Open access • Journal Article • DOI:10.1137/18M1206849

On Uniqueness and Computation of the Decomposition of a Tensor into Multilinear Rank-$(1,L_r,L_r)$ Terms — Source link

Ignat Domanov, Lieven De Lathauwer

Published on: 05 May 2020 - SIAM Journal on Matrix Analysis and Applications (Society for Industrial and Applied Mathematics)

Topics: Tensor (intrinsic definition), Rank (linear algebra), Multilinear map, Multilinear algebra and Uniqueness

ON UNIQUENESS AND COMPUTATION OF THE DECOMPOSITION OF A TENSOR INTO MULTILINEAR RANK-$(1, L_r, L_r)$ TERMS*

IGNAT DOMANOV† AND LIEVEN DE LATHAUWER†

Abstract. Canonical Polyadic Decomposition (CPD) represents a third-order tensor as the minimal sum of rank-1 terms. Because of its uniqueness properties the CPD has found many concrete applications in telecommunication, array processing, machine learning, etc. On the other hand, in several applications the rank-1 constraint on the terms is too restrictive. A multilinear rank-(M, N, L) constraint (where a rank-1 term is the special case for which $M = N = L = 1$) could be more realistic, while it still yields a decomposition with attractive uniqueness properties.

In this paper we focus on the decomposition of a tensor $\mathcal{T}$ into a sum of multilinear rank-$(1, L_r, L_r)$ terms, $r = 1, \dots, R$. This particular decomposition type has already found applications in wireless communication, chemometrics and the blind signal separation of signals that can be modelled as exponential polynomials and rational functions. We find conditions on the terms which guarantee that the decomposition is unique and can be computed by means of the eigenvalue decomposition of a matrix even in the cases where none of the factor matrices has full column rank. We consider both the case where the decomposition is exact and the case where the decomposition holds only approximately. We show that in both cases the number of the terms R and their "sizes" $L_1, \dots, L_R$ do not have to be known a priori and can be estimated as well. The conditions for uniqueness are easy to verify, especially for terms that can be considered "generic". In particular, we obtain the following two generalizations of a well known result on generic uniqueness of the CPD (i.e., the case $L_1 = \dots = L_R = 1$): we show that the multilinear rank-$(1, L_r, L_r)$ decomposition of an $I \times J \times K$ tensor is generically unique if i) $L_1 = \dots = L_R =: L$ and $R \le \min((J - L)(K - L), I)$ or if ii) $\sum L_R \le \min((I - 1)(J - 1), K)$ and $J \ge \max(L_i + L_j)$.

Key words. multilinear algebra, third-order tensor, block term decomposition, multilinear rank, signal separation, factor analysis, eigenvalue decomposition, uniqueness

AMS subject classifications. 15A23, 15A69

1. Introduction.

1.1. Terminology and problem setting.

Throughout the paper $\mathbb{F}$ denotes the field of real or complex numbers.

By definition, a tensor $\mathcal{T} = (t_{ijk}) \in \mathbb{F}^{I\times J\times K}$ is *multiLinear rank-$(1, L, L)$ (ML rank-$(1, L, L)$)* if it equals the outer product of a nonzero vector $\mathbf{a} \in \mathbb{F}^I$ and a rank-L matrix $\mathbf{E} = (e_{ij}) \in \mathbb{F}^{J\times K}$: $\mathcal{T} = \mathbf{a} \circ \mathbf{E}$, which means that $t_{ijk} = a_i e_{jk}$ for all values of indices. If it is only known that the rank of $\mathbf{E}$ is bounded by L, then we say that $\mathcal{T} = \mathbf{a} \circ \mathbf{E}$ is ML rank at most $(1, L, L)$ and write "$\mathcal{T}$ is max ML rank-$(1, L, L)$ ".

In this paper we study the *decomposition* of $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ into a sum of such terms

*Submitted to the editors DATE.

Funding: This work was funded by (1) Research Council KU Leuven: C1 project c16/15/059-nD; (2) the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme; (3) F.W.O.: project G.0830.14N, G.0881.14N, G.0F67.18N (EOS SeLMA); (4) EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no. 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

† Group Science, Engineering and Technology, KU Leuven - Kulak, E. Sabbelaan 53, 8500 Kortrijk, Belgium and Dept. of Electrical Engineering ESAT/STADIUS KU Leuven, Kasteelpark Arenberg 10, bus 2446, B-3001 Leuven-Heverlee, Belgium (ignat.domanov@kuleuven.be, lieven.delathauwer@kuleuven.be).

of max ML rank-$(1, L_r, L_r)$ [1]:

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{E}_r, \qquad \mathbf{a}_r \in \mathbb{F}^I \setminus \{\mathbf{0}\}, \qquad \mathbf{E}_r \in \mathbb{F}^{J\times K}, \qquad r_{\mathbf{E}_r} \le L_r, \tag{1.1}$$

where $\mathbf{0}$ denotes the zero vector and $r_{\mathbf{E}_r}$ denotes the rank of $\mathbf{E}_r$. If exactly $r_{\mathbf{E}_r} = L_r$ for all r, then we call (1.1) "the decomposition of $\mathcal{T}$ into a sum of 'ML rank-$(1, L_r, L_r)$ terms" or, briefly, its "ML rank-$(1, L_r, L_r)$ decomposition".

In this paper we study the uniqueness and computation of (1.1). For uniqueness we use the following basic definition.

Definition 1.1. *Let $L_1, \dots, L_R$ be fixed positive integers. The decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is* unique *if for any two decompositions of the form (1.1) one can be obtained from another by a permutation of summands.*

Thus, the uniqueness is not affected by the trivial ambiguities in (1.1): permutation of the max ML rank-$(1, L_r, L_r)$ terms and (nonzero) scaling/counterscaling $\lambda\mathbf{a}_r$ and $\lambda^{-1}\mathbf{E}_r$. Definition 1.1 implies that if the decomposition is unique, then it is necessarily minimal, that is, if (1.1) holds with $r_{\mathbf{E}_r} = L_r$, then a decomposition of the form (1.1) with smaller L_r does not exist, in particular, a decomposition with smaller number of terms does not exist.

We will not only investigate the "global" uniqueness of decomposition (1.1) but also particular instances of "partial" uniqueness. Let us call the matrix

$$\mathbf{A} = [\mathbf{a}_1 \ \dots \ \mathbf{a}_R]$$

the first factor matrix of the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms. For uniqueness of $\mathbf{A}$, we will resort to the following definition.

Definition 1.2. *Let $L_1, \dots, L_R$ be fixed positive integers. The first factor matrix of the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is* unique *if for any two decompositions of the form (1.1) their first factor matrices coincide up to column permutation and (nonzero) scaling.*

It follows from Definition 1.2 that if $\mathcal{T}$ admits a decomposition of the form (1.1) with fewer than R terms, then the first factor matrix is not unique. On the other hand, as a preview of one result, Example 2.15 will illustrate that the first factor matrix may be unique without the overall ML rank decomposition being unique.

Definitions 1.1 and 1.2 concern deterministic forms of uniqueness. We will also develop generic uniqueness results. To make the rank constraints $r_{\mathbf{E}_r} \le L_r$ in (1.1) easier to handle and to present the definition of generic uniqueness, we factorize $\mathbf{E}_r$ as $\mathbf{B}_r\mathbf{C}_r^T$, where the matrices $\mathbf{B}_r \in \mathbb{F}^{J\times L_r}$ and $\mathbf{C}_r \in \mathbb{F}^{K\times L_r}$ are rank at most L_r. Thus, (1.1) can be rewritten as

$$\begin{gathered}\mathcal{T} = \sum_{r=1}^{R} \mathbf{a}_r \circ (\mathbf{B}_r\mathbf{C}_r^T), \\ \mathbf{a}_r \in \mathbb{F}^I \setminus \{\mathbf{0}\},\ \mathbf{B}_r \in \mathbb{F}^{J\times L_r},\ \mathbf{C}_r \in \mathbb{F}^{K\times L_r},\ r_{\mathbf{B}_r} \le L_r,\ r_{\mathbf{C}_r} \le L_r,\ r = 1, \dots, R.\end{gathered} \tag{1.2}$$

[1] The results of this paper can also be applied for the decomposition into a sum of max ML rank-$(L_r, 1, L_r)$ (resp. -$(L_r, L_r, 1)$) terms by switching the first and second (resp. third) dimensions of $\mathcal{T}$.

Throughout the paper, we set

$$\mathbf{B} = [\mathbf{B}_1 \ \dots \ \mathbf{B}_R] \in \mathbb{F}^{J\times \sum L_r}, \quad \mathbf{B}_r = [\mathbf{b}_{1,r} \ \dots \ \mathbf{b}_{L_r,r}] = (b_{jl,r})_{j,l=1}^{J,L_r}$$
$$\mathbf{C} = [\mathbf{C}_1 \ \dots \ \mathbf{C}_R] \in \mathbb{F}^{K\times \sum L_r}, \quad \mathbf{C}_r = [\mathbf{c}_{1,r} \ \dots \ \mathbf{c}_{L_r,r}] = (c_{kl,r})_{k,l=1}^{K,L_r}.$$

We call the matrices $\mathbf{B}$ and $\mathbf{C}$ *the second and third factor matrix* of $\mathcal{T}$, respectively. Decomposition (1.2) can then be represented in matrix form as

$$\mathbf{T}_{(1)} := [\text{vec}(\mathbf{H}_1) \ \dots \ \text{vec}(\mathbf{H}_I)] = [\text{vec}(\mathbf{E}_1) \ \dots \ \text{vec}(\mathbf{E}_R)]\mathbf{A}^T, \tag{1.3}$$

$$\mathbf{T}_{(2)} := [\mathbf{H}_1 \ \dots \ \mathbf{H}_I]^T = [\mathbf{a}_1 \otimes \mathbf{C}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{C}_R]\mathbf{B}^T = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{E}_r^T, \tag{1.4}$$

$$\mathbf{T}_{(3)} := [\mathbf{H}_1^T \ \dots \ \mathbf{H}_I^T]^T = [\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R]\mathbf{C}^T = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{E}_r, \tag{1.5}$$

where $\mathbf{H}_1, \dots, \mathbf{H}_I \in \mathbb{F}^{J\times K}$ denote the horizontal slices of $\mathcal{T}$, $\mathbf{H}_i := (t_{ijk})_{j,k=1}^{J,K}$, $\text{vec}(\mathbf{H}_i)$ denotes the $JK \times 1$ column vector obtained by stacking the columns of the matrix $\mathbf{H}_i$ on top of one another, and "$\otimes$" denotes the Kronecker product. The matrices $\mathbf{T}_{(1)} \in \mathbb{F}^{JK\times I}$, $\mathbf{T}_{(2)} \in \mathbb{F}^{IK\times J}$, and $\mathbf{T}_{(3)} \in \mathbb{F}^{IJ\times K}$ are called *the matrix unfoldings*[2] of $\mathcal{T}$. One can easily verify that $\mathcal{T}$ is ML rank-$(1, L, L)$ if and only if $r_{\mathbf{T}_{(1)}} = 1$ and $r_{\mathbf{T}_{(2)}} = r_{\mathbf{T}_{(3)}} = L$.

We have now what we need to formally define generic uniqueness.

Definition 1.3. *Let $L_1, \dots, L_R$ be fixed positive integers and let μ be a measure on $\mathbb{F}^{I\times R} \times \mathbb{F}^{J\times \sum L_r} \times \mathbb{F}^{K\times \sum L_r}$ that is absolutely continuous with respect to the Lebesgue measure. The decomposition of an $I \times J \times K$ tensor into a sum of max ML rank-$(1, L_r, L_r)$ terms is* generically unique *if*

$$\mu\{(\mathbf{A}, \mathbf{B}, \mathbf{C}) : \quad \textit{decomposition (1.2) is not unique}\} = 0.$$

Thus, if the entries of the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are randomly sampled from an absolutely continuous distribution, then generic uniqueness means uniqueness that holds with probability one.

If $L_1 = \dots = L_R = 1$, then the minimal decomposition of the form (1.1) is known as the Canonical Polyadic Decomposition (CPD) (aka CANDECOMP/PARAFAC). Because of their uniqueness properties both CPD and decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms have found many concrete applications in telecommunication, array processing, machine learning, etc. [25, 9, 10, 31]. For the decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms we mention in particular applications in wireless communication [14], chemometrics [4] and blind signal separation of signals that can be modeled as exponential polynomials [13] and rational functions [15]. Some advantages of a blind separation method that relies on decomposition of the form (1.1) over the methods that rely on PCA, ICA, and CPD are discussed in [9, 31]. As a matter of fact, it is a profound advantage of the tensor setting over the common vector/matrix setting that data components do not need to be rank-1 to admit a unique recovery, i.e., terms such as the ones in (1.1) allow us to model more general contributions to observed data. It is also worth noting that if $R \leq I$, then (1.1) can

[2] Some papers, e.g., [25], define the matrix unfoldings as the transposed matrices $\mathbf{T}_{(1)}^T$, $\mathbf{T}_{(2)}^T$, and $\mathbf{T}_{(3)}^T$.

reformulated as a problem of finding a basis consisting of low-rank matrices, namely the basis $\{\mathbf{E}_1, \dots, \mathbf{E}_R\}$ of the matrix subspace spanned by the horizontal slices of $\mathcal{T}$, $\text{span}\{\mathbf{H}_1, \dots, \mathbf{H}_I\}$ [28].

In this paper we find conditions on the factor matrices which guarantee that the decomposition of a tensor into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique (in the deterministic or in the generic sense). We also derive conditions under which, perhaps surprisingly, the decomposition can essentially be computed by means of a matrix eigenvalue decomposition (EVD). This will be possible even in cases where none of the factor matrices has full column rank. The main results are formulated in Theorems 2.5, 2.6, 2.13, 2.16 and 2.17 below. Table 1.1 summarizes known and new[3] results for generic decompositions. By way of comparison, the known results guarantee that the decomposition of an $8 \times 8 \times 50$ tensor into a sum of $R-1$ ML rank-$(1, 1, 1)$ terms and one ML rank-$(1, 2, 2)$ term is generically unique up to $R \leq 8$ (row 3) and can be computed by means of EVD up to $R \leq 7$ (rows 1 and 2), while the results obtained in the paper imply that generic uniqueness holds up to $R \leq 48$ (row 8) and that computation is possible up to $R \leq 39$ (row 6).

A final word of caution is in order. It may happen that a tensor admits more than one decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms among which only one is exactly ML rank-$(1, L_r, L_r)$ (see Example 2.8 below). In this case one can thus say that the ML rank-$(1, L_r, L_r)$ decomposition of the tensor is unique. In this paper however, we will always present conditions for uniqueness of the decomposition into a sum of *max* ML rank-$(1, L_r, L_r)$ terms. It is clear that such conditions imply also uniqueness of the (exactly) ML rank-$(1, L_r, L_r)$ decomposition.

Throughout the paper $\mathbf{O}$, $\mathbf{I}$, and $\mathbf{I}_n$ denote the zero matrix, the identity matrix, and the specific identity matrix of size $n \times n$, respectively; $\text{Null}(\cdot)$ denotes the null space of a matrix; "^{T}", "^{H}", and "†" denote the transpose, hermitian transpose, and pseudo-inverse, respectively. We will also use the shorthand notations $\sum L_r$, $\sum d_r$, and $\min L_r$ for $\sum\limits_{r=1}^{R} L_r$, $\sum\limits_{r=1}^{R} d_r$, and $\min\limits_{1 \leq r \leq R} L_r$, respectively.

All numerical experiments in the paper were performed in MATLAB R2018b. To make the results reproducible, the random number generator was initialized using the built-in function `rng('default')` (the Mersenne Twister with seed 0).

1.2. Organization of the paper. In subsection 1.3 we remind known results on the decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms (subsection 1.3.1) and introduce auxiliary results on uniqueness and computation of the special case of the (approximate) symmetric joint block diagonalization problem (subsection 1.3.2). The results of subsection 1.3.2 are essential for understanding the algorithm for computation of the decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms (Algorithm 2.1). The reader who is interested only in results on uniqueness, and not in the computation of the decomposition, can safely skip subsection 1.3.2. The main results of the paper are presented in section 2: subsections 2.1 to 2.4 are preparatory and contain, respectively, necessary conditions for uniqueness, explanation of the key idea behind our derivation, some technical notations, and a technical convention that facilitates the presentation; the actual main results are formulated in subsection 2.5 and subsection 2.6 (see Table 1.1(b)). To make the paper easier to follow some technical notations were moved to a dedicated section 3. For the same reason, long proofs we moved to a dedicated section 4 and appendixes. We conclude the paper in section 5.

[3] One of the new results, namely, the part of statement 4) in Theorem 2.13 that relies on the assumption $I \geq R$, is not mentioned in the table because its presentation requires additional notations.

Table 1.1

Known and some of the new bounds on R and $L_1, \dots, L_R$ under which the decomposition of an $I \times J \times K$ tensor into a sum of max ML rank-$(1, L_r, L_r)$ terms is generically unique, where $\min(I, J, K, R) \geq 2$. Additional bounds can be obtained by switching J and K in rows 2, 5, 6, and 8. The boxed line in each cell with bounds indicates which factor matrices are required to have full column rank (f.c.r). (Since we are in the generic setting, full column rank of the first, second, and third factor matrix is equivalent to $I \geq R$, $J \geq \sum L_r$, and $K \geq \sum L_r$, respectively.) The check mark in the "λ"-column indicates that the result on uniqueness comes with an EVD based algorithm. The bounds in rows 4 and 6 hold upon verification that a particular matrix has full column rank. For row 4 no exceptions have been reported. We have verified the bounds in row 6 for $\max(I, J) \leq 5$. For the case where not all L_r are identical we found three exceptions in which the matrix does not have full column rank; for the case $L_1 = \dots = L_R = L$ we haven't found exceptions. (For more details on the bounds in row 6 see Appendix A). The bounds in row 8 imply that generic uniqueness does hold for two of three exceptions.

(a) Known bounds (subsection 1.3.1)

#	ref	$L_1 \leq \dots \leq L_R$	$L_1 = \dots = L_R =: L$	λ
1	[12]	$\boxed{J \geq \sum L_r,\ K \geq \sum L_r}$	$\boxed{J \geq RL,\ K \geq RL}$	✓
2	[21]	$\boxed{I \geq R,\ J \geq \sum L_r}$ $K \geq L_R + 1$	$\boxed{I \geq R,\ J \geq RL}$ $K \geq L + 1$	✓
3	[12]	$\boxed{I \geq R}$ $J \geq L_p + \dots + L_R$ and $K \geq L_q + \dots + L_R$,	$\boxed{I \geq R}$ $\min(\lfloor \frac{J}{L} \rfloor, R) + \min(\lfloor \frac{K}{L} \rfloor, R) \geq R + 2$, where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x	
4	[32]	*not applicable*	*(upon verification)* $\boxed{I \geq R}$ $C_J^{L+1} C_K^{L+1} \geq C_{R+L}^{L+1} - R$	✓

(b) New bounds (subsection 2.6)

#	ref	$L_1 \leq \dots \leq L_R$	$L_1 = \dots = L_R =: L$	λ
5	Theorem 2.12	$\boxed{\text{no f.c.r. assumptions}}$ $K \geq L_2 + \dots + L_R + 1$ and $J \geq L_{\min(I,R)-1} + \dots + L_R$	$\boxed{\text{no f.c.r. assumptions}}$ $K \geq (R-1)L + 1$ and $J \geq (R - \min(R, I) + 2)L$	✓
6	Theorem 2.13 4)	*(upon verification)* $\boxed{K \geq \sum L_r}$ $J \geq L_{R-1} + L_R$ and $C_I^2 C_J^2 \geq \sum_{r_1 \leq r_2} L_{r_1} L_{r_2}$	*(upon verification)* $\boxed{K \geq RL}$ $J \geq 2L$ and $C_I^2 C_J^2 \geq C_R^2 L^2$	✓
	verification mechanism is explained in Appendix A	exceptions for $\max(I, J) \leq 5$: 3 tuples $(I, J, R, L_1, \dots, L_R)$ with $L_1 = \dots, L_{R-1} = 1$, $L_R = 4$, $J = 5$, and $(I, R) \in \{(2, 3), (4, 9), (5, 12)\}$	there are no exceptions for $\max(I, J) \leq 5$	
7	Theorem 2.16	*not applicable*	$\boxed{I \geq R}$ $(J - L)(K - L) \geq R$	
8	Theorem 2.17	$\boxed{K \geq \sum L_r}$ $J \geq L_{R-1} + L_R$ and $(I-1)(J-1) \geq \sum L_r$	$\boxed{K \geq RL}$ $J \geq 2L$ and $(I-1)(J-1) \geq RL$	

1.3. Previous results.

1.3.1. Results on decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms. In the following two theorems it is assumed that at least two factor matrices have full column rank. The first result is well-known. Its proof is essentially obtained by picking two generic mixtures of slices of $\mathcal{T}$ and computing their generalized EVD. The values $L_1, \dots, L_R$ need not be known in advance and can be found as multiplicities of the eigenvalues.

THEOREM 1.4. *[12, Theorem 4.1] Let $\mathcal{T}$ admit decomposition (1.2). Assume that any two columns of $\mathbf{A}$ are linearly independent and that the matrices $\mathbf{B}$ and $\mathbf{C}$ have full column rank. Then the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique and can be computed by means of EVD. Moreover, any decomposition of $\mathcal{T}$ into a sum of $\hat{R}$ terms of max ML rank-$(1, \hat{L}_{\hat{r}}, \hat{L}_{\hat{r}})$ for which $\sum\limits_{\hat{r}=1}^{\hat{R}} \hat{L}_{\hat{r}} = \sum\limits_{r=1}^{R} L_r$ should necessarily coincide with decomposition (1.2).*

THEOREM 1.5. *[21, Corollary 1.4] Let $\mathcal{T}$ admit ML rank-$(1, L_r, L_r)$ decomposition (1.2) and let at least one of the following assumptions hold:*
- *a) $\mathbf{A}$ and $\mathbf{B}$ have full column rank and $r_{[\mathbf{C}_i\ \mathbf{C}_j]} \geq \max(L_i, L_j) + 1$ for all $1 \leq i < j \leq R$;*
- *b) $\mathbf{A}$ and $\mathbf{C}$ have full column rank and $r_{[\mathbf{B}_i\ \mathbf{B}_j]} \geq \max(L_i, L_j) + 1$ for all $1 \leq i < j \leq R$.*

Then the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique and can be computed by means of EVD.

The uniqueness and computation of the decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms, where $L_1 = \dots = L_R := L$, was also studied in [32, Subsection 5.2] and [29]. We do not reproduce the results from [32] (resp. [29]) here because this would require many specific notations. We just mention that one of the assumptions in [32] (resp. [29]) is that the first factor matrix (resp. the second or third factor matrix) has full column rank and another assumption implies that the dimensions of $\mathcal{T}$ satisfy the inequality $C^{L+1}_{\min(J,RL)} C^{L+1}_{\min(K,RL)} \geq C^{L+1}_{R+L} - R$ (resp. the inequality $C^2_{\min(I,R)} C^2_{\min(J,K,LR)} \geq C^2_R L^2$), where C^k_n denotes the binomial coefficient

$$C^k_n := \frac{n!}{k!(n-k)!}.$$

To present the next result we need the definitions of k-rank of a matrix ("k" refers to J.B. Kruskal) and k'-rank of a block matrix.

DEFINITION 1.6. *The k-rank of the matrix $\mathbf{A} = [\mathbf{a}_1\ \dots\ \mathbf{a}_R]$ is the largest number $k_{\mathbf{A}}$ such that any $k_{\mathbf{A}}$ columns of $\mathbf{A}$ are linearly independent.*

DEFINITION 1.7. *[12, Definition 3.2] The k'-rank of the matrix $\mathbf{B} = [\mathbf{B}_1\ \dots\ \mathbf{B}_R]$ is the largest number $k'_{\mathbf{B}}$ such that any set $\{\mathbf{B}_i\}$ of $k'_{\mathbf{B}}$ blocks of $\mathbf{B}$ yields a set of linearly independent columns.*

In the following theorem none of the factor matrices is required to have full column rank.

THEOREM 1.8. *[12, Lemma 4.2] Let $\mathcal{T}$ admit ML rank-$(1, L_r, L_r)$ decomposition (1.2) with $L_1 = \dots = L_R$. Assume that*

$$k_{\mathbf{A}} + k'_{\mathbf{B}} + k'_{\mathbf{C}} \geq 2R + 2.$$

Then the first factor matrix in the max ML rank-$(1, L_r, L_r)$ decomposition of $\mathcal{T}$ is unique. If additionally, $r_{\mathbf{A}} = R$, then the overall max ML rank-$(1, L_r, L_r)$ decomposition of $\mathcal{T}$ is unique.

In the following theorem we summarize the known results on generic uniqueness of the decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms. Statements 1), 2)-3), and 4) are just generic counterparts of Theorem 1.4, Theorem 1.5, and Theorem 1.8, respectively. Some of the statements have also appeared in [12, 21, 37, 38].

THEOREM 1.9. *Let $L_1 \leq \cdots \leq L_R$. Then each of the following conditions implies that the decomposition of an $I \times J \times K$ tensor into a sum of max ML rank-$(1, L_r, L_r)$ terms is generically unique:*

1) $I \geq 2$, $J \geq \sum L_r$, and $K \geq \sum L_r$;

2) $I \geq R$, $J \geq \sum L_r$, and $K \geq L_R + 1$;

3) $I \geq R$, $J \geq L_R + 1$, and $K \geq \sum L_r$;

4) $I \geq R$ and $k'_{\mathbf{B},gen} + k'_{\mathbf{C},gen} \geq R + 2$, where

$$k'_{\mathbf{B},gen} := \max\{p : \ L_{R-p+1} + \cdots + L_R \leq J\},$$
$$k'_{\mathbf{C},gen} := \max\{q : \ L_{R-q+1} + \cdots + L_R \leq K\}.$$

1.3.2. An auxiliary result on symmetric joint block diagonalization problem. In subsection 2.5 we will establish a link between decomposition (1.1) and a special case of the Symmetric Joint Block Diagonalization (S-JBD) problem introduced in this subsection. In particular, we will show in subsection 2.5 that uniqueness and computation of the first factor matrix in (1.1) follow from uniqueness and computation of a solution of the S-JBD problem. We will consider both the cases where decomposition (1.1) is exact and the case where the decomposition holds only approximately. In the latter case, decomposition (1.1) is just fitted to the given tensor $\mathcal{T}$. Thus, in this subsection, we also consider both the cases where the S-JBD is exact and the case where the S-JBD holds approximately.

Exact S-JBD. Let $\mathbf{V}_1, \ldots, \mathbf{V}_Q$ be $K \times K$ symmetric matrices that can be jointly block diagonalized as

$$\begin{gathered} \mathbf{V}_q = \mathbf{N}\mathbf{D}_q\mathbf{N}^T, \quad \mathbf{N} = [\mathbf{N}_1 \ \ldots \ \mathbf{N}_R], \quad \mathbf{N}_r \in \mathbb{F}^{K \times d_r}, \\ \mathbf{D}_q = \operatorname{blockdiag}(\mathbf{D}_{1,q}, \ldots, \mathbf{D}_{R,q}), \quad \mathbf{D}_{r,q} = \mathbf{D}_{r,q}^T \in \mathbb{F}^{d_r \times d_r}, \quad q = 1, \ldots, Q, \end{gathered} \tag{1.6}$$

where $d_1, \ldots, d_R, Q$ are positive integers, and $\operatorname{blockdiag}(\mathbf{D}_{1,q}, \ldots, \mathbf{D}_{R,q})$ denotes a block-diagonal matrix with the matrices $\mathbf{D}_{1,q}, \ldots, \mathbf{D}_{R,q}$ on the diagonal. It is worth noting that the columns of $\mathbf{N}$ are not required to be orthogonal and that we deal with the non-hermitian transpose in (1.6) even if $\mathbb{F} = \mathbb{C}$. Let $\mathbf{\Pi}$ be a $\sum d_r \times \sum d_r$ permutation matrix such that $\mathbf{N}\mathbf{\Pi}$ admits the same block partitioning as $\mathbf{N}$ and let $\mathbf{D}$ be a nonsingular symmetric block diagonal matrix whose diagonal blocks have dimensions $d_1, \ldots, d_R$. Then obviously $\mathbf{V}_1, \ldots, \mathbf{V}_Q$ can also be jointly block diagonalized as

$$\mathbf{V}_q = (\mathbf{N}\mathbf{D}\mathbf{\Pi})(\mathbf{\Pi}^T\mathbf{D}^{-1}\mathbf{D}_q\mathbf{D}^{-T}\mathbf{\Pi})(\mathbf{N}\mathbf{D}\mathbf{\Pi})^T =: \tilde{\mathbf{N}}\tilde{\mathbf{D}}_q\tilde{\mathbf{N}}^T, \quad q = 1, \ldots, Q.$$

We say that the solution of the S-JBD problem (1.6) *is unique*, if for any two solutions

$$\mathbf{V}_q = \mathbf{N}\mathbf{D}_q\mathbf{N}^T = \tilde{\mathbf{N}}\tilde{\mathbf{D}}_q\tilde{\mathbf{N}}^T, \qquad q = 1, \ldots, Q$$

there exist matrices $\mathbf{D}$ and $\mathbf{\Pi}$ such that

$$\tilde{\mathbf{N}} = \mathbf{N}\mathbf{D}\mathbf{\Pi}, \quad \tilde{\mathbf{D}}_q = \mathbf{\Pi}^T\mathbf{D}^{-1}\mathbf{D}_q\mathbf{D}^{-T}\mathbf{\Pi}, \quad q = 1, \ldots, Q.$$

Thus, if the solution of (1.6) is unique, then the number of blocks R in (1.6) is minimal and the column spaces of $\mathbf{N}_1, \dots, \mathbf{N}_R$ (as well as their dimensions $d_1, \dots, d_R$) can be identified up to permutation. For a thorough study of JBD we refer to [5] and the references therein.

