinear factorization (3) into the coupled BTD of the form (24) reviewed in Section 2.1.

3.1. Representation of monomial structure via low-rank matrix

In this section we will propose to encode a monomial equality constraint of the form (35) via the rank deficiency of a matrix, which is to the best of our knowledge a novel contribution of this paper. Before presenting the low-rank matrix used to represent a monomial equality constraint, we need to introduce some notation. Consider again the factorization of $\mathbf{X}$ given by (3), consisting of R rank-one terms $\mathbf{a}_1\mathbf{s}_1^T,\ldots,\mathbf{a}_R\mathbf{s}_R^T$. Recall that we say that column $\mathbf{a}_r$ is subject to a monomial equality constraint of degree L if there exist L entries $a_{p_1,r}, \ldots, a_{p_L,r}$ and L entries $a_{s_1,r}, \ldots, a_{s_L,r}$ such that relation (4) is satisfied. We assume that every column $\mathbf{a}_1, \ldots, \mathbf{a}_r$ enjoys N such monomial equality constraints of degree L, each denoted by the subscript 'n', i.e., $a_{p_{1,n},r} \cdots a_{p_{L,n},r} - a_{s_{1,n},r} \cdots a_{s_{L,n},r} = 0$. For notational convenience, the scalars $a_{p_{1,n},r}, \ldots, a_{p_{L,n},r}$ and $a_{s_{1,n},r}, \ldots, a_{s_{L,n},r}$ will be viewed as coordinates of the vectors

$$\begin{cases} \mathbf{a}_r^{(+,n)} = [a_{1r}^{(+,n)}, \ldots, a_{Lr}^{(+,n)}]^T = \left[a_{p_{1,n},r}, \ldots, a_{p_{L,n},r}\right]^T \in \mathbb{C}^L, \\ \mathbf{a}_r^{(-,n)} = [a_{1r}^{(-,n)}, \ldots, a_{Lr}^{(-,n)}]^T = \left[a_{s_{1,n},r}, \ldots, a_{s_{L,n},r}\right]^T \in \mathbb{C}^L, \end{cases} \tag{34}$$

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 $a_{lr}^{(-,n)}$). To summarize, we assume that the bilinear rank-R factorization of $\mathbf{X}$ is subject to N monomial equality constraints involving monomials of degree L:

$$\prod_{l=1}^{L} a_{lr}^{(+,n)} - \prod_{l=1}^{L} a_{lr}^{(-,n)} = 0, \quad r \in \{1, \ldots, R\},\ n \in \{1, \ldots, N\}. \tag{35}$$

Define the structured matrix $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) \in \mathbb{C}^{L \times L}$:

$$\mathbf{A}_L\left(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}\right) := \begin{bmatrix} a_{1r}^{(+,n)} & 0 & \cdots & 0 & (-1)^L \cdot a_{1r}^{(-,n)} \\ a_{2r}^{(-,n)} & a_{2r}^{(+,n)} & \ddots & & 0 \\ 0 & a_{3r}^{(-,n)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{Lr}^{(-,n)} & a_{Lr}^{(+,n)} \end{bmatrix}. \tag{36}$$

The low-rank property of matrix $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ stated in Lemma 3.1 will be used to translate a bilinear matrix factorization subject to monomial equality constraints into a coupled BTD.

Lemma 3.1. *Consider the vectors $\boldsymbol{a}_r^{(+,n)} \in \mathbb{C}^L$ and $\boldsymbol{a}_r^{(-,n)} \in \mathbb{C}^L$ with property* (35)*. Then rank($\boldsymbol{A}_L(\boldsymbol{a}_r^{(+,n)}, \boldsymbol{a}_r^{(-,n)})$) $\leq L-1$. Furthermore, if $\prod_{l=1}^{L} a_{lr}^{(+,n)} \neq 0$ or $\prod_{l=1}^{L} a_{lr}^{(-,n)} \neq 0$, then rank($\boldsymbol{A}_L(\boldsymbol{a}_r^{(+,n)}, \boldsymbol{a}_r^{(-,n)})$) $= L-1$.*

Proof. From the cofactor expansion of $|\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})|$ along the first row, the connection between (35) and (36) becomes clear:

$$|\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})| = a_{1r}^{(+,n)}\cdot\begin{vmatrix} a_{2r}^{(+,n)} & 0 & \cdots & 0 \\ a_{3r}^{(-,n)} & \ddots & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ & & a_{Lr}^{(-,n)} & a_{Lr}^{(+,n)} \end{vmatrix} \tag{37}$$

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

where we exploited that the two involved $(L-1)\times(L-1)$ minors in (37) are triangular. The determinant property (37) also explains that $\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})$ is low-rank under the condition (35). More precisely, since $|\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})| = 0$, $\text{rank}(\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})) \leq L-1$. Furthermore, if $\prod_{m=1}^{L} a_{p_m} \neq 0$ or $\prod_{n=1}^{L} a_{s_n} \neq 0$, then the minors in (37) do not vanish and consequently $\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})$ is a rank-$(L-1)$ matrix. □

To summarize, a monomial relation of the form (35) can be represented via the rank deficiency of the matrix in (36). Consequently, the structure of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})$ can be relaxed without dropping the monomial equality constraint.

3.2. Bilinear factorizations subject to monomial equality constraints via coupled BTD

Consider the bilinear factorization (3) in which the columns of $\mathbf{A}$ satisfy N monomial relations of the form (35). The bilinear property of the matrix factorization $\mathbf{X} = \mathbf{A}\mathbf{S}^T$ together with the low-rank property of matrix (36) enables us to transform (3) 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 $\mathcal{X}^{(\cdot\cdot 1,n)} \in \mathbb{C}^{L\times L},\ldots,\mathcal{X}^{(\cdot\cdot K,n)} \in \mathbb{C}^{L\times L}$, (cf. Eq. (18) with $\mathbf{E}_r = \mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})$):

$$\mathcal{X}^{(\cdot\cdot k,n)} = \mathbf{A}_L(\mathbf{x}_k^{(+,n)},\mathbf{x}_k^{(-,n)}) = \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})s_{kr},\ k \in \{1,\ldots,K\}, \tag{38}$$

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. (34)):

$$\begin{cases} \mathbf{x}_k^{(+,n)} = [x_{1k}^{(+,n)},\ldots,x_{Lk}^{(+,n)}]^T = \left[x_{p_{1,n},k},\ldots,x_{p_{L,n},k}\right]^T, \\ \mathbf{x}_k^{(-,n)} = [x_{1k}^{(-,n)},\ldots,x_{Lk}^{(-,n)}]^T = \left[x_{s_{1,n},k},\ldots,x_{s_{L,n},k}\right]^T. \end{cases} \tag{39}$$

The *key observation* is that since $\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})$ defined by (36) is low-rank, the tensor $\mathcal{X}^{(n)}$ with matrix slices (38) is a BTD. The collection of all tensors $\{\mathcal{X}^{(1)},\ldots,\mathcal{X}^{(N)}\}$ yields the coupled BTD (cf. Eq. (24) with $\mathbf{E}_r^{(n)} = \mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})$):

$$\mathbb{C}^{L\times L\times K} \ni \mathcal{X}^{(n)} = \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)}) \circ \mathbf{s}_r, \quad n \in \{1,\ldots,N\}. \tag{40}$$

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 $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) = \mathbf{E}_r^{(n)}$ in which (cf. Eq. (25) with $I_n = J_n = L$ and $P = L-1$):

$$\mathbf{E}_r^{(n)} = \mathbf{M}^{(n,r)}\mathbf{N}^{(n,r)T} = [\widetilde{\mathbf{M}}^{(n,r)}, \mathbf{0}_{L,L-1-L_{r,n}}][\widetilde{\mathbf{N}}^{(n,r)}, \mathbf{0}_{L,L-1-L_{r,n}}]^T, \tag{41}$$

where $\widetilde{\mathbf{M}}^{(n,r)} \in \mathbb{C}^{L\times L_{r,n}}$ and $\widetilde{\mathbf{N}}^{(n,r)} \in \mathbb{C}^{L\times L_{r,n}}$ are rank-$L_{r,n}$ matrices and $\mathbf{0}_{m,n}$ denotes an $(m\times n)$ zero matrix. Note that any $\widetilde{\mathbf{M}}^{(n,r)}$ and $\widetilde{\mathbf{N}}^{(n,r)}$ obtained via a rank factorization of $\mathbf{E}_r^{(n)}$ can be used (e.g., via the singular value decomposition of $\mathbf{E}_r^{(n)}$). Note also 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 3.1. We can now conclude that if for all $r \in \{1,\ldots,R\}$ there exists an $n \in \{1,\ldots,N\}$ such that $L_{r,n} = L-1$ so that condition (26) with $P = L-1$ is satisfied, then the bilinear matrix factorization (3) subject to the monomial equality constraints of the form (4) can be turned into the coupled BTD (40). Theorem 3.2 below summarizes the uniqueness result based on the link between a bilinear matrix factorization subject to the monomial equality constraints and the coupled BTD.

Theorem 3.2. *Consider the coupled BTD of $\mathcal{X}^{(n)} \in \mathbb{C}^{I_n\times J_n\times K}$, $n \in \{1,\ldots,N\}$ in (40). If*

$$\begin{cases} \mathbf{S} \text{ has full column rank,} \\ \boldsymbol{G}_{BTD}^{(N,L)} \text{ has full column rank,} \end{cases} \tag{42}$$

then the coupled BTD rank of $\{\mathcal{X}^{(n)}\}$ is R, the coupled BTD of $\{\mathcal{X}^{(n)}\}$ is unique, the bilinear factorization of $\boldsymbol{X}$ in (3) is unique, and $\boldsymbol{A}$ in (3) has full column rank.

Proof. The result is an immediate consequence of Theorem 2.1 and Lemma 3.1. □

Note that in Theorem 3.2 we state that if condition (42) is satisfied, then $\mathbf{A}$ in (3) 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 structure on $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ and instead used the low-rank factorization $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) = \mathbf{E}_r^{(n)} = \mathbf{M}^{(n,r)}\mathbf{N}^{(n,r)T}$ in the coupled BTD of $\{\mathcal{X}^{(n)}\}$.

4. Identifiability conditions for bilinear matrix factorizations subject to the monomial equality constraints

By exploiting the properties of the mixed discriminant reviewed in Section 4.1, we will in this section explain how to explicitly take the structure of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ into account. More precisely, instead of considering the matrix $\mathbf{G}_{\text{BTD}}^{(N,L)}$, we will work with a matrix $\mathbf{G}_{\text{MEC}}^{(N,L)}$ derived in Section 4.2 that explicitly takes the structure of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ into account. Using the matrix $\mathbf{G}_{\text{MEC}}^{(N,L)}$, we will in Section 4.3 derive a uniqueness condition for bilinear matrix factorizations subject to monomial equality constraints. We will also explain that the obtained uniqueness condition is a generalization of the uniqueness condition stated in Theorem 1.2 for CPD to bilinear matrix factorizations subject to monomial equality constraints. Finally, in Section 4.4 we explain that the obtained uniqueness condition based on $\mathbf{G}_{\text{MEC}}^{(N,L)}$ is in fact equivalent to the uniqueness condition stated in Theorem 3.2, which is based on the matrix $\mathbf{G}_{\text{BTD}}^{(N,L)}$ that does not explicitly take the structure of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ into account.

4.1. Mixed discriminants

In Theorem 4.5 we present a uniqueness condition for the bilinear factorization of $\mathbf{X}$. 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 $\sum_{k=1}^{K} \mathbf{A}_L(\mathbf{x}_r^{(+,n)}, \mathbf{x}_r^{(-,n)}) w_{kr} = \sum_{s=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_{sr} = \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ is a matrix with rank strictly less than L. The latter property can be used to derive a condition that ensures the uniqueness of $\mathbf{W}$. In this section we will provide a derivation based on mixed discriminants, defined next.

4.1.1. Definition

Let $\mathbf{H}^{(r)} \in \mathbb{C}^{L\times L}$ and $d_r \in \mathbb{C}$. The mixed discriminants of the sum of R matrices $\mathbf{H}^{(1)}d_1 + \cdots + \mathbf{H}^{(R)}d_R$ correspond to the coefficients of the homogeneous polynomial

$$\left|\sum_{r=1}^{R} \mathbf{H}^{(r)} d_r\right| = \sum_{r_1,\ldots,r_L=1}^{R} \mathcal{D}(\mathbf{H}^{(r_1)}, \ldots, \mathbf{H}^{(r_L)}) \cdot d_{r_1} \cdots d_{r_L}. \tag{43}$$

The coefficients $\{\mathcal{D}(\mathbf{H}^{(r_1)}, \ldots, \mathbf{H}^{(r_L)})\}$ in (43) are known as mixed discriminants and are given by

$$\mathcal{D}(\mathbf{H}^{(r_1)}, \ldots, \mathbf{H}^{(r_L)}) = \frac{\partial^L \left|\mathbf{H}^{(r_1)} d_{r_1} + \cdots + \mathbf{H}^{(r_L)} d_{r_L}\right|}{\partial d_{r_1} \cdots \partial d_{r_L}}. \tag{44}$$

It can be verified that [40]:

$$\mathcal{D}\left(\mathbf{H}^{(r_1)}, \ldots, \mathbf{H}^{(r_L)}\right) = \frac{1}{L!} \sum_{\sigma \in S_L} \operatorname{sgn}(\sigma) \left|\left[\mathbf{h}_{\sigma(1)}^{(r_1)}, \mathbf{h}_{\sigma(2)}^{(r_2)}, \ldots, \mathbf{h}_{\sigma(L)}^{(r_L)}\right]\right|, \tag{45}$$

where $\mathbf{h}_{\sigma(l)}^{(r_l)}$ 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.

4.1.2. Properties

From (45) 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 (45) reduces to the determinant

$$\mathcal{D}\left(\mathbf{H}, \ldots, \mathbf{H}\right) = \sum_{\sigma \in S_L} \operatorname{sgn}(\sigma) \prod_{l=1}^{L} \mathbf{h}_{l,\sigma(l)} = \left|\mathbf{H}\right|. \tag{46}$$

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 (45) we obtain (a scaled version of) the permanent

$$\mathcal{D}\left(\mathbf{D}^{(1)},\ldots,\mathbf{D}^{(L)}\right) = \frac{1}{L!}\sum_{\sigma\in S_L}\prod_{l=1}^{L} d_{l,l}^{(\sigma(l))} = \frac{1}{L!}\overset{+}{|}\mathbf{B}\overset{+}{|}, \tag{47}$$

where $\mathbf{B}\in\mathbb{C}^{L\times L}$ is given by $(\mathbf{B})_{il} = d_{ii}^{(l)}$ and $\overset{+}{|}\mathbf{B}\overset{+}{|}$ denotes the permanent of $\mathbf{B}$. 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

$$\mathcal{D}\left(\mathbf{D}^{(1)},\ldots,\mathbf{D}^{(L)}\right) = \mathcal{D}\left(\mathbf{D}^{(\sigma(1))},\ldots,\mathbf{D}^{(\sigma(L))}\right) = \frac{1}{L!}\overset{+}{|}\mathbf{B}\overset{+}{|}, \quad \forall\sigma\in S_L, \tag{48}$$

which follows from the column permutation invariance property of the permanent, i.e., $\overset{+}{|}\mathbf{B}\overset{+}{|} = \overset{+}{|}\mathbf{B\Pi}\overset{+}{|}$ 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 (46) when $\mathrm{sgn}(\sigma)$ is dropped). This directly explains the permutation invariance property of the permanent. The three properties (46)–(48) of the mixed discriminant will be used in the derivation of Theorem 4.5. A further discussion of the mixed discriminant and its properties can be found in [40, 41]. A discussion of the properties of the permanent can be found in [36, 37].

4.2. Construction of $\boldsymbol{G}_{MEC}^{(N,L)}$ *and its properties*

The proof of the uniqueness condition stated in Theorem 4.5 will make use of a compact expression of the mixed discriminants associated with the expansion of the expression $\left|\sum_{r=1}^{R}\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})d_r\right|$ in terms of the scalars $d_1,\ldots,d_R$. Observe that

$$\begin{aligned}\left|\sum_{r=1}^{R}\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})d_r\right| &= \sum_{\sigma\in S_L}\mathrm{sgn}(\sigma)\prod_{l=1}^{L}\left(\sum_{r=1}^{R} d_r\cdot(\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})_{l\sigma(l)}\right)\\ &= \prod_{l=1}^{L}\left(\sum_{r=1}^{R} d_r a_{lr}^{(+,n)}\right) - \prod_{l=1}^{L}\left(\sum_{r=1}^{R} d_r a_{lr}^{(-,n)}\right),\end{aligned} \tag{49}$$

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 (49) directly follows from the patterned structure of $\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})$. (See also equations (36) and (37).) In terms of the matrices and vectors defined below a compact expression of (49) will be introduced in Lemma 4.2 below. 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

$$\mathbf{A}_{(l_1,\ldots,l_R)}^{(+,n)} = \left[\mathbf{1}_{l_1}^T\otimes\mathbf{a}_1^{(+,n)},\ldots,\mathbf{1}_{l_R}^T\otimes\mathbf{a}_R^{(+,n)}\right]\in\mathbb{C}^{L\times L}, \tag{50}$$

$$\mathbf{A}_{(l_1,\ldots,l_R)}^{(-,n)} = \left[\mathbf{1}_{l_1}^T\otimes\mathbf{a}_1^{(-,n)},\ldots,\mathbf{1}_{l_R}^T\otimes\mathbf{a}_R^{(-,n)}\right]\in\mathbb{C}^{L\times L}. \tag{51}$$

From the matrices in (50) and (51), 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:

$$\mathbf{g}_{+}^{(n,L)} = \left[{}^{+}\!\left|\mathbf{A}^{(+,n)}_{(L-1,1,0,0,\ldots,0)}\right|^{+}, {}^{+}\!\left|\mathbf{A}^{(+,n)}_{(L-1,0,1,0,\ldots,0)}\right|^{+}, \ldots, {}^{+}\!\left|\mathbf{A}^{(+,n)}_{(0,\ldots,0,1,L-1)}\right|^{+} \right], \tag{52}$$

$$\mathbf{g}_{-}^{(n,L)} = \left[{}^{+}\!\left|\mathbf{A}^{(-,n)}_{(L-1,1,0,0,\ldots,0)}\right|^{+}, {}^{+}\!\left|\mathbf{A}^{(-,n)}_{(L-1,0,1,0,\ldots,0)}\right|^{+}, \ldots, {}^{+}\!\left|\mathbf{A}^{(-,n)}_{(0,\ldots,0,1,L-1)}\right|^{+} \right]. \tag{53}$$

Based on (52) and (53) we in turn build the row vector, whose entries correspond to the mixed discriminants of $|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r|$, as will be made clear in the proof of Lemma 4.2:

$$\mathbf{g}_{\text{MEC}}^{(n,L)} = \left(\mathbf{g}_{+}^{(n,L)} - \mathbf{g}_{-}^{(n,L)}\right) \mathbf{D}_{\mathbf{W}}^{(L)} \in \mathbb{C}^{1\times(C_{R+L-1}^{L}-R)}, \tag{54}$$

in which the subscript 'MEC' stands for Monomial Equality Constraint 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

$$\mathbf{D}_{\mathbf{W}}^{(L)} = \operatorname{diag}\left(w^{(L)}_{(L-1,1,0,0,\ldots,0)}, w^{(L)}_{(L-1,0,1,0,\ldots,0)}, \ldots, w^{(L)}_{(0,\ldots,0,1,L-1)}\right), \tag{55}$$

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, ${}^{+}\!\left|\mathbf{A}^{(+,n)}_{(l_1,l_2,\ldots,l_R)}\right|^{+}$ and ${}^{+}\!\left|\mathbf{A}^{(-,n)}_{(l_1,l_2,\ldots,l_R)}\right|^{+}$ 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 (47)). Stacking yields

$$\mathbf{G}_{\text{MEC}}^{(N,L)} = \begin{bmatrix} \mathbf{g}_{\text{MEC}}^{(1,L)} \\ \mathbf{g}_{\text{MEC}}^{(2,L)} \\ \vdots \\ \mathbf{g}_{\text{MEC}}^{(N,L)} \end{bmatrix} \in \mathbb{C}^{N\times(C_{R+L-1}^{L}-R)}. \tag{56}$$

It can be verified that (56) is an extension of (12) to the monomial case, i.e., if $\mathbf{X}$ satisfies the CPD factorization (10) with full column rank $\mathbf{S}$, then $\mathbf{G}_{\text{MEC}}^{(N,L)}$ reduces to $\mathbf{G}_{\text{CPD}}^{(2)}$. 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 $N = C_I^2 C_J^2$ and $L = 2$. It will be shown in the proof of Lemma 4.2 that $\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r\right| = \mathbf{g}_{\text{MEC}}^{(n,L)} \cdot \mathbf{f}^{(L)}(\mathbf{d})$, where

$$\mathbf{f}^{(L)}(\mathbf{d}) = [d_1^{L-1}d_2, d_1^{L-1}d_3, \ldots, d_{R-1}d_R^{L-1}]^T \in \mathbb{C}^{(C_{R+L-1}^{L}-R)}. \tag{57}$$

Comparing (13) with (57), it is clear that the latter is also an extension of the former. More precisely, $\mathbf{f}^{(L)}(\mathbf{d})$ consists of all C_{R+L-1}^{L} distinct entries of $\mathbf{d}\otimes\cdots\otimes\mathbf{d}$ minus the R entries $d_1^L, \ldots, d_R^L$. The vector $\mathbf{f}^{(L)}(\mathbf{d})$ has the following two properties.

Lemma 4.1. *Consider a vector* $\boldsymbol{f}^{(L)}(\boldsymbol{d}) \in \mathbb{C}^{(C_{R+L-1}^{L}-R)}$ *of the form* (57)*. Then*

$$\omega(\boldsymbol{d}) \geq 2 \Rightarrow \boldsymbol{f}^{(L)}(\boldsymbol{d}) \neq \boldsymbol{0}, \tag{58}$$

$$\boldsymbol{f}^{(L)}(\boldsymbol{d}) = \boldsymbol{0} \Rightarrow \omega(\boldsymbol{d}) \leq 1. \tag{59}$$

Proof. Property (58) follows from the fact that if $\omega(\mathbf{d}) \geq 2$, then $d_i d_j^{L-1} \neq 0$ for some $i \neq j$. Similarly, $\mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}$ implies that $d_i d_j^{L-1} = 0$ for all $i \neq j$, necessitating that $\omega(\mathbf{d}) \leq 1$. □

Lemmas 4.2 and 4.3 relate $\mathbf{g}_{\mathrm{MEC}}^{(n,L)}$ and $\mathbf{G}_{\mathrm{MEC}}^{(N,L)}$ to $|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r|$ and $\mathbf{A}$, respectively.