In subsection 2.5 we will rework (1.2) into a problem of the form (1.6). In the case $d_1 = \cdots = d_R = 1$ the S-JBD problem (1.6) is reduced to a special case of the classical symmetric joint diagonalization (S-JD) problem (a.k.a. simultaneous diagonalization by congruence), where "special" means that the number of matrices Q equals the size R of the diagonal matrices. It is well known and can easily be derived from [24, Theorem 4.5.17] that if there exists a rank-R linear combination of $\mathbf{V}_1, \dots, \mathbf{V}_Q$, then the solution of S-JD is unique and can be computed by means of (simultaneous) EVD. The following theorem states that a similar result also holds for S-JBD problem (1.6).

THEOREM 1.10. *Let $Q := C^2_{d_1+1} + \cdots + C^2_{d_R+1}$, $\min(d_1, \dots, d_R) \geq 2$ and let $\mathbf{V}_1, \dots, \mathbf{V}_Q$ be $K \times K$ symmetric matrices that can be jointly block diagonalized as in (1.6). Assume that*

a) $\mathbf{N}$ has full column rank;

b) the matrices $\mathbf{D}_1, \dots, \mathbf{D}_Q$ are linearly independent.

Then the solution of S-JBD problem (1.6) is unique and can be computed by means of (simultaneous) EVD[4].

Proof. Let $\lambda_1, \dots, \lambda_Q \in \mathbb{F}$ be generic. Since Q is equal to the dimension of the subspace of all $\sum d_r \times \sum d_r$ symmetric block diagonal matrices, the block diagonal matrix $\sum \lambda_q \mathbf{D}_q$ in $\sum \lambda_q \mathbf{V}_q = \mathbf{N}(\sum \lambda_q \mathbf{D}_q)\mathbf{N}^T$ is also generic. Thus, replacing each equation in (1.6) by a (known) generic linear combination of all equations, we can assume without loss of generality (w.l.o.g.) that the matrices $\mathbf{D}_q$ are generic. By [21, Theorem 1.10], the solution of the obtained S-JBD problem is unique and can be computed by means of (simultaneous) EVD if we have at least 3 equations, which is the case since $Q \geq C^2_{2+1} = 3$. ∎

The algebraic procedure related to Theorem 1.10 is summarized in Algorithm 1.1 (see [5, Subsection 2.3] and [21, Algorithm 1 and Theorem 1.10]), where we assume w.l.o.g. that $K = \sum d_r$. The value R and the matrices $\mathbf{U}_1, \dots, \mathbf{U}_R$ in step 1 can be computed as follows. Vectorizing the matrix equation $\mathbf{O} = \mathbf{U}\mathbf{V}_q - \mathbf{V}_q\mathbf{U}^T$, we obtain that $\mathbf{0} = (\mathbf{V}_q^T \otimes \mathbf{I}) \operatorname{vec}(\mathbf{U}) - (\mathbf{I} \otimes \mathbf{V}_q) \operatorname{vec}(\mathbf{U}^T) = (\mathbf{V}_q^T \otimes \mathbf{I} - (\mathbf{I} \otimes \mathbf{V}_q)\mathbf{P}) \operatorname{vec}(\mathbf{U})$, where $\mathbf{P}$ denotes the $K^2 \times K^2$ permutation matrix that transforms the vectorized form of a $K \times K$ matrix into the vectorized form of its transpose. Let $\mathbf{M}$ denote the $K^2Q \times K^2$ matrix formed by the rows of $\mathbf{V}_q^T \otimes \mathbf{I} - (\mathbf{I} \otimes \mathbf{V}_q)\mathbf{P}$, $q = 1, \dots, Q$. Then we obtain $R = \dim \operatorname{Null}(\mathbf{M})$ and choose $\mathbf{U}_1, \dots, \mathbf{U}_R$ such that $\operatorname{vec}(\mathbf{U}_1), \dots \operatorname{vec}(\mathbf{U}_R)$ form a basis of $\operatorname{Null}(\mathbf{M})$.

It is worth noting that the computations in steps 1 and 2 can be simplified as follows. From the proof of Theorem 1.10 it follows that the matrices $\mathbf{V}_1, \dots, \mathbf{V}_Q$ in step 1 can be replaced by three generic linear combinations. It was also proved in [5] that the simultaneous EVD in step 2 can be replaced by the EVD of a single matrix $\mathbf{Z}$, namely, a generic linear combination of $\mathbf{U}_1, \dots, \mathbf{U}_R$. Then the values $d_1, \dots, d_R$ can be computed as the multiplicities of R (distinct) eigenvalues of $\mathbf{Z}$.

Approximate S-JBD. Optimization based schemes for the approximate S-JBD problem are discussed in the recent paper [6] (see also [5, 21, 35] and references therein). The authors of [5] proposed a variant of Algorithm 1.1 in which the null

[4]The simultaneous EVD problem consists of finding a similarity transform that reduces a set of (commuting) matrices to diagonal form.

Algorithm 1.1 Computation of S-JBD problem (1.6) under the conditions in Theorem 1.10

Input: $K \times K$ symmetric matrices $\mathbf{V}_1, \dots, \mathbf{V}_Q$ with the property that there exist matrices $\mathbf{N}$ and $\mathbf{D}_1, \dots, \mathbf{D}_Q$ such that $\mathbf{V}_1, \dots, \mathbf{V}_Q$ can be factorized as in (1.6), the assumptions in Theorem 1.10 hold and $K = \sum d_r$

1: Find R and the matrices $\mathbf{U}_1, \dots, \mathbf{U}_R$ that form a basis of the subspace
$$\{\mathbf{U} \in \mathbb{F}^{K\times K} : \ \mathbf{U}\mathbf{V}_q = \mathbf{V}_q\mathbf{U}^T, \ q = 1, \dots, Q\}$$

2: Find $\mathbf{N}$ and the values $d_1, \dots, d_R$ from the simultaneous EVD
$$\mathbf{U}_r = \mathbf{N}\,\text{blockdiag}(\lambda_{1r}\mathbf{I}_{d_1}, \dots, \lambda_{Rr}\mathbf{I}_{d_R})\mathbf{N}^{-1}, \qquad r = 1, \dots, R$$

3: For each $q = 1, \dots, Q$ compute $\mathbf{D}_q = \mathbf{N}^{-1}\mathbf{V}_q\mathbf{N}^{-T}$

Output: Matrices $\mathbf{N}$, $\mathbf{D}_1, \dots, \mathbf{D}_Q$ and the values R, $d_1, \dots, d_R$ such that (1.6) holds

space of $\mathbf{M}$ in step 1 is replaced[5] by the subspace spanned by the $\tilde{R} \le R$ smallest right singular vectors of $\mathbf{M}$, $\text{vec}(\mathbf{U}_1), \dots, \text{vec}(\mathbf{U}_{\tilde{R}})$, and the simultaneous EVD problem in step 2 is replaced by the EVD of single matrix $\mathbf{Z}$, where $\mathbf{Z}$ is a generic linear combination of $\mathbf{U}_1, \dots, \mathbf{U}_{\tilde{R}}$. The block-diagonal matrices $\mathbf{D}_q$ in step 3 can be found without explicitly computing the inverse of $\mathbf{N}$ by solving the linear set of equations $\mathbf{N}\mathbf{D}_q\mathbf{N}^T = \mathbf{V}_q$ in the least squares sense. Although the simultaneous EVD in step 2 is replaced by the EVD of a single matrix $\mathbf{Z}$, the experiments in [5] show that the proposed variant of Algorithm 1.1 may outperform optimization based algorithms. On the other hand, it is clear that the loss of "diversity" when replacing the $\tilde{R}$ matrices in step 2 by a single generic linear combination may result in a poor estimate of $\mathbf{N}$ and also in a wrong detection of $d_1, \dots, d_R$ (cf. also the discussion for CPD in [2]). That is why in this paper we will use the following (still simple but more robust) procedure to compute an approximate solution of the simultaneous EVD in step 2. (Note that the simultaneous EVD is (obviously) a new concept by itself, for which no dedicated numerical algorithms are available yet and their derivation is outside the scope of this paper.) First, we stack the matrices $\mathbf{U}_1, \dots, \mathbf{U}_{\tilde{R}}$ into an $\tilde{R} \times K \times K$ tensor $\mathcal{U}$ and interpret the simultaneous EVD in step 2 as a structured decomposition of $\mathcal{U}$ into a sum of ML rank-$(1, 1, 1)$ terms (i.e., just rank-1 terms):

$$(1.7) \qquad \mathcal{U} = \sum_{k=1}^{K} \mathbf{a}_k \circ (\mathbf{b}_k\mathbf{c}_k^T) \ \text{ or } \mathbf{U}_r = \mathbf{C}\,\text{diag}(a_{r1}, \dots, a_{rK})\mathbf{B}^T, \quad r = 1, \dots, \tilde{R},$$

where $\mathbf{B}^T = \mathbf{P}^T\mathbf{N}^{-1}$, $\mathbf{C} = \mathbf{N}\mathbf{P}$ (implying that $\mathbf{B} = \mathbf{C}^{-T}$),

$$(1.8) \qquad \text{diag}(a_{r1}, \dots, a_{rK}) = \mathbf{P}^T\,\text{blockdiag}(\lambda_{1r}\mathbf{I}_{d_1}, \dots, \lambda_{Rr}\mathbf{I}_{d_R})\mathbf{P}, \quad r = 1, \dots, \tilde{R}.$$

and $\mathbf{P}$ is an arbitrary permutation matrix. If $\mathbf{P} = \mathbf{I}_K$, then, by (1.8),

$$(1.9) \quad \mathbf{a}_1 = \dots = \mathbf{a}_{d_1} = [\lambda_{11} \ \dots \ \lambda_{1\tilde{R}}]^T, \mathbf{a}_{d_1+1} = \dots = \mathbf{a}_{d_1+d_2} = [\lambda_{21} \ \dots \ \lambda_{2\tilde{R}}]^T, \dots$$

If $\mathbf{P}$ is not the identity, then the vectors $\mathbf{a}_1, \dots, \mathbf{a}_K$ can be permuted such that (1.9) holds. It can easily be shown that, in the exact case, decomposition (1.7) is minimal, that is, (1.7) is a CPD of $\mathcal{U}$, and that the constraint $\mathbf{B} = \mathbf{C}^{-T}$ holds for any solution of (1.7).

[5] In noisy cases, the exact null space of $\mathbf{M}$ is always one-dimensional and spanned by the vectorized identity matrix.

There exist many optimization based algorithms that can compute the CPD of $\mathcal{U}$ in the least squares sense (see, for instance, [36]). Recall from Footnote 5 that, also in the noisy case, $\mathbf{U}_{\tilde{R}}$ can be taken equal to a scalar multiple of the identity matrix. This actually allows us to enforce the constraint $\mathbf{B} = \mathbf{C}^{-T}$ by setting $\mathbf{U}_{\tilde{R}} = \omega\mathbf{I}_K$, where ω is a weight coefficient chosen by the user. Finally, clustering the K vectors $\mathbf{a}_k \in \mathbb{F}^{\tilde{R}}$ into R clusters (modulo sign and scaling) we obtain the values $d_1, \dots, d_R$ as the sizes of clusters and also the permutation matrix $\mathbf{P}$. Then we set $\mathbf{N} = \mathbf{C}\mathbf{P}^T$.

2. Our contribution.

Before stating the main results (subsections 2.5 and 2.6), we present necessary conditions for uniqueness (subsection 2.1), explain the key idea behind our derivation (subsection 2.2), introduce some notations (subsection 2.3) and a convention (subsection 2.4).

2.1. Necessary conditions for uniqueness.

Let $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ admit the ML rank-$(1, L_r, L_r)$ decomposition (1.1). It was shown in [13, Theorem 2.4] that if the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique, then $\mathbf{A}$ does not have proportional columns (trivial) and the following condition holds:

$$\begin{gathered}\text{for every vector } \mathbf{w} \in \mathbb{F}^R \text{ that has at least two nonzero entries,}\\ \text{the rank of the matrix } \sum_{r=1}^{R} w_r\mathbf{E}_r \text{ is greater than } \max_{\{r: w_r\neq 0\}} L_r.\end{gathered} \tag{2.1}$$

In the following theorem we generalize well-known necessary conditions for uniqueness of the CPD (see [16] and references therein) to the decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms. The condition in statement 1) is more restrictive than (2.1) but is easier to check.

THEOREM 2.1. *Let $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ admit the ML rank-$(1, L_r, L_r)$ decomposition (1.2), i.e., $r_{\mathbf{B}_r} = r_{\mathbf{C}_r} = L_r$ for all r. If the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique, then the following statements hold:*

1) *the matrix $[\operatorname{vec}(\mathbf{E}_1)\ \dots\ \operatorname{vec}(\mathbf{E}_R)]$ has full column rank, where $\mathbf{E}_r := \mathbf{B}_r\mathbf{C}_r^T$ for all r;*
2) *the matrix $[\mathbf{a}_1 \otimes \mathbf{B}_1\ \dots\ \mathbf{a}_R \otimes \mathbf{B}_R]$ has full column rank;*
3) *the matrix $[\mathbf{a}_1 \otimes \mathbf{C}_1\ \dots\ \mathbf{a}_R \otimes \mathbf{C}_R]$ has full column rank.*

Proof. The three statements come from the three matrix representations (1.3), (1.5), and (1.4). The details of the proof are given in Appendix B. ∎

2.2. The key idea.

Let $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ admit the ML rank-$(1, L_r, L_r)$ decomposition (1.1), and let $\mathbf{T}_1, \dots, \mathbf{T}_K \in \mathbb{F}^{I\times J}$ denote the frontal slices of $\mathcal{T}$, $\mathbf{T}_k := (t_{ijk})_{i,j=1}^{I,J}$. It is clear that

$$f_1\mathbf{T}_1 + \dots + f_K\mathbf{T}_K = \sum_{k=1}^{K} f_k \sum_{r=1}^{R} \mathbf{a}_r\mathbf{e}_{k,r}^T = \sum_{r=1}^{R} \mathbf{a}_r \sum_{k=1}^{K} \mathbf{e}_{k,r}^T f_k = \sum_{r=1}^{R} \mathbf{a}_r(\mathbf{E}_r\mathbf{f})^T, \tag{2.2}$$

where $\mathbf{e}_{k,r}$ denotes the kth column of $\mathbf{E}_r$. Thus, if $\mathbf{f}$ belongs to the null space of all but one of the matrices $\mathbf{E}_1, ..., \mathbf{E}_R$, then $f_1\mathbf{T}_1 + \dots + f_K\mathbf{T}_K$ is rank-1 and its column space is spanned by a column of $\mathbf{A}$. We will make assumptions on $\mathbf{A}$ and $\mathbf{E}_1, \dots, \mathbf{E}_R$ that guarantee that the identity $f_1\mathbf{T}_1 + \dots + f_K\mathbf{T}_K = \mathbf{z}\mathbf{y}^T$ holds if and only if $\mathbf{z}$ is proportional to a column of $\mathbf{A}$ and $\mathbf{f}$ belongs to the null space of all matrices $\mathbf{E}_1, \ldots, \mathbf{E}_R$ but one:

$$(2.3)\quad f_1\mathbf{T}_1 + \cdots + f_K\mathbf{T}_K = \mathbf{z}\mathbf{y}^T \Leftrightarrow \exists r \text{ such that } \mathbf{z} = c\mathbf{a}_r,\ \mathbf{Z}_r\mathbf{f} = \mathbf{0} \text{ and } \mathbf{E}_r\mathbf{f} \neq \mathbf{0},$$
$$\text{where } \mathbf{Z}_r := [\mathbf{E}_1^T\ \ldots\ \mathbf{E}_{r-1}^T\ \mathbf{E}_{r+1}^T\ \ldots\ \mathbf{E}_R^T]^T.$$

In our algorithm we use $\mathcal{T}$ to construct a $C_I^2 C_J^2 \times K^2$ matrix $\mathbf{R}_2(\mathcal{T})$ such that the following equivalence holds true:

$$(2.4)\quad \mathbf{f} \in \mathbb{F}^K \text{ is a solution of } \mathbf{R}_2(\mathcal{T})(\mathbf{f} \otimes \mathbf{f}) = \mathbf{0} \quad \Leftrightarrow \quad r_{f_1\mathbf{T}_1+\cdots+f_K\mathbf{T}_K} \leq 1.$$

By (2.2)–(2.4), the set of all solutions of

$$(2.5)\qquad \mathbf{R}_2(\mathcal{T})(\mathbf{f} \otimes \mathbf{f}) = \mathbf{0}$$

is the union of the subspaces $\text{Null}\,(\mathbf{Z}_1), \ldots, \text{Null}\,(\mathbf{Z}_R)$ and any nonzero solution of (2.5) gives us a column of $\mathbf{A}$. We establish a link between (2.5) and S-JBD problem (1.6). By solving the S-JBD problem we will be able to find the subspaces $\text{Null}\,(\mathbf{Z}_1), \ldots, \text{Null}\,(\mathbf{Z}_R)$ and the entire factor matrix $\mathbf{A}$, which will then be used to recover the overall decomposition.

2.3. Construction of the matrix $\mathbf{R}_2(\mathcal{T})$ and its submatrix $\mathbf{Q}_2(\mathcal{T})$. In this subsection we present the explicit construction of the matrix $\mathbf{R}_2(\mathcal{T})$ in (2.4). In fact, the construction follows directly from (2.4). It is clear that

$$(2.6)\quad r_{f_1\mathbf{T}_1+\cdots+f_K\mathbf{T}_K} \leq 1 \quad \Leftrightarrow \quad \text{all } 2 \times 2 \text{ minors of } f_1\mathbf{T}_1 + \cdots + f_K\mathbf{T}_K \text{ are zero.}$$

Since there are $C_I^2 C_J^2$ minors and since each minor is a weighted sum of K^2 monomials $f_i f_j$, $1 \leq i, j \leq K$, the condition in the RHS of (2.6) can be rewritten as $\mathbf{R}_2(\mathcal{T})(\mathbf{f} \otimes \mathbf{f}) = \mathbf{0}$, where $\mathbf{R}_2(\mathcal{T})$ is a $C_I^2 C_J^2 \times K^2$ matrix whose entries are the second degree polynomials in the entries of $\mathcal{T}$. Variants of the following explicit construction of $\mathbf{R}_2(\mathcal{T})$ can be found in [11, 18, 32].

Definition 2.2. *The*

$$(2.7)\qquad \left((i_1 + C_{i_2-1}^2 - 1)C_J^2 + j_1 + C_{j_2-1}^2, (k_2 - 1)K + k_1\right)\text{-th}$$

entry of the $C_I^2 C_J^2 \times K^2$ matrix $\mathbf{R}_2(\mathcal{T})$ equals

$$(2.8)\qquad t_{i_1j_1k_1}t_{i_2j_2k_2} + t_{i_1j_1k_2}t_{i_2j_2k_1} - t_{i_1j_2k_1}t_{i_2j_1k_2} - t_{i_1j_2k_2}t_{i_2j_1k_1},$$

where

$$1 \leq i_1 < i_2 \leq I,\ 1 \leq j_1 < j_2 \leq J,\ 1 \leq k_1, k_2 \leq K.$$

Since the expression in (2.8) is invariant under the permutation $(k_1, k_2) \to (k_2, k_1)$, the $((k_2-1)K+k_1)$-th column of the matrix $\mathbf{R}_2(\mathcal{T})$ coincides with its $((k_1-1)K+k_2)$-th column. In other words, the rows of $\mathbf{R}_2(\mathcal{T})$ are vectorized $K \times K$ symmetric matrices, implying that C_{K-1}^2 columns of $\mathbf{R}_2(\mathcal{T})$ are repeated twice. Hence $\mathbf{R}_2(\mathcal{T})$ is of the form

$$(2.9)\qquad \mathbf{R}_2(\mathcal{T}) = \mathbf{Q}_2(\mathcal{T})\mathbf{P}_K^T,$$

where $\mathbf{Q}_2(\mathcal{T})$ holds the C_{K+1}^2 unique columns of $\mathbf{R}_2(\mathcal{T})$ and $\mathbf{P}_K^T \in \mathbb{F}^{C_{K+1}^2 \times K^2}$ is a binary (0/1) matrix with exactly one element equal to "1" per column. Formally, $\mathbf{Q}_2(\mathcal{T})$ is defined as follows.

Definition 2.3. *$\mathbf{Q}_2(\mathcal{T})$ denotes the $C_I^2C_J^2 \times C_{K+1}^2$ submatrix of $\mathbf{R}_2(\mathcal{T})$ formed by the columns with indices $(k_2-1)K+k_1$, where $1 \le k_1 \le k_2 \le K$.*

It can be easily checked that (2.9) holds for $\mathbf{P}_K$ defined by

$$(\mathbf{P}_K)_{(k_1-1)K+k_2,j} = \begin{cases} 1, & \text{if } j = \min(k_1,k_2) + C^2_{\max(k_1,k_2)}, \\ 0, & \text{otherwise,} \end{cases} \tag{2.10}$$

where $1 \le k_1, k_2 \le K$.

In our algorithm we will work with the smaller matrix $\mathbf{Q}_2(\mathcal{T})$ while in the theoretical development we will use $\mathbf{R}_2(\mathcal{T})$. More specifically, a vector $\mathbf{f} \in \mathbb{F}^K$ is a solution of (2.5) if and only if $\mathbf{f} \otimes \mathbf{f}$ belongs to the intersection of the null space of $\mathbf{R}_2(\mathcal{T})$ and the subspace of vectorized $K \times K$ symmetric matrices,

$$\operatorname{vec}(\mathbb{F}^{K\times K}_{sym}) := \{\operatorname{vec}(\mathbf{M}) : \ \mathbf{M} \in \mathbb{F}^{K\times K}, \ \mathbf{M} = \mathbf{M}^T\}, \quad \dim(\operatorname{vec}(\mathbb{F}^{K\times K}_{sym})) = C^2_{K+1}. \tag{2.11}$$

By (2.9), the intersection can actually be recovered from the null space of $\mathbf{Q}_2(\mathcal{T})$ as

$$\operatorname{Null}(\mathbf{R}_2(\mathcal{T})) \cap \operatorname{vec}(\mathbb{F}^{K\times K}_{sym}) = \mathbf{P}_K(\mathbf{P}_K^T\mathbf{P}_K)^{-1} \operatorname{Null}(\mathbf{Q}_2(\mathcal{T})). \tag{2.12}$$

It is worth noting that the matrix $\mathbf{D} := \mathbf{P}_K(\mathbf{P}_K^T\mathbf{P}_K)^{-1}$ in (2.12) has the following simple form

$$(\mathbf{D})_{(k_1-1)K+k_2,j} = \begin{cases} 1, & \text{if } j = k_1 + C^2_{k_1} \text{ and } k_1 = k_2, \\ \frac{1}{2}, & \text{if } j = \min(k_1,k_2) + C^2_{\max(k_1,k_2)} \text{ and } k_1 \ne k_2, \\ 0, & \text{otherwise.} \end{cases} \tag{2.13}$$

2.4. Convention $r_{\mathbf{T}_{(3)}} = K$.

The results of this paper rely on equivalence (2.3), which does not hold if the frontal slices $\mathbf{T}_1, \dots, \mathbf{T}_K$ of the tensor $\mathcal{T}$ are linearly dependent. One can easily verify that $\mathbf{T}_{(3)} = [\operatorname{vec}(\mathbf{T}_1) \ \dots \ \operatorname{vec}(\mathbf{T}_K)]$, implying that linear independence of $\mathbf{T}_1, \dots, \mathbf{T}_K$ is equivalent to full column rank of $\mathbf{T}_{(3)}$, i.e., to the condition $r_{\mathbf{T}_{(3)}} = K$.

Thus, to apply the results of the paper for tensors with $r_{\mathbf{T}_{(3)}} < K$, one should first "compress" $\mathcal{T}$ to an $I \times J \times \tilde{K}$ tensor $\tilde{\mathcal{T}}$ such that $r_{\tilde{\mathbf{T}}_{(3)}} = \tilde{K}$. Such a compression can, for instance, be done by taking $\tilde{\mathcal{T}}$ with $\tilde{\mathbf{T}}_{(3)}$ equal to the "U" factor in the compact SVD of $\mathbf{T}_{(3)} = \mathbf{U}\mathbf{S}\mathbf{V}^H$. In this case, by (1.5),

$$\tilde{\mathbf{T}}_{(3)} := \mathbf{U} = \mathbf{T}_{(3)}\mathbf{V}\mathbf{S}^{-1} = [\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R](\mathbf{S}^{-1}\mathbf{V}^T\mathbf{C})^T,$$

implying that $\tilde{\mathcal{T}}$ and $\mathcal{T}$ share the first two factor matrices and that the slices of $\tilde{\mathcal{T}}$ are obtained from linear mixtures of the $I \times J$ matrix slices of $\mathcal{T}$. If the decomposition of $\tilde{\mathcal{T}}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique, then, by statement 2) of Theorem 2.1, the matrix $[\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R]$ has full column rank. Thus, when the matrices $\mathbf{A}$ and $\mathbf{B}$ are obtained from $\tilde{\mathcal{T}}$, the remaining matrix $\mathbf{C}$ can be found from (1.5) as $\mathbf{C} = \left([\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R]^\dagger \mathbf{T}_{(3)}\right)^T$. For future reference, we summarize the above discussion in statement 1) of the following theorem. Statement 2) is the generic version of statement 1) and can be proved in a similar way.

Theorem 2.4.

1) *Let $\mathcal{T}$ be an $I \times J \times K$ tensor and let $\tilde{\mathcal{T}}$ be an $I \times J \times \tilde{K}$ tensor formed by $\tilde{K}$ linearly independent mixtures of the $I \times J$ matrix slices of $\mathcal{T}$. If the*

decomposition of $\tilde{\mathcal{T}}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms i) is unique or, moreover, ii) is unique and can be computed by means of (simultaneous) EVD, then the same holds true for $\mathcal{T}$.

2) If the decomposition of an $I \times J \times \tilde{K}$ tensor into a sum of max ML rank-$(1, L_r, L_r)$ terms i) is generically unique or, moreover, ii) is generically unique and can generically be computed by means of (simultaneous) EVD, then the same holds true for tensors with dimensions $I \times J \times K$, where $K \geq \tilde{K}$.

Thus, in the cases where the assumption $r_{\mathbf{T}_{(3)}} = K$ (resp. the assumptions $IJ \geq \sum L_r \geq K$) allows us to simplify the presentation, namely, in Theorems 2.5 and 2.6 (resp. in Theorem 2.13), we will assume w.l.o.g. that $r_{\mathbf{T}_{(3)}} = K$ (resp. $\sum L_r \geq K$).

2.5. Main uniqueness results and algorithm. In subsection 2.5.1 we present results on uniqueness and computation of the exact ML rank-$(1, L_r, L_r)$ decomposition (1.1). In subsection 2.5.2 we explain how to compute an approximate solution in the case where the decomposition is not exact. In subsection 2.5.3 we illustrate our results by examples.

2.5.1. Exact ML rank-$(1, L_r, L_r)$ decomposition. In the following theorem both assumptions (2.14), (2.15) need to hold, and at least one of the assumptions (2.16) and (2.17). In statement 4) of Lemma 3.1 below we will show that (2.16) actually implies (2.17).

By itself, Theorem 2.5 can be used to show uniqueness of a decomposition, but not only that. As we will explain later, the theorem comes with an algorithm for the actual computation of the decomposition (namely, Algorithm 2.1). In this respect, another comment is in order. If one wishes to use Theorem 2.5 to show uniqueness, and if one wishes to do so via (2.16), then there is no need to construct the matrix $\mathbf{Q}_2(\mathcal{T})$ in (2.17). On the other hand, Theorem 2.5 comes with Algorithm 2.1 for the actual computation of the decomposition. In this algorithm we work via the null space of $\mathbf{Q}_2(\mathcal{T})$ (and not just its dimension as in (2.17)), i.e., matrix $\mathbf{Q}_2(\mathcal{T})$ is constructed, also in cases where the uniqueness by itself follows from (2.16).

THEOREM 2.5. *Let $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ admit the ML rank-$(1, L_r, L_r)$ decomposition (1.1), i.e., $r_{\mathbf{E}_r} = L_r$ for all r. Assume that*

$$r_{\mathbf{T}_{(3)}} = K \text{ and} \tag{2.14}$$

$$d_r := \dim \operatorname{Null}(\mathbf{Z}_r) \geq 1, \qquad r = 1, \dots, R, \tag{2.15}$$

where $\mathbf{T}_{(3)}$ is defined in (1.5) and $\mathbf{Z}_r := [\mathbf{E}_1^T \ \dots \ \mathbf{E}_{r-1}^T \ \mathbf{E}_{r+1}^T \ \dots \ \mathbf{E}_R^T]^T$. Assume also that

$$k_{\mathbf{A}} \geq 2 \text{ and rank of } \mathbf{F} := [\mathbf{E}_{r_1} \ \mathbf{E}_{r_2} \ \dots \ \mathbf{E}_{r_{R-r_{\mathbf{A}}+2}}] \text{ is } L_{r_1} + \dots + L_{r_{R-r_{\mathbf{A}}+2}} \text{ for all } 1 \leq r_1 < \dots < r_{R-r_{\mathbf{A}}+2} \leq R \tag{2.16}$$

or

$$\dim \operatorname{Null}(\mathbf{Q}_2(\mathcal{T})) = \sum_{r=1}^{R} C_{d_r+1}^2 =: Q, \tag{2.17}$$

where $\mathbf{Q}_2(\mathcal{T})$ is constructed by Definition 2.3. Consider the following conditions:

a) $K \geq \sum L_r - \min L_r + 1$ and $k_{\mathbf{A}} \geq 2$;

b) the matrix $\mathbf{A}$ has full column rank, i.e., $r_{\mathbf{A}} = R$;

c) $k_{\mathbf{A}} = r_{\mathbf{A}} < R$, assumption (2.16) holds and

$$\text{rank of } \mathbf{G} := [\mathbf{E}_{r_1}^T\ \mathbf{E}_{r_2}^T\ \dots\ \mathbf{E}_{r_{R-r_{\mathbf{A}}+2}}^T] \text{ is } L_{r_1} + \cdots + L_{r_{R-r_{\mathbf{A}}+2}} \quad \text{for all } 1 \le r_1 < \cdots < r_{R-r_{\mathbf{A}}+2} \le R; \tag{2.18}$$

d) the matrix $[\mathbf{E}_1^T\ \dots\ \mathbf{E}_R^T]^T$ has maximum possible rank, namely, $\sum L_r$;

e) the inequality

$$C_{K+1}^2 - Q > -\tilde{L}_1\tilde{L}_2 + \sum_{1\le r_1<r_2\le R} L_{r_1}L_{r_2}$$

holds, where $\tilde{L}_1$ and $\tilde{L}_2$ denote the two smallest values in $\{L_1, \dots, L_R\}$.

The following statements hold.

1. *The matrix $\mathbf{A}$ in the ML rank-$(1, L_r, L_r)$ decomposition (1.1) can be computed by means of (simultaneous) EVD up to column permutation and scaling.*
2. *If either condition b) or condition c) holds, then the overall ML rank-$(1, L_r, L_r)$ decomposition (1.1) can be computed by means of (simultaneous) EVD.*
3. *If condition a) holds, then any decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms has R nonzero terms and its first factor matrix can be chosen as $\mathbf{AP}$, where every column of $\mathbf{P} \in \mathbb{F}^{R\times R}$ contains precisely a single 1 with zeros everywhere else.*
4. *If conditions a) and e) hold, then the first factor matrix of the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique and can be computed by means of (simultaneous) EVD.*
5. *If conditions a) and b) hold, or conditions a) and c) hold, or condition d) holds, then the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique and can be computed by means of (simultaneous) EVD.*

Proof. See section 4. □

We make the following comments on the assumptions, conditions, and statements in Theorem 2.5.

1) Statement 1) says that $\mathbf{A}$ can be computed by means of EVD. On the other hand, statement 4) says that the first factor matrix is unique and can be computed by means of EVD, under a more restrictive condition. A similar observation can be made for the computation of the entire decomposition in statements 2) and 3), respectively. What we mean is the following. All assumptions and conditions in Theorem 2.5, except (2.14), are formulated in terms of a specific ML rank-$(1, L_r, L_r)$ decomposition of $\mathcal{T}$, namely, in terms of the matrices $\mathbf{A}$ and $\mathbf{E}_1, \dots, \mathbf{E}_R$. There is a subtlety in the sense that $\mathcal{T}$ may admit alternative decompositions for which the assumptions (2.15) and (2.17) and conditions b) and c) do not all hold and which cannot necessarily be (partially) found by means of EVD. The more restrictive conditions in statements 4) and 5) exclude the existence of such alternative decompositions. Statement 3) is a "transition statement" in which the alternatives for the first factor matrix are restricted. Thus, statements 1) and 2) are mainly meant to cover cases where the first factor matrix and the overall decomposition, respectively, are not unique in the sense that there may be alternatives for which the assumptions/conditions do not hold. See Example 2.8 below for an illustration.

2) The matrix $\mathbf{P}$ in statement 3) is a column selection matrix, possibly with repeated columns. Thus, statement 3) says that the first factor matrix of any decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms can be obtained by

selecting columns of $\mathbf{A}$, where column repetition is allowed but the total number of columns should be equal to R.

3) The assumptions in Theorem 1.4, Theorem 1.5, and Theorem 1.8 are symmetric with respect to the last two dimensions while the assumptions and conditions in Theorem 2.5 are not. To get another set of conditions on uniqueness and computation one can just permute the last two dimensions of $\mathcal{T}$.

4) As in Theorem 1.4 and Theorem 1.5, the number of ML rank-$(1, L_r, L_r)$ terms and the values of L_r are not required to be known in advance; they are found by the algorithm.

5) Assumption (2.17) means that we require the subspace $\dim \operatorname{Null}(\mathbf{Q}_2(\mathcal{T}))$ to have the minimal possible dimension (see statement 3) of Lemma 3.1 below).

6) It can be shown that Statement 5) is a criterion that is "effective" in the sense of [8].

Instead of the matrices $\mathbf{A}$ and $\mathbf{E}_1, \dots, \mathbf{E}_R$, Theorem 2.5 can also be given in terms of the factor matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ (cf. Theorems 1.4, 1.5 and 1.8). Namely, substituting $\mathbf{E}_r = \mathbf{B}_r\mathbf{C}_r^T$ and $\mathcal{T} = \sum \mathbf{a}_r \circ (\mathbf{B}_r\mathbf{C}_r^T)$, in the expressions for $\mathbf{Z}_r$, $\mathbf{F}$, $\mathbf{G}$, $[\mathbf{E}_1^T \ \dots \ \mathbf{E}_R^T]^T$ and $\mathbf{Q}_2(\mathcal{T})$, respectively, we obtain the following result.

THEOREM 2.6. *Let $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ admit the ML rank-$(1, L_r, L_r)$ decomposition (1.2), i.e., $r_{\mathbf{B}_r} = r_{\mathbf{C}_r} = L_r$ for all r. Assume that*

$$\text{the matrix } [\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R]\mathbf{C}^T \text{ has full column rank and} \tag{2.19}$$

$$d_r := \dim \operatorname{Null}(\mathbf{Z}_{r,\mathbf{C}}) \geq 1, \qquad r = 1, \dots, R, \tag{2.20}$$

where $\mathbf{Z}_{r,\mathbf{C}} := [\mathbf{C}_1 \ \dots \ \mathbf{C}_{r-1} \ \mathbf{C}_{r+1} \ \dots \ \mathbf{C}_R]^T$. Assume also that

$$k_{\mathbf{A}} \geq 2 \text{ and } k'_{\mathbf{B}} \geq R - r_{\mathbf{A}} + 2 \tag{2.21}$$

or[6]

$$\dim \operatorname{Null}\left(\Phi(\mathbf{A}, \mathbf{B})\mathbf{S}_2(\mathbf{C})^T\right) = \sum_{r=1}^{R} C^2_{d_r+1} =: Q, \tag{2.22}$$

where the matrices $\Phi(\mathbf{A}, \mathbf{B})$ and $\mathbf{S}_2(\mathbf{C})$ are defined in (3.2) and (3.3) below[7]*. Consider the following conditions:*

a) $K \geq \sum L_r - \min L_r + 1$ and $k_{\mathbf{A}} \geq 2$;

b) the matrix $\mathbf{A}$ has full column rank, i.e., $r_{\mathbf{A}} = R$;

c) $k_{\mathbf{A}} = r_{\mathbf{A}} < R$, (2.21) holds and $k'_{\mathbf{C}} \geq R - r_{\mathbf{A}} + 2$;

d) $K = \sum\limits_{r=1}^{R} L_r$ (implying that $\mathbf{C}$ is $K \times K$ nonsingular and that $d_r = L_r$ for all r);

e) the inequality

$$C^2_{K+1} - Q > -\tilde{L}_1\tilde{L}_2 + \sum_{1\leq r_1<r_2\leq R} L_{r_1}L_{r_2}$$

holds, where $\tilde{L}_1$ and $\tilde{L}_2$ denote the two smallest values in $\{L_1, \dots, L_R\}$.

Then statements 1) to 5) in Theorem 2.5 hold.

[6] In statement 4) of Lemma 3.1 below we show that (2.21) implies (2.22).

[7] The definitions of $\Phi(\mathbf{A}, \mathbf{B})$ and $\mathbf{S}_2(\mathbf{C})$ require additional notations and are postponed to section 3 for the sake of readability. Here we just mention that each entry of $\Phi(\mathbf{A}, \mathbf{B})$ is a product of a 2×2 minor of $\mathbf{A}$ and a 2×2 minor of $\mathbf{B}$ and that each entry of $\mathbf{S}_2(\mathbf{C})$ is of the form $c_{i_1j_1}c_{i_2j_2} + c_{i_1j_2}c_{i_2j_1}$.

Proof. The proof is given in Appendix B. ∎

Statement 5) in Theorem 2.6/Theorem 2.5 allows us to trade full column rank of the factor matrices $\mathbf{B}$ and $\mathbf{C}$ for a higher k-rank of $\mathbf{A}$ than in Theorem 1.4. In particular the following result can be used in cases where none of the factor matrices has full column rank.

COROLLARY 2.7. *Let $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ admit the ML rank-$(1, L_r, L_r)$ decomposition (1.2), i.e., $r_{\mathbf{B}_r} = r_{\mathbf{C}_r} = L_r$ for all r. Assume that*

$$r_{\mathbf{C}} \geq \sum L_r - \min L_r + 1, \qquad k'_{\mathbf{B}} \geq R - r_{\mathbf{A}} + 2 \quad \text{and} \quad k_{\mathbf{A}} \geq 2. \tag{2.23}$$

Then the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique and can be computed by means of (simultaneous) EVD if

$$\text{either}\ \ r_{\mathbf{A}} = R \qquad \text{or} \qquad k_{\mathbf{A}} = r_{\mathbf{A}} < R \ \ \text{and}\ \ k'_{\mathbf{C}} \geq R - r_{\mathbf{A}} + 2. \tag{2.24}$$

Proof. The proof is given in Appendix B. ∎

The algebraic procedure that will result from Theorem 2.5 (or Theorem 2.6) is summarized in Algorithm 2.1. In this subsection we explain how Algorithm 2.1 computes the exact ML rank-$(1, L_r, L_r)$ decomposition (1.1). In subsection 2.5.2 we will explain how the steps in Algorithm 2.1 can be modified to compute an approximate ML rank-$(1, L_r, L_r)$ decomposition of $\mathcal{T}$.

In Phase I we recover the first factor matrix. In steps $1-3$ we compute a basis $\mathbf{v}_1, \dots, \mathbf{v}_Q$ of the subspace $\operatorname{Null}(\mathbf{R}_2(\mathcal{T})) \cap \operatorname{vec}(\mathbb{F}^{K\times K}_{sym})$. The computation relies on identity (2.12): we construct the smaller matrix $\mathbf{Q}_2(\mathcal{T})$, compute a basis of $\operatorname{Null}(\mathbf{Q}_2(\mathcal{T}))$ and map it to a basis of $\operatorname{Null}(\mathbf{R}_2(\mathcal{T})) \cap \operatorname{vec}(\mathbb{F}^{K\times K}_{sym})$. In steps 4 and 5 we construct S-JBD problem (1.6) and solve it by Algorithm 1.1.

It will be proved (see proof of the first statement of Theorem 2.5) that submatrix $\mathbf{N}_r \in \mathbb{F}^{K\times d_r}$ of the matrix $\mathbf{N} = [\mathbf{N}_1 \ \dots \ \mathbf{N}_R]$ computed in step 5 holds a basis of $\operatorname{Null}(\mathbf{Z}_r)$, $r = 1, \dots, R$. In addition, it can be easily verified that $\operatorname{Null}(\mathbf{Z}_r) = \operatorname{Null}(\mathbf{Z}_{r,\mathbf{C}})$, so we have that

$$\mathbf{N}_r^T[\mathbf{C}_1 \ \dots \ \mathbf{C}_{r-1} \ \mathbf{C}_{r+1} \ \dots \ \mathbf{C}_R] = \mathbf{O}, \qquad r = 1, \dots, R. \tag{2.25}$$

In step 6 we use (2.25) to compute the columns of $\mathbf{A}$: since by (2.25) and (1.5),

$$\begin{aligned}
[\mathbf{N}_r^T\mathbf{H}_1^T \ \dots \ \mathbf{N}_r^T\mathbf{H}_I^T] =& \mathbf{N}_r^T\mathbf{T}_{(3)}^T = \mathbf{N}_r^T\mathbf{C}[\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R]^T = \\
& \mathbf{N}_r^T\mathbf{C}_r(\mathbf{a}_r^T \otimes \mathbf{B}_r^T) = (1 \otimes \mathbf{N}_r^T\mathbf{C}_r)(\mathbf{a}_r^T \otimes \mathbf{B}_r^T) = \\
& \mathbf{a}_r^T \otimes (\mathbf{N}_r^T\mathbf{C}_r\mathbf{B}_r^T) = \mathbf{a}_r^T \otimes (\mathbf{N}_r^T\mathbf{E}_r^T), \qquad r = 1, \dots, R,
\end{aligned} \tag{2.26}$$

it follows that

$$[\operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_1^T) \ \dots \ \operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_I^T)] = \operatorname{vec}(\mathbf{N}_r^T\mathbf{E}_r^T)\mathbf{a}_r^T, \qquad r = 1, \dots, R, \tag{2.27}$$

implying that $\mathbf{a}_r$ is the vector that generates the row space of only right singular vector of $[\operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_1^T) \ \dots \ \operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_I^T)]$ that corresponds to a nonzero singular value.

In Phase II we recover the overall decomposition. Since, by Theorem 2.5 (or Theorem 2.6), the computation is possible if at least one of the conditions d), b), or c) holds, we consider three cases.

Case 1: condition d) in Theorem 2.6 implies that $\mathbf{C}$ is a $K \times K$ nonsingular matrix and that $K = \sum d_r = \sum L_r$. Since the $K \times \sum d_r$ matrix $\mathbf{N}$ computed in step 5 has full column rank, it follows that $\mathbf{N}$ is also $K \times K$ nonsingular. Since, by (2.25),

$$\mathbf{N}^T\mathbf{C} = [\mathbf{N}_1 \ \dots \mathbf{N}_R]^T[\mathbf{C}_1 \ \dots \ \mathbf{C}_R] = \operatorname{blockdiag}(\mathbf{N}_1^T\mathbf{C}_1, \dots, \mathbf{N}_R^T\mathbf{C}_R),$$

Algorithm 2.1 Computation of ML rank-$(1, L_r, L_r)$ decomposition (1.1) under various conditions expressed in Theorem 2.5

Input: tensor $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ admitting decomposition (1.1)

Phase I (computation of $\mathbf{A}$)

Construct the $C_I^2 C_J^2$-by-C_{K+1}^2 matrix $\mathbf{Q}_2(\mathcal{T})$ as in Definition 2.3

Find $\mathbf{g}_q \in \mathbb{F}^{C_{K+1}^2}$, $q = 1, \dots, Q$ that form a basis of $\text{Null}\,(\mathbf{Q}_2(\mathcal{T}))$, where $Q = C_{d_1+1}^2 + \cdots + C_{d_R+1}^2$

Compute $\mathbf{v}_q := \mathbf{D}\mathbf{g}_q \in \mathbb{F}^{K^2}$, $q = 1, \dots, Q$, where $\mathbf{D}$ is defined in (2.13)

For each $q = 1, \dots, Q$ reshape $\mathbf{v}_q$ into the $K \times K$ symmetric matrix $\mathbf{V}_q$

Compute $\mathbf{N}$ and the values R, $d_1, \dots, d_R$ in S-JBD problem (1.6) by Algorithm 1.1

For each $r = 1, \dots, R$ take $\mathbf{a}_r$ equal to the vector that generates the row space of $[\text{vec}(\mathbf{N}_r^T\mathbf{H}_1^T)\ \dots\ \text{vec}(\mathbf{N}_r^T\mathbf{H}_I^T)]$, where $\mathbf{H}_i := (t_{ijk})_{j,k=1}^{J,K}$

Phase II (computation of the overall decomposition under one of the conditions d), b), or c))

Case 1: condition d) *in Theorem* 2.5 *holds*

For each $r = 1, \dots, R$ compute the vector that generates the column space of $[\text{vec}(\mathbf{N}_r^T\mathbf{H}_1^T)\ \dots\ \text{vec}(\mathbf{N}_r^T\mathbf{H}_I^T)]$ and reshape it into the matrix $\mathbf{B}_r$

Compute $\mathbf{C}$ from the set of linear equations
$$\mathbf{T}_{(3)} = [\mathbf{a}_1 \otimes \mathbf{B}_1\ \dots\ \mathbf{a}_R \otimes \mathbf{B}_R]\mathbf{C}^T$$

For each $r = 1, \dots, R$ set $\mathbf{E}_r = \mathbf{B}_r\mathbf{C}_r^T$

Case 2: condition b) *in Theorem* 2.5 *holds*

Compute $\mathbf{E}_1, \dots, \mathbf{E}_R$ by solving the set of linear equations
$$\mathbf{T}_{(1)} = [\text{vec}(\mathbf{E}_1)\ \dots\ \text{vec}(\mathbf{E}_R)]\mathbf{A}^T$$

Case 3: condition c) *in Theorem* 2.5 *holds*

Choose (possibly overlapping) subsets $\Omega_1, \dots, \Omega_M \subset \{1, \dots, R\}$ such that $\text{card}(\Omega_1) = \cdots = \text{card}(\Omega_M) = R - r_{\mathbf{A}} + 2$ and $\{1, \dots, R\} = \Omega_1 \cup \cdots \cup \Omega_M$

for each $m = 1, \dots, M$ **do**

Find linearly independent vectors $\mathbf{h}_1, \mathbf{h}_2 \in \mathbb{F}^I$ that belong to the column space of $\mathbf{A}$ and satisfy
$$\mathbf{a}_r^T\mathbf{h}_1 = \mathbf{a}_r^T\mathbf{h}_2 = 0 \text{ for all } r \in \{1, \dots, R\} \setminus \Omega_m$$

Compute the $2 \times J \times K$ tensor $\mathcal{Q}^{(m)}$ with $\mathbf{Q}_{(1)}^{(m)} = \mathbf{T}_{(1)}[\mathbf{h}_1\ \mathbf{h}_2]$

Compute the ML rank-$(1, L_r, L_r)$ decomposition of $\mathcal{Q}^{(m)}$ by the EVD in Theorem 1.4:
$$\mathcal{Q}^{(m)} = \sum_{r\in\Omega_m} \hat{\mathbf{a}}_r \circ \hat{\mathbf{E}}_r \qquad \text{(the vectors } \hat{\mathbf{a}}_r \text{ are a by-product)}$$

end for

Compute $\mathbf{x}$ from the linear equation
$$[\mathbf{a}_1 \otimes \text{vec}(\hat{\mathbf{E}}_1)\ \dots\ \mathbf{a}_r \otimes \text{vec}(\hat{\mathbf{E}}_R)]\mathbf{x} = \text{vec}(\mathbf{T}_{(1)})$$

For each $r = 1, \dots, R$ set $\mathbf{E}_r = x_r\hat{\mathbf{E}}_r$

Output: Matrices $\mathbf{A} \in \mathbb{F}^{I\times R}$, $\mathbf{E}_1, \dots, \mathbf{E}_R \in \mathbb{F}^{J\times K}$ such that (1.1) holds we have that $\mathbf{C}=\mathbf{N}^{-T}\operatorname{blockdiag}(\mathbf{N}_1^T\mathbf{C}_1,\dots,\mathbf{N}_R^T\mathbf{C}_R)$. Since $\mathbf{C}$ and $\mathbf{N}$ are nonsingular, the matrices $\mathbf{N}_r^T\mathbf{C}_r\in\mathbb{F}^{L_r\times L_r}$ are also nonsingular. To compute $\mathbf{B}_1,\dots,\mathbf{B}_R$ we use identity (2.27). In step 7 we compute $\operatorname{vec}(\mathbf{N}_r^T\mathbf{E}_r^T)$ as the vector that generates the column space of the left singular vector of $[\operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_1^T)\ \dots\ \operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_I^T)]$ corresponding to the only nonzero singular value. In addition, $(\mathbf{N}_r^T\mathbf{E}_r^T)^T=\mathbf{B}_r(\mathbf{N}_r^T\mathbf{C}_r)^T$ by definition of $\mathbf{E}_r$. W.l.o.g. we set $\mathbf{B}_r$ equal to $(\mathbf{N}_r^T\mathbf{E}_r^T)^T$, as the nonsingular factor $(\mathbf{N}_r^T\mathbf{C}_r)^T$ can be compensated for in the factor $\mathbf{C}$. As such, in step 8 we finally recover $\mathbf{C}$ from (1.5).

It is worth noting that the vectors $\mathbf{a}_r$ in step 6 and the matrices $\mathbf{B}_r$ in step 7 can be computed simultaneously. Indeed, by (2.27), $\mathbf{B}_r$ and $\mathbf{a}_r$, can be found from $\operatorname{vec}(\mathbf{B}_r)\mathbf{a}_r^T=[\operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_1^T)\ \dots\ \operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_I^T)]$.

Case 2: condition b) implies that $\mathbf{A}$ has full column rank. Hence, by (1.3), $[\operatorname{vec}(\mathbf{E}_1)\ \dots\ \operatorname{vec}(\mathbf{E}_R)]=\mathbf{T}_{(1)}(\mathbf{A}^T)^\dagger$.

Case 3: We assume that condition c) holds. In steps $11-18$ we use the matrix $\mathbf{A}$ estimated in Phase I and the tensor $\mathcal{T}$ to recover the matrices $\mathbf{E}_1,\dots,\mathbf{E}_R$. There exist $C_R^{R-r_{\mathbf{A}}+2}$ subsets of $\{1,\dots,R\}$ of cardinality $R-r_{\mathbf{A}}+2$. In principle, one can choose any M of them that cover the set $\{1,\dots,R\}$. (One can, for instance, choose $M=\lceil\frac{R}{R-r_{\mathbf{A}}+2}\rceil$ and set $\Omega_m=\{(m-1)(R-r_{\mathbf{A}}+2)+1,\dots,m(R-r_{\mathbf{A}}+2)\}$ for $m=1,\dots,M-1$ and $\Omega_M=\{r_{\mathbf{A}}-1,\dots,R\}$, where $\lceil x\rceil$ denotes the least integer greater than or equal to x.) To explain steps $12-16$ we assume for simplicity that, in step 11, $\Omega_1=\{1,\dots,R-r_{\mathbf{A}}+2\}$. In steps 13 and 14 we project out the last $r_{\mathbf{A}}-2$ terms in the ML rank-$(1,L_r,L_r)$ decomposition of $\mathcal{T}$. It can be shown that the tensor $\mathcal{Q}^{(1)}$ constructed in step 14 admits the ML rank-$(1,L_r,L_r)$ decomposition $\mathcal{Q}^{(1)}=\sum\limits_{r=1}^{R-r_{\mathbf{A}}+2}\hat{\mathbf{a}}_r\circ\hat{\mathbf{E}}_r$, where $\hat{\mathbf{a}}_r=[\mathbf{h}_1\ \mathbf{h}_2]^T\mathbf{a}_r\in\mathbb{F}^2$ and $\hat{\mathbf{E}}_r$ is proportional to $\mathbf{E}_r$, $r=1,\dots,R-r_{\mathbf{A}}+2$. By condition c), $\mathcal{Q}^{(1)}$ satisfies the assumptions in Theorem 1.4. Thus, the ML rank-$(1,L_r,L_r)$ decomposition $\mathcal{Q}^{(1)}$ is unique and can be computed by means of (simultaneous) EVD. The remaining matrices $\mathbf{E}_{R-r_{\mathbf{A}}+3},\dots,\mathbf{E}_R$ can be estimated up to scaling factors in a similar way by choosing other subsets Ω_m. In step 17 we use (1.3) to compute the scaling factors $x_1,\dots,x_R$ such that $\mathcal{T}=\sum\limits_{r=1}^{R}\mathbf{a}_r\circ(x_r\hat{\mathbf{E}}_r)$.

One may wonder what to do if several of conditions b), c) or d) hold together. Conditions b) and c) are mutually exclusive. If conditions b) and d) hold, then uniqueness and computation follow already from Theorem 1.5. Indeed, conditions b) and d) in Theorem 2.6 imply that the matrices $\mathbf{A}$ and $\mathbf{C}$ have full column rank, and, by Corollary 3.2, assumption (2.22) is more restrictive than the assumption $r_{[\mathbf{B}_i\ \mathbf{B}_j]}\geq\max(L_i,L_j)+1$ for all $1\leq i<j\leq R$. It is less clear if Algorithm 2.1 can further be simplified if conditions c) and d) hold together. Since the computation in Case 1 consists basically of step 8 (it was explained above that step 7 can be integrated into step 6) we give priority to Case 1 over the more cumbersome Case 3 when conditions c) and d) hold together.

The number of ML rank-$(1,L_r,L_r)$ terms R and their "sizes" $L_1,\dots,L_R$ do not have to be known a priori as they are found in Phase 1 and Phase 2, respectively. Namely, Algorithm 1.1 in step 5 estimates R as the number of blocks of $\mathbf{N}$ and estimates d_r as the number of columns in the rth block. If condition d) in Theorem 2.5 holds, then we set $L_r:=d_r$. If condition b) or c) in Theorem 2.5 holds, then we just set $L_r=r_{\mathbf{E}_r}$.

It is worth noting that if condition c) in Theorem 2.5 holds and if the sets Ω_m in step 11 are chosen in a particular way, then the "sizes" $r_{\hat{\mathbf{E}}_r}=L_r$ of the ML rank-

$(1, L_r, L_r)$ terms of the tensors $\mathcal{Q}^{(m)}$, constructed in step 14, can be computed by solving an overdetermined system of linear equations. That is, the values $L_1, \dots, L_R$ can be found without executing step 15. Indeed, one can easily verify that condition c) in Theorem 2.5 implies that the equalities

$$\sum_{r\in\Omega_m} r_{\hat{\mathbf{E}}_r} = r_{\mathbf{Q}^{(m)}_{(2)}} = r_{\mathbf{Q}^{(m)}_{(3)}} \tag{2.28}$$

hold for any Ω_m, $m = 1, \dots, M$. If M has the maximum possible value, i.e., $M = C_R^{R-r_{\mathbf{A}}+2}$, then the M identities in (2.28) can be rewritten as the system of linear equations $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}$, where $\tilde{\mathbf{A}}$ is a binary (0/1) $M \times R$ matrix such that none of the rows are proportional and each row of $\tilde{\mathbf{A}}$ has exactly $R - r_{\mathbf{A}} + 2$ ones. The vectors $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{b}}$ consist of the values $r_{\hat{\mathbf{E}}_r}$, $1 \le r \le R$ and $r_{\mathbf{Q}^{(m)}_{(2)}}$, $1 \le m \le M$, respectively. One can easily verify that $\tilde{\mathbf{A}}$ has full column rank, i.e., the unique solution of (2.28) yields the values $L_1, \dots, L_R$.