Lemma 4.2. *Let $\boldsymbol{A}_L(\boldsymbol{a}_r^{(+,n)}, \boldsymbol{a}_r^{(-,n)}) \in \mathbb{C}^{L \times L}$ be of the form (36) and let $d_1, \dots, d_R \in \mathbb{C}$. Then*

$$\left| \sum_{r=1}^{R} \boldsymbol{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r \right| = \boldsymbol{g}_{MEC}^{(n,L)} \cdot \boldsymbol{f}^{(L)}(\boldsymbol{d}), \tag{60}$$

where $\boldsymbol{g}_{MEC}^{(n,L)} \in \mathbb{C}^{1 \times (C_{R+L-1}^{L} - R)}$ is given by (54) and $\boldsymbol{f}^{(L)}(\boldsymbol{d}) \in \mathbb{C}^{(C_{R+L-1}^{L} - R)}$ is given by (57).

Proof. Define

$$\mathbf{A}^{(+,r)} = \begin{bmatrix} \mathbf{a}_r^{(+,1)T} \\ \vdots \\ \mathbf{a}_r^{(+,N)T} \end{bmatrix} \in \mathbb{C}^{N \times L} \text{ and } \mathbf{A}^{(-,r)} = \begin{bmatrix} \mathbf{a}_r^{(-,1)T} \\ \vdots \\ \mathbf{a}_r^{(-,N)T} \end{bmatrix} \in \mathbb{C}^{N \times L}. \tag{61}$$

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

$$[L]_R = \{(l_1, \dots, l_R) \mid l_1 + \dots + l_R = L \text{ and } l_1, \dots, l_R \geq 0\}. \tag{62}$$

Note that the cardinality of $[L]_R$ is C_{R+L-1}^{L}. 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 $d_1, \dots, d_R$ yields the homogeneous polynomial

$$\left| \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r \right| = \prod_{l=1}^{L} \left(\sum_{r=1}^{R} a_{lr}^{(+,n)} d_r \right) - \prod_{l=1}^{L} \left(\sum_{r=1}^{R} a_{lr}^{(-,n)} d_r \right) \tag{63}$$

$$= \left| \sum_{r=1}^{R} D_n(\mathbf{A}^{(+,r)}) d_r \right| - \left| \sum_{r=1}^{R} D_n(\mathbf{A}^{(-,r)}) d_r \right| \tag{64}$$

$$= \sum_{r_1, \dots, r_L = 1}^{R} \Big[\mathcal{D}\left(D_n(\mathbf{A}^{(+,r_1)}), \dots, D_n(\mathbf{A}^{(+,r_L)}) \right) \tag{65}$$

$$- \mathcal{D}\left(D_n(\mathbf{A}^{-,r_1)}), \dots, D_n(\mathbf{A}^{(-,r_L)}) \right) \Big] d_{r_1} \cdots d_{r_L}$$

$$= \sum_{(l_1, \dots, l_R) \in [L]_R} \frac{L!}{l_1! \cdots l_R!} \Big[\mathcal{D}\big(\underbrace{D_n(\mathbf{A}^{(+,1)}), \dots, D_n(\mathbf{A}^{(+,1)})}_{l_1 \text{ times}}, \dots, \underbrace{D_n(\mathbf{A}^{(+,R)}), \dots, D_n(\mathbf{A}^{(+,R)})}_{l_R \text{ times}} \big) \tag{66}$$

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

$$= \sum_{(l_1, \dots, l_R) \in [L]_R} \frac{1}{l_1! \cdots l_R!} \left(\overset{+}{|} \mathbf{A}_{(l_1, \dots, l_R)}^{(+,n)} \overset{+}{|} - \overset{+}{|} \mathbf{A}_{(l_1, \dots, l_R)}^{(-,n)} \overset{+}{|} \right) d_1^{l_1} \cdots d_R^{l_R}, \tag{67}$$

where (64) follows from the definition (61), (65) follows from the definition of the mixed discriminant (43), (66) follows from the permutation invariance property (48) and (67) follows from property (47).

Due to property (46), we also know that if

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

then $\overset{+}{|}\mathbf{A}^{(+,n)}_{(l_1,\dots,l_R)}\overset{+}{|} - \overset{+}{|}\mathbf{A}^{(-,n)}_{(l_1,\dots,l_R)}\overset{+}{|} = \prod_{l=1}^{L} a_{lr}^{(+,n)} - \prod_{l=1}^{L} a_{lr}^{(-,n)} = 0$. Consequently, (67) can be written as

$$\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})d_r\right| = \mathbf{g}_{\text{MEC}}^{(n,L)} \cdot \mathbf{f}^{(L)}(\mathbf{d}), \tag{68}$$

where $\mathbf{g}_{\text{MEC}}^{(n,L)}$ and $\mathbf{f}^{(L)}(\mathbf{d})$ are given by (54) and (57), respectively. □

Lemma 4.3. *If $\boldsymbol{G}_{MEC}^{(N,L)} \in \mathbb{C}^{N\times(C_{R+L-1}^{L}-R)}$ given by* (56) *has full column rank, then $\boldsymbol{A} \in \mathbb{C}^{I\times R}$ in* (3) *has full column rank.*

Proof. Assume that $\mathbf{G}_{\text{MEC}}^{(N,L)}$ has full column rank. Suppose that $\mathbf{A}$ does not have full column rank. Then there exists a vector $\mathbf{d} \in \mathbb{C}^R$ with property $\omega(\mathbf{d}) \geq 2$ such that $\mathbf{A}\mathbf{d} = \mathbf{0}$. This also means that

$$\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})d_r\right| = 0, \quad n \in \{1,\dots,N\}, \tag{69}$$

where $\mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)}) \in \mathbb{C}^{L\times L}$ is given by (36) and N denotes the number of involved monomial equality constraints of the form (35). Due to relation (60) in Lemma 4.2, (69) can be written more compactly as

$$\mathbf{g}_{\text{MEC}}^{(n,L)} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = \left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)},\mathbf{a}_r^{(-,n)})d_r\right| = 0, \quad n \in \{1,\dots,N\}, \tag{70}$$

Stacking yields

$$\mathbf{G}_{\text{MEC}}^{(N,L)} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}\,, \tag{71}$$

where $\mathbf{G}_{\text{MEC}}^{(N,L)}$ is given by (56). Since $\omega(\mathbf{d}) \geq 2$ we know from property (58) in Lemma 4.1 that $\mathbf{f}^{(L)}(\mathbf{d}) \neq \mathbf{0}$. This property together with relation (71) in turn implies that $\mathbf{G}_{\text{MEC}}^{(N,L)}$ cannot have full column rank, which is a contradiction. □

4.3. Uniqueness condition based on $\boldsymbol{G}_{MEC}^{(N,L)}$

Using Kruskal's permutation lemma stated in Lemma 4.4 we will derive the sufficient uniqueness condition stated in Theorem 4.5 for bilinear factorizations subject to monomial equality constraints.

Lemma 4.4. *[35, 42]. Consider two matrices* $\boldsymbol{S} \in \mathbb{C}^{K \times R}$ *and* $\widehat{\boldsymbol{S}} \in \mathbb{C}^{K \times \widehat{R}}$ *with no zero columns and* $\widehat{R} \leq R$. *Let* $r_{\widehat{\mathbf{S}}}$ *denote the rank of* $\widehat{\boldsymbol{S}}$. *If for every* $\boldsymbol{z} \in \mathbb{C}^K$, *we have that*

$$\omega(\widehat{\boldsymbol{S}}^T \boldsymbol{z}) \leq R - r_{\widehat{\mathbf{S}}} + 1 \Rightarrow \omega(\boldsymbol{S}^T \boldsymbol{z}) \leq \omega(\widehat{\boldsymbol{S}}^T \boldsymbol{z}), \tag{72}$$

then $\widehat{R} = R$ *and* $\boldsymbol{S} = \widehat{\boldsymbol{S}}\boldsymbol{\Pi}\Delta_{\boldsymbol{S}}$, *where* $\boldsymbol{\Pi}$ *is an* $(R \times R)$ *column permutation matrix and* $\Delta_{\boldsymbol{S}}$ *is an* $(R \times R)$ *nonsingular diagonal matrix.*

Theorem 4.5. *Consider an R-term bilinear factorization of* $\boldsymbol{X}$ *in* (3) *subject to* N *monomial equality constraints of the form* (4). *If*

$$\begin{cases} \boldsymbol{S} \text{ has full column rank,} \\ \boldsymbol{G}^{(N,L)}_{MEC} \text{ has full column rank,} \end{cases} \tag{73}$$

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

Proof. Let the pair $(\widehat{\mathbf{A}}, \widehat{\mathbf{S}})$ be an alternative decomposition of (3) with $\widehat{R} \leq R$ terms so that

$$\mathbf{X} = \mathbf{A}\mathbf{S}^T = \widehat{\mathbf{A}}\widehat{\mathbf{S}}^T. \tag{74}$$

We first establish uniqueness of $\mathbf{S}$, i.e., we provide a condition that ensures that $\mathbf{S} = \widehat{\mathbf{S}}\boldsymbol{\Pi}\Delta_{\mathbf{S}}$, where $\boldsymbol{\Pi}$ is an $(R \times R)$ column permutation matrix and $\Delta_{\mathbf{S}}$ is an $(R \times R)$ nonsingular diagonal matrix. Lemma 4.4 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 $\omega(\widehat{\mathbf{S}}^T\mathbf{z}) \leq 1$. Lemma 4.3 together with the full column rank assumption of $\mathbf{G}^{(N,L)}_{\mathrm{MEC}}$ stated in condition (73) implies that $\mathbf{A}$ has full column rank. 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 $r_{\widehat{\mathbf{S}}}$ 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 $\omega(\widehat{\mathbf{d}}) = 0 \Leftrightarrow \widehat{\mathbf{S}}^T\mathbf{z} = \mathbf{0}$. Since $\mathbf{A}$ has full column rank, we know from (74) that $\mathbf{A}\mathbf{S}^T\mathbf{z} = \widehat{\mathbf{A}}\widehat{\mathbf{S}}^T\mathbf{z} = \mathbf{0} \Leftrightarrow \mathbf{S}^T\mathbf{z} = \mathbf{0}$, 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 (74) we obtain

$$\mathbf{X}\mathbf{z} = \mathbf{A}\mathbf{S}^T\mathbf{z} = \widehat{\mathbf{A}}\widehat{\mathbf{S}}^T\mathbf{z}. \tag{75}$$

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 $\omega(\widehat{\mathbf{d}}) = \omega(\widehat{\mathbf{C}}^T\mathbf{z}) = 1$. Due to relation (38), relation (75) can be expressed in terms of $(L \times L)$ matrices:

$$\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r = \sum_{r=1}^{R} \mathbf{A}_L(\widehat{\mathbf{a}}_r^{(+,n)}, \widehat{\mathbf{a}}_r^{(-,n)}) \widehat{d}_r\,, \quad n \in \{1, \ldots, N\}. \tag{76}$$