Algorithm 2.1 should be seen as an algebraic computational proof-of-concept. It opens a new line of research of numerical aspects and strategies; the development of such dedicated numerical strategies is out of the scope of this paper.

In the given form, the computational cost of Algorithm 2.1 is dominated by steps 1, 2, and 5. Since each entry of the $C_I^2C_J^2$-by-C_{K+1}^2 matrix $\mathbf{Q}_2(\mathcal{T})$ is of the form (2.8), step 1 requires at most $7C_I^2C_J^2C_{K+1}^2$ flops, i.e. 4 multiplications and 3 additions per entry (note that no distinction between complex and real data is made). The cost of finding a basis $\mathbf{g}_1, \dots, \mathbf{g}_Q$ via the SVD is of order $6C_I^2C_J^2(C_{K+1}^2)^2 + 20(C_{K+1}^2)^3$ when the SVD is implemented via the R-SVD method [22]. The cost of step 5 is dominated by step 1 in Algorithm 1.1. This cost is of order $6(K^2Q)^2(K^2)^2 + 20(K^2)^3 = (6Q^2 + 20)K^6$ (cost of the SVD of a $K^2Q \times K^2$ matrix[8]). Thus, the total computational cost of Algorithm 2.1 is of order $\mathcal{O}(I^2J^2K^4 + K^6)$. Paper [32, Section S.1] explains an indirect technique to reduce the total cost of the steps 1 and 2 to $\mathcal{O}(\max(IJ^2K^2, J^2K^4))$. In this case, the total computational cost of Algorithm 2.1 will be of order $\mathcal{O}(\max(IJ^2K^2 + K^6, J^2K^4 + K^6))$.

2.5.2. Approximate ML rank-$(1, L_r, L_r)$ decomposition. Now we discuss noisy variants of the steps in Algorithm 2.1. We consider two scenarios.

I. In the exact case the matrix $\mathbf{Q}_2(\mathcal{T})$ has exactly Q nonzero singular values, the matrices $\mathbf{V}_q$ obtained in step 6 are at most rank-$\sum d_r$ and the matrix $\mathbf{M}$ constructed in subsection 1.3.2 has exactly R nonzero singular values. In the *first scenario* we assume that the perturbation of the tensor is "small enough" to recover the correct values of Q, R and $d_1, \dots, d_R$ in Phase I. In this case we proceed as follows. In step 2 we set $\mathbf{g}_q$ equal to the qth smallest right singular vector of $\mathbf{Q}_2(\mathcal{T})$. In step 5 we use the noisy variant of Algorithm 1.1 (see the end of subsection 1.3.2) which gives us R and the values $d_1, \dots, d_R$. In steps 6 and 7 we choose $\mathbf{a}_r$ and $\mathbf{B}_r$ such that $\operatorname{vec}(\mathbf{B}_r)\mathbf{a}_r^T$ is the best rank-1 approximation of the matrix $[\operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_1^T) \ \dots \ \operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_I^T)]$. After steps 10 and 18 we replace the matrices $\mathbf{E}_1, \dots, \mathbf{E}_R$ by their truncated SVDs. Assuming the values of $d_1, \dots, d_R$ computed in step 5 are correct, the truncation ranks can

[8]Recall that the vectorized matrices $\mathbf{U}_1, \dots, \mathbf{U}_R$ in step 1 of Algorithm 1.1 can be found from the SVD of the $K^2Q \times K^2$ matrix $\mathbf{M}$ formed by the rows of $\mathbf{V}_q^T \otimes \mathbf{I} - (\mathbf{I} \otimes \mathbf{V}_q)\mathbf{P}$, $q = 1, \dots, Q$, where $\mathbf{P}$ denotes the $K^2 \times K^2$ permutation matrix that transforms the vectorized form of a $K \times K$ matrix into the vectorized form of its transpose.

generically be determined as

$$L_r = d_r + \frac{K - \sum d_r}{R-1}, \qquad r = 1, \dots, R. \tag{2.29}$$

Indeed, if the matrices $\mathbf{Z}_{1,\mathbf{C}}, \dots, \mathbf{Z}_{R,\mathbf{C}}$ have full column rank, then, by (2.20), $d_r = K - \sum\limits_{k=1}^{R} L_k + L_r$. Hence $\sum d_r = RK - R\sum\limits_{k=1}^{R} L_k + \sum\limits_{k=1}^{R} L_k$, implying that $\sum\limits_{k=1}^{R} L_k = \frac{RK-\sum d_r}{R-1}$. Thus, $L_r = d_r - K + \sum\limits_{k=1}^{R} L_k = d_r - K + \frac{RK-\sum d_r}{R-1} = d_r + \frac{K-\sum d_r}{R-1}$. In steps 8, 10, and 17 we solve the linear systems in the least squares sense.

An approximate ML rank-$(1, L_r, L_r)$ decomposition of the tensor $\mathcal{Q}^{(m)}$ in step 15 can be computed in the least squares sense using optimization based techniques. In this case the values $L_1, \dots, L_R$ should be known in advance. They can be estimated as follows. First the values $r_{\mathbf{Q}^{(m)}_{(2)}}$ and $r_{\mathbf{Q}^{(m)}_{(3)}}$ in (2.28) should be replaced by their numerical ranks (with respect to some threshold). Then the system of linear equations (2.28) should be solved in the least squares sense, subject to positive integer constraints on $r_{\hat{\mathbf{E}}_r} = L_r$.

II. In the *second scenario* we assume that the perturbation of the tensor is not "small enough" to guess the values of Q, R and $d_1, \dots, d_R$ in Phase 1. We explain how we proceed if (only) the values of R and $\sum L_r$ are known. Since, generically, $d_r = K - \sum\limits_{k=1}^{R} L_k + L_r$, we obtain that $\sum d_r = RK - (R-1)\sum L_r$. In step 2, we replace Q by its lower bound

$$Q_{min} := \underset{\sum \hat{d}_r = \sum d_r}{\operatorname{argmin}} \left(C^2_{\hat{d}_1+1} + \dots + C^2_{\hat{d}_R+1} \right).$$

In the first scenario, the matrix $\mathbf{N}$ was estimated as the third factor matrix in CPD (1.7) and the partition of $\mathbf{N}$ into blocks $\mathbf{N}_1, \dots, \mathbf{N}_R$ (and, in particular, the values of $d_1, \dots, d_R$) was obtained by clustering the columns of the first factor matrix in the CPD. In the second scenario, we compute only matrix $\mathbf{N}$ in step 5, without estimating the values of $d_1, \dots, d_R$. Since, by (2.26), $\mathbf{T}_{(3)}\mathbf{N}_r = \mathbf{a}_r \otimes (\mathbf{E}_r\mathbf{N}_r)$, it follows that $\mathbf{T}_{(3)}\mathbf{N}$ coincides up to permutation of columns with the matrix $[\mathbf{a}_1 \otimes (\mathbf{E}_1\mathbf{N}_1)\ \dots\ \mathbf{a}_R \otimes (\mathbf{E}_R\mathbf{N}_R)]$. So, clustering the columns of $\mathbf{T}_{(3)}\mathbf{N}$ into R clusters (modulo sign and scaling) we obtain the values $d_1, \dots, d_R$ as the sizes of clusters and the columns of $\mathbf{A}$ as their centers. The noisy variants of the remaining steps are the same as in the first scenario.

2.5.3. Examples.

Example 2.8. In this example we illustrate how to apply statement 2) of Theorem 2.5 for the computation of a decomposition that is not unique but does satisfy (2.15). Let $R \geq 2$. We consider an $R \times (R+2) \times (R+2)$ tensor $\mathcal{T}$ generated by (1.2) in which

$$\mathbf{A} = [\mathbf{a}_1\ \dots\ \mathbf{a}_R],$$
$$\mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2\ \mathbf{b}_3\ \mathbf{b}_1\ \mathbf{b}_2\ \mathbf{b}_4\ \mathbf{b}_5\ \dots \mathbf{b}_{3R-2}], \text{ and } \mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3\ \mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_4\ \dots\ \mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_{R+2}],$$

where the entries of $\mathbf{a}_1, \dots, \mathbf{a}_R$, $\mathbf{b}_1, \dots, \mathbf{b}_{3R-2}$, and $\mathbf{c}_1, \dots, \mathbf{c}_{R+2}$ are independently drawn from the standard normal distribution $N(0,1)$. Thus, $\mathcal{T}$ is a sum of R ML rank-$(1, 3, 3)$ terms (i.e., $L_1 = \cdots = L_R = 3$):

$$
\begin{aligned}
&\mathcal{T} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{E}_r, \quad \text{where} \\
&\mathbf{E}_1 = [\mathbf{b}_1\ \mathbf{b}_2\ \mathbf{b}_3][\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3]^T, \qquad \mathbf{E}_2 = [\mathbf{b}_1\ \mathbf{b}_2\ \mathbf{b}_4][\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_4]^T, \text{ and} \\
&\mathbf{E}_r = [\mathbf{b}_{3r-4}\ \mathbf{b}_{3r-3}\ \mathbf{b}_{3r-2}][\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_{r+2}]^T \qquad \text{for } r \geq 3.
\end{aligned}
\tag{2.30}
$$

Nonuniqueness. Let us show that the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, 3, 3)$ terms is not unique. Let $\mathcal{T}_2$ equal the sum of the first two ML rank-$(1, L_r, L_r)$ terms:

$$
\mathcal{T}_2 = \mathbf{a}_1 \circ (\mathbf{b}_1\mathbf{c}_1^T + \mathbf{b}_2\mathbf{c}_2^T + \mathbf{b}_3\mathbf{c}_3^T) + \mathbf{a}_2 \circ (\mathbf{b}_1\mathbf{c}_1^T + \mathbf{b}_2\mathbf{c}_2^T + \mathbf{b}_4\mathbf{c}_4^T).
\tag{2.31}
$$

It can be proved that $\mathcal{T}_2$ admits exactly three decompositions into a sum of max ML rank-$(1, L_r, L_r)$ terms, namely (2.31) itself and the decompositions

$$
\begin{aligned}
\mathcal{T}_2 = \mathbf{a}_1 \circ (\mathbf{b}_3\mathbf{c}_3^T - \mathbf{b}_4\mathbf{c}_4^T) + (\mathbf{a}_1 + \mathbf{a}_2) \circ (\mathbf{b}_1\mathbf{c}_1^T + \mathbf{b}_2\mathbf{c}_2^T + \mathbf{b}_4\mathbf{c}_4^T) = \\
(\mathbf{a}_1 + \mathbf{a}_2) \circ (\mathbf{b}_1\mathbf{c}_1^T + \mathbf{b}_2\mathbf{c}_2^T + \mathbf{b}_3\mathbf{c}_3^T) - \mathbf{a}_2 \circ (\mathbf{b}_3\mathbf{c}_3^T - \mathbf{b}_4\mathbf{c}_4^T).
\end{aligned}
\tag{2.32}
$$

Since $\mathcal{T}_2$ admits three decompositions it follows that $\mathcal{T}$ admits at least three decompositions for $R \geq 2$. In other words, the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is not unique.

Computation for $R \geq 3$. Now we show that, by statement 2) of Theorem 2.5, decomposition (2.30) can be computed by means of (simultaneous) EVD, at least for $R = 3, \ldots, 20$ (which are the values of R we have tested). First we show that assumptions (2.14), (2.15), (2.17), and condition b) hold. Assumption (2.14) and condition b) are trivial. The values of $d_1, \ldots, d_R$ in (2.15) can be computed by (2.20), which easily gives $d_1 = \cdots = d_R = 1$. It can also be verified that $\mathbf{Q}_2(\mathcal{T})$ is a $C_R^2 C_{R+2}^2 \times C_{R+3}^2$ matrix and that (at least for $R = 3, \ldots, 20$) $\dim \operatorname{Null}(\mathbf{Q}_2(\mathcal{T})) = R = \sum C_{d_r+1}^2$, i.e., (2.17) holds as well. (To compute the null space we used the MATLAB built-in function `null`.)

Let us now illustrate how Algorithm 2.1 recovers the matrices $\mathbf{A}$, $\mathbf{E}_1, \ldots, \mathbf{E}_R$. As has been mentioned before, since the matrix $\mathbf{N}$ computed in step 5 consists of the blocks $\mathbf{N}_1 \in \mathbb{F}^{K \times d_1}, \ldots, \mathbf{N}_R \in \mathbb{F}^{K \times d_R}$ which hold, respectively, bases of the subspaces $\operatorname{Null}(\mathbf{Z}_1) = \operatorname{Null}(\mathbf{Z}_{1,\mathbf{C}}), \ldots, \operatorname{Null}(\mathbf{Z}_R) = \operatorname{Null}(\mathbf{Z}_{R,\mathbf{C}})$, it follows that (2.25) holds. Since $d_1 = \cdots = d_R = 1$, the S-JBD problem in step 5 is actually a symmetric joint diagonalization problem. Thus, in step 5, we obtain an $(R+2) \times R$ matrix $\mathbf{N} = [\mathbf{n}_1\ \ldots\ \mathbf{n}_R]$ and (2.25) takes the following form :

$$
\mathbf{n}_r^T [\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3\ \ldots\ \mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_{r+1}\ \mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_{r+3}\ \ldots \mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_{R+2}] = \mathbf{0}, \qquad r = 1, \ldots, R.
$$

Then in step 6 we compute $\mathbf{a}_r$, by (2.27), i.e., as the vector that generates the row space of only right singular vector of $[\mathbf{H}_1\mathbf{n}_r\ \ldots\ \mathbf{H}_I\mathbf{n}_r]$:

$$
[\mathbf{H}_1\mathbf{n}_r\ \ldots\ \mathbf{H}_I\mathbf{n}_r] = [\operatorname{vec}(\mathbf{n}_r^T\mathbf{H}_1^T)\ \ldots\ \operatorname{vec}(\mathbf{n}_r^T\mathbf{H}_I^T)] = \operatorname{vec}(\mathbf{n}_r^T\mathbf{E}_r^T)\mathbf{a}_r^T = (\mathbf{E}_r\mathbf{n}_r)\mathbf{a}_r^T.
$$

Finally, in step 12 we reshape the columns of $\mathbf{T}_{(1)}(\mathbf{A}^T)^\dagger$ into the matrices $\mathbf{E}_1$ and $\mathbf{E}_2$.

It is worth noting that none of the three decompositions of $\mathcal{T}_2$ can be computed by Theorem 2.5 while for $R = 3, \ldots, 20$ decomposition (2.30) of $\mathcal{T}$, involving additional terms, can be computed by Theorem 2.5. Let us explain. First, one can easily verify that the third matrix unfolding of $\mathcal{T}_2 \in \mathbb{F}^{R\times(R+2)\times(R+2)}$ is rank-4, so, as it was explained in subsection 2.4, for investigating properties of $\mathcal{T}_2$, we can w.l.o.g. focus on $\mathcal{T}_2 \in \mathbb{F}^{R\times(R+2)\times 4}$. It can be verified that $\mathbf{Q}_2(\mathcal{T}_2)$ is a $C_R^2 C_{R+2}^2 \times 10$ matrix, that $\dim \operatorname{Null}(\mathbf{Q}_2(\mathcal{T}_2)) = 5$, and that for all decompositions in (2.31) and (2.32) we have $(d_1, d_2) \in \{(1,1),(2,1),(1,2)\}$. Thus, $C_{d_1+1}^2 + C_{d_2+1}^2 \leq 4 < 5 = \dim \operatorname{Null}(\mathbf{Q}_2(\mathcal{T}_2))$, implying that assumption (2.17) does not hold.

To explain why (2.17) does hold for $\mathcal{T}$ while it does not hold for $\mathcal{T}_2$, we refer to equivalence (2.3). From (2.2) and (2.30) it follows that

$$
\begin{aligned}
(2.33)\quad f_1\mathbf{T}_1 + \cdots + f_{R+2}\mathbf{T}_{R+2} = &\left((\mathbf{a}_1+\mathbf{a}_2)\mathbf{b}_1^T + \sum_{r=3}^{R}\mathbf{a}_r\mathbf{b}_{3r-4}^T\right)\mathbf{f}^T\mathbf{c}_1 + \\
&\left((\mathbf{a}_1+\mathbf{a}_2)\mathbf{b}_2^T + \sum_{r=3}^{R}\mathbf{a}_r\mathbf{b}_{3r-3}^T\right)\mathbf{f}^T\mathbf{c}_2 + (\mathbf{a}_1\mathbf{b}_3^T)\mathbf{f}^T\mathbf{c}_3 + (\mathbf{a}_2\mathbf{b}_4^T)\mathbf{f}^T\mathbf{c}_4 + \\
&\sum_{r=3}^{R}(\mathbf{a}_r\mathbf{b}_{3r-2}^T)\mathbf{f}^T\mathbf{c}_{r+2}.
\end{aligned}
$$

Above, we have numerically verified that $\dim \operatorname{Null}(\mathbf{Q}_2(\mathcal{T})) = R = \sum C_{d_r+1}^2$, which guarantees that (2.3) holds for $\mathcal{T}$, i.e., $f_1\mathbf{T}_1 + \cdots + f_{R+2}\mathbf{T}_{R+2}$ is rank-1 if and only if $\mathbf{f}$ belongs to the null spaces of all matrices $[\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3]^T, \dots, [\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_{R+3}]^T$ but one. On the other hand, in the case of $\mathcal{T}_2$, one can easily find a counterexample to the implication "⇒" in (2.3). Indeed, for $\mathcal{T}_2$ the linear combination in the LHS of (2.33) of the frontal slices of $\mathcal{T}_2$ can be rewritten as the RHS without the terms under the summation signs. Then the implication "⇒" in (2.3) does not hold for a vector $\mathbf{f}$ such that $\mathbf{c}_3^T\mathbf{f} = \cdots = \mathbf{c}_{R+2}^T\mathbf{f} = 0$ but $|\mathbf{c}_1^T\mathbf{f}| + |\mathbf{c}_2^T\mathbf{f}| \neq 0$.

Example 2.9. We consider a $3 \times J \times 15$ tensor generated by (1.2) in which the entries of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are independently drawn from the standard normal distribution $N(0,1)$ and $L_1 = L_2 = L_3 = 2$, $L_4 = L_5 = 3$, and $L_6 = 4$. Thus, $\mathcal{T}$ is a sum of $R = 6$ terms. For $J \geq 9$, one can easily check that $d_r = L_r - 1$ and that (2.14) and condition a) in Theorem 2.5 hold. We illustrate statements 4) and 5) of Theorem 2.5 by considering J in the sets $\{9, 10, 11, 12, 13\}$ and $\{14, 15\}$, respectively.

1. Let $J \in \{9, \dots, 12, 13\}$. Computations indicate that for $J = 9$ the null space of the 108×120 matrix $\mathbf{Q}_2(\mathcal{T})$ has dimension 15. (To compute the null space we used the MATLAB built-in function `null`.) Since $\sum C_{d_r+1}^2 = C_2^2 + C_2^2 + C_2^2 + C_3^2 + C_3^2 + C_4^2 = 15$, it follows that (2.17) holds. It is clear that (2.17) will also hold for $J > 9$. Since
$$C_{K+1}^2 - Q = 105 > 101 = -\tilde{L}_1\tilde{L}_2 + \sum_{1\leq r_1<r_2\leq R} L_{r_1}L_{r_2},$$
it follows that condition e) also holds. Hence, by statement 4) of Theorem 2.5, the first factor matrix of $\mathcal{T}$ is unique and can be computed in Phase I of Algorithm 2.1.
2. Let $J \in \{14, 15\}$. Then condition c) in Theorem 2.5 holds. Hence, by statement 5) of Theorem 2.5, the overall decomposition is unique and can be computed by Algorithm 2.1. In step 11 we can, for instance, set $M = 2$ and choose $\Omega_1 = \{1, 2, 3, 4, 5\}$ and $\Omega_2 = \{1, 2, 3, 4, 6\}$. In this case the loop in steps $12 - 16$ is executed twice which yields matrices $\hat{\mathbf{E}}_1, \dots, \hat{\mathbf{E}}_4, \hat{\mathbf{E}}_5$ and

matrices $\alpha_1\hat{\mathbf{E}}_1, \dots, \alpha_4\hat{\mathbf{E}}_4, \hat{\mathbf{E}}_6$, respectively, where $\alpha_1, \dots, \alpha_4$ are nonzero values. The computed matrices $\hat{\mathbf{E}}_1, \dots, \hat{\mathbf{E}}_6$ necessarily coincide with the matrices $\mathbf{E}_1, \dots, \mathbf{E}_6$ in decomposition (1.1) up to permutation of indices and scaling factors. Note that neither R nor $L_1, \dots, L_R$ should be known a priori.

In the following two examples we assume that the decomposition in (1.1) is perturbed with a random additive term. The examples demonstrate the computation of the approximate ML rank-$(1, L_r, L_r)$ decomposition (1.1).

Example 2.10. In this example we illustrate the computation of $L_1, \dots, L_R$ and the computation of the approximate ML rank-$(1, L_r, L_r)$ decomposition assuming that the exact decomposition satisfies condition b) in Theorem 2.5 (i.e., Case 2 in Algorithm 2.1).

First we consider the case where the decomposition is exact. We consider a $3\times8\times8$ tensor generated by (1.2) in which the entries of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are independently drawn from the standard normal distribution $N(0, 1)$ and $L_1 = 2$, $L_2 = 3$, $L_3 = 4$. Thus, $\mathcal{T}$ is a sum of $R = 3$ terms. It can be numerically verified that $d_1 = 1$, $d_2 = 2$, $d_3 = 3$ and that the null space of the 84×36 matrix $\mathbf{Q}_2(\mathcal{T})$ has dimension $10 = C^2_{d_1+1} + C^2_{d_2+1} + C^2_{d_3+1}$. Hence, by statement 5) of Theorem 2.5, the overall decomposition is unique and can be computed by Algorithm 2.1 (Case 2). Note that if the third dimension is decreased by 1, then condition a) in Theorem 2.5 does not hold. It can also be shown that if the first dimension is decreased by 1, then assumption (2.17) in Theorem 2.5 does not hold.

Now we consider a noisy variant. Since the problem is already challenging we exclude to some extent random tensors that may pose additional numerical difficulties[9] by limiting the condition numbers of the matrix unfoldings $\mathbf{T}_{(1)}$ and $\mathbf{T}_{(3)}$. More concretely, we select 100 random tensors with $\max(cond(\mathbf{T}_{(1)}), cond(\mathbf{T}_{(3)})) \le 10$, where $cond(\cdot)$ denotes the condition number of a matrix, i.e., the ratio of the largest and smallest singular value. We estimate the ML rank values and the factor matrices from $T + c\mathcal{N}$, where $\mathcal{N}$ is a perturbation tensor and c controls the signal-to-noise level. The entries of $\mathcal{N}$ are independently drawn from the standard normal distribution $N(0, 1)$ and the following Signal-to-Noise Ratio (SNR) measure is used: $SNR\ [dB] = 10\log(\|\mathcal{T}\|_F^2/c^2\|\mathcal{N}\|_F^2)$, where $\|\cdot\|_F$ denotes the Frobenius norm of a tensor. To compute the decomposition of $\mathcal{T} + c\mathcal{N}$ we use the noisy version of Algorithm 2.1 explained in subsection 2.5.2 (the second scenario). We assume that $R = 3$ and $\sum L_r = 9$ are known. Since we are in a generic setting, $\sum d_r = RK - (R-1)\sum L_r = 6$. Assuming that $d_1 \le d_2 \le d_3$, this implies that the triplet (d_1, d_2, d_3) coincides with one of the triplets $(1, 1, 4)$, $(1, 2, 3)$, $(2, 2, 2)$. The respective values for $C^2_{d_1+1} + C^2_{d_2+1} + C^2_{d_3+1}$ are 8, 10, and 9. Consequently, in our computations we replace Q by $Q_{min} = \min(8, 10, 9) = 8$.

The matrix $\mathbf{A}$ and the values of d_1, d_2, and d_3 are estimated as in subsection 2.5.2 (the second scenario). The matrix $\mathbf{N}$ in the simultaneous EVD in step 2 of Algorithm 1.1 was found in two ways: i) from the EVD of a single generic linear combination of $\mathbf{U}_1, \dots, \mathbf{U}_R$ and ii) by computing CPD (1.7). Since we are in a generic setting, the values of L_1, L_2, and L_3 can be found from the values of d_1, d_2, and d_3 by (2.29). This means that if $L_1 \le L_2 \le L_3$, then the triplet (L_1, L_2, L_3) necessarily coincides with one of the triplets $(2, 2, 5)$, $(2, 3, 4)$, $(3, 3, 3)$. Table 2.1 shows the frequencies with which each triplet occurs as a function of the SNR. To measure the

[9] Note that, if the first or third matrix unfolding has a large condition number, we are approaching, as explained above, a situation in which the conditions in Theorem 2.5 and hence the working assumptions in Algorithm 2.1 are not satisfied.

performance we compute the relative error on the estimates of the first factor matrix $\mathbf{A}$ and on the estimates of the matrix formed by the vectorised multilinear terms, $[\mathbf{a}_1 \otimes \mathrm{vec}(\mathbf{E}_1)\ \ldots\ \mathbf{a}_R \otimes \mathrm{vec}(\mathbf{E}_R)]$. (We compensate for scaling and permutation ambiguities.) The results are shown in Figure 2.1. Note that the accuracy of the estimates is of about the same order as the accuracy of the given tensors.

Table 2.1

Frequencies with which the ML rank values have been estimated correctly (second row) or incorrectly (first and third row) (see Example 2.10)

L_1, L_2, L_3	SNR (dB)							
	15	20	25	30	35	40	45	50
2, 2, 5	21	12	8	-	-	-	-	-
2, 3, 4	63	79	89	96	100	99	100	100
3, 3, 3	16	9	3	4	-	1	-	-

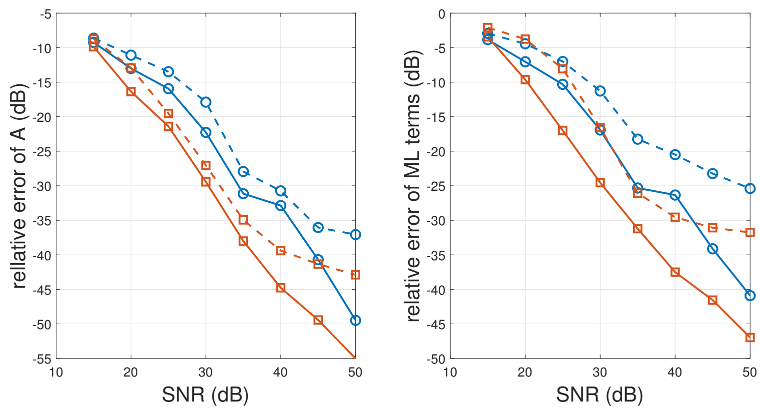

Fig. 2.1. *Mean (○) and median (□) curves for the relative errors on the first factor matrix* $\mathbf{A}$ *(left plot) and the matrix formed by the vectorized ML terms* $[\mathbf{a}_1 \otimes \mathrm{vec}(\mathbf{E}_1)\ \ldots\ \mathbf{a}_R \otimes \mathrm{vec}(\mathbf{E}_R)]$ *(right plot). The dashed and solid line correspond to the version of Algorithm 1.1 where the solution* $\mathbf{N}$ *of the simultaneous EVD in step 2 is obtained from the EVD of a single generic linear combination and from the CPD (1.7), respectively (see Example 2.10).*

Example 2.11. In this example we illustrate the computation of $L_1, \ldots, L_R$ and the computation of the approximate ML rank-$(1, L_r, L_r)$ decomposition assuming that the exact decomposition satisfies condition d) in Theorem 2.5 (i.e., Case 1 in Algorithm 2.1).

We consider a $3 \times 9 \times 10$ tensor generated by (1.2) in which the entries of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are independently drawn from the standard normal distribution $N(0, 1)$ and $L_1 = 1$, $L_2 = 2$, $L_3 = 3$, and $L_4 = 4$. Thus, $\mathcal{T}$ is a sum of $R = 4$ terms. We find numerically that $d_1 = 1$, $d_2 = 2$, $d_3 = 3$, $d_4 = 4$ and that the null space of the 216×55 matrix $\mathbf{Q}_2(\mathcal{T})$ has dimension $20 = C^2_{d_1+1} + C^2_{d_2+1} + C^2_{d_3+1} + C^2_{d_4+1}$. Hence, by statement 5) of Theorem 2.5, the overall decomposition is unique and can be computed by Algorithm 2.1 (Case 1). It can be shown that in this example we are again in a bordering case with respect to working assumptions in Algorithm 2.1, i.e., if the first or third dimension is decreased by 1, then the decomposition cannot be computed by Algorithm 2.1. As in Example 2.10, we use the noisy version of Algorithm 2.1

explained in subsection 2.5.2 (the second scenario). We assume that $R = 4$ and $\sum L_r = 10$ are known. Since we are in a generic setting, $\sum d_r = RK - (R-1)\sum L_r = 10$. One can easily verify that there exist exactly 9 tuples (d_1, d_2, d_3, d_4) such that $d_1 \le d_2 \le d_3 \le d_4$ and $\sum d_r = 10$. Since $K = \sum L_r$ we have that $L_r = d_r$. The possible tuples (L_1, L_2, L_3, L_4) $(= (d_1, d_2, d_3, d_4))$ are shown in the first column of Table 2.2. The respective 9 values for $C^2_{d_1+1} + C^2_{d_2+1} + C^2_{d_3+1} + C^2_{d_4+1}$ are 31, 26, 23, 22, 22, 20, 19, 19 and 18. Consequently, in our computations we replace Q by $Q_{min} = 18$. The matrix $\mathbf{N}$ was found in two ways: i) from the EVD of a single generic linear combination of $\mathbf{U}_1, \dots, \mathbf{U}_R$ and ii) by computing CPD (1.7). In the latter case the last frontal slice of $\mathcal{U}$ in (1.7), i.e., the matrix $\mathbf{U}_R$, was replaced by $\omega\mathbf{U}_R$ with $\omega = 2$ (see explanation at the end of subsection 1.3.2). The results are shown in Table 2.2 and Figure 2.2. Again, despite the difficulty of the problem the accuracy of the estimates is of about the same order as the accuracy of the given tensors.

TABLE 2.2
Frequencies with which the ML rank values have been estimated correctly (sixth row) or incorrectly (remaining rows) (see Example 2.11)

L_1, L_2, L_3, L_4	SNR (dB)							
	15	20	25	30	35	40	45	50
1, 1, 1, 7	1	-	-	-	-	-	-	-
1, 1, 2, 6	5	1	-	-	-	-	-	-
1, 1, 3, 5	8	2	2	-	-	-	-	-
1, 1, 4, 4	4	4	1	3	-	1	-	-
1, 2, 2, 5	13	10	5	-	-	-	-	-
1, 2, 3, 4	54	73	88	96	100	99	100	100
1, 3, 3, 3	6	3	2	-	-	-	-	-
2, 2, 2, 4	3	2	2	-	-	-	-	-
2, 2, 3, 3	6	5	-	1	-	-	-	-

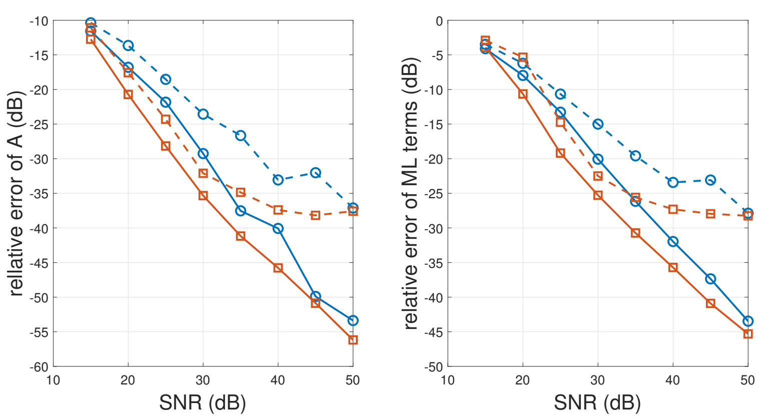

FIG. 2.2. *Mean (○) and median (□) curves for the relative errors on the first factor matrix $\mathbf{A}$ (left plot) and the matrix formed by the vectorized ML terms $[\mathbf{a}_1 \otimes \operatorname{vec}(\mathbf{E}_1) \ \dots \ \mathbf{a}_R \otimes \operatorname{vec}(\mathbf{E}_R)]$ (right plot). The dashed and solid line correspond to the version of Algorithm 1.1 where the solution $\mathbf{N}$ of the simultaneous EVD in step 2 is obtained from the EVD of a single generic linear combination and from the CPD (1.7), respectively (see Example 2.11).*

2.6. Results for generic decompositions. The main results of this subsection are summarized in Table 1.1(b). The results in subsection 2.6.1 are generic counterparts of Corollary 2.7 and Theorem 2.5 and therefore are sufficient for generic uniqueness and guarantee that a generic decomposition can be computed by means of EVD. In subsection 2.6.2 we discuss a necessary condition for generic uniqueness that is more restrictive than generic versions of the conditions in Theorem 2.1 at least for $\mathbb{F} = \mathbb{C}$. In subsection 2.6.3 we present two results on generic uniqueness of decompositions with a factor matrix that has full column rank. These results are generalizations of Strassen's result on generic uniqueness of the CPD. The conditions are very mild are and easy to verify but they do not imply an algorithm.

2.6.1. Generic counterparts of the results from subsection 2.5.1. The first two results of this subsection are the generic counterparts of Corollary 2.7 and Theorem 2.5 (or Theorem 2.6). To simplify the presentation and w.l.o.g. we assume that $L_1 \leq \cdots \leq L_R$. It is clear that the assumptions $J \geq L_{\min(I,R)-1} + \cdots + L_R$ and $I \geq 2$ in Theorem 2.12 are, respectively, the generic version of the assumption $k'_{\mathbf{B}} \geq R - r_{\mathbf{A}} + 2$ and $k_{\mathbf{A}} \geq 2$ in (2.23). The generic version of the condition $k'_{\mathbf{C}} \geq R - r_{\mathbf{A}} + 2$ in (2.24) coincides with $K \geq L_{\min(I,R)-1} + \cdots + L_R$, which always holds because of the assumption $K \geq L_2 + \cdots + L_R + 1$ in (2.34). Hence, in the generic setting, the conditions in (2.24) can be dropped. Thus, we have the following result.

THEOREM 2.12. *Let* $L_1 \leq \cdots \leq L_R \leq \min(J, K)$ *and let* $\mathcal{T} \in \mathbb{F}^{I \times J \times K}$ *admit decomposition* (1.2), *where the entries of the matrices* $\mathbf{A} \in \mathbb{F}^{I \times R}$, $\mathbf{B} \in \mathbb{F}^{J \times \sum L_r}$, *and* $\mathbf{C} \in \mathbb{F}^{K \times \sum L_r}$ *are randomly sampled from an absolutely continuous distribution. Assume that*

$$K \geq L_2 + \cdots + L_R + 1, \tag{2.34}$$

$$J \geq L_{\min(I,R)-1} + \cdots + L_R, \quad \text{and } I \geq 2. \tag{2.35}$$

Then the decomposition of $\mathcal{T}$ *into a sum of max ML rank-*$(1, L_r, L_r)$ *terms is unique and can be computed by means of (simultaneous) EVD.*

In the following theorem, assumptions (2.36), (2.37), (2.38), conditions (2.39)–(2.41) and statements 1) to 4) correspond, respectively, to assumptions (2.14), (2.15), (2.17), conditions e), b), d) and statements 1), 3), 4), 5) in Theorem 2.5. The convention $L_1 \leq \cdots \leq L_R$ implies that $d_1 := K - \sum\limits_{k=1}^{R} L_k + L_1 \leq \cdots \leq d_R := K - \sum\limits_{k=1}^{R} L_k + L_R$. Thus, the R constraints in (2.15) are replaced by the single constraint $d_1 \geq 1$ in (2.37), which moreover coincides with condition a) in Theorem 2.5. Hence, in a generic setting, statement 2) in Theorem 2.5 becomes the part of statement 5) that relies on condition a). That is why the following result contains fewer statements than Theorem 2.5.

THEOREM 2.13. *Let* $L_1 \leq \cdots \leq L_R \leq \min(J, K)$ *and let* $\mathcal{T} \in \mathbb{F}^{I \times J \times K}$ *admit decomposition* (1.2), *where the entries of the matrices* $\mathbf{A} \in \mathbb{F}^{I \times R}$, $\mathbf{B} \in \mathbb{F}^{J \times \sum L_r}$, *and* $\mathbf{C} \in \mathbb{F}^{K \times \sum L_r}$ *are randomly sampled from an absolutely continuous distribution.*

Assume that[10]

$$IJ \geq \sum_{r=1}^{R} L_r \geq K, \tag{2.36}$$

$$d_1 := K - \sum_{r=1}^{R} L_r + L_1 \geq 1, \tag{2.37}$$

and that there exist vectors $\tilde{\mathbf{a}}_r \in \mathbb{F}^I$, *and matrices* $\tilde{\mathbf{B}}_r \in \mathbb{F}^{J\times L_r}$, $\tilde{\mathbf{C}}_r \in \mathbb{F}^{K\times L_r}$ *such that*

$$\dim \operatorname{Null}\left(\mathbf{Q}_2(\tilde{\mathcal{T}})\right) = \sum_{r=1}^{R} C^2_{d_r+1}, \tag{2.38}$$

where $\tilde{\mathcal{T}} = \sum \tilde{\mathbf{a}}_r \circ (\tilde{\mathbf{B}}_r \tilde{\mathbf{C}}_r^T)$ *and* $d_r := K - \sum\limits_{k=1}^{R} L_k + L_r$, $r = 1, \dots, R$. *The following statements hold generically.*

1) *The matrix* $\mathbf{A}$ *in* (1.2) *can be computed by means of (simultaneous) EVD.*
2) *Any decomposition of* $\mathcal{T}$ *into a sum of max ML rank-*$(1, L_r, L_r)$ *terms has* R *nonzero terms and its first factor matrix is equal to* $\mathbf{AP}$*, where every column of* $\mathbf{P} \in \mathbb{F}^{R\times R}$ *contains precisely a single* 1 *with zeros everywhere else.*
3) *If*

$$K \geq -\frac{1}{2} - \sqrt{\frac{1}{4} + \frac{2L_1L_2}{R-1}} + \sum_{r=1}^{R} L_r, \tag{2.39}$$

then the first factor matrix of the decomposition of $\mathcal{T}$ *into a sum of max ML rank-*$(1, L_r, L_r)$ *terms is unique.*

4) *The decomposition of* $\mathcal{T}$ *into a sum of max ML rank-*$(1, L_r, L_r)$ *terms is unique and can be computed by means of (simultaneous) EVD if any of the following two conditions holds:*

$$I \geq R, \tag{2.40}$$

$$K = \sum_{r=1}^{R} L_r. \tag{2.41}$$

Proof. The proof is given in Appendix B. ∎

To verify the uniqueness and EVD-based computability of a generic decomposition in the case $I \geq R$, one can use Theorem 2.12 (i.e., verify the assumptions $K - \sum L_r + L_1 \geq 1$ and $J \geq L_{\min(I,R)-1} + \cdots + L_R = L_{R-1} + L_R$) or Theorem 2.13 (i.e., verify the assumptions $IJ \geq \sum L_r$, $K - \sum L_r + L_1 \geq 1$, and (2.38)). Let us briefly comment on these two options. From statement 4) of Lemma 3.1 below, it follows that for $I \geq R$, the assumptions in Theorem 2.13 are at least as relaxed as the assumptions

[10] The inequality $\sum L_r \geq K$ in (2.36) is added for notational purposes; it simplifies the formulation of (2.37) and (2.38). By statement 2) of Theorem 2.4, uniqueness and computation of a generic decomposition of an $I \times J \times K$ tensor with $K \geq \sum L_r$ follow from uniqueness and computation of a generic decomposition of an $I \times J \times \sum L_r$ tensor. In other words, the assumption $\sum L_r \geq K$ in (2.36) is not a constraint: if $K \geq \sum L_r$, then the assumptions and conditions in Theorem 2.13 should be verified for $K = \sum L_r$.

in Theorem 2.12. On one hand, the assumption $J \geq L_{R-1} + L_R$ in Theorem 2.12 is easy to verify; on the other hand, it can be more restrictive than assumption (2.38) in Theorem 2.13. For instance, it can be verified that uniqueness and EVD-based computability of a generic decomposition of a $3 \times 6 \times 8$ tensor into a sum of max ML rank-$(1, L_r, L_r)$ terms with $L_1 = L_2 = 3$ and $L_3 = 4$ follow from Theorem 2.13 but do not follow from Theorem 2.12 (indeed, $6 = J \geq L_{R-1} + L_R = 3 + 4$ does not hold).

We now explain how to verify assumption (2.38).

In the proof of Theorem 2.13 we explain that if assumption (2.38) holds for one triplet of matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{C}}$, then (2.38) holds also for a generic triplet. The other way around, it suffices to verify (2.38) for a generic triplet, where some care needs to be taken that the algebraic situation is not obfuscated by numerical effects. Hence one possibility to verify (2.38) is to randomly select matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{C}}$, construct $\mathbf{Q}_2(\tilde{\mathcal{T}})$ and estimate its rank numerically. Because of the rounding errors such computations cannot be considered as a formal proof of (2.38), unless it is clear that the rounding did not affect the rank of $\mathbf{Q}_2(\tilde{\mathcal{T}})$. To have a formal proof of (2.38) one can chose matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{C}}$ such that the entries of $\mathbf{Q}_2(\tilde{\mathcal{T}})$ are integers and, possibly, such that $\mathbf{Q}_2(\tilde{\mathcal{T}})$ is sparse, so the identity in (2.38) becomes easy to prove. Both possibilities are illustrated in the upcoming Example 2.14. Another possibility to have a formal proof of (2.38) is to perform all computations over a finite field. This approach is explained in Appendix A. Note that both approaches can be quite expensive and may require a third-party implementation.

Example 2.14. Let $\mathcal{T}$ be $3 \times 3 \times 5$ tensor generated by (1.2) in which the entries of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are independently drawn from the standard normal distribution $N(0, 1)$ and $L_1 = L_2 = L_3 = 1$, $L_4 = 2$. To prove that the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique and can be computed by means of (simultaneous) EVD we verify assumptions (2.36), (2.37), (2.38) and condition (2.41) in Theorem 2.13. Assumptions (2.36), (2.37) and condition (2.41) obviously hold. Let us now illustrate two possibilities to verify (2.38).

I. The matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, *and* $\tilde{\mathbf{C}}$ *are generic.* For 5 randomly generated triplets $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}})$ in Example 2.14, we have obtained that the condition number of the 9×15 matrix $\mathbf{Q}_2(\tilde{\mathcal{T}})$ took values 223.12, 75.46, 681.37, 2832.9, and 147.65 which clearly suggests that $\mathbf{Q}_2(\tilde{\mathcal{T}})$ is a full-rank matrix (i.e., $r_{\mathbf{Q}_2(\tilde{\mathcal{T}})} = 9$). Hence, by the rank-nullity theorem, $\dim \operatorname{Null}\left(\mathbf{Q}_2(\tilde{\mathcal{T}})\right) = 15 - 9 = 6$. Since (2.41) holds, it follows that $d_r = K - \sum\limits_{k=1}^{R} L_k + L_r = L_r$, implying that $C^2_{d_1+1} + \cdots + C^2_{d_4+1} = 1 + 1 + 1 + 3 = 6$. Thus, assumption (2.38) holds if we can trust our impression that $\mathbf{Q}_2(\tilde{\mathcal{T}})$ has full rank generically.

II. The matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, *and* $\tilde{\mathbf{C}}$ *have integer entries.* We set

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \qquad \tilde{\mathbf{B}} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 & 1 \end{bmatrix}, \qquad \tilde{\mathbf{C}} = \mathbf{I}_5$$

and compute $\tilde{\mathcal{T}} = \sum \tilde{\mathbf{a}}_r \circ (\tilde{\mathbf{B}}_r \tilde{\mathbf{C}}_r^T)$. It can be easily verified that

$$\mathbf{Q}_2(\tilde{\mathcal{T}}) = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 & 0 & 3 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -3 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and that the nine nonzero columns of $\mathbf{Q}_2(\tilde{\mathcal{T}})$ are linearly independent. Hence, again, by the rank-nullity theorem, $\dim \operatorname{Null}\Big(\mathbf{Q}_2(\tilde{\mathcal{T}})\Big) = 15 - 9 = 6$. Thus, assumption (2.38) holds with certainty. Note that the matrix $\mathbf{Q}_2(\tilde{\mathcal{T}})$ is sparse and the identity in (2.38) is easy to prove because we paid attention to the choice of the entries of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$.

It is worth noting that the decomposition of a $3 \times 3 \times 5$ tensor into a sum of 5 generic rank-1 terms is not unique. More precisely, it is known that such tensors admit exactly six decompositions [34]. Our example demonstrates that if two of the rank-1 terms are forced to share the same vector in the first mode, and hence together form an ML rank-$(1, 2, 2)$ term, then the decomposition becomes unique.

2.6.2. Necessary condition for generic uniqueness. The necessity of the conditions

$$R \leq JK, \quad \sum L_r \leq IJ, \quad \sum L_r \leq IK \tag{2.42}$$

follows trivially from Theorem 2.1. Next, counting the number of parameters on each side of (1.1), one would expect that uniqueness does not hold if the LHS of (1.1) contains fewer parameters than the RHS:

$$IJK < S := \sum_{r=1}^{R} (I - 1 + (J + K - L_r)L_r), \tag{2.43}$$

where the value S is an upper bound on the number of parameters needed to parameterize[11] a sum of R generic ML rank-$(1, L_r, L_r)$ terms in the LHS of (1.1) and IJK is equal to the dimension of the space of $I \times J \times K$ tensors. In fact it is known [37] and follows from the fiber dimension theorem [30, Theorem 3.7, p. 78] that the reverse of inequality (2.43), that is

$$S = \sum_{r=1}^{R} (I - 1 + (J + K - L_r)L_r) \leq IJK, \tag{2.44}$$

is necessary for generic uniqueness if $\mathbb{F} = \mathbb{C}$. It can be verified that condition (2.44) is more restrictive than (2.42) and, thus, is more interesting at least for $\mathbb{F} = \mathbb{C}$.

[11] The number of parameters can be computed as follows. Using, for instance, the LDU factorization we obtain that a generic $J \times K$ rank-L_r matrix involves $(JL_r - \frac{L_r(L_r+1)}{2}) + L_r + (KL_r - \frac{L_r(L_r+1)}{2}) = (J + K - L_r)L_r$ parameters, where we obviously assume that $\max L_r \leq \min(J, K)$. Hence, the rth term in (1.1) can be parameterized with $I - 1 + (J + K - L_r)L_r$ parameters.

Recall that for $L_1 = \cdots = L_R = 1$ the minimal decomposition of form (1.2) corresponds to CPD. It has been shown in [7] that, for CPD, *the condition* $S < IJK \leq 15000$ *is also sufficient* for generic uniqueness, with a few known exceptions. The following example demonstrates that for the decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms the bound is $S < IJK$ *not* sufficient. However, in the example the first factor matrix is generically unique, i.e., the decomposition is generically partially unique.

Example 2.15. We consider a $2 \times 8 \times 7$ tensor generated as the sum of three random ML rank-$(1, 3, 3)$ tensors. More precisely, the tensors are generated by (1.2) in which the entries of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are independently drawn from the standard normal distribution $N(0, 1)$. Since $S = 3(2-1+(8+7-3)3) = 111$ and $IJK = 112$, the inequality $S < IJK$ holds. In this example first we show that tensors generated in this way admit infinitely many decompositions, namely, we show that there exists at least a two-parameter family of decompositions. Second, we prove generic uniqueness of the first factor matrix.

Nonuniqueness of the generic decomposition. Let $\mathcal{T}$ admit decomposition (1.2) with generic factor matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. Then the matrices $\mathbf{U} := [\mathbf{a}_2\ \mathbf{a}_3] \in \mathbb{F}^{2\times 2}$, $\mathbf{V} := [\mathbf{b}_2\ \dots\ \mathbf{b}_9] \in \mathbb{F}^{8\times 8}$, and $\mathbf{W} := [\mathbf{c}_1\ \dots\ \mathbf{c}_5\ \mathbf{c}_7\ \mathbf{c}_8] \in \mathbb{F}^{7\times 7}$ are nonsingular. Let $\widehat{\mathcal{T}}$ denote a tensor such that $\widehat{\mathbf{T}}_{(3)} = (\mathbf{U}^{-1} \otimes \mathbf{V}^{-1})\mathbf{T}_{(3)}\mathbf{W}^{-T}$. Then, by (1.5), $\widehat{\mathcal{T}}$ admits the decomposition of the form (1.2), where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are replaced by

$$\mathbf{U}^{-1}\mathbf{A} = \begin{bmatrix} d_1 & 1 & 0 \\ d_2 & 0 & 1 \end{bmatrix}, \quad \mathbf{V}^{-1}\mathbf{B} = [\mathbf{f}\ \mathbf{I}_8], \quad \text{and} \quad \mathbf{W}^{-1}\mathbf{C} = [\mathbf{e}_1\ \mathbf{e}_2\ \mathbf{e}_3\ \mathbf{e}_4\ \mathbf{e}_5\ \mathbf{g}\ \mathbf{e}_6\ \mathbf{e}_7\ \mathbf{h}],$$

respectively. It is clear that a decomposition of $\widehat{\mathcal{T}}$ with factor matrices $\widehat{\mathbf{A}}$, $\widehat{\mathbf{B}}$, and $\widehat{\mathbf{C}}$ generates a decomposition of $\mathcal{T}$ with factor matrices $\mathbf{U}\widehat{\mathbf{A}}$, $\mathbf{V}\widehat{\mathbf{B}}$, and $\mathbf{W}\widehat{\mathbf{C}}$. In particular, if the decomposition of $\widehat{\mathcal{T}}$ is not unique, then the decomposition of $\mathcal{T}$ is not unique either. Below we present a procedure to construct a two-parameter family of decompositions of $\widehat{\mathcal{T}}$. First we choose parameters $p_1, p_2 \in \mathbb{F}$ and compute the values α, β, γ, and δ:

$$\begin{aligned}
&\alpha = (f_1g_2 - g_1 + f_2g_3)p_1 + (f_1h_2 - h_1 + f_2h_3)p_2 + 1,\\
&\beta = (f_3g_4 - f_5 + f_4g_5)d_1p_1 + (f_3h_4 + f_4h_5)d_1p_2,\\
&\gamma = (f_6g_6 + f_7g_7)d_2p_1 + (f_6h_6 - f_8 + f_7h_7)d_2p_2,\\
&\delta = \beta + \alpha - \gamma\alpha.
\end{aligned}$$

Second, if α and δ are nonzero, we also compute the values:

$$\begin{aligned}
&\tau_1 = -p_1\gamma/\delta, && \tau_2 = -p_2\beta/\delta, && \tau_3 = (p_2 + \tau_2)/\alpha, && \tau_4 = \alpha\tau_1 - p_1,\\
&q_1 = h_1\tau_3 + g_1\tau_1 + 1, && q_2 = h_1\tau_2 + g_1\tau_4 + 1, && r_1 = h_2\tau_3 + g_2\tau_1, && r_2 = h_2\tau_2 + g_2\tau_4,\\
&s_1 = h_3\tau_3 + g_3\tau_1, && s_2 = h_3\tau_2 + g_3\tau_4, && &&\\
&t = h_4p_2/\delta, && u = h_5p_2/\delta, && v = -g_6p_1/\delta, && w = -g_7p_1/\delta.
\end{aligned}$$

Third, we construct matrices $\tilde{\mathbf{E}}_1$, $\tilde{\mathbf{E}}_2$, and $\tilde{\mathbf{E}}_3$ as

$$\tilde{\mathbf{E}}_1 := \begin{bmatrix} f_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ f_2 & 0 & 1 & 0 & 0 & 0 & 0 \\ f_3q_1 & f_3r_1 & f_3s_1 & f_3t & f_3u & f_3v & f_3w \\ f_4q_1 & f_4r_1 & f_4s_1 & f_4t & f_4u & f_4v & f_4w \\ f_5q_1 & f_5r_1 & f_5s_1 & f_5t & f_5u & f_5v & f_5w \\ f_6q_2 & f_6r_2 & f_6s_2 & f_6t\alpha & f_6u\alpha & f_6v\alpha & f_6w\alpha \\ f_7q_2 & f_7r_2 & f_7s_2 & f_7t\alpha & f_7u\alpha & f_7v\alpha & f_7w\alpha \\ f_8q_2 & f_8r_2 & f_8s_2 & f_8t\alpha & f_8u\alpha & f_8v\alpha & f_8w\alpha \end{bmatrix},$$

$$\tilde{\mathbf{E}}_2 := \widehat{\mathbf{H}}_1 - d_1\tilde{\mathbf{E}}_1, \qquad \tilde{\mathbf{E}}_3 := \widehat{\mathbf{H}}_2 - d_2\tilde{\mathbf{E}}_1, \tag{2.45}$$

where $\widehat{\mathbf{H}}_1 \in \mathbb{F}^{8\times 7}$ and $\widehat{\mathbf{H}}_2 \in \mathbb{F}^{8\times 7}$ denote the horizontal slices of $\widehat{\mathcal{T}}$. The identities in (2.45) mean that $\widehat{\mathcal{T}} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \circ \tilde{\mathbf{E}}_1 + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \circ \tilde{\mathbf{E}}_2 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \circ \tilde{\mathbf{E}}_3$, i.e., $\widehat{\mathcal{T}}$ admits a two-parameter family of decompositions, as indicated above. By symbolic computations in MATLAB we have also verified that all 4×4 minors of $\tilde{\mathbf{E}}_1$, $\tilde{\mathbf{E}}_2$, and $\tilde{\mathbf{E}}_3$ are identically zero, that is $\tilde{\mathbf{E}}_1$, $\tilde{\mathbf{E}}_2$, and $\tilde{\mathbf{E}}_3$ are at most rank-3 matrices.

Generic uniqueness of the first factor matrix.

Let $\tilde{\mathcal{T}} := \sum \tilde{\mathbf{a}}_r \circ (\tilde{\mathbf{B}}_r \tilde{\mathbf{C}}_r^T)$ with

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \qquad \tilde{\mathbf{B}}_1\tilde{\mathbf{C}}_1^T = [\mathbf{e}_5 + \mathbf{e}_7\ \mathbf{e}_1\ \mathbf{e}_2\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}],$$

$$\tilde{\mathbf{B}}_2\tilde{\mathbf{C}}_2^T = [\mathbf{0}\ \mathbf{0}\ \mathbf{e}_5\ \mathbf{e}_3\ \mathbf{e}_4\ \mathbf{e}_5\ \mathbf{e}_5], \qquad \tilde{\mathbf{B}}_3\tilde{\mathbf{C}}_3^T = [\mathbf{e}_8\ \mathbf{0}\ \mathbf{e}_8\ \mathbf{0}\ \mathbf{e}_8\ \mathbf{e}_6\ \mathbf{e}_7],$$

where $\mathbf{e}_1, \dots, \mathbf{e}_8$ denote the vectors of the canonical basis of $\mathbb{F}^8$.

Generic uniqueness of the first factor matrix follows from statement 3) of Theorem 2.13. Indeed, (2.36), (2.37), and (2.39) are trivial: $7 = K < IJ = 16$, $K - \sum L_r + \min L_r = 7 - 9 + 3 = 1$, $7 = K \geq -\frac{1}{2} - \sqrt{\frac{1}{4} + \frac{2L_1L_2}{R-1}} + \sum\limits_{r=1}^{R} L_r = -\frac{1}{2} - \sqrt{\frac{1}{4} + 9} + 9 \approx 5.5$. Condition (2.38) can be verified exactly, i.e., without round-off errors for the specific $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{C}}$ given above. (For this particular choice of $\tilde{\mathcal{T}}$, the 28×28 matrix $\mathbf{Q}_2(\tilde{\mathcal{T}})$ is sparse and its nonzero entries belong to the set $\{-2, -1, 0, 1, 2\}$). Moreover, the first factor matrix can be computed in Phase I of Algorithm 2.1. Since $d_r = K - (\sum_{p=1}^{R} L_p - L_r) = 7 - (9 - 3) = 1$, it follows that the S-JBD in step 5 reduces to joint diagonalization.

2.6.3. Strassen type results: decompositions with a factor matrix that has full column rank. In this subsection we narrow the investigation of generic uniqueness to the situation where one of the factor matrices has full column rank. Put the other way around, we generalize the famous Strassen result for generic uniqueness of the CPD for situations in which a factor matrix has full column rank to the decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms. While CPD is symmetric in $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, in the decomposition into a sum of ML rank-$(1, L_r, L_r)$ terms factor matrix $\mathbf{A}$ plays a role that is different from the role of $\mathbf{B}$ and $\mathbf{C}$. Consequently, we will consider two cases. In the first case we assume that $R \leq I$, i.e., that the first factor matrix has full column rank (see Theorem 2.16). In the second case we assume that $\sum L_r \leq K$, i.e., that the third factor matrix has full column rank (see Theorem 2.17). The result for $\sum L_r \leq J$, i.e., for the case where the second factor matrix has full column rank then follows from Theorem 2.17 by symmetry.