Since $\omega(\widehat{\mathbf{d}}) = 1$ and $\mathbf{A}_L(\widehat{\mathbf{a}}_r^{(+,n)}, \widehat{\mathbf{a}}_r^{(-,n)})$ is a matrix with rank strictly less than L, 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 (76) must vanish as well:

$$\left| \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r \right| = 0, \quad n \in \{1, \ldots, N\}. \tag{77}$$

Thanks to Lemma 4.2 (see also equations (69)–(71) in the proof of Lemma 4.3) we know that identity (77) can be expressed more compactly as

$$\mathbf{G}_{\text{MEC}}^{(N,L)} \cdot \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}, \tag{78}$$

where $\mathbf{G}_{\text{MEC}}^{(N,L)}$ is given by (56). Since $\mathbf{G}_{\text{MEC}}^{(N,L)}$ has full column rank by assumption, we know that $\mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}$. Due to property (59) in Lemma 4.1 this implies that $\omega(\mathbf{d}) \leq 1$. Hence, 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 Lemma 4.4 is satisfied. We conclude that condition (73) is sufficient for the uniqueness of $\mathbf{S}$. This fact together with the full column rank assumption of $\mathbf{S}$ also implies the uniqueness of $\mathbf{A} = \mathbf{X}(\mathbf{S}^T)^\dagger$. $\square$

4.4. Equivalence between Theorem 3.2 and 4.5 ($\boldsymbol{G}_{MEC}^{(N,L)} = \boldsymbol{G}_{BTD}^{(N,L)}$)

Proposition 4.8 below explains that the uniqueness condition (73) in Theorem 4.5 is equivalent to the uniqueness condition (42) in Theorem 3.2. The proof of Proposition 4.8 will be based on the following lemmas related to symmetric tensors.

Lemma 4.6. *[43, Proposition 3.4] Let $Sym^L(\mathbb{C}^R)$ denote the vector space of all symmetric L-th order tensors on vector space $\mathbb{C}^R$. The dimension of $Sym^L(\mathbb{C}^R)$ is $\binom{L+R-1}{L}$. Furthermore, since $\{\boldsymbol{e}_1^{(R)}, \ldots, \boldsymbol{e}_R^{(R)}\}$ is a basis for $\mathbb{C}^R$, the set of vectors*

$$\sum_{\sigma \in S_L} \boldsymbol{e}_{i_{\sigma(1)}}^{(R)} \circ \boldsymbol{e}_{i_{\sigma(2)}}^{(R)} \circ \cdots \circ \boldsymbol{e}_{i_{\sigma(L)}}^{(R)} \in Sym^L(\mathbb{C}^R), \quad 1 \leq i_1 \leq i_2 \leq \cdots \leq i_L \leq R \tag{79}$$

is a basis for $Sym^L(\mathbb{C}^R)$, where S_L denotes the symmetric group of permutations on $\{1, \ldots, L\}$.

Note that we will work with vectorized symmetric tensors. In that case the basis vectors (79) can be expressed as

$$\sum_{\sigma \in S_L} \mathbf{e}_{i_{\sigma(1)}}^{(R)} \otimes \mathbf{e}_{i_{\sigma(2)}}^{(R)} \otimes \cdots \otimes \mathbf{e}_{i_{\sigma(L)}}^{(R)}, \quad 1 \leq i_1 \leq i_2 \leq \cdots \leq i_L \leq R. \tag{80}$$

Lemma 4.7. *[43, Lemma 4.2] Let $\mathcal{A} \in Sym^L(\mathbb{C}^R)$ be a symmetric tensor. Then there exist vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s \in \mathbb{C}^R$ such that $\mathcal{A} = \sum_{i=1}^{s} \boldsymbol{y}_i \circ \cdots \circ \boldsymbol{y}_i$.*

In words, Lemma 4.7 states that every symmetric tensor admits a decomposition as a sum of symmetric rank-one tensors.

Proposition 4.8. *Consider the bilinear factorization of $\boldsymbol{X}$ in (3) subject to N monomial equality constraints of the form (4). Let $\boldsymbol{G}_{MEC}^{(N,L)}$ be the matrix given by (56) and let $\boldsymbol{G}_{BTD}^{(N,L)}$ be the matrix given by (31). Then*

$$\boldsymbol{G}_{MEC}^{(N,L)} = \boldsymbol{G}_{BTD}^{(N,L)}. \tag{81}$$

Proof. Consider the low-rank factorization $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) = \mathbf{M}^{(r,n)}\mathbf{N}^{(r,n)T}$, where $\mathbf{M}^{(r,n)} \in \mathbb{C}^{L\times(L-1)}$ and $\mathbf{N}^{(r,n)} \in \mathbb{C}^{L\times(L-1)}$ are matrices with rank strictly less than L. Recall that $\mathbf{G}_{\text{MEC}}^{(N,L)}$ was obtained from the expansion of the $|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r|$. Using the low-rank factorization $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) = \mathbf{M}^{(r,n)}\mathbf{N}^{(r,n)T}$, we obtain

$$\begin{aligned}\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r\right| &= \left|\sum_{r=1}^{R} \mathbf{M}^{(r,n)}\mathbf{N}^{(r,n)T}d_r\right| \\ &= \left|\mathbf{N}^{(n)}\text{Diag}\left(\mathbf{d}^{(\text{ext})}\right)\mathbf{M}^{(n)T}\right|, \quad n \in \{1,\ldots,N\},\end{aligned} \tag{82}$$

where we exploited that $|\mathbf{A}| = |\mathbf{A}^T|$, $\mathbf{M}^{(n)}$ and $\mathbf{N}^{(n)}$ are of the form (27) and (28), respectively, and $\text{Diag}(\mathbf{d}^{(\text{ext})})$ is a diagonal matrix that holds the column vector

$$\mathbf{d}^{(\text{ext})} = \mathbf{d} \otimes \mathbf{1}_{L-1} = \left[d_1\mathbf{1}_{L-1}^T, \ldots, d_R\mathbf{1}_{L-1}^T\right]^T \in \mathbb{C}^{(L-1)R} \tag{83}$$

on its diagonal. Relation (82) can be expressed in terms of compound matrices as follows

$$\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r\right| = \mathcal{C}_L\left(\mathbf{N}^{(n)}\right)\mathcal{C}_L\left(\text{Diag}\left(\mathbf{d}^{(\text{ext})}\right)\right)\mathcal{C}_L\left(\mathbf{M}^{(n)T}\right), \ n \in \{1,\ldots,N\}. \tag{84}$$

Consider the set $S = \{(i_1, i_2, \ldots, i_L) \mid 1 \le i_1 < i_2 < \cdots < i_L \le R(L-1)\}$, in which the L-tuples are ordered lexicographically and indexed by $S(1), \ldots, S(C_{R(L-1)}^L)$. Let[1]

$$\mathbf{d}^{(\text{ext},L)} = \left[d_{S(1)}^{(\text{ext})}, d_{S(2)}^{(\text{ext})}, \ldots, d_{S(C_{R(L-1)}^L)}^{(\text{ext})}\right]^T \in \mathbb{C}^{C_{R(L-1)}^L}, \tag{85}$$

where $d_{S(j)}^{(\text{ext})} = d_{i_1}^{(\text{ext})}d_{i_2}^{(\text{ext})}\cdots d_{i_L}^{(\text{ext})}$ with $S(j) = (i_1, i_2, \ldots, i_L)$. Relation (84) can now also be expressed as

$$\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})d_r\right| = \left(\mathcal{C}_L\left(\mathbf{M}^{(n)}\right) \odot \mathcal{C}_L\left(\mathbf{N}^{(n)}\right)\right)\mathbf{d}^{(\text{ext},L)}, \ n \in \{1,\ldots,N\}. \tag{86}$$

[1] The vector $\mathbf{d}^{(\text{ext},L)}$ corresponds to the diagonal part of the diagonal compound matrix $\mathcal{C}_L(\text{Diag}(\mathbf{d}^{(\text{ext})})) = \mathcal{C}_L(\text{Diag}(\mathbf{d} \otimes \mathbf{1}_{L-1}))$ in (84), i.e., $\text{Diag}(\mathbf{d}^{(\text{ext},L)}) = \mathcal{C}_L(\text{Diag}(\mathbf{d}^{(\text{ext})}))$.

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

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

where $j_1 = \lceil \frac{i_1}{L-1} \rceil$, $j_2 = \lceil \frac{i_2}{L-1} \rceil$, …, $j_L = \lceil \frac{i_L}{L-1} \rceil$. Using the "compression" matrix $\mathbf{P}_{\text{BTD}} \in \mathbb{C}^{C^L_{R(L-1)} \times (C^L_{R+L-1}-R)}$, defined by (32), we obtain the following compact version of (86):

$$\left| \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}^{(+,n)}_r, \mathbf{a}^{(-,n)}_r) d_r \right| = \left(\mathcal{C}_L \left(\mathbf{M}^{(n)} \right) \odot \mathcal{C}_L \left(\mathbf{N}^{(n)} \right) \right) \mathbf{P}_{\text{BTD}} \cdot \mathbf{f}^{(L)}(\mathbf{d}),\ n \in \{1, \ldots, N\}, \tag{87}$$

where $\mathbf{f}^{(L)}(\mathbf{d})$ is given by (57). Stacking yields

$$\mathbf{G}^{(N,L)}_{\text{BTD}} \cdot \mathbf{f}^{(L)}(\mathbf{d}), \tag{88}$$

where $\mathbf{G}^{(N,L)}_{\text{BTD}}$ is of the form (31). We also know from the proof of Lemma 4.3 that the stacking of $\left| \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}^{(+,n)}_r, \mathbf{a}^{(-,n)}_r) d_r \right|$, $n \in \{1, \ldots, N\}$, yields $\mathbf{G}^{(N,L)}_{\text{MEC}} \mathbf{f}^{(L)}(\mathbf{d})$. Hence, from (88) we can conclude that for any $\mathbf{d} \in \mathbb{C}^R$, we have that

$$\mathbf{G}^{(N,L)}_{\text{BTD}} \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{G}^{(N,L)}_{\text{MEC}} \mathbf{f}^{(L)}(\mathbf{d}) \Leftrightarrow \left(\mathbf{G}^{(N,L)}_{\text{BTD}} - \mathbf{G}^{(N,L)}_{\text{MEC}} \right) \mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}.$$

This in turn implies that

$$\left(\mathbf{G}^{(N,L)}_{\text{BTD}} - \mathbf{G}^{(N,L)}_{\text{MEC}} \right) \sum_{j=1}^{s_i} \mathbf{f}^{(L)}(\mathbf{d}^{(i)}_j) = \mathbf{0}, \tag{89}$$

where $\mathbf{d}^{(i)}_j \in \mathbb{C}^R$ and $s_i \in \mathbb{N}$. According to Lemma 4.6 there exist $C^L_{L+R-1} - R$ symmetric tensors $\mathcal{A}_1, \ldots, \mathcal{A}_{C^L_{L+R-1}-R} \in \text{Sym}^L(\mathbb{C}^R)$ with zero diagonal elements, i.e., $(a_i)_{k,\ldots,k} = 0$ for all $i \in \{1, \ldots, C^L_{L+R-1} - R\}$ and $k \in \{1, \ldots, R\}$. In particular, $\mathcal{A}_1, \ldots, \mathcal{A}_{C^L_{L+R-1}-R}$ can be chosen to be the basis vectors (79) with $1 \leq i_1 \leq i_2 \leq \cdots \leq i_L \leq R$ and $i_1 \neq i_L$. Let

$$\mathbf{f}^{(L)}(\mathcal{A}_j) := [(a_i)_{1,1,\ldots,2}, (a_i)_{1,1,\ldots,3}, \ldots, (a_i)_{R-1,R,\ldots,R}] \in \mathbb{C}^{C^L_{L+R-1}-R}$$

be a vector that consists of all distinct elements of $\mathcal{A}_i$ minus the R zero diagonal elements $(a_i)_{1,1,\ldots,1}, \ldots, (a_i)_{R,R,\ldots,R}$. Due to Lemma 4.7 there exist vectors $\mathbf{f}^{(L)}(\mathbf{d}^{(i)}_1), \ldots, \mathbf{f}^{(L)}(\mathbf{d}^{(i)}_{s_i})$ such that

$$\mathbf{f}^{(L)}(\mathcal{A}_j) = \sum_{j=1}^{s_i} \mathbf{f}^{(L)}(\mathbf{d}^{(i)}_j) \in \mathbb{C}^{C^L_{L+R-1}-R}.$$

Since $(\mathbf{G}^{(N,L)}_{\text{BTD}} - \mathbf{G}^{(N,L)}_{\text{MEC}}) \mathbf{f}^{(L)}(\mathcal{A}_j) = (\mathbf{G}^{(N,L)}_{\text{BTD}} - \mathbf{G}^{(N,L)}_{\text{MEC}}) \sum_{j=1}^{s_i} \mathbf{f}^{(L)}(\mathbf{d}^{(i)}_j) = \mathbf{0}$ and the vectors $\mathbf{f}^{(L)}(\mathcal{A}_1), \ldots, \mathbf{f}^{(L)}(\mathcal{A}_{C^L_{L+R-1}-R})$ are linearly independent, we conclude from (89) that

$$\dim \left(\ker \left(\mathbf{G}^{(N,L)}_{\text{BTD}} - \mathbf{G}^{(N,L)}_{\text{MEC}} \right) \right) = C^L_{L+R-1} - R.$$

We can now conclude that

$$\mathbf{G}^{(N,L)}_{\text{BTD}} - \mathbf{G}^{(N,L)}_{\text{MEC}} = \mathbf{0} \Leftrightarrow \mathbf{G}^{(N,L)}_{\text{BTD}} = \mathbf{G}^{(N,L)}_{\text{MEC}}.$$

□

5. Algorithm for bilinear factorization subject to monomial equality constraints

In this section we will present an algebraic algorithm tailored for the bilinear factorization of $\mathbf{X}$. The overall idea is that since $\mathbf{S}$ is assumed to have full column rank, we know that $\mathbf{a}_r \in \text{range}(\mathbf{X})$. Hence, there exists a vector $\mathbf{w}_r$ such that

$$\mathbf{X}\mathbf{w}_r = \mathbf{a}_r, \quad r \in \{1, \ldots, R\}. \tag{90}$$

Th goal is now to look for a matrix $\mathbf{W} \in \mathbb{C}^{R\times R}$ whose columns $\mathbf{w}_1, \ldots, \mathbf{w}_R$ have the property (90) so that $\mathbf{X}\mathbf{w}_r$ obeys the N monomial constraints associated with column $\mathbf{a}_r$ and that the R rank-1 terms $\mathbf{a}_1\mathbf{s}_1^T, \cdots, \mathbf{a}_R\mathbf{s}_R^T$ in (3) becomes separated, i.e.,

$$\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_R] = \mathbf{S}^T\mathbf{W} \quad \text{and} \quad \mathbf{W} = \mathbf{S}^{-T}\mathbf{\Pi}\mathbf{\Delta}, \tag{91}$$

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 $\omega(\mathbf{d}_r) = 1$. By exploiting the monomial equality constraints on $\mathbf{a}_r$, $\mathbf{w}_r$ can, under certain conditions, be obtained, observing only $\mathbf{X}$. Algorithms based on Theorem 4.5 can be derived. However, it turns out to be more convenient to work with an alternative null space formulation of Theorem 4.5 presented in Section 5.1. Based on this null space formulation, we will in Section 5.2 present an algebraic algorithm for bilinear factorization subject to monomial equality constraints.

5.1. Uniqueness condition in terms of dimension of null space

Theorem 5.3 below provides an alternative formulation of Theorem 4.5, which may be more easy to comprehend. It makes use of a matrix $\mathbf{\Psi}^{(N,L)} \in \mathbb{C}^{N\times R^L}$, defined as

$$\mathbf{\Psi}^{(N,L)} = \begin{bmatrix} \boldsymbol{\psi}^{(1,L)} \\ \vdots \\ \boldsymbol{\psi}^{(N,L)} \end{bmatrix} = \begin{bmatrix} (\widetilde{\mathbf{a}}_1^{(+,1)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(+,1)})^T \\ \vdots \\ (\widetilde{\mathbf{a}}_1^{(+,N)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(+,N)})^T \end{bmatrix} - \begin{bmatrix} (\widetilde{\mathbf{a}}_1^{(-,1)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(-,1)})^T \\ \vdots \\ (\widetilde{\mathbf{a}}_1^{(-,N)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(-,N)})^T \end{bmatrix}, \tag{92}$$

where

$$\begin{cases} \widetilde{\mathbf{a}}_l^{(+,n)} = [a_{l1}^{(+,n)}, \ldots, a_{lR}^{(+,n)}]^T = \left[a_{p_{l,n},1}, \ldots, a_{p_{l,n},R}\right]^T = \mathbf{e}_{p_{l,n}}^{(I)T}\mathbf{A} \in \mathbb{C}^R, \\ \widetilde{\mathbf{a}}_l^{(-,n)} = [a_{l1}^{(-,n)}, \ldots, a_{lR}^{(-,n)}]^T = \left[a_{s_{l,n},1}, \ldots, a_{s_{l,n},R}\right]^T = \mathbf{e}_{s_{l,n}}^{(I)T}\mathbf{A} \in \mathbb{C}^R. \end{cases} \tag{93}$$

In words, $\widetilde{\mathbf{a}}_l^{(+,n)}$ is the $p_{l,n}$-th row of $\mathbf{A}$ and $\widetilde{\mathbf{a}}_l^{(-,n)}$ is the $s_{l,n}$-th row of $\mathbf{A}$. Theorem 5.3 will also make use of the subspace

$$\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}, \tag{94}$$

where we recall that $\pi_S^{(L)}$ denotes the subspace of vectorized R^L symmetric tensors. The link between Theorem 4.5 and Theorem 5.3 follows from Lemmas 5.1 and 5.2 below and the following relation (as will be explained in the proof of Lemma 5.1):

$$\begin{aligned} \mathbf{g}_{\text{MEC}}^{(n,L)} \cdot \mathbf{f}^{(L)}(\mathbf{d}) &= \left| \sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_r \right| \\ &= \left(\widetilde{\mathbf{a}}_1^{(+,n)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(+,n)} - \widetilde{\mathbf{a}}_1^{(-,n)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(-,n)}\right)^T (\mathbf{d} \otimes \cdots \otimes \mathbf{d}) \\ &= \boldsymbol{\psi}^{(n,L)}(\mathbf{d} \otimes \cdots \otimes \mathbf{d}), \end{aligned} \tag{95}$$

where $\mathbf{d} = [d_1, \ldots, d_R]^T \in \mathbb{C}^R$.

Lemma 5.1. *Consider the bilinear factorization of $\boldsymbol{X}$ in* (3) *subject to N monomial equality constraints of the form* (4). *Let $\boldsymbol{G}_{MEC}^{(N,L)} \in \mathbb{C}^{N \times (C_{R+L-1}^{L} - R)}$ be given by* (56) *and let the matrix $\boldsymbol{\Psi}^{(N,L)} \in \mathbb{C}^{N \times R^L}$ be given by* (92). *Then*

$$\boldsymbol{G}_{MEC}^{(N,L)} \left(\sum_s \alpha_s \boldsymbol{f}^{(L)}(\boldsymbol{d}_s) \right) = \boldsymbol{\Psi}^{(N,L)} \left(\sum_s \alpha_s \boldsymbol{d}_s \otimes \cdots \otimes \boldsymbol{d}_s \right). \tag{96}$$

where $\alpha_s \in \mathbb{C}$ and $\boldsymbol{d}_s \in \mathbb{C}^R$.

Proof. From the definition of $\boldsymbol{\psi}^{(n,L)}$, we obtain

$$\begin{aligned}
\boldsymbol{\psi}^{(n,L)} \left(\sum_s \alpha_s \mathbf{d}_s \otimes \cdots \otimes \mathbf{d}_s \right) &= (\widetilde{\mathbf{a}}_1^{(+,n)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(+,n)})^T \left(\sum_s \alpha_s \mathbf{d}_s \otimes \cdots \otimes \mathbf{d}_s \right) \\
&\quad - (\widetilde{\mathbf{a}}_1^{(-,n)} \otimes \cdots \otimes \widetilde{\mathbf{a}}_L^{(-,n)})^T \left(\sum_s \alpha_s \mathbf{d}_s \otimes \cdots \otimes \mathbf{d}_s \right) \\
&= \left(\sum_s \alpha_s \prod_{l=1}^{L} (\widetilde{\mathbf{a}}_l^{(+,n)})^T \mathbf{d}_s \right) - \left(\sum_s \alpha_s \prod_{l=1}^{L} (\widetilde{\mathbf{a}}_l^{(-,n)})^T \mathbf{d}_s \right) \\
&= \sum_s \alpha_s \left(\prod_{l=1}^{L} \left(\sum_{r=1}^{R} a_{lr}^{(+,n)} d_{rs} \right) - \prod_{l=1}^{L} \left(\sum_{r=1}^{R} a_{lr}^{(-,n)} d_{rs} \right) \right) \\
&= \mathbf{g}_{\mathrm{MEC}}^{(n,L)} \cdot \left(\sum_s \alpha_s \mathbf{f}^{(L)}(\mathbf{d}_s) \right),
\end{aligned} \tag{97}$$

where the last identity follows from relation (49) and the left-hand side of (68). □