First factor matrix has full column rank. First we recall the corresponding result for the CPD. One can easily verify that if $L_1 = \cdots = L_R = 1$ and $R \le I$, then the bound $S \le IJK$ in (2.44) is equivalent to $R \le (J-1)(K-1)+1$. In [3] it was shown that generically for $R = (J-1)(K-1)+1$ and $R \le I$ a tensor admits more than one decomposition. Hence, if $R \le I$ and $\mathbb{F} = \mathbb{C}$, for generic uniqueness of the CPD it is necessary to have that

$$R \le (J-1)(K-1). \tag{2.46}$$

If $R \le I$ and $\mathbb{F} = \mathbb{R}$, then, in general, condition (2.46) is not necessary for generic uniqueness of CPD [1]. On the other hand, it is well-known [33] (see also [19, Corollary 1.7], [3] and references therein) that if $R \le I$, then condition (2.46) is sufficient for generic uniqueness of the CPD for both $\mathbb{F} = \mathbb{R}$ and $\mathbb{F} = \mathbb{C}$. Thus, under the assumption $R \le I$, condition (2.46) is sufficient if $\mathbb{F} = \mathbb{R}$ and condition (2.46) is necessary and sufficient if $\mathbb{F} = \mathbb{C}$. The following theorem generalizes this "Strassen-type" CPD result for the decomposition into a sum of ML rank-$(1, L, L)$ terms. (One can easily verify that if $R \le I$, then the condition $R \le (J-L)(K-L)$ in (2.47) is equivalent to the bound $S < IJK$ in (2.44)).

THEOREM 2.16. *Let $\mathcal{T}$ admit decomposition (1.2), where*

$$L_1 = \cdots = L_R =: L \le \min(J, K), \qquad R \le I$$

and the entries of the matrices $\mathbf{A}$*,* $\mathbf{B}$*, and* $\mathbf{C}$ *are randomly sampled from an absolutely continuous distribution. For both $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, if*

$$R \le (J-L)(K-L), \tag{2.47}$$

then the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique. If $\mathbb{F} = \mathbb{C}$ and $R \ge (J-L)(K-L)+2$, then the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is not unique.[12]

Proof. The proof is given in Appendix C. ∎

Second or third factor matrix has full column rank. Permuting I, J and K in the Strassen condition (2.46), we have that generic uniqueness of the CPD holds if

$$R \le (I-1)(J-1) \ \text{ and } \ R \le K. \tag{2.48}$$

While Theorem 2.16 extended CPD condition (2.46), the following theorem generalizes (2.48) for the decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms.

THEOREM 2.17. *Let $L_1 \le \cdots \le L_R \le \min(J, K)$ and let $\mathcal{T}$ admit decomposition (1.2), where the entries of the matrices* $\mathbf{A}$*,* $\mathbf{B}$*, and* $\mathbf{C}$ *are randomly sampled from an absolutely continuous distribution. If*

$$2 \le I, \quad L_{R-1} + L_R \le J, \qquad \sum_{r=1}^{R} L_r \le (I-1)(J-1), \qquad \textit{and} \qquad \sum_{r=1}^{R} L_r \le K, \tag{2.49}$$

then the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique.

Proof. The proof is given in Appendix H. ∎

[12]The remaining case $\mathbb{F} = \mathbb{C}$, $R \le I$, and $R \ge (J-L)(K-L)+1$ requires further investigation.

Recall that if $\mathbb{F} = \mathbb{C}$, then condition (2.47) in Theorem 2.16 is both necessary and sufficient for generic uniqueness. Apparently, condition $\sum\limits_{r=1}^{R} L_r \leq (I-1)(J-1)$ in Theorem 2.17 is only sufficient. Indeed, one can easily verify that if $\sum L_r \leq K$, then the necessary bound $S \leq IJK$ in (2.44) is equivalent to $\sum L_r \leq (I-1)(J-1) + (I-1)\frac{\sum L_r - R}{\sum L_r} + \frac{\sum L_r^2}{\sum L_r}$. Thus, the gap between the necessary bound $S \leq IJK$ in (2.44) and the sufficient bound $\sum L_r \leq (I-1)(J-1)$ in Theorem 2.17 is equal to $(I-1)\frac{\sum L_r - R}{\sum L_r} + \frac{\sum L_r^2}{\sum L_r}$.

2.7. Constrained decompositions. In many applications the factor matrices $\mathbf{A}$, $\mathbf{B}$, and/or $\mathbf{C}$ in decomposition (1.2) are subject to constraints like non-negativity [4], partial symmetry [27], Vandermonde structure of columns [26], etc.

In this subsection we briefly explain how the results from previous sections can be applied to constrained decompositions.

It is clear that Theorem 2.5 can be applied as is. Indeed, if, for instance, assumptions (2.14)–(2.16) and conditions a) and b) in Theorem 2.5 hold for a constrained decomposition of $\mathcal{T}$, then, by statement 5), the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique and can be computed by means of (simultaneous) EVD. This result also implies that Algorithm 2.1 will find the constrained decomposition.

Now we discuss variants for generic uniqueness. We assume that the factor matrices in the constrained decomposition depend analytically on some complex or real parameters, which is the case in all instances above. More specifically, we assume that the entries of $\mathbf{A}(\mathbf{z})$, $\mathbf{B}(\mathbf{z})$, and $\mathbf{C}(\mathbf{z})$ are analytic functions of $\mathbf{z} \in \mathbb{F}^n$ and that the matrix functions $\mathbf{A}(\mathbf{z})$, $\mathbf{B}(\mathbf{z})$, $\mathbf{C}(\mathbf{z})$ are known. One can define generic uniqueness of a constrained decomposition similar to the unconstrained case: the decomposition of an $I \times J \times K$ tensor into a sum of constrained max ML rank-$(1, L_r, L_r)$ terms is generically unique if

$$\mu_n\{\mathbf{z} : \text{ decomposition } \mathcal{T} = \sum_{r=1}^{R} \mathbf{a}_r(\mathbf{z}) \circ (\mathbf{B}_r(\mathbf{z})\mathbf{C}_r(\mathbf{z})^T) \text{ is not unique}\} = 0,$$

where μ_n denotes a measure on $\mathbb{F}^n$ that is absolutely continuous with respect to the Lebesgue measure. It is clear that Definition 1.3 corresponds to the case $n = IR + J\sum L_r + K\sum L_r$. Note that depending on structure of the factor matrices, the bounds in the statements of Theorems 2.16 and 2.17 may not hold or can be further improved. Also, Theorems 2.12 and 2.13 cannot be used as is; instead one should verify that the conditions of Theorem 2.5 hold for generic $\mathbf{z}$. Note that, because of the analytical dependency of the factor matrices on $\mathbf{z}$, it is sufficient to verify the assumptions and conditions in Theorem 2.5 for a single triplet of constrained factor matrices.

Example 2.18. In the decomposition considered in [26], $\mathbf{B}$ and $\mathbf{C}$ are Vandermonde structured matrices, namely,

$$\mathbf{b}_p = [1 \;\; \exp(jC_1 z_p) \;\; \dots \;\; (\exp(jC_1 z_p)^{J-1})]^T, \;\; p = 1, \dots, s$$
$$\mathbf{c}_q = [1 \;\; \exp(jC_2 \sin(z_{s+q})) \;\; \dots \;\; \exp(jC_2 \sin(z_{s+q}))^{K-1}]^T, \;\; q = 1, \dots, s,$$

where C_1 and C_2 are known real values, $s := \sum L_r$, and $z_1, \dots, z_{2s}$ are unknown real values. No structure is assumed on $\mathbf{A}$, so it can parameterized with IR parameters

$z_{2s+1},\dots,z_{2s+IR}$ which we will also assume real. Thus, the overall constrained decomposition can be parameterized with $n = 2s + IR$ real parameters. W.l.o.g. we assume that $L_1 \le \dots \le L_R$. We claim that if

$$\text{(2.50)}\quad IJ \ge \sum_{r=1}^{R} L_r, \quad K \ge L_2 + \dots + L_R + 1, \quad R \ge I \ge 3, \quad J \ge L_{I-1} + \dots + L_R,$$

then the constrained decomposition is generically unique. Indeed, generically the matrices $\mathbf{B}$ and $\mathbf{C}$ have maximal k'-rank and the matrix $\mathbf{A}$ has maximal k-rank. The assumptions in (2.50) just express the fact that assumptions (2.14)–(2.16) and conditions a) and c) in Theorem 2.5 hold generically. Thus, the generic uniqueness of the constrained decomposition follows from statement 5) of Theorem 2.5.

3. Expression of $\mathbf{R}_2(\mathcal{T})$ and $\mathbf{Q}_2(\mathbf{T})$ in terms of A, B, and C.

In this section we explain construction of the matrices $\Phi(\mathbf{A},\mathbf{B})$ and $\mathbf{S}_2(\mathbf{C})$ that have appeared in Theorem 2.6. The results of this section will also be used later in the proof of statement 4) of Theorem 2.5.

Let $\mathbf{x},\mathbf{y} \in \mathbb{F}^n$. Then $\mathbf{x}\wedge\mathbf{y}$ denotes a $C_n^2 \times 1$ vector formed by all 2×2 minors of $[\mathbf{x}\ \mathbf{y}]$ and $\mathbf{x}\cdot\mathbf{y}$ denotes a $C_{n+1}^2 \times 1$ vector formed by all 2×2 permanents of $[\mathbf{x}\ \mathbf{y}]$. More specifically,

$$\text{the } (n_1 + C_{n_2-1}^2)\text{-th entry of}\quad \mathbf{x}\wedge\mathbf{y} \text{ equals } x_{n_1}y_{n_2} - x_{n_2}y_{n_1},\quad 1 \le n_1 < n_2 \le n,$$
$$\text{the } (n_1 + C_{n_2}^2)\text{-th entry of}\quad \mathbf{x}\cdot\mathbf{y} \text{ equals } x_{n_1}y_{n_2} + x_{n_2}y_{n_1},\quad 1 \le n_1 \le n_2 \le n.$$

It can easily be verified that $\mathbf{x}\wedge\mathbf{y}$ and $\mathbf{x}\cdot\mathbf{y}$ coincide with the vectorized strictly upper triangular part of $\mathbf{x}\mathbf{y}^T - \mathbf{y}\mathbf{x}^T$ and with the vectorized upper triangular part of $\mathbf{x}\mathbf{y}^T + \mathbf{y}\mathbf{x}^T$, respectively.

We extend the definitions of "$\wedge$" and "$\cdot$" to matrices as follows. If $\mathbf{B}_{r_1} \in \mathbb{F}^{J\times L_{r_1}}$ and $\mathbf{B}_{r_2} \in \mathbb{F}^{J\times L_{r_2}}$ are submatrices of $\mathbf{B}$, then $\mathbf{B}_{r_1}\wedge\mathbf{B}_{r_2}$ is the $C_J^2 \times L_{r_1}L_{r_2}$ matrix that has columns $\mathbf{b}_{l_1,r_1}\wedge\mathbf{b}_{l_2,r_2}$, where $1 \le l_1 \le L_{r_1}$ and $1 \le l_2 \le L_{r_2}$, i.e.,

$$\mathbf{B}_{r_1}\wedge\mathbf{B}_{r_2} := [\mathbf{b}_{1,r_1}\wedge\mathbf{b}_{1,r_2}\ \dots \mathbf{b}_{1,r_1}\wedge\mathbf{b}_{L_2,r_2}\ \dots\ \mathbf{b}_{L_1,r_1}\wedge\mathbf{b}_{1,r_2}\ \dots \mathbf{b}_{L_1,r_1}\wedge\mathbf{b}_{L_2,r_2}].$$

If $\mathbf{C}_{r_1} \in \mathbb{F}^{K\times L_{r_1}}$ and $\mathbf{C}_{r_2} \in \mathbb{F}^{K\times L_{r_1}}$ are submatrices of $\mathbf{C}$, then $\mathbf{C}_{r_1}\cdot\mathbf{C}_{r_2}$ is the $C_{K+1}^2 \times L_{r_1}L_{r_2}$ matrix that has columns $\mathbf{c}_{l_1,r_1}\cdot\mathbf{c}_{l_2,r_2}$, where $1 \le l_1 \le L_{r_1}$ and $1 \le l_2 \le L_{r_2}$, i.e.,

$$\mathbf{C}_{r_1}\cdot\mathbf{C}_{r_2} := [\mathbf{c}_{1,r_1}\cdot\mathbf{c}_{1,r_2}\ \dots \mathbf{c}_{1,r_1}\cdot\mathbf{c}_{L_2,r_2}\ \dots\ \mathbf{c}_{L_1,r_1}\cdot\mathbf{c}_{1,r_2}\ \dots \mathbf{c}_{L_1,r_1}\cdot\mathbf{c}_{L_2,r_2}].$$

Let $\mathbf{P}_n$ denote the $n^2 \times C_{n+1}^2$ matrix defined on all vectors of the form $\mathbf{x}\cdot\mathbf{y}$ by

$$\text{(3.1)}\qquad \mathbf{P}_n(\mathbf{x}\cdot\mathbf{y}) = \mathbf{x}\otimes\mathbf{y} + \mathbf{y}\otimes\mathbf{x}$$

and extended by linearity. It can be easily checked that for $n = K$ the matrix $\mathbf{P}_n$ can be constructed as in (2.10), so $\mathbf{P}_n^T$ is a column selection matrix.

Lemma 3.1. *Let $\mathcal{T}$ admit decomposition (1.2), $r_{\mathbf{C}} = K$, and let the values d_r be defined in (2.20). Define the $C_I^2C_J^2 \times \sum\limits_{r_1<r_2} L_{r_1}L_{r_2}$ matrix $\Phi(\mathbf{A},\mathbf{B})$ and $C_{K+1}^2 \times \sum\limits_{r_1<r_2} L_{r_1}L_{r_2}$ matrix $\mathbf{S}_2(\mathbf{C})$ as*

$$\text{(3.2)}\qquad \Phi(\mathbf{A},\mathbf{B}) := [(\mathbf{a}_1\wedge\mathbf{a}_2)\otimes(\mathbf{B}_1\wedge\mathbf{B}_2)\ \dots\ (\mathbf{a}_{R-1}\wedge\mathbf{a}_R)\otimes(\mathbf{B}_{R-1}\wedge\mathbf{B}_R)],$$
$$\text{(3.3)}\qquad \mathbf{S}_2(\mathbf{C}) := [\mathbf{C}_1\cdot\mathbf{C}_2\ \dots\ \mathbf{C}_{R-1}\cdot\mathbf{C}_R].$$

Then

1) $\mathbf{Q}_2(\mathcal{T}) = \Phi(\mathbf{A}, \mathbf{B})\mathbf{S}_2(\mathbf{C})^T$*;*
2) $\mathbf{R}_2(\mathcal{T}) = \Phi(\mathbf{A}, \mathbf{B})\mathbf{S}_2(\mathbf{C})^T\mathbf{P}_K^T$*, where* $\mathbf{P}_K$ *is defined as in* (3.1)*;*
3) $\dim \operatorname{Null}(\mathbf{Q}_2(\mathcal{T})) \geq \dim \operatorname{Null}\left(\mathbf{S}_2(\mathbf{C})^T\right) = \sum C^2_{d_r+1}$*;*
4) if $r_{\mathbf{A}} + k'_{\mathbf{B}} \geq R + 2$ *and* $k_{\mathbf{A}} \geq 2$*, then the matrix* $\Phi(\mathbf{A}, \mathbf{B})$ *has full column rank and* $\dim \operatorname{Null}\left(\Phi(\mathbf{A}, \mathbf{B})\mathbf{S}_2(\mathbf{C})^T\right) = \sum C^2_{d_r+1}$*, i.e.,* (2.21) *implies* (2.22)*; similarly,* (2.16) *implies* (2.17)*;*
5) If $\Phi(\mathbf{A}, \mathbf{B})$ *has full column rank, then* $[\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R]$ *also has full column rank;*
6) If $\Phi(\mathbf{A}, \mathbf{B})$ *has full column rank, then* $k'_{\mathbf{B}} \geq 2$.

Proof. The proofs of statements 1), 2) and 6) follow from the construction of the matrices $\mathbf{Q}_2(\mathcal{T})$, $\Phi(\mathbf{A}, \mathbf{B})$, $\mathbf{S}_2(\mathbf{C})$ and are therefore grouped in Appendix D. The proof of statement 3) consists of several steps and is given in a dedicated Appendix E. The proofs of statements 4) and 5) rely on Lemma F.1, which contains auxiliary results on compound matrices. Lemma F.1 and statements 4), 5) are proved in Appendix F. ◻

COROLLARY 3.2. *Let* $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ *admit the ML rank-*$(1, L_r, L_r)$ *decomposition* (1.2)*. Let also the matrices* $\mathbf{A}$ *and* $\mathbf{C}$ *have full column rank and assumptions* (2.19)*,* (2.20)*, and* (2.22) *in Theorem* 2.6 *hold. Then the matrices* $[\mathbf{B}_i \ \mathbf{B}_j]$ *have full column rank for all* $1 \leq i < j \leq R$*. In particular, assumption* b) *in Theorem* 1.5 *holds.*

Proof. The proof is given in Appendix D. ◻

4. Proof of Theorem 2.5 .

We will need the following two lemmas.

LEMMA 4.1. *Let* $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ *admit the ML rank-*$(1, L_r, L_r)$ *decomposition* (1.1)*. Assume that conditions* (2.14) *and* (2.15) *hold. Let* $\mathbf{N}_r$ *be a* $K \times d_r$ *matrix whose columns form a basis of* $\operatorname{Null}(\mathbf{Z}_r)$ *and let* $\mathbf{M}_r$ *be a* $d_r^2 \times C^2_{d_r+1}$ *matrix whose columns form a basis of the subspace* $\operatorname{vec}(\mathbb{F}^{d_r\times d_r}_{sym})$ *(see* (2.11)*),* $r = 1, \dots, R$*. By definition, set*

$$\mathbf{N} := [\mathbf{N}_1 \ \dots \ \mathbf{N}_R], \qquad \mathbf{W} := [(\mathbf{N}_1 \otimes \mathbf{N}_1)\mathbf{M}_1 \ \dots \ (\mathbf{N}_R \otimes \mathbf{N}_R)\mathbf{M}_R].$$

The following statements hold.
1) The $K \times \sum d_r$ *matrix* $\mathbf{N}$ *has full column rank.*
2) The $K^2 \times Q$ *matrix* $\mathbf{W}$ *has full column rank, where* $Q = C^2_{d_1+1} + \cdots + C^2_{d_R+1}$*.*
3) The matrices $\mathbf{E}_1, \dots, \mathbf{E}_R$ *are linearly independent.*

Proof. The proof is given in Appendix G. ◻

LEMMA 4.2. *Let* $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ *admit the ML rank-*$(1, L_r, L_r)$ *decomposition* (1.1) *in which the matrices* $\mathbf{E}_1, \dots, \mathbf{E}_R$ *are linearly independent and such that either condition* b) *or condition* c) *in Theorem* 2.5 *holds. Then the following statements hold.*
1) If the matrix $\mathbf{A}$ *is known, then the matrices* $\mathbf{E}_1, \dots, \mathbf{E}_R$ *can be computed by means of EVD.*
2) Any decomposition of $\mathcal{T}$ *of the form*

$$\mathcal{T} = \sum_{r=1}^{\tilde{R}} \tilde{\mathbf{a}}_r \circ \tilde{\mathbf{E}}_r, \quad \tilde{\mathbf{a}}_r \text{ is a column of } \mathbf{A}, \quad \tilde{\mathbf{E}}_r \in \mathbb{F}^{J\times K}, \quad 1 \leq r_{\tilde{\mathbf{E}}_r} \leq L_r, \quad \tilde{R} \leq R$$

coincides with decomposition (1.1).

Proof. The proof is given in Appendix G. ◻

Proof of Theorem 2.5*. Proof of statement* 1)*.* Let $\mathbf{T}_1, \dots, \mathbf{T}_K$ denote the frontal slices of $\mathcal{T}$, $\mathbf{T}_k := (t_{ijk})_{i,j=1}^{I,J}$ and let $\mathbf{N}_r$ be a $K \times d_r$ matrix whose columns form a basis of $\operatorname{Null}(\mathbf{Z}_r)$. If $\mathbf{f} = \mathbf{N}_r\mathbf{x}$ for some nonzero $\mathbf{x} \in \mathbb{F}^{d_r}$, then

$$
\begin{aligned}
f_1\mathbf{T}_1 + \dots + f_K\mathbf{T}_K = \sum_{k=1}^{K} f_k \sum_{q=1}^{R} \mathbf{a}_q\mathbf{e}_{k,q}^T = \sum_{q=1}^{R}\mathbf{a}_q \sum_{k=1}^{K}\mathbf{e}_{k,q}^T f_k = \\
\sum_{q=1}^{R}\mathbf{a}_q(\mathbf{E}_q\mathbf{f})^T = \sum_{q=1}^{R}\mathbf{a}_q(\mathbf{E}_q\mathbf{N}_r\mathbf{x})^T = \mathbf{a}_r(\mathbf{E}_r\mathbf{N}_r\mathbf{x})^T,
\end{aligned} \tag{4.1}
$$

where $\mathbf{e}_{k,q}$ denotes the kth column of $\mathbf{E}_q$. Thus,

$$
r_{f_1\mathbf{T}_1+\dots+f_K\mathbf{T}_K} \leq 1 \text{ for all } \mathbf{f} = \mathbf{N}_r\mathbf{x}, \text{ where } \mathbf{x} \in \mathbb{F}^{d_r},\ r = 1,\dots,R. \tag{4.2}
$$

In subsection 2.3 we have explained that the condition $r_{f_1\mathbf{T}_1+\dots+f_K\mathbf{T}_K} \leq 1$ is equivalent to the condition $\mathbf{R}_2(\mathcal{T})(\mathbf{f}\otimes\mathbf{f}) = \mathbf{0}$, where the matrix $\mathbf{R}_2(\mathcal{T})$ is constructed in Definition 2.2, i.e., that equality (2.4) holds. Hence from (4.2), (2.4) and the identity

$$
\mathbf{R}_2(\mathcal{T})(\mathbf{f}\otimes\mathbf{f}) = \mathbf{R}_2(\mathcal{T})((\mathbf{N}_r\mathbf{x})\otimes(\mathbf{N}_r\mathbf{x})) = \mathbf{R}_2(\mathcal{T})(\mathbf{N}_r\otimes\mathbf{N}_r)(\mathbf{x}\otimes\mathbf{x}),
$$

it follows that

$$
\mathbf{R}_2(\mathcal{T})(\mathbf{N}_r\otimes\mathbf{N}_r)(\mathbf{x}\otimes\mathbf{x}) = \mathbf{0}, \quad \text{for all } \mathbf{x}\in\mathbb{F}^{d_r} \quad \text{and } r = 1,\dots,R. \tag{4.3}
$$

Since

$$
\operatorname{vec}(\mathbb{F}_{sym}^{d_r\times d_r}) = \operatorname{span}\{\mathbf{x}\otimes\mathbf{x}:\ \mathbf{x}\in\mathbb{F}^{d_r}\},
$$

it follows that (4.3) is equivalent to

$$
\mathbf{R}_2(\mathcal{T})(\mathbf{N}_r\otimes\mathbf{N}_r)\mathbf{m}_r = \mathbf{0}, \quad \text{for all } \mathbf{m}_r \in \operatorname{vec}(\mathbb{F}_{sym}^{d_r\times d_r}) \quad \text{and } r = 1,\dots,R.
$$

In other words,

$$
\mathbf{R}_2(\mathcal{T})(\mathbf{N}_r\otimes\mathbf{N}_r)\mathbf{M}_r = \mathbf{O}, \qquad r = 1,\dots,R, \tag{4.4}
$$

where $\mathbf{M}_r$ is a $d_r^2 \times C_{d_r+1}^2$ matrix whose columns form a basis of $\operatorname{vec}(\mathbb{F}_{sym}^{d_r\times d_r})$. By statement 2) of Lemma 4.1 and (4.4), $\mathbf{R}_2(\mathcal{T})\mathbf{W} = \mathbf{O}$. Since the columns of $\mathbf{W}$ belong to $\operatorname{vec}(\mathbb{F}_{sym}^{K\times K})$, it follows that

$$
\text{column space of } \mathbf{W} \subseteq \operatorname{Null}(\mathbf{R}_2(\mathcal{T})) \cap \operatorname{vec}(\mathbb{F}_{sym}^{K\times K}). \tag{4.5}
$$

By statement 2) of Lemma 4.1, the column space of $\mathbf{W}$ has dimension Q. On the other hand, from (2.12) and (2.17) it follows that the dimension of $\operatorname{Null}(\mathbf{R}_2(\mathcal{T})) \cap \operatorname{vec}(\mathbb{F}_{sym}^{K\times K})$ is also Q. Hence, by (4.5),

$$
\text{column space of } \mathbf{W} = \operatorname{Null}(\mathbf{R}_2(\mathcal{T})) \cap \operatorname{vec}(\mathbb{F}_{sym}^{K\times K}). \tag{4.6}
$$

Let $\mathbf{v}_1,\dots,\mathbf{v}_Q$ be a basis of $\operatorname{Null}(\mathbf{R}_2(\mathcal{T})) \cap \operatorname{vec}(\mathbb{F}_{sym}^{K\times K})$. Then there exists a nonsingular $Q\times Q$ matrix $\mathbf{M}$ such that

$$
\begin{aligned}
[\mathbf{v}_1\ \dots\ \mathbf{v}_Q] = \mathbf{W}\mathbf{M} = [(\mathbf{N}_1\otimes\mathbf{N}_1)\mathbf{M}_1\ \dots\ (\mathbf{N}_R\otimes\mathbf{N}_R)\mathbf{M}_R]\mathbf{M} = \\
[\mathbf{N}_1\otimes\mathbf{N}_1 \dots\ \mathbf{N}_R\otimes\mathbf{N}_R]\operatorname{blockdiag}(\mathbf{M}_1,\dots,\mathbf{M}_R)\mathbf{M} =: [\mathbf{N}_1\otimes\mathbf{N}_1\dots\ \mathbf{N}_R\otimes\mathbf{N}_R]\tilde{\mathbf{M}},
\end{aligned} \tag{4.7}
$$

where

$$
\tilde{\mathbf{M}} = \operatorname{blockdiag}(\mathbf{M}_1,\dots,\mathbf{M}_R)\mathbf{M} \in \mathbb{F}^{\sum d_r^2\times Q}.
$$

Let

$$\mathbf{D}_q := \operatorname{blockdiag}(\mathbf{D}_{1,q}, \dots, \mathbf{D}_{R,q}) \in \mathbb{F}^{\sum q_r \times \sum q_r},$$

where the blocks $\mathbf{D}_{1,q}, \dots, \mathbf{D}_{R,q}$ are defined as

$$\begin{bmatrix} \operatorname{vec}(\mathbf{D}_{1,q}) \\ \vdots \\ \operatorname{vec}(\mathbf{D}_{R,q}) \end{bmatrix} = \text{ the } q\text{-th column of } \tilde{\mathbf{M}}$$

and let $\mathbf{V}_q$ denote the $K \times K$ matrix such that $\mathbf{v}_q = \operatorname{vec}(\mathbf{V}_q)$, $q = 1, \dots, Q$. Thus, we can rewrite (4.7) as

$$\mathbf{V}_q = [\mathbf{N}_1 \ \dots \ \mathbf{N}_R]\mathbf{D}_q[\mathbf{N}_1 \ \dots \ \mathbf{N}_R]^T = \mathbf{N}\mathbf{D}_q\mathbf{N}^T, \qquad q = 1, \dots, Q. \tag{4.8}$$

Since $\mathbf{V}_1, \dots, \mathbf{V}_Q$ are symmetric and since, by statement 1) of Lemma 4.1, the matrix $\mathbf{N}$ has full column rank, it follows easily that the matrices $\mathbf{D}_1, \dots, \mathbf{D}_Q$ are also symmetric. Besides, since $\mathbf{V}_1, \dots, \mathbf{V}_Q$ are linearly independent, the same holds for $\mathbf{D}_1, \dots, \mathbf{D}_Q$. Thus, (4.8) is the S-JBD problem of the form (1.6). By Theorem 1.10, the solution of (4.8) is unique and can be computed by means of (simultaneous) EVD. Now we can use the matrices $\mathbf{N}_r$ to recover the columns of $\mathbf{A}$. Recall that the matrix $\mathbf{N}_r$ holds a basis of $\operatorname{Null}(\mathbf{Z}_r)$, so we can repeat the derivation in (2.25)–(2.27) and obtain that the column $\mathbf{a}_r$ is proportional to the right singular vector of the matrix $[\operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_1^T) \ \dots \ \operatorname{vec}(\mathbf{N}_r^T\mathbf{H}_1^T)]$ corresponding to the only nonzero singular value.

Proof of statement 2). By statement 3) of Lemma 4.1, the matrices $\mathbf{E}_1, \dots, \mathbf{E}_R$ are linearly independent and, by statement 1), we can assume that the matrix $\mathbf{A}$ is known. Thus, the result follows from statement 1) of Lemma 4.2.

Proof of statement 3). We assume that $\mathcal{T}$ admits an alternative decomposition of the form (1.1):

$$\mathcal{T} = \sum_{r=1}^{\tilde{R}} \tilde{\mathbf{a}}_r \circ \tilde{\mathbf{E}}_r, \quad \tilde{\mathbf{a}}_r \in \mathbb{F}^I \setminus \{\mathbf{0}\}, \quad \tilde{\mathbf{E}}_r \in \mathbb{F}^{J \times K}, \quad 1 \le r_{\tilde{\mathbf{E}}_r} \le L_r,$$

in which we obviously assume that $\tilde{R} \le R$. First we show that $\tilde{R} = R$. From condition a) and (2.14) it follows that

$$\sum_{k=1}^{R} L_k - \min_{1 \le k \le R} L_k + 1 \le K = r_{\mathbf{T}_{(3)}} \le \sum_{k=1}^{\tilde{R}} r_{\tilde{\mathbf{E}}_k} \le \sum_{k=1}^{\tilde{R}} L_k. \tag{4.9}$$

Assuming that $\tilde{R} < R$, we obtain, by (4.9), the contradiction

$$L_R = L_R + \sum_{k=1}^{\tilde{R}} L_k - \sum_{k=1}^{\tilde{R}} L_k \le \sum_{k=1}^{R} L_k - \sum_{k=1}^{\tilde{R}} L_k \le \min_{1 \le k \le R} L_k - 1 < L_R.$$

Thus $\tilde{R} = R$.

Now we prove that each $\tilde{\mathbf{a}}_r$ is proportional to a column of $\mathbf{A}$. By definition, set

$$\tilde{d}_r := \dim \operatorname{Null}\left(\tilde{\mathbf{Z}}_r\right), \text{ where } \tilde{\mathbf{Z}}_r := [\tilde{\mathbf{E}}_1^T \ \dots \ \tilde{\mathbf{E}}_{r-1}^T \ \tilde{\mathbf{E}}_{r+1}^T \ \dots \ \tilde{\mathbf{E}}_R^T]^T, \qquad r = 1, \dots, R.$$

Since $r_{\tilde{\mathbf{Z}}_r} \le \min(\sum L_r - \min L_r, K)$, it follows from condition a) that $\tilde{d}_r \ge 1$. Let $\tilde{\mathbf{N}}_r$ be a $K \times \tilde{d}_r$ matrix whose columns form a basis of $\operatorname{Null}\left(\tilde{\mathbf{Z}}_r\right)$. If $\mathbf{f} = \tilde{\mathbf{N}}_r\mathbf{x}$ for some nonzero $\mathbf{x} \in \mathbb{F}^{\tilde{d}_r}$, then we obtain (see (4.1)) that

$$f_1\mathbf{T}_1 + \cdots + f_K\mathbf{T}_K = \tilde{\mathbf{a}}_r(\tilde{\mathbf{E}}_r\tilde{\mathbf{N}}_r\mathbf{x})^T, \qquad r = 1, \ldots, R.$$

By (2.14), the linear combination $f_1\mathbf{T}_1 + \cdots + f_K\mathbf{T}_K$ is not zero for any $f_1, \ldots, f_K$ such that $\mathbf{f} \neq \mathbf{0}$. Hence, for any column $\tilde{\mathbf{a}}_r$ there exist $f_1, \ldots, f_K$ such that the column space of the linear combination $f_1\mathbf{T}_1 + \cdots + f_K\mathbf{T}_K$ is one-dimensional and is spanned by $\tilde{\mathbf{a}}_r$. Thus, to prove that each $\tilde{\mathbf{a}}_r$ is proportional to a column of $\mathbf{A}$, it is sufficient to show that the following implication holds:

$$f_1\mathbf{T}_1 + \cdots + f_K\mathbf{T}_K = \mathbf{z}\mathbf{y}^T \ \Rightarrow \ \text{there exists } \ r \text{ such that } \mathbf{z} = c\mathbf{a}_r. \tag{4.10}$$

If $r_{f_1\mathbf{T}_1+\cdots+f_K\mathbf{T}_K} = 1$, then, by (2.4), $\mathbf{R}_2(\mathcal{T})(\mathbf{f}\otimes\mathbf{f}) = \mathbf{0}$. Hence, by (4.6), $\mathbf{f}\otimes\mathbf{f}$ belongs to the column space of the matrix $\mathbf{W}$. Hence, there exists a block diagonal matrix $\mathbf{D}$ such that $\mathbf{f}\mathbf{f}^T = \mathbf{N}\mathbf{D}\mathbf{N}^T$. Since, by statement 1) of Lemma 4.1, $\mathbf{N}$ has full column rank, the matrix $\mathbf{D}$ contains exactly one nonzero block and its rank is one. In other words, $\mathbf{f}$ belongs to the null space of $\mathbf{N}_r$ for some $r = 1, \ldots, R$. Hence implication (4.10) follows from (4.1).

Proof of statement 4). Let $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{C}}$ denote the factor matrices of an alternative decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms. By statement 3), it is sufficient to show that $\tilde{\mathbf{A}}$ does not have repeated columns. We argue by contradiction. If $\tilde{\mathbf{a}}_i = \tilde{\mathbf{a}}_j$ for some $i \neq j$, then $\tilde{\mathbf{a}}_i \wedge \tilde{\mathbf{a}}_j = \mathbf{0}$. Hence, the matrix $\Phi(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ defined in (3.2), has at least L_iL_j zero columns, implying that $r_{\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})} \leq \sum\limits_{1\leq r_1<r_2\leq R} L_{r_1}L_{r_2} - L_iL_j$. Hence, by statement 1) of Lemma 3.1,

$$\begin{aligned} r_{\mathbf{Q}_2(\mathcal{T})} = r_{\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})\mathbf{S}_2(\tilde{\mathbf{C}})^T} \leq r_{\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})} \leq \\ \sum_{1\leq r_1<r_2\leq R} L_{r_1}L_{r_2} - L_iL_j \leq \sum_{1\leq r_1<r_2\leq R} L_{r_1}L_{r_2} - \tilde{L}_1\tilde{L}_2. \end{aligned} \tag{4.11}$$

On the other hand, from the rank-nullity theorem and condition e) it follows that

$$r_{\mathbf{Q}_2(\mathcal{T})} = C_{K+1}^2 - Q > \sum_{1\leq r_1<r_2\leq R} L_{r_1}L_{r_2} - \tilde{L}_1\tilde{L}_2,$$

which is a contradiction with (4.11).

Proof of statement 5). If conditions a) and b) hold or conditions a) and c) hold, then the result follows from statement 3) and Lemma 4.2.

Let condition d) hold. Then the matrices $\mathbf{C}$ and $\mathbf{N}$ are square nonsingular and, by (2.25), $\mathbf{C}^T\mathbf{N} = \operatorname{blockdiag}(\mathbf{C}_1^T\mathbf{N}_1, \ldots, \mathbf{C}_R^T\mathbf{N}_R)$. Hence

$$\mathbf{C} = \mathbf{N}^{-T} \operatorname{blockdiag}(\mathbf{N}_1^T\mathbf{C}_1, \ldots, \mathbf{N}_R^T\mathbf{C}_R)$$

in which the matrices $\mathbf{N}_r^T\mathbf{C}_r \in \mathbb{F}^{L_r\times L_r}$ are also nonsingular. Thus, w.l.o.g. we can set $\mathbf{C} = \mathbf{N}^{-T}$. Finally, by (1.4), the matrix $\mathbf{B}$ can be uniquely recovered from the set of linear equations $[\mathbf{a}_1 \otimes \mathbf{C}_1 \ \ldots \ \mathbf{a}_R \otimes \mathbf{C}_R]\mathbf{B}^T = \mathbf{T}_{(2)}$. We can also avoid the computation of $\mathbf{N}^{-T}$ and proceed as in steps $8-9$ of Algorithm 2.1 (for details we refer to "Case 1" after Theorem 2.6).

To prove the uniqueness it is sufficient to show that assumptions (2.14), (2.15), and (2.17) and condition d) hold for any decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms. Assume that $\mathcal{T}$ admits an alternative decomposition with factor matrices $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}_1\ \dots\ \tilde{\mathbf{a}}_{\tilde{R}}]$, $\tilde{\mathbf{B}} = [\tilde{\mathbf{B}}_1\ \dots\ \tilde{\mathbf{B}}_{\tilde{R}}]$, and $\tilde{\mathbf{C}} = [\tilde{\mathbf{C}}_1\ \dots\ \tilde{\mathbf{C}}_{\tilde{R}}]$, where $\tilde{R} \le R$, the matrices $\tilde{\mathbf{B}}_r \in \mathbb{F}^{J\times\tilde{L}_r}$ and $\tilde{\mathbf{C}}_r \in \mathbb{F}^{K\times\tilde{L}_r}$ have full column rank, and $\tilde{L}_r \le L_r$ for $1 \le r \le \tilde{R}$. Then, by (1.5),

$$\mathbf{T}_{(3)} = [\mathbf{a}_1 \otimes \mathbf{B}_1\ \dots\ \mathbf{a}_R \otimes \mathbf{B}_R]\mathbf{C}^T = [\tilde{\mathbf{a}}_1 \otimes \tilde{\mathbf{B}}_1\ \dots\ \tilde{\mathbf{a}}_{\tilde{R}} \otimes \tilde{\mathbf{B}}_{\tilde{R}}]\tilde{\mathbf{C}}^T. \tag{4.12}$$

Since $r_{\mathbf{T}_{(3)}} = K$ and $\mathbf{C}$ is $K\times K$ nonsingular, it readily follows from (4.12) that $\tilde{R} = R$, that $\tilde{L}_r = L_r$ for all r and that $\tilde{\mathbf{C}}$ is $K \times K$ nonsingular. Hence, the values $d_1, \dots, d_R$ in (2.20) and the values $d_1, \dots, d_R$ computed for the alternative decomposition are equal to $L_1, \dots, L_R$, respectively. Thus, assumptions (2.14), (2.15), and (2.17) and condition d) hold for the alternative decomposition. ∎

5. Conclusion.

In this paper we have studied the decomposition of a third-order tensor into a sum of ML rank-$(1, L_r, L_r)$ terms. We have obtained conditions for uniqueness of the first factor matrix and for uniqueness of the overall decomposition. We have also presented an algorithm that computes the decomposition, estimates the number of ML rank-$(1, L_r, L_r)$ terms R and their "sizes" $L_1, \dots, L_R$. All steps of the algorithm rely on conventional linear algebra. In the case where the decomposition is not exact, a noisy version of the algorithm can compute an approximate ML rank-$(1, L_r, L_r)$ decomposition. In our examples the accuracy of the estimates was of about the same order as the accuracy of the tensor.

The ML rank-$(1, L_r, L_r)$ decomposition takes an intermediate place between the little studied decomposition into a sum of ML rank-(M_r, N_r, L_r) terms and the well studied CPD (the special case where $M_r = N_r = L_r = 1$). Namely, the ML rank-$(1, L_r, L_r)$ decomposition is the special case where $M_r = 1$ and $N_r = L_r$. The results in this paper may be used as stepping stones towards a better understanding of the ML rank-(M_r, N_r, L_r) decomposition.

Acknowledgments. The authors would like to thank Yang Qi (The University of Chicago) for his comments on subsection 2.6.

Appendix A. On testing (2.38) over a finite field.

In this appendix we explain how to verify assumption (2.38) over a finite field. We also explain how to test whether the decomposition of an $I \times J \times K$ tensor into a sum of max ML rank-$(1, L_r, L_r)$ terms is generically unique under the assumptions in row 6 of Table 1.1.

We rely on an idea proposed in [7]. The idea is to generate random integer matrices $\tilde{\mathbf{A}}_r$, $\tilde{\mathbf{B}}_r$, $\tilde{\mathbf{C}}_r$ and then to perform all computations over a finite field $GF(p^k)$, where p is prime. Obviously, if (2.38) holds for $\tilde{\mathbf{A}}_r$, $\tilde{\mathbf{B}}_r$ and $\tilde{\mathbf{C}}_r$ considered over $GF(p^k)$, then it will necessarily hold for $\tilde{\mathbf{A}}_r$, $\tilde{\mathbf{B}}_r$ and $\tilde{\mathbf{C}}_r$ considered over $\mathbb{F}$[13]. On the other hand, if (2.38) does not hold for $\tilde{\mathbf{A}}_r$, $\tilde{\mathbf{B}}_r$, $\tilde{\mathbf{C}}_r$ over $GF(p^k)$, then no conclusion can be drawn. In this case one can try to repeat the computations for other random integer matrices $\tilde{\mathbf{A}}_r$, $\tilde{\mathbf{B}}_r$, $\tilde{\mathbf{C}}_r$ or increment k, or choose another prime p. If (2.38) does not hold for several such trials, this can be an indication that (2.38) does not hold for any $\tilde{\mathbf{A}}_r$, $\tilde{\mathbf{B}}_r$ and $\tilde{\mathbf{C}}_r$. Note that, by the rank-nullity theorem, the computation of the null space can be reduced to the computation of the rank. Although the computation of the rank over the finite field is more expensive than the numerical estimation of the rank, it has the advantage that the dimension in (2.38) is computed exactly, i.e., without roundoff errors.

[13] In the proof of Theorem 2.13 we have explained that this will in turn apply that (2.38) holds over $\mathbb{F}$ for generic $\tilde{\mathbf{A}}_r$, $\tilde{\mathbf{B}}_r$, $\tilde{\mathbf{C}}_r$.

Now we explain how to test whether the bounds in row 6 of Table 1.1 guarantee generic uniqueness of the decomposition. By Lemma 3.1, $\mathbf{Q}_2(\tilde{\mathcal{T}})$ can be factorized as $\mathbf{Q}_2(\tilde{\mathcal{T}}) = \Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})\mathbf{S}_2(\tilde{\mathbf{C}})$, where $\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})$ is an $C_I^2C_J^2 \times \sum\limits_{r_1<r_2} L_{r_1}L_{r_2}$ matrix and $\mathbf{S}_2(\tilde{\mathbf{C}})$ is an $C_{K+1}^2 \times \sum\limits_{r_1<r_2} L_{r_1}L_{r_2}$ matrix. Also, by statement 3) of Lemma 3.1, $\dim\operatorname{Null}\left(\mathbf{S}_2(\tilde{\mathbf{C}})^T\right) = \sum C_{d_r+1}^2$ for generic $\tilde{\mathbf{C}}$. It is clear now that if $\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})$ has full column rank, then (2.38) holds for $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ and generic $\tilde{\mathbf{C}}$.

We claim that the assumptions $C_I^2C_J^2 \geq \sum\limits_{r_1<r_2} L_{r_1}L_{r_2}$ and $J \geq L_{R-1}+L_R$ in row 6 of Table 1.1 are necessary for $\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})$ to have full column rank. Indeed, the former expresses the fact that the number of columns of $\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})$ does not exceed the number of its rows. The latter means that $k'_{\tilde{\mathbf{B}}} \geq 2$ holds for generic $\tilde{\mathbf{B}}$, which, by statement 6) of Lemma 3.1, is necessary for full column rank of $\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})$. To verify that $\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})$ has full column rank for some $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ we performed computations over $GF(2^{15})$ as explained above. The computations were done in MATLAB R2018b, where $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ were generated using the built-in function `gf` (Galois field arrays) and the rank of $\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})$ was computed with the built-in function `rank`. We limited ourselves to the cases where $\min(I,J) \geq 2$ and $\max(I,J) \leq 5$. Together with the assumptions $J \geq L_{R-1}+L_R$ and $C_I^2C_J^2 \geq \sum\limits_{r_1<r_2} L_{r_1}L_{r_2}$ we ended up with 435 tuples $(I,J,R,L_1,\dots,L_R)$. The matrix $\Phi(\tilde{\mathbf{A}},\tilde{\mathbf{B}})$ did not have full column rank in three cases: $(I,R) \in \{(2,3),(4,9),(5,12)\}$, $J=5$, $L_1=\dots,L_{R-1}=1$, and $L_R=4$.

To show that in the remaining 432 cases generic uniqueness and computation follow from statement 4) of Theorem 2.13, we need to verify assumptions (2.36),(2.37) and condition (2.41). The assumption $\sum L_r = K$ in row 6 of Table 1.1 coincides with condition (2.41) and implies assumption (2.37). From statement 5) of Lemma 3.1 it follows that $[\tilde{\mathbf{a}}_1 \otimes \tilde{\mathbf{B}}_1\ \dots\ \tilde{\mathbf{a}}_R \otimes \tilde{\mathbf{B}}_R]$ has full column rank, and in particular, that $IJ \geq \sum L_r$. Hence, since $\sum L_r = K$, we obtain that assumption (2.36) also holds.

Appendix B. Proofs of Theorems 2.1, 2.6, Corollary 2.7 and Theorem 2.13.

Proof of Theorem 2.1. Proof of statement 1). Assume to the contrary that the matrix $[\operatorname{vec}(\mathbf{E}_1)\ \dots\ \operatorname{vec}(\mathbf{E}_R)]$ does not have full column rank. Then the matrices $\mathbf{E}_1,\dots,\mathbf{E}_R$ are linearly dependent. We assume w.l.o.g. that $\mathbf{E}_1 = \alpha_2\mathbf{E}_2+\dots+\alpha_R\mathbf{E}_R$. Then $\mathcal{T}$ admits a decomposition into a sum of $R-1$ terms:

$$\mathcal{T} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{E}_r = \mathbf{a}_1 \circ (\sum_{r=2}^R \alpha_r\mathbf{E}_r) + \sum_{r=2}^R \mathbf{a}_r \circ \mathbf{E}_r = \sum_{r=2}^R (\alpha_r\mathbf{a}_1+\mathbf{a}_r)\circ\mathbf{E}_r,$$

which is a contradiction.

Proof of statement 2). Assume to the contrary that the matrix $[\mathbf{a}_1 \otimes \mathbf{B}_1\ \dots\ \mathbf{a}_R \otimes \mathbf{B}_R]$ does not have full column rank. Then there exists $\mathbf{f} = [\mathbf{f}_1^T\ \dots\ \mathbf{f}_R^T]^T \in \mathbb{F}^{\sum L_r}\setminus\{\mathbf{0}\}$ such that $\sum(\mathbf{a}_r \otimes \mathbf{B}_r)\mathbf{f}_r = \mathbf{0}$. We assume w.l.o.g. that the first entry of $\mathbf{f}$ is nonzero and partition $\mathbf{f}_1$, $\mathbf{B}_1$, and $\mathbf{C}_1$ as

$$\mathbf{f} = \begin{bmatrix} f_1 \\ \bar{\mathbf{f}}_1 \end{bmatrix}, \qquad \mathbf{B}_1 = [\mathbf{b}_1\ \bar{\mathbf{B}}_1], \qquad \mathbf{C}_1 = [\mathbf{c}_1\ \bar{\mathbf{C}}_1].$$

Since $\sum(\mathbf{a}_r \otimes \mathbf{B}_r)\mathbf{f}_r = \mathbf{0}$, it follows that

$$\text{(B.1)} \quad \mathbf{a}_1 \otimes \mathbf{b}_1 = -\frac{1}{f_1}\left[(\mathbf{a}_1 \otimes \bar{\mathbf{B}}_1)\bar{\mathbf{f}}_1 + \sum_{r=2}^{R}(\mathbf{a}_r \otimes \mathbf{B}_r)\mathbf{f}_r\right] = -\frac{1}{f_1}\left[\mathbf{a}_1 \otimes (\bar{\mathbf{B}}_1\bar{\mathbf{f}}_1) + \sum_{r=2}^{R}\mathbf{a}_r \otimes (\mathbf{B}_r\mathbf{f}_r)\right].$$

Hence, by (1.5) and (B.1),

$$\mathbf{T}_{(3)} = \sum_{r=1}^{R}(\mathbf{a}_r \otimes \mathbf{B}_r)\mathbf{C}_r^T = (\mathbf{a}_1 \otimes \mathbf{b}_1)\mathbf{c}_1^T + (\mathbf{a}_1 \otimes \bar{\mathbf{B}}_1)\bar{\mathbf{C}}_1^T + \sum_{r=2}^{R}(\mathbf{a}_r \otimes \mathbf{B}_r)\mathbf{C}_r^T =$$
$$-\frac{1}{f_1}\left[\mathbf{a}_1 \otimes (\bar{\mathbf{B}}_1\bar{\mathbf{f}}_1) + \sum_{r=2}^{R}\mathbf{a}_r \otimes (\mathbf{B}_r\mathbf{f}_r)\right]\mathbf{c}_1^T + (\mathbf{a}_1 \otimes \bar{\mathbf{B}}_1)\bar{\mathbf{C}}_1^T + \sum_{r=2}^{R}(\mathbf{a}_r \otimes \mathbf{B}_r)\mathbf{C}_r^T =$$
$$\mathbf{a}_1 \otimes \left[-\frac{1}{f_1}\bar{\mathbf{B}}_1\bar{\mathbf{f}}_1\mathbf{c}_1^T + \bar{\mathbf{B}}_1\bar{\mathbf{C}}_1^T\right] + \sum_{r=2}^{R}\mathbf{a}_r \otimes \left[-\frac{1}{f_1}\mathbf{B}_r\mathbf{f}_r\mathbf{c}_1^T + \mathbf{B}_r\mathbf{C}_r^T\right] =: \sum_{r=1}^{R}\mathbf{a}_r \otimes \tilde{\mathbf{E}}_r,$$

where $r_{\tilde{\mathbf{E}}_1} \leq r_{\bar{\mathbf{B}}_1} = L_1 - 1$ and $r_{\tilde{\mathbf{E}}_r} \leq r_{\mathbf{B}_r} = L_r$ for $r \geq 2$. Thus, $\mathcal{T}$ admits an alternative decomposition into a sum of max ML rank-$(1, L_r, L_r)$ terms $\mathcal{T} = \sum \mathbf{a}_r \circ \tilde{\mathbf{E}}_r$ with $r_{\tilde{\mathbf{E}}_1} < r_{\mathbf{E}_1}$ and $r_{\tilde{\mathbf{E}}_r} \leq r_{\mathbf{E}_r}$ for $r \geq 2$. This contradiction completes the proof.

Proof of statement 3). The proof is similar to the proof of statement 2) . ∎

Proof of Theorem 2.6. By (1.5), assumption (2.19) is equivalent to assumption (2.14). Substituting $\mathbf{E}_r = \mathbf{B}_r\mathbf{C}_r^T$ in the expressions for $\mathbf{Z}_r$, $\mathbf{F}$, $\mathbf{G}$, and $[\mathbf{E}_1^T \ \dots \ \mathbf{E}_R^T]^T$, we obtain that

$$\mathbf{Z}_r = \operatorname{blockdiag}(\mathbf{B}_1, \dots, \mathbf{B}_{r-1}, \mathbf{B}_{r+1}, \dots, \mathbf{B}_R)[\mathbf{C}_1 \ \dots \ \mathbf{C}_{r-1} \ \mathbf{C}_{r+1} \ \dots \ \mathbf{C}_R]^T,$$
$$\mathbf{F} = [\mathbf{B}_{r_1} \ \mathbf{B}_{r_2} \ \dots \ \mathbf{B}_{r_{R-r_{\mathbf{A}}+2}}] \operatorname{blockdiag}(\mathbf{C}_{r_1}^T, \mathbf{C}_{r_2}^T, \dots, \mathbf{C}_{r_{R-r_{\mathbf{A}}+2}}^T),$$
$$\mathbf{G} = [\mathbf{C}_{r_1} \ \mathbf{C}_{r_2} \ \dots \ \mathbf{C}_{r_{R-r_{\mathbf{A}}+2}}] \operatorname{blockdiag}(\mathbf{B}_{r_1}^T, \mathbf{B}_{r_2}^T, \dots, \mathbf{B}_{r_{R-r_{\mathbf{A}}+2}}^T),$$
$$[\mathbf{E}_1^T \ \dots \ \mathbf{E}_R^T]^T = \operatorname{blockdiag}(\mathbf{B}_1, \dots, \mathbf{B}_R)\mathbf{C}^T.$$

Since the matrices $\mathbf{B}_r$ and $\mathbf{C}_r$ have full column rank, it follows that

$$\text{(B.2)} \quad d_r = \dim \operatorname{Null}(\mathbf{Z}_r) = \dim \operatorname{Null}\left([\mathbf{C}_1 \ \dots \ \mathbf{C}_{r-1} \ \mathbf{C}_{r+1} \ \dots \ \mathbf{C}_R]^T\right) = \dim \operatorname{Null}(\mathbf{Z}_{r,\mathbf{C}}),$$

that (2.16) and (2.18) are equivalent to (2.21) and $k'_{\mathbf{C}} \geq R - r_{\mathbf{A}} + 2$, respectively, and that condition d) in Theorem 2.5 is equivalent to $r_{\mathbf{C}^T} = \sum L_r$. Since, by (2.14) and (1.5), $K = r_{\mathbf{T}_{(3)}} \leq r_{\mathbf{C}^T} \leq K$, it follows that $r_{\mathbf{C}} = r_{\mathbf{C}^T} = K = \sum L_r$. Hence $\mathbf{C}$ is a nonsingular $K \times K$ matrix. This in turn, by (B.2), implies that $d_r = L_r$. Thus, condition d) in Theorem 2.5 is equivalent to condition d) in Theorem 2.6. ∎

Proof of Corollary 2.7. We consider two cases $r_{\mathbf{C}} = K$ and $r_{\mathbf{C}} < K$.

i) Let $r_{\mathbf{C}} = K$. Together the assumptions in (2.23) and conditions in (2.24) imply that assumption (2.21) and condition a) in Theorem 2.6 hold. In turn, condition a) implies that assumption (2.20) holds. The two conditions in (2.24) coincide with condition b) and condition c) in Theorem 2.6, respectively. Thus, to apply statement 5) in Theorem 2.6 it only remains to verify that assumption (2.19) holds. Since $r_{\mathbf{C}} = K$, it is sufficient to prove that the matrix $[\mathbf{a}_1 \otimes \mathbf{B}_1 \ \ldots \ \mathbf{a}_R \otimes \mathbf{B}_R]$ has full column rank. This follows from statements 4) and 5) of Lemma 3.1.

ii) If $r_{\mathbf{C}} < K$, then the result follows from i) and statement 1) of Theorem 2.4. □

Proof of Theorem 2.13. We show that statements 1) to 4) in Theorem 2.13 correspond, respectively, to statements 1), 3), 4), and 5) in Theorem 2.5. One can easily check that assumptions (2.36), (2.37), and conditions (2.40), (2.41) in Theorem 2.13 are, respectively, the generic versions of assumptions (2.14), (2.15) and conditions b), d) in Theorem 2.5. Hence, to prove statements 1), 2), and 4), it is sufficient to show that assumption (2.38) implies that (2.17) holds generically. To prove statement 3) we should additionally show that (2.39) implies that condition e) holds generically.

1) We show that assumption (2.38) implies that (2.17) holds generically. We will make use of [17, Lemma 6.3] which states the following: if the entries of a matrix $\mathbf{F}(\mathbf{x})$ depend analytically on $\mathbf{x} \in \mathbb{F}^n$ and if $\mathbf{F}(\mathbf{x}_0)$ has full column rank for at least one $\mathbf{x}_0$, then $\mathbf{F}(\mathbf{x})$ has full column rank for generic $\mathbf{x}$. Let the vectors $\mathbf{x}$ and $\mathbf{x}_0$ be formed by the entries of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{C}}$ respectively. We construct $\mathbf{F}(\mathbf{x})$ as follows. By Lemma 3.1, each entry of $\mathbf{Q}_2(\mathcal{T})$ is a polynomial in $\mathbf{x}$. By the rank-nullity theorem and assumption (2.38),

$$\text{(B.3)} \qquad r_{\mathbf{Q}_2(\tilde{\mathcal{T}})} = C_{K+1}^2 - \sum_{r=1}^{R} C^2_{K-(L_1+\cdots+L_{r-1}+L_{r+1}+\cdots+L_R)+1} =: P,$$

implying that P columns of $\mathbf{Q}_2(\tilde{\mathcal{T}})$ are linearly independent. We define $\mathbf{F}(\mathbf{x})$ as the submatrix formed by the corresponding columns[14] of $\mathbf{Q}_2(\mathcal{T})$. Then (B.3) implies that $\mathbf{F}(\mathbf{x}_0)$ has full column rank. Now, by [17, Lemma 6.3], $\mathbf{F}(\mathbf{x})$ has full column rank for generic $\mathbf{x}$. Hence $r_{\mathbf{Q}_2(\mathcal{T})} \geq P$. Hence, by the rank-nullity theorem, $\dim \operatorname{Null}(\mathbf{Q}_2(\mathcal{T})) = C^2_{K+1} - r_{\mathbf{Q}_2(\mathcal{T})} \leq C^2_{K+1} - P = \sum\limits_{r=1}^{R} C^2_{d_r+1}$. On the other hand, since, by statement 3) of Lemma 3.1, $\dim \operatorname{Null}(\mathbf{Q}_2(\mathcal{T})) \geq \sum\limits_{r=1}^{R} C^2_{d_r+1}$ we obtain that (2.17) in Theorem 2.5 holds.