Note that $\mathbf{f}^{(L)}(\mathbf{d}) \in \mathbb{C}^{(C_{R+L-1}^{L} - R)}$ given by (57) can be interpreted as a vector that holds the distinct entries of a symmetric rank-1 tensor $\mathbf{d} \circ \cdots \circ \mathbf{d} \in \mathrm{Sym}^L(\mathbb{C}^R)$ minus the R diagonal entries $d_{1,\ldots,1}, \ldots, d_{R,\ldots,R}$. In Lemma 5.2 below we will consider a vector $\mathbf{f}^{(L)}(\mathcal{D}) \in \mathbb{C}^{(C_{R+L-1}^{L} - R)}$ that can be interpreted as a vector that holds the distinct entries of a symmetric tensor $\mathcal{D} \in \mathrm{Sym}^L(\mathbb{C}^R)$ minus the R diagonal entries $d_{1,\ldots,1}, \ldots, d_{R,\ldots,R}$. More precisely, the coordinates of any vector $\mathbf{x} \in \mathbb{C}^{(C_{R+L-1}^{L} - R)}$ can be interpreted as the distinct off-diagonal entries of a symmetric tensor $\mathcal{D} \in \mathrm{Sym}^L(\mathbb{C}^R)$. Due to Lemma 4.7 we know that there exist vectors $\mathbf{d}_1, \ldots, \mathbf{d}_s \in \mathbb{C}^R$ such that $\mathcal{D} = \sum_{i=1}^{s} \mathbf{d}_i \circ \cdots \circ \mathbf{d}_i$. Hence, we obtain the decomposition $\mathbf{x} = \sum_{i=1}^{s} \mathbf{f}^{(L)}(\mathbf{d}_i)$, where $\mathbf{d}_i \in \mathbb{C}^R$ is associated with a symmetric rank-one term in the symmetric tensor decomposition $\mathcal{D} = \sum_{i=1}^{s} \mathbf{d}_i \circ \cdots \circ \mathbf{d}_i$. This will be denoted by

$$\mathbf{x} = \mathbf{f}^{(L)}(\mathcal{D}), \quad \mathbf{f}^{(L)}(\mathcal{D}) = \sum_{i=1}^{s} \mathbf{f}^{(L)}(\mathbf{d}_i). \tag{98}$$

Lemma 5.2. *Consider the bilinear factorization of $\boldsymbol{X}$ in* (3) *subject to N monomial equality constraints of the form* (4). *If dim(ker($\boldsymbol{\Psi}^{(N,L)}$) $\cap\, \pi_S^{(L)}$) $= R$, then $\boldsymbol{A}$ in* (3) *has full column rank.*

PROOF. Since $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ has rank strictly less than L (i.e., $|\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})| = 0$), we have that

$$\left|\sum_{r=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})\mathbf{e}_r^{(R)}\right| = \boldsymbol{\psi}^{(n,L)}(\mathbf{e}_r^{(R)} \otimes \cdots \otimes \mathbf{e}_r^{(R)}) = 0, \quad r \in \{1, \ldots, R\},$$

where relations (49) and (96) were used. Since we assume that the subspace $\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$ is R-dimensional, the linearly independent vectors in the set $\{\mathbf{e}_r^{(R)} \otimes \cdots \otimes \mathbf{e}_r^{(R)}\}_{r=1}^{R}$ form a basis for $\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$, i.e., any $\mathbf{d} \in \ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$ can be written as

$$\mathbf{d} = \sum_{r=1}^{R} \alpha_r \mathbf{e}_r^{(R)} \otimes \cdots \otimes \mathbf{e}_r^{(R)},$$

where $\alpha_r \in \mathbb{C}$. Due to relation (96) in Lemma 5.1 and the fact that $\mathbf{\Psi}^{(N,L)}\mathbf{d} = \mathbf{0}$, we know that

$$\boldsymbol{\psi}^{(n,L)}\mathbf{d} = \boldsymbol{\psi}^{(n,L)} \left(\sum_{r=1}^{R} \alpha_r \mathbf{e}_r^{(R)} \otimes \cdots \otimes \mathbf{e}_r^{(R)}\right) = \mathbf{g}_{\text{MEC}}^{(n,L)} \cdot \left(\sum_{r=1}^{R} \alpha_r f^{(L)}(\mathbf{e}_r^{(R)})\right) = 0,$$

Since $f^{(L)}(\mathbf{e}_r^{(R)}) = \mathbf{0}$ for any $r \in \{1, \ldots, R\}$, we have $\sum_{r=1}^{R} \alpha_r f^{(L)}(\mathbf{e}_r^{(R)}) = \mathbf{0}$. The other way around, if $\mathbf{G}_{\text{MEC}}^{(N,L)}\mathbf{x} = \mathbf{G}_{\text{MEC}}^{(N,L)}\mathbf{f}^{(L)}(\mathcal{D}) = \mathbf{0}$, where relation (98) was used, then $\boldsymbol{\psi}^{(n,L)}(\sum_{i=1}^{s} \mathbf{d}_i \otimes \cdots \otimes \mathbf{d}_i) = 0$, which by assumption $\dim(\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$ implies that $\mathcal{D} = \sum_{i=1}^{s} \mathbf{d}_i \circ \cdots \circ \mathbf{d}_i$ is a diagonal tensor. This in turn implies that $\mathbf{x} = \mathbf{f}^{(L)}(\mathcal{D}) = \mathbf{0}$. Hence, when $\dim(\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$, then $\mathbf{G}_{\text{MEC}}^{(N,L)}\mathbf{x} = \mathbf{0}$ implies that $\mathbf{x} = \mathbf{0}$, i.e., $\ker(\mathbf{G}_{\text{MEC}}^{(N,L)}) = \{\mathbf{0}\}$. This in turn implies that $\mathbf{G}_{\text{MEC}}^{(N,L)}$ has full column rank. Lemma 4.3 now tells us that $\mathbf{A}$ has full column rank. □

Theorem 5.3. *Consider the bilinear factorization of* $\boldsymbol{X}$ *in* (3) *subject to* N *monomial equality constraints of the form* (4)*. If*

$$\begin{cases} \boldsymbol{S} \text{ has full column rank,} \\ ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)} \text{ is an } R\text{-dimensional subspace,} \end{cases} \tag{99}$$

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

PROOF. Due to Lemma 5.2 we know that $\dim(\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$ implies that $\mathbf{A}$ has full column rank. Since $\mathbf{S}$ has full column rank by assumption, we can conclude that $\mathbf{X}$ has rank R. W.l.o.g. we can now assume that $\mathbf{S}$ is square ($K = R$) and nonsingular. Since $\mathbf{X} = \mathbf{A}\mathbf{S}^T$ and $\mathbf{S}$ is nonsingular, there exists a nonsingular matrix $\mathbf{W} = \mathbf{S}^{-T}\mathbf{\Pi}\mathbf{\Delta} \in \mathbb{C}^{R \times R}$, for some $\mathbf{\Pi} \in \mathbb{C}^{R \times R}$ permutation matrix and nonsingular diagonal matrix $\mathbf{\Delta} \in \mathbb{C}^{R \times R}$, with property

$$\mathbf{X}\mathbf{W} = \mathbf{A}\mathbf{S}^T\mathbf{W} = \mathbf{A}\mathbf{D}, \tag{100}$$

where $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_R] = \mathbf{S}^T\mathbf{W} = \mathbf{S}^T\mathbf{S}^{-T}\mathbf{\Pi}\mathbf{\Delta} = \mathbf{\Pi}\mathbf{\Delta} \in \mathbb{C}^{R \times R}$ is a column permuted nonsingular diagonal matrix. We will now argue that $\mathbf{D}$ and therefore also $\mathbf{W}$ is unique (up to the intrinsic column scaling and permutation ambiguities). Using (95), we obtain

$$\left|\sum_{s=1}^{R} \mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)}) d_{sr}\right| = \boldsymbol{\psi}^{(n,L)}(\mathbf{d}_r \otimes \cdots \otimes \mathbf{d}_r) = 0, \quad r \in \{1, \ldots, R\},$$

where exploited that $\omega(\mathbf{d}_r) = 1$ for all $r \in \{1, \ldots, R\}$ and that $\mathbf{A}_L(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ has rank strictly less than L. Overall, we obtain

$$\boldsymbol{\Psi}^{(N,L)}(\mathbf{D} \odot \cdots \odot \mathbf{D}) = \mathbf{0}.$$

Since $\dim(\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$, the columns of $\mathbf{D} \odot \cdots \odot \mathbf{D}$ form a basis for $\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$. Consequently, if the columns of $\mathbf{B} \in \mathbb{C}^{R^L \times R}$ constitute an alternative basis for $\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$, then there exists a nonsingular matrix $\mathbf{F} \in \mathbb{C}^{R \times R}$ such that

$$\mathbf{B} = (\mathbf{D} \odot \cdots \odot \mathbf{D})\mathbf{F}^T. \tag{101}$$

Due to, e.g., Theorem 1.2 we can conclude from relation (101) that $\mathbf{D}$ is unique (up to the intrinsic column scaling and permutation ambiguities). This implies that when $\dim(\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$, then $\mathbf{w}_1, \ldots, \mathbf{w}_R$ are the only vectors (up to scaling ambiguities) with the property that $\mathbf{X}\mathbf{W}\boldsymbol{\Delta}^{-1}\boldsymbol{\Pi} = \mathbf{A}$, where $\boldsymbol{\Delta}$ is an arbitrary nonsingular diagonal matrix and $\boldsymbol{\Pi}$ is an arbitrary permutation matrix. We can now conclude that $\mathbf{A} = \mathbf{X}\mathbf{W}\boldsymbol{\Delta}\boldsymbol{\Pi}$ and $\mathbf{S}^T = \boldsymbol{\Pi}\boldsymbol{\Delta}\mathbf{W}^{-1}$, implying the uniqueness of the bilinear factorization of $\mathbf{X}$. □

Note that Theorem 4.5 is based on the assumption that $\dim(\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$. Proposition 5.4 below explains that this is equivalent to $\mathbf{G}_{\text{MEC}}^{(N,L)}$ having full column rank.

Proposition 5.4. *The subspace $\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$ is R-dimensional if and only if $\boldsymbol{G}_{MEC}^{(N,L)}$ has full column rank.*

PROOF. Assume that $\mathbf{G}_{\text{MEC}}^{(N,L)}$ has full column rank. Then $\mathbf{G}_{\text{MEC}}^{(N,L)}\mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}$ implies that $\mathbf{f}^{(L)}(\mathbf{d}) = \mathbf{0}$. Due to Lemma 4.1 we know that $\omega(\mathbf{d}) \leq 1$. Lemma 5.1 implies that $\mathbf{d} \otimes \cdots \otimes \mathbf{d} \in \ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$. This fact together with the fact that there are R linearly independent vectors $\mathbf{d}_1, \ldots, \mathbf{d}_R$ with property $\omega(\mathbf{d}_r) = 1$ implies that $\dim(\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) \geq R$. We will now argue that $\dim(\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) \leq R$. Assume on the contrary that $\dim(\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) > R$. Note that any vector in $\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$ corresponds to a vectorized L-th order symmetric tensor, e.g., $\mathbf{d} \in \ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$ implies that there exists a symmetric tensor $\mathcal{D} \in \text{Sym}^L(\mathbb{C}^R)$ such that $\mathbf{d} = [d_{1,1,\ldots,1}, d_{1,1,\ldots,2}, \ldots, d_{R,R,\ldots,R}]^T$. Since $\dim(\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) > R$ we can assume that the symmetric tensor $\mathcal{D}$ has a nonzero off-diagonal element, i.e., $d_{i_1,\ldots,i_L} \neq 0$ for some $i_m \neq i_n$ (if not, then $\dim(\ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$). According to Lemma 4.7 there exist vectors $\mathbf{d}_1, \ldots, \mathbf{d}_s \in \mathbb{C}^R$ such that $\mathcal{D} = \sum_{i=1}^{s} \mathbf{d}_i \circ \cdots \circ \mathbf{d}_i$. Let $\sum_{i=1}^{s} \mathbf{d}_i \otimes \cdots \otimes \mathbf{d}_i$ denote the vectorized version of $\mathcal{D} = \sum_{i=1}^{s} \mathbf{d}_i \circ \cdots \circ \mathbf{d}_i$. Then, since $\sum_{i=1}^{s} \mathbf{d}_i \otimes \cdots \otimes \mathbf{d}_i \in \ker(\boldsymbol{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$, we have

$$0 = \boldsymbol{\psi}^{(n,L)}\left(\sum_{i=1}^{s} \mathbf{d}_i \otimes \cdots \otimes \mathbf{d}_i\right) = \mathbf{g}_{\text{MEC}}^{(n,L)} \cdot \left(\sum_{i=1}^{s} f^{(L)}(\mathbf{d}_i)\right),$$

where relation (96) in Lemma 5.1 was used. Since $\mathbf{G}_{\text{MEC}}^{(N,L)}$ has full column rank, we must have that $\sum_{i=1}^{s} f^{(L)}(\mathbf{d}_i) = \mathbf{0}$. The latter implies that $d_{i_1,\ldots,i_L} = 0$ whenever $i_m \neq i_n$, which is a contradiction. We can now conclude that if $\mathbf{G}_{\text{MEC}}^{(N,L)}$ has full column rank, then $\dim(\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$.

Conversely, assume that $\dim(\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$. Then, as explained in the proof of Lemma 5.2, $\mathbf{G}_{\text{MEC}}^{(N,L)}$ has full column rank. □

5.2. Algorithm based on the null space formulation

In this section an algebraic algorithm for the bilinear factorization problem will be outlined that is based on the null space formulation discussed in Section 5.1.

Step 1: Construction of matrix $\boldsymbol{P}^{(N,L)}$. Assume that condition (99) in Theorem 5.3 is satisfied. W.l.o.g. we can assume that $\mathbf{S}$ is square nonsingular ($K = R$). Note that from (95) there exist R linearly independent vectors $\mathbf{w}_1, \ldots, \mathbf{w}_R$, each with property $\omega(\mathbf{S}^T\mathbf{w}_r) = \omega(\mathbf{d}_r) = 1$, such that for every $n \in \{1, \ldots, N\}$ we have that

$$\left| \sum_{q=1}^{R} \mathbf{A}_L(\mathbf{a}_q^{(+,n)}, \mathbf{a}_q^{(-,n)}) d_{qr} \right| = \left(\prod_{l=1}^{L} a_{ls}^{(+,n)} - \prod_{l=1}^{L} a_{ls}^{(-,n)} \right) d_{sr}^L \;= 0, \tag{102}$$

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 $\mathbf{d}_r = \mathbf{e}_r^{(R)}$. From $\mathbf{X}\mathbf{w}_r = \mathbf{A}\mathbf{S}^T\mathbf{w}_r = \mathbf{A}\mathbf{d}_r = \mathbf{a}_r$ we conclude that

$$a_{lr}^{(+,n)} = \mathbf{e}_{p_{l,n}}^{(I)T}\mathbf{X}\mathbf{w}_r \quad \text{and} \quad a_{lr}^{(-,n)} = \mathbf{e}_{s_{l,n}}^{(I)T}\mathbf{X}\mathbf{w}_r. \tag{103}$$

Plugging (103) into (35) yields

$$a_{1r}^{(+,n)} \cdots a_{Lr}^{(+,n)} - a_{1r}^{(-,n)} \cdots a_{Lr}^{(-,n)} = 0 \Leftrightarrow \mathbf{p}_L^{(n)T} \cdot \big(\underbrace{\mathbf{w}_r \otimes \cdots \otimes \mathbf{w}_r}_{L \text{ times}}\big) = 0, \tag{104}$$

where

$$\begin{aligned}
\mathbf{p}_L^{(n)T} &:= ((\mathbf{e}_{p_{1,n}}^{(I)T}\mathbf{X}) \otimes \cdots \otimes (\mathbf{e}_{p_{L,n}}^{(I)T}\mathbf{X})) - ((\mathbf{e}_{s_{1,n}}^{(I)T}\mathbf{X}) \otimes \cdots \otimes (\mathbf{e}_{s_{L,n}}^{(I)T}\mathbf{X})) \\
&= \Big((\mathbf{e}_{p_{1,n}}^{(I)T}\mathbf{A}) \otimes \cdots \otimes (\mathbf{e}_{p_{L,n}}^{(I)T}\mathbf{A}) - (\mathbf{e}_{s_{1,n}}^{(I)T}\mathbf{A}) \otimes \cdots \otimes (\mathbf{e}_{s_{L,n}}^{(I)T}\mathbf{A})\Big) (\mathbf{S}^T \otimes \cdots \otimes \mathbf{S}^T) \\
&= \boldsymbol{\psi}^{(n,L)}(\mathbf{S}^T \otimes \cdots \otimes \mathbf{S}^T) \in \mathbb{C}^{R^L},
\end{aligned}$$

in which relations (92) and (93) were used. Stacking yields

$$\mathbf{P}^{(N,L)} \cdot (\mathbf{w}_r \otimes \cdots \otimes \mathbf{w}_r) = \mathbf{0}\,, \tag{105}$$

where

$$\mathbf{P}^{(N,L)} = \begin{bmatrix} \mathbf{p}_L^{(1)} \\ \vdots \\ \mathbf{p}_L^{(N)} \end{bmatrix} = \mathbf{\Psi}^{(N,L)}(\mathbf{S}^T \otimes \cdots \otimes \mathbf{S}^T) \in \mathbb{C}^{N \times R^L}, \tag{106}$$

where $\mathbf{\Psi}^{(N,L)}$ is given by (92).

Step 2: Computation of $\boldsymbol{Q}$ whose columns form a basis for ker($\boldsymbol{P}^{(N,L)}) \cap \pi_S^{(L)}$. 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 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

$$\mathbf{Q} = (\mathbf{W} \odot \cdots \odot \mathbf{W})\mathbf{F}^T. \tag{107}$$

Lemma 5.5 below states that $\dim(\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$ or equivalently $\mathbf{G}_{\text{MEC}}^{(N,L)}$ has full column rank, then $\dim(\ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}) = R$ and the columns of $\mathbf{Q}$ in (107) indeed form a basis for $\ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}$.

Lemma 5.5. *Assume that $\boldsymbol{S}$ in* (3) *is nonsingular. If condition* (99) *in Theorem 5.3 is satisfied (or equivalently $\boldsymbol{G}_{MEC}^{(N,L)}$ has full column rank), then dim(ker($\boldsymbol{P}^{(N,L)}) \cap \pi_S^{(L)}) = R$.*

Proof. Since we assume that $\dim(\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$ (or equivalently $\mathbf{G}_{\text{MEC}}^{(N,L)}$ has full column rank; see Proposition 5.4), we know that the columns of $\mathbf{D} \odot \cdots \odot \mathbf{D}$ with $\mathbf{D} = \mathbf{S}^T\mathbf{W}$ form a basis for $\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$. (See Eqs. (100)–(101).) Let the columns of $\mathbf{B}$ form an alternative basis for $\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$, then there exists a nonsingular matrix $\mathbf{F} \in \mathbb{C}^{R\times R}$ such that $\mathbf{B} = (\mathbf{D} \odot \cdots \odot \mathbf{D})\mathbf{F}^T$. Since $\mathbf{S}$ is nonsingular, we know from (106) that $\mathbf{x} \in \ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}$ if and only if $(\mathbf{S}^{-T} \otimes \cdots \otimes \mathbf{S}^{-T})\mathbf{x} \in \ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}$. Hence, if the columns of $\mathbf{Q}$ form a basis for $\ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}$, then

$$\begin{aligned}\mathbf{Q} &= (\mathbf{S}^{-T} \otimes \cdots \otimes \mathbf{S}^{-T})\mathbf{B} \\ &= (\mathbf{S}^{-T} \otimes \cdots \otimes \mathbf{S}^{-T})(\mathbf{D} \odot \cdots \odot \mathbf{D})\mathbf{F}^T \\ &= (\mathbf{W} \odot \cdots \odot \mathbf{W})\mathbf{F}^T,\end{aligned}$$

where $\mathbf{W} = \mathbf{S}^{-T}\mathbf{D}$. Since $\text{rank}(\mathbf{Q}) = \text{rank}(\mathbf{W} \odot \cdots \odot \mathbf{W}) = \text{rank}(\mathbf{F}) = R$, we conclude that $\dim(\ker(\mathbf{\Psi}^{(N,L)}) \cap \pi_S^{(L)}) = R$ implies that $\dim(\ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}) = R$ when $\mathbf{S}$ is nonsingular. □

Step 3: Build tensor $\mathcal{Q}$ from $\boldsymbol{Q}$. It is clear that relation (107) corresponds to a matrix representation of the CPD of a tensor $\mathcal{Q} \in \mathbb{C}^{R\times\cdots\times R}$ of order $L+1$:

$$\mathcal{Q} = \sum_{r=1}^{R} \mathbf{w}_r \circ \cdots \circ \mathbf{w}_r \circ \mathbf{f}_r\,. \tag{108}$$

Step 4: Obtain $\boldsymbol{W}$ from CPD of $\mathcal{Q}$. Since all the factor matrices of the CPD in (108) have full column rank, we know that $\mathbf{W}$ can be recovered from (108) via an EVD (e.g., [3, 38]).

Step 5: Obtain $\boldsymbol{A}$ and $\boldsymbol{S}$. Once $\mathbf{W}$ has been obtained, we can immediately compute $\mathbf{A} = \mathbf{X}\mathbf{W}$ and $\mathbf{S} = \mathbf{W}^{-T}$. This also implies the uniqueness of $\mathbf{A} = \mathbf{X}\mathbf{W}$ and $\mathbf{S} = \mathbf{W}^{-T}$.

Theorem 5.6 below summarizes the above uniqueness result for bilinear factorizations, which is based on a constructive interpretation of Theorem 5.3.

Theorem 5.6. *Consider the bilinear factorization of* $\boldsymbol{X}$ *in* (3) *with* N *monomial relations of the form* (4)*. If*

$$\begin{cases} \boldsymbol{S} \text{ has full column rank,} \\ ker(\boldsymbol{P}^{(N,L)}) \cap \pi_S^{(L)} \text{ is an } R\text{-dimensional subspace,} \end{cases} \tag{109}$$

then $\boldsymbol{A}$ *and* $\boldsymbol{S}$ *in* (3) *are unique,* $\boldsymbol{A}$ *has full column rank, and the bilinear factorization of* $\boldsymbol{X}$ *can be reduced to the CPD of* $\mathcal{Q}$ *in* (108)*. The latter decomposition can be computed algebraically via an EVD.*

Proof. The result follows directly from the preceding Steps 1–5 and Lemma 5.5. □

The constructive use of Theorem 5.6 when used to compute the bilinear factorization of $\mathbf{X}$ in (3) is summarized as Algorithm 1.

Algorithm 1 Summary of constructive use of Theorem 5.6.

1. Build $\mathbf{P}^{(N,L)}$ in (106).
2. Obtain matrix $\mathbf{Q}$ whose columns form a basis for $\ker(\mathbf{P}^{(N,L)}) \cap \pi_S^{(L)}$.
3. Build tensor $\mathcal{Q}$ in (108) from $\mathbf{Q}$.
4. Obtain $\mathbf{W}$ from CPD of $\mathcal{Q}$.
5. Compute $\mathbf{A} = \mathbf{X}\mathbf{W}$ and $\mathbf{S} = \mathbf{W}^{-T}$.

6. Applications

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

A nice property of bilinear factorizations subject to monomial equality constraints is that they allow us to extend the CPD model (2) to binary weighted CPD (7) in which $\mathbf{E}^{(r)}$ in (8) 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 (8) extends to

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}^{(r)} \circ \mathbf{s}_r = \sum_{r=1}^{R} (\mathbf{D}^{(r)} * (\mathbf{a}_r \mathbf{b}_r^T)) \circ \mathbf{s}_r \tag{110}$$

for binary weighted CPD. We stress again that the binary "connectivity" matrices $\mathbf{D}^{(r)}$ are not known (fixed) *a priori*. From (10) and (110) it is clear that (7) 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 and a 2-by-2 submatrix of $\mathbf{E}^{(r)}$ can have rank two, 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

$$e_{i_1 j_1}^{(r)} e_{i_2 j_2}^{(r)} e_{i_1 j_2}^{(r)} e_{i_2 j_1}^{(r)} \cdot (e_{i_1 j_1}^{(r)} e_{i_2 j_2}^{(r)} - e_{i_1 j_2}^{(r)} e_{i_2 j_1}^{(r)}) = 0. \tag{111}$$

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 (35) with $L = 6$ and

$$\begin{cases} \mathbf{a}_r^{(+,n)} = [e_{i_1 j_1}^{(r)}, e_{i_2 j_2}^{(r)}, e_{i_1 j_2}^{(r)}, e_{i_2 j_1}^{(r)}, e_{i_1 j_1}^{(r)}, e_{i_2 j_2}^{(r)}]^T, \\ \mathbf{a}_r^{(-,n)} = [e_{i_1 j_1}^{(r)}, e_{i_2 j_2}^{(r)}, e_{i_1 j_2}^{(r)}, e_{i_2 j_1}^{(r)}, e_{i_1 j_2}^{(r)}, e_{i_2 j_1}^{(r)}]^T, \end{cases} \tag{112}$$

in which the superscript $'n'$ is associated with the tuple (i_1, i_2, j_1, j_2). From (111) it is also clear that the rank of the matrix $\mathbf{A}_6(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ will depend on the $(0,1)$-binary pattern of $\mathbf{D}^{(r)}$. The maximal rank of $\mathbf{A}_6(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ as a function of $(0,1)$-binary pattern of $\mathbf{D}^{(r)}$ is listed in Table 1.

$e_{i_1 j_1}^{(r)}$	x	0	x	x	x	x	0	x	0	0	x	x	0	0	0	0
$e_{i_2 j_2}^{(r)}$	x	x	0	x	x	x	0	0	x	x	0	0	x	0	0	0
$e_{i_1 j_2}^{(r)}$	x	x	x	0	x	0	x	x	0	x	0	0	0	x	0	0
$e_{i_2 j_1}^{(r)}$	x	x	x	x	0	0	x	0	x	0	x	0	0	0	x	0
Max rank of $\mathbf{A}_6(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$	5	5	5	5	5	4	4	3	3	3	3	2	2	2	2	0

Table 1: Max rank of $\mathbf{A}_6(\mathbf{a}_r^{(+,n)}, \mathbf{a}_r^{(-,n)})$ as a function of the zero-pattern of $[e_{i_1 j_1}^{(r)}, e_{i_2 j_2}^{(r)}, e_{i_1 j_2}^{(r)}, e_{i_2 j_1}^{(r)}]^T$. An '$x$' means that the corresponding entry is nonzero, e.g., $[0\ x\ 0\ 0]^T$ means that $e_{i_2 j_2}^{(r)}$ is nonzero.

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 (7) different from the uniqueness properties the CPD of $\mathcal{X}$ in (8). 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 (73) in Theorem 4.5 is satisfied, implying the uniqueness of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{S}$.

Remark. 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 (110) 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.

6.2. Application: Binary Matrix Factorization (BMF)

Consider the BMF (6). 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 (6).

6.2.1. Nonhomogeneous formulation

Let us first explain how Theorem 5.6 can be adapted to solve the BMF problem using a variant of the ACMA method [7] for blind separation of constant modulus signals. The property $a_{i,r} \in \{0,1\}$ can be expressed via the nonhomogeneous (polynomial) relation

$$a_{i,r}(a_{i,r} - 1) = a_{i,r}a_{i,r} - a_{i,r} = 0. \tag{113}$$

Following the procedure in Section 5.2, we obtain

$$a_{i,r}(a_{i,r} - 1) = a_{i,r}a_{i,r} - a_{i,r} = 0 \Leftrightarrow \left[\mathbf{p}_1^{(i)}, \mathbf{p}_2^{(i)}\right] \begin{bmatrix} \mathbf{w}_r \\ \mathbf{w}_r \otimes \mathbf{w}_r \end{bmatrix} = 0, \quad 1 \leq i \leq I, \tag{114}$$

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

$$\mathbf{P}_{\text{BMF}}^{(I)} \cdot \begin{bmatrix} \mathbf{w}_r \\ \mathbf{w}_r \otimes \mathbf{w}_r \end{bmatrix} = \mathbf{0}, \quad \text{where } \mathbf{P}_{\text{BMF}}^{(I)} = \begin{bmatrix} \mathbf{p}_1^{(1)} & \mathbf{p}_2^{(1)} \\ \vdots & \vdots \\ \mathbf{p}_1^{(I)} & \mathbf{p}_2^{(I)} \end{bmatrix} \in \mathbb{C}^{I \times (R+R^2)}. \tag{115}$$

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 (115) 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 $\mathbf{A} = \mathbf{X}\mathbf{W}$.

6.2.2. Homogeneous formulation

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 (114), we propose the following homogeneous (monomial) formulation of the binary constraint of $a_{ir} \in \{0,1\}$:

$$a_{i_1,r}a_{i_1,r}a_{i_2,r} - a_{i_1,r}a_{i_2,r}a_{i_2,r} = 0, \quad 1 \leq i_1 < i_2 \leq I. \tag{116}$$

From (116) it is clear that the BMF (6) can also be interpreted as a bilinear factorization of the form (3) involving $N = C_I^2$ monomial relations of degree $L = 3$,

$\mathbf{a}_r^{(+,n)} = [a_{i_1,r}, a_{i_1,r}, a_{i_2,r}]^T$ 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 (114), an interesting property of the homogeneous formulation (116) 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 (109) in Theorem 5.6 is satisfied when $\mathbf{P}^{(N,L)}$ is built from the C_I^2 equations associated with (116). In contrast, the nonhomogeneous approach based on (113) together with Theorem 5.6 cannot guarantee uniqueness for this BMF.

6.2.3. Numerical experiments

Let us end the section with an illustrative example. Consider (6) 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 $\mathrm{P}(\mathbf{S}) = \min_{\mathbf{\Pi\Lambda}} \|\mathbf{S} - \widehat{\mathbf{S}}\mathbf{\Pi\Lambda}\|_F / \|\mathbf{S}\|_F$, where $\mathbf{\Pi}$ and $\mathbf{\Lambda}$ denote a permutation matrix and a diagonal matrix, respectively. We compare the two algebraic methods associated with Theorem 5.6 when the nonhomogeneous (114) and homogeneous (116) approaches are used. The mean $\mathrm{P}(\mathbf{S})$ 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.

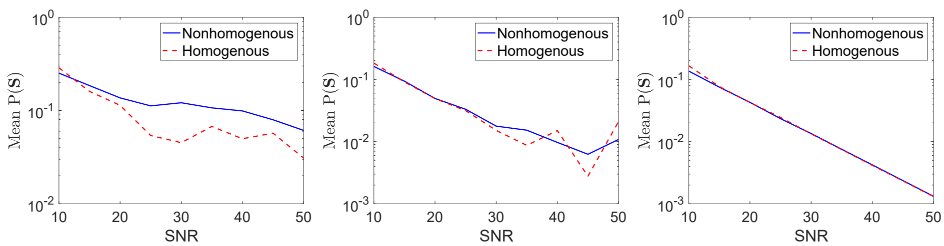

Figure 1: Mean $\mathrm{P}(\mathbf{S})$ values over 100 Monte Carlo runs when (left) $I = 20$, (middle) $I = 30$ and (right) $I = 40$.

7. Conclusion

In this paper we studied bilinear factorizations of which one of the factor matrices is subject to monomial equality constraints and the other has full column rank. We have explained that such bilinear 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 bilinear factorizations that allowed us to relax the monomial equality constraint into a low-rank constraint. By taking into account the specific structure that captures the monomial constraint, we have even reduced the bilinear factorization to a CPD. We demonstrated that BMFs can be interpreted as bilinear factorizations subject to monomial equality constraints. This led to an algorithm and relatively easy-to-check uniqueness conditions for BMF. Finally, the framework of bilinear factorizations subject to monomial equality constraints 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) Flemish Government: (i) This work was supported by the Fonds de la Recherche Scientifique–FNRS and the Fonds Wetenschappelijk Onderzoek–Vlaanderen under EOS Project no 30468160 (SeLMA), (ii) funding from the Flemish Government under the “Onderzoeksprogramma Artificiele Intelligentie (AI) Vlaanderen” programme, (2) KU Leuven Internal Funds C16/15/059 and ID-N project no 3E190402, (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.

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] C. Bocci, L. Chiantini, G. Ottaviani, Refined methods for the identifiability of tensors, Annali di Matematica 193 (2014) 1691–1702.
[26] I. Domanov, L. De Lathauwer, Generic uniqueness of a structured matrix factorization and applications in blind source separation, IEEE Journal of Selected Topics in Signal Processing 10 (4) (2016) 701–711.
[27] P. Comon, Y. Qi, K. Usevich, Identifiability of an X-rank decomposition of polynomial maps, SIAM Journal on Applied Algebra and Geometry 1 (1) (2017) 388–414.
[28] M. Sørensen, N. D. Sidiropoulos, L. De Lathauwer, Canonical polyadic decomposition of a tensor that has missing fibers: A monomial factorization approach, in: Proc. of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 12-17, Brighton, United Kingdom, 2019.
[29] 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.
[30] 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.
[31] 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.
[32] 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.
[33] 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.
[34] 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.
[35] 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.
[36] M. Marcus, H. Minc, On the relation between the determinant and the permanent, Illinois J. Math. 5 (1961) 376–381.
[37] H. Minc, Permanents, Cambridge University Press, 1984.
[38] S. E. Leurgans, R. T. Ross, R. B. Abel, A decomposition of three-way arrays, SIAM J. Matrix Anal. Appl. 14 (1993) 1064–1083.
[39] 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.
[40] R. B. Bapat, Mixed discriminants of positive semidefinite matrices, Linear Algebra and its Applications 126 (1989) 107–124.
[41] R. B. Bapat, T. E. S. Raghavan, Nonnegative matrices and applications, Cambridge University Press, 1997.
[42] A. Stegeman, N. D. Sidiropoulos, On Kruskal's uniqueness condition for the CANDECOMP/PARAFAC decomposition, Linear Algebra and its Applications 420 (2007) 540–552.
[43] P. Comon, G. Golub, L.-H. Lim, B. Mourrain, Symmetric tensors and symmetric tensor rank, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 1254–1279.