2) We show that assumption (2.39) implies that condition e) holds generically. Let $S = \sum L_r$. Then $d_r = K - \sum\limits_{k=1}^{R} L_k + L_r = K - S + L_r$. Since $L_1 \leq \cdots \leq L_R$, the inequality in condition e) takes the form

$$\text{(B.4)} \quad C^2_{K+1} - \sum_{r=1}^{R} C^2_{K-S+L_r+1} > \sum_{1 \leq r_1 < r_2 \leq R} L_{r_1} L_{r_2} - L_1 L_2 = \frac{S^2 - \sum L_r^2}{2} - L_1 L_2.$$

Using simple algebraic manipulations one can rewrite (B.4) as

$$\text{(B.5)} \qquad K^2 + K(1-2S) + S^2 - S - \frac{2L_1L_2}{R-1} < 0.$$

One can easily check that K is a solution of (B.5) if and only if

$$S - \frac{1}{2} - \sqrt{\frac{1}{4} + \frac{2L_1L_2}{R-1}} < K < S - \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2L_1L_2}{R-1}},$$

implying that (2.39) is a generic version of condition e). □

[14] The column selection depends only on the fixed $\mathbf{x}_0$.

Appendix C. Proof of Theorem 2.16. First we recall a result on the generic uniqueness of the decomposition of a matrix into rank-1 terms that admit a particular structure [20]. Let $p_1, \dots, p_N$ be known polynomials in l variables and let $\mathbf{Y} \in \mathbb{F}^{I \times N}$ admit a decomposition of the form

$$\text{(C.1)} \qquad \mathbf{Y} = \sum_{r=1}^{R} \mathbf{a}_r [p_1(\mathbf{z}_r) \ \dots \ p_N(\mathbf{z}_r)], \qquad \mathbf{a}_r \in \mathbb{F}^I, \quad \mathbf{z}_r \in \mathbb{F}^l, \quad r = 1, \dots, R.$$

Decomposition (C.1) can be interpreted as a matrix factorization $\mathbf{Y} = \mathbf{A}\mathbf{P}^T$ that is structured in the sense that the columns of $\mathbf{P}$ are in

$$\text{(C.2)} \qquad V := \{[p_1(\mathbf{z}) \ \dots \ p_N(\mathbf{z})]^N : \ \mathbf{z} \in \mathbb{F}^l\} \subset \mathbb{F}^N.$$

We say that the decomposition *is unique* if any two decompositions of the form (C.1) are the same up to permutation of summands. We say that the decomposition into a sum of structured rank-1 matrices is *generically unique* if

$$\mu\{(\mathbf{a}_1, \dots, \mathbf{a}_R, \mathbf{z}_1, \dots, \mathbf{z}_R) : \text{ decomposition (C.1) is not unique}\} = 0,$$

where μ denotes a measure on $\mathbb{F}^{(I+l)R}$ that is absolutely continuous with respect to the Lebesgue measure. We will need the following result.

THEOREM C.1. *(a corollary of [20, Theorem 1]) Assume that*
a) $R \le I$*;*
b) $\dim \operatorname{span}\{V\} \ge \hat{N}$*;*
c) the set V *is invariant under complex scaling, i.e.,* $\lambda V = V$ *for all* $\lambda \in C$*;*
d) the dimension of the Zariski closure of V *is less than or equal to* $\hat{l}$*;*
e) $R \le \hat{N} - \hat{l}$*.*
Then decomposition (C.1) is generically unique.

Proof of Theorem 2.16. (i) First we rewrite (1.2) in the form of the structured matrix decomposition (C.1). In step (ii) we will apply Theorem C.1 to (C.1). By (1.3), decomposition (1.2) can be rewritten as

$$\mathbf{Y} := \mathbf{T}_{(1)}^T = \mathbf{A}[\operatorname{vec}(\mathbf{B}_1\mathbf{C}_1^T) \ \dots \ \operatorname{vec}(\mathbf{B}_R\mathbf{C}_R^T)]^T =: \mathbf{A}\mathbf{P}^T.$$

So, the columns of $\mathbf{P}$ are of the form

$$\operatorname{vec}([\mathbf{b}_1 \ \dots \ \mathbf{b}_L][\mathbf{c}_1 \ \dots \ \mathbf{c}_L]^T) = \mathbf{c}_1 \otimes \mathbf{b}_1 + \dots + \mathbf{c}_L \otimes \mathbf{b}_L =: [p_1(\mathbf{z}) \ \dots \ p_N(\mathbf{z})]^T,$$

where

$$\mathbf{z} = [\mathbf{b}_1^T \ \dots \ \mathbf{b}_L^T \ \mathbf{c}_1^T \ \dots \ \mathbf{c}_L^T]^T, \quad l = JL + KL, \quad N = JK.$$

Hence the set V in (C.2) consists of vectorized $J \times K$ matrices whose rank does not exceed L.

(ii) Now we check assumptions a) to e) in Theorem C.1. Assumption a) holds by (2.47). Since V contains, in particular, all vectorized rank-1 matrices, it spans the entire $\mathbb{F}^N$. Hence we can choose $\hat{N} = N = JK$ in assumption b). Assumption c) is trivial. It is well-known that the set V is an algebraic variety of dimension $(J + K - L)L$, so assumption d) holds for $\hat{l} = (J + K - L)L$. Finally, assumption e) holds by (2.47): $R \le (J - L)(K - L) = JK - (J + K - L)L = \hat{N} - \hat{l}$. ∎

Appendix D. Proofs of statements 1), 2) and 6) of Lemma 3.1 and proof of Corollary 3.2.

Proofs of statements 1), 2) and 6) of Lemma 3.1. 1) Since $\mathcal{T}=\sum\limits_{r=1}^{R}\mathbf{a}_r\circ(\mathbf{B}_r\mathbf{C}_r^T)$, it follows that $t_{ijk}=\sum\limits_{r=1}^{R}a_{ir}\sum\limits_{l=1}^{L_r}b_{jl,r}c_{kl,r}$. Hence

$$t_{i_1j_1k_1}t_{i_2j_2k_2}=\sum_{r_1=1}^{R}\sum_{r_2=1}^{R}a_{i_1r_1}a_{i_2r_2}\sum_{l_1=1}^{L_{r_1}}\sum_{l_2=1}^{L_{r_2}}b_{j_1l_1,r_1}b_{j_2l_2,r_2}c_{k_1l_1,r_1}c_{k_2l_2,r_2}. \tag{D.1}$$

By Definition 2.3, the entry of $\mathbf{Q}_2(\mathcal{T})$ with the index in (2.7) is equal to (2.8), where $1\le i_1<i_2\le I$, $1\le j_1<j_2\le J$, and $1\le k_1\le k_2\le K$. Applying (D.1) to each term in (2.8) and making simple algebraic manipulations we obtain that the expression in (2.8) is equal to

$$\sum_{1\le r_1<r_1\le R}\Big[(a_{i_1r_1}a_{i_2r_2}-a_{i_2r_1}a_{i_1r_2})\times$$
$$\sum_{l_1=1}^{L_{r_1}}\sum_{l_2=1}^{L_{r_2}}(b_{j_1l_1,r_1}b_{j_2l_2,r_2}-b_{j_2l_1,r_1}b_{j_1l_2,r_2})(c_{k_1l_1,r_1}c_{k_2l_2,r_2}+c_{k_2l_1,r_1}c_{k_1l_2,r_2})\Big]=$$
$$\sum_{1\le r_1<r_1\le R}(\mathbf{a}_{r_1}\wedge\mathbf{a}_{r_2})_{i_1+C^2_{i_2-1}}\sum_{l_1=1}^{L_{r_1}}\sum_{l_2=1}^{L_{r_2}}(\mathbf{b}_{l_1,r_1}\wedge\mathbf{b}_{l_2,r_2})_{j_1+C^2_{j_2-1}}(\mathbf{c}_{l_1,r_1}\cdot\mathbf{c}_{l_2,r_2})_{k_1+C^2_{k_2}},$$

which, by the definition of $\Phi(\mathbf{A},\mathbf{B})$ and $\mathbf{S}_2(\mathbf{C})$, is the entry of $\Phi(\mathbf{A},\mathbf{B})\mathbf{S}_2(\mathbf{C})^T$ with the index in (2.7).

2) follows from the identity $\mathbf{R}_2(\mathcal{T})=\mathbf{Q}_2(\mathcal{T})\mathbf{P}_K^T$ and 1).

6) We assume that $\Phi(\mathbf{A},\mathbf{B})$ has full column rank. It is sufficient to prove that the identities $\mathbf{h}=\mathbf{B}_{r_1}\mathbf{f}_1=\mathbf{B}_{r_1}\mathbf{f}_2$ are valid only for $\mathbf{h}=\mathbf{0}$. From the definition of the operation "$\wedge$" it follows that $(\mathbf{B}_{r_1}\wedge\mathbf{B}_{r_2})(\mathbf{f}_1\otimes\mathbf{f}_2)=(\mathbf{B}_{r_1}\mathbf{f}_1)\wedge(\mathbf{B}_{r_2}\mathbf{f}_2)=\mathbf{h}\wedge\mathbf{h}=\mathbf{0}$. Hence $[(\mathbf{a}_{r_1}\wedge\mathbf{a}_{r_2})\otimes(\mathbf{B}_{r_1}\wedge\mathbf{B}_{r_2})](\mathbf{f}_1\otimes\mathbf{f}_2)=(\mathbf{a}_{r_1}\wedge\mathbf{a}_{r_2})\otimes[(\mathbf{B}_{r_1}\wedge\mathbf{B}_{r_2})(\mathbf{f}_1\otimes\mathbf{f}_2)]=\mathbf{0}$. Now, since $(\mathbf{a}_{r_1}\wedge\mathbf{a}_{r_2})\otimes(\mathbf{B}_{r_1}\wedge\mathbf{B}_{r_2})$ is formed by the columns of the full column rank matrix $\Phi(\mathbf{A},\mathbf{B})$, it follows that $\mathbf{f}_1\otimes\mathbf{f}_2=\mathbf{0}$, which easily implies that $\mathbf{h}=\mathbf{0}$. ∎

Proof of Corollary 3.2. W.l.o.g. we assume that $i=1$ and $j=2$. Since $\mathbf{C}$ has full column rank, and, by (2.19), $\mathbf{C}^T$ has full column rank, it follows that $\mathbf{C}$ is $K\times K$ nonsingular and that $K=\sum L_r$. This readily implies that $d_r=L_r$ for all r. From the rank-nullity theorem and (2.22) it follows that

$$r_{\Phi(\mathbf{A},\mathbf{B})}\ge r_{\Phi(\mathbf{A},\mathbf{B})\mathbf{S}_2(\mathbf{C})^T}=C^2_{K+1}-\dim\operatorname{Null}\left(\Phi(\mathbf{A},\mathbf{B})\mathbf{S}_2(\mathbf{C})^T\right)=$$
$$C^2_{\sum L_r+1}-\sum C^2_{L_r+1}=\sum_{r_1<r_2}L_{r_1}L_{r_2}.$$

Since $\Phi(\mathbf{A},\mathbf{B})$ is a $C^2_{K+1}\times\sum\limits_{r_1<r_2}L_{r_1}L_{r_2}$ matrix, it follows that $\Phi(\mathbf{A},\mathbf{B})$ has full column rank. In particular, the submatrix $(\mathbf{a}_1\wedge\mathbf{a}_2)\otimes(\mathbf{B}_1\wedge\mathbf{B}_2)$ has full column rank, implying that the same holds true for the matrix $\mathbf{B}_1\wedge\mathbf{B}_2$. Assume that $[\mathbf{B}_1\ \mathbf{B}_2][\mathbf{f}_1^T\ \mathbf{f}_2^T]^T=\mathbf{0}$ for some $\mathbf{f}_1\in\mathbb{F}^{L_1}$ and $\mathbf{f}_2\in\mathbb{F}^{L_2}$. Then $\mathbf{B}_2\mathbf{f}_2=-\mathbf{B}_1\mathbf{f}_1$. One can easily verify that $(\mathbf{B}_1\wedge\mathbf{B}_2)(\mathbf{f}_1\otimes\mathbf{f}_2)=\mathbf{B}_1\mathbf{f}_1\wedge\mathbf{B}_2\mathbf{f}_2=-\mathbf{B}_1\mathbf{f}_1\wedge\mathbf{B}_1\mathbf{f}_1=\mathbf{0}$. Hence $\mathbf{f}_1\otimes\mathbf{f}_2=\mathbf{0}$. Thus, $\mathbf{f}_1=\mathbf{0}$ or $\mathbf{f}_2=\mathbf{0}$, implying that $\mathbf{B}_1\mathbf{f}_1=\mathbf{0}$ or $\mathbf{B}_2\mathbf{f}_2=\mathbf{0}$. Since $\mathbf{B}_1$ and $\mathbf{B}_2$ have full column rank and $\mathbf{B}_2\mathbf{f}_2=-\mathbf{B}_1\mathbf{f}_1$, it follows that both $\mathbf{f}_1$ and $\mathbf{f}_2$ are the zero vectors. Hence the matrix $[\mathbf{B}_1\ \mathbf{B}_2]$ has full column rank. ∎

Appendix E. Proof of statement 3) of Lemma 3.1.

Proofs of statement 3) *of Lemma* 3.1. The inequality in statement 3) follows immediately from statement 1). We prove the identity $\dim\operatorname{Null}\left(\mathbf{S}_2(\mathbf{C})^T\right) = \sum C^2_{d_r+1}$. Throughout the proof, $\operatorname{col}(\cdot)$ denotes the column space of a matrix.

Obviously, $\dim\operatorname{Null}\left(\mathbf{S}_2(\mathbf{C})^T\right) = \dim\operatorname{Null}\left(\mathbf{S}_2(\mathbf{C})^H\right)$. Since $\operatorname{vec}(\mathbb{F}^{K\times K}_{sym})$ is the orthogonal sum of the subspaces $\operatorname{Null}\left(\mathbf{S}_2(\mathbf{C})^H\right)$ and $\operatorname{col}(\mathbf{S}_2(\mathbf{C}))$, it is sufficient to show that there exists a subspace S such that

$$\operatorname{vec}(\mathbb{F}^{K\times K}_{sym}) = \operatorname{span}\{S, \operatorname{col}(\mathbf{S}_2(\mathbf{C}))\}, \tag{E.1}$$

$$S \cap \operatorname{col}(\mathbf{S}_2(\mathbf{C})) = \{\mathbf{0}\}, \tag{E.2}$$

$$\dim S = \sum C^2_{d_r+1}. \tag{E.3}$$

We explicitly construct a possible S and show that (E.1)–(E.3) hold.

(i) *Construction of S.* Since $r_{\mathbf{C}} = K$ and $\dim\operatorname{Null}(\mathbf{Z}_{r,\mathbf{C}}) = d_r$, it follows that $r_{\mathbf{Z}^T_{r,\mathbf{C}}} = r_{\mathbf{Z}_{r,\mathbf{C}}} = K - d_r$. Let $W_r = \operatorname{col}(\mathbf{Z}^T_{r,\mathbf{C}}) \cap \operatorname{col}(\mathbf{C}_r)$ and let V_r denote the orthogonal complement of W_r in $\operatorname{col}(\mathbf{C}_r)$. Then

$$\begin{aligned}
\dim W_r &= \dim\operatorname{col}(\mathbf{Z}^T_{r,\mathbf{C}}) + \dim\operatorname{col}(\mathbf{C}_r)\\
&\quad - \dim\operatorname{col}([\mathbf{C}_1\ \dots \mathbf{C}_{r-1}\ \mathbf{C}_{r+1}\ \dots\ \mathbf{C}_R\ \mathbf{C}_r]) = K - d_r + L_r - K = L_r - d_r,\\
\dim V_r &= \dim\operatorname{col}(\mathbf{C}_r) - \dim W_r = L_r - (L_r - d_r) = d_r.
\end{aligned}$$

Let $\mathbf{V}_r \in \mathbb{F}^{K\times d_r}$ be a matrix whose columns form a basis of V_r. We set

$$S = \operatorname{col}([\mathbf{V}_1 \cdot \mathbf{V}_1\ \dots\ \mathbf{V}_R \cdot \mathbf{V}_R]).$$

(ii) *Proof of* (E.1). Let $\mathbf{W}_r \in \mathbb{F}^{K\times(L_r-d_r)}$ be a matrix whose columns form a basis of W_r. Since $r_{\mathbf{C}} = K$ and $\operatorname{col}(\mathbf{C}_r) = \operatorname{col}([\mathbf{V}_r\ \mathbf{W}_r])$, it follows that

$$\begin{aligned}
\operatorname{vec}(\mathbb{F}^{K\times K}_{sym}) &= \operatorname{col}([\mathbf{C}\cdot\mathbf{C}]) = \operatorname{span}\{\operatorname{col}(\mathbf{C}_{r_1}\cdot\mathbf{C}_{r_2}) : 1 \le r_1, r_2 \le R\}\\
&= \operatorname{span}\{\operatorname{col}(\mathbf{S}_2(\mathbf{C})), \operatorname{col}(\mathbf{C}_r\cdot\mathbf{C}_r) : 1 \le r \le R\}\\
&= \operatorname{span}\{\operatorname{col}(\mathbf{S}_2(\mathbf{C})), \operatorname{col}(\mathbf{V}_r\cdot\mathbf{V}_r), \operatorname{col}(\mathbf{V}_r\cdot\mathbf{W}_r), \operatorname{col}(\mathbf{W}_r\cdot\mathbf{W}_r) : 1 \le r \le R\}\\
&= \operatorname{span}\{\operatorname{col}(\mathbf{S}_2(\mathbf{C})), S, \operatorname{col}(\mathbf{V}_r\cdot\mathbf{W}_r), \operatorname{col}(\mathbf{W}_r\cdot\mathbf{W}_r) : 1 \le r \le R\}.
\end{aligned} \tag{E.4}$$

From the construction of $\mathbf{W}_r$, $\mathbf{V}_r$ and $\mathbf{S}_2(\mathbf{C})$ it follows that

$$\operatorname{span}\{\operatorname{col}(\mathbf{V}_r\cdot\mathbf{W}_r), \operatorname{col}(\mathbf{W}_r\cdot\mathbf{W}_r)\} \subseteq \operatorname{col}(\mathbf{C}_r\cdot\mathbf{Z}^T_{r,\mathbf{C}}) \subseteq \operatorname{col}(\mathbf{S}_2(\mathbf{C})), \quad 1 \le r \le R. \tag{E.5}$$

Now, (E.1) follows from (E.4) and (E.5).

(iii) *Proof of* (E.2). From the construction of V_r it follows that

$$\operatorname{col}(\mathbf{V}_r) \text{ is orthogonal to } \operatorname{col}(\mathbf{C}_1), \dots, \operatorname{col}(\mathbf{C}_{r-1}), \operatorname{col}(\mathbf{C}_{r+1}), \dots, \operatorname{col}(\mathbf{C}_R). \tag{E.6}$$

Let $\mathbf{P}_K$ be defined as in (3.1). Then

$$\begin{aligned}
\operatorname{col}(\mathbf{P}_K(\mathbf{V}_r\cdot\mathbf{V}_r)) &= \operatorname{span}\{\mathbf{x}_r \otimes \mathbf{y}_r + \mathbf{y}_r \otimes \mathbf{x}_r :\ \mathbf{x}_r, \mathbf{y}_r \in V_r\},\\
\operatorname{col}(\mathbf{P}_K(\mathbf{C}_{r_1}\cdot\mathbf{C}_{r_2})) &= \operatorname{span}\{\mathbf{x}_{r_1} \otimes \mathbf{y}_{r_2} + \mathbf{y}_{r_2} \otimes \mathbf{x}_{r_1} :\ \mathbf{x}_{r_1} \in \operatorname{col}(\mathbf{C}_{r_1}), \mathbf{y}_{r_2} \in \operatorname{col}(\mathbf{C}_{r_2})\}.
\end{aligned} \tag{E.7}$$

It now easily follows from (E.6) that

$$\operatorname{col}(\mathbf{P}_K(\mathbf{V}_r\cdot\mathbf{V}_r)) \text{ is orthogonal to } \operatorname{col}(\mathbf{P}_K(\mathbf{C}_{r_1}\cdot\mathbf{C}_{r_2})),\ 1 \le r \le R,\ 1 \le r_1 < r_2 \le R.$$

Hence $\mathbf{P}_K S$ is orthogonal to $\mathbf{P}_K\,\mathrm{col}(\mathbf{S}_2(\mathbf{C}))$. Since $\mathbf{P}_K$ is a bijective linear map from $\mathbb{F}^{C^2_{K+1}}$ to $\mathrm{vec}\,(\mathbb{F}^{K\times K}_{sym})$, it follows that the subspaces S and $\mathrm{col}(\mathbf{S}_2(\mathbf{C}))$ are linearly independent, that is, (E.2) holds.

(iii) *Proof of* (E.3). Since $\mathbf{P}_K$ is a bijective linear map, it is sufficient to prove that $\dim \mathbf{P}_K S = \sum C^2_{d_r+1}$. From the construction of V_r it follows that $\mathrm{col}(\mathbf{V}_{r_1})$ is orthogonal to $\mathrm{col}(\mathbf{V}_{r_2})$ for $r_1 \neq r_2$. Hence, by (E.7), $\mathrm{col}(\mathbf{P}_K(\mathbf{V}_{r_1}\cdot\mathbf{V}_{r_1}))$ is orthogonal to $\mathrm{col}(\mathbf{P}_K(\mathbf{V}_{r_2}\cdot\mathbf{V}_{r_2}))$ for $r_1 \neq r_2$. Since $\mathbf{P}_K S = \mathrm{span}\{\mathrm{col}(\mathbf{P}_K(\mathbf{V}_r\cdot\mathbf{V}_r)) :\ 1 \le r \le R\}$, it follows that $\mathbf{P}_K S$ is the orthogonal sum of the subspaces $\mathrm{col}(\mathbf{P}_K(\mathbf{V}_r\cdot\mathbf{V}_r))$. Hence $\dim \mathbf{P}_K S = \sum \dim\mathrm{col}(\mathbf{P}_K(\mathbf{V}_r\cdot\mathbf{V}_r))$. To prove that $\dim\mathrm{col}(\mathbf{P}_K(\mathbf{V}_r\cdot\mathbf{V}_r)) = C^2_{d_r+1}$ we show that the $C^2_{d_r+1}$ columns $\mathbf{v}_i\otimes\mathbf{v}_j+\mathbf{v}_j\otimes\mathbf{v}_i$, $1\le i\le j\le d_r$ of $\mathbf{P}_K(\mathbf{V}_r\cdot\mathbf{V}_r)$ are linearly independent, where $\mathbf{v}_1,\dots,\mathbf{v}_{d_r}$ denote the columns of $\mathbf{V}_r$. Indeed, assume that there exist values λ_{ij} , $1 \le i \le j \le d_r$ such that $\mathbf{0} = \sum\limits_{1\le i\le j\le d_r} \lambda_{ij}(\mathbf{v}_i\otimes\mathbf{v}_j+\mathbf{v}_j\otimes\mathbf{v}_i)$.

Then

$$\begin{aligned}\mathbf{0} &= \sum_{1\le i\le d_r}\mathbf{v}_i\otimes\sum_{i\le j\le d_r}\lambda_{ij}\mathbf{v}_j + \sum_{1\le j\le d_r}\mathbf{v}_j\otimes\sum_{1\le i\le j}\lambda_{ij}\mathbf{v}_i\\ &= \sum_{1\le i\le d_r}\mathbf{v}_i\otimes\left(\sum_{i<j\le d_r}\lambda_{ij}\mathbf{v}_j+\sum_{1\le j<i}\lambda_{ji}\mathbf{v}_j+2\lambda_{ii}\mathbf{v}_{ii}\right).\end{aligned} \tag{E.8}$$

Since the vectors $\mathbf{v}_1,\dots,\mathbf{v}_{d_r}$ are linearly independent, it follows from (E.8) that $\lambda_{ij} = 0$ for all values of indices. ∎

Appendix F. Proof of statements 4) and 5) of Lemma 3.1. By definition, set

$$\mathcal{C}_2(\mathbf{A}) := [\mathbf{a}_1\wedge\mathbf{a}_2\ \dots\ \mathbf{a}_{R-1}\wedge\mathbf{a}_R] \in \mathbb{F}^{C^2_I\times C^2_R}, \tag{F.1}$$

$$\mathcal{C}'_2(\mathbf{B}) := [\mathbf{B}_1\wedge\mathbf{B}_2\ \dots\ \mathbf{B}_{R-1}\wedge\mathbf{B}_R] \in \mathbb{F}^{C^2_J\times\sum\limits_{r_1<r_2}L_{r_1}L_{r_2}}. \tag{F.2}$$

The matrix $\mathcal{C}_2(\mathbf{A})$ is called the second compound matrix of $\mathbf{A}$. We will need the following properties of $\mathcal{C}_2(\cdot)$ and $\mathcal{C}'_2(\cdot)$.

Lemma F.1. *Let $\mathbf{Y}$ be a matrix such that $\mathcal{C}_2(\mathbf{Y})$, and $\mathcal{C}'_2(\mathbf{YB})$ are defined. Then the following statements hold.*

1) If $\mathbf{A}$ has full column rank, then $\mathcal{C}_2(\mathbf{A})$ also has full column rank;
2) $\mathcal{C}_2(\mathbf{A}^T) = \mathcal{C}_2(\mathbf{A})^T$;
3) $\mathcal{C}_2(\mathbf{Y})\mathcal{C}_2(\mathbf{B}) = \mathcal{C}_2(\mathbf{YB})$ (Binet-Cauchy formula);
4) $\mathcal{C}_2(\mathbf{Y})\mathcal{C}'_2(\mathbf{B}) = \mathcal{C}'_2(\mathbf{YB})$.

Proof. Statements 1) to 3) are classical properties of the compound matrices (see, for instance, [24, pp. 21–22]). Statement 4) follows from statement 3). Indeed, from the definition of $\mathcal{C}_2(\mathbf{B})$ and $\mathcal{C}'_2(\mathbf{B})$ it follows that there exists a column selection matrix $\mathbf{P}$ such that $\mathcal{C}'_2(\mathbf{B}) = \mathcal{C}_2(\mathbf{B})\mathbf{P}$. Moreover, for any matrix $\mathbf{Y}$ such that $\mathcal{C}_2(\mathbf{Y})$, and $\mathcal{C}'_2(\mathbf{YB})$ are defined, the identity $\mathcal{C}'_2(\mathbf{YB}) = \mathcal{C}_2(\mathbf{YB})\mathbf{P}$ holds with the same $\mathbf{P}$. Hence, by statement 3), $\mathcal{C}_2(\mathbf{Y})\cdot\mathcal{C}'_2(\mathbf{B}) = \mathcal{C}_2(\mathbf{Y})\cdot\mathcal{C}_2(\mathbf{B})\mathbf{P} = \mathcal{C}_2(\mathbf{YB})\mathbf{P} = \mathcal{C}'_2(\mathbf{YB})$. ∎

Proof of statement 4) of Lemma 3.1. First we prove that condition (2.21) implies that $\Phi(\mathbf{A},\mathbf{B})$ has full column rank. In the case $k'_{\mathbf{B}} = 2$, we have $r_{\mathbf{A}} = R$. Hence, by statement 1) of Lemma F.1 the $C^2_I \times C^2_R$ matrix $C_2(\mathbf{A})$ has full column rank. The fact that $k'_{\mathbf{B}} = 2$ further implies that $[\mathbf{B}_{r_1}\ \mathbf{B}_{r_2}]$ has full column rank for all $r_1 \le r_2$. Hence, by statement 1) of Lemma F.1, the matrix $\mathcal{C}_2([\mathbf{B}_{r_1}\ \mathbf{B}_{r_2}])$ also has full column rank. Since $\mathbf{B}_{r_1} \wedge \mathbf{B}_{r_2}$ is formed by columns of $\mathcal{C}_2([\mathbf{B}_{r_1}\ \mathbf{B}_{r_2}])$, it also has full column rank. One can easily prove that full column rank of $C_2(\mathbf{A})$ and the matrices $\mathbf{B}_{r_1} \wedge \mathbf{B}_{r_2}$, $r_1 \leq r_2$ implies full column rank of $\Phi(\mathbf{A}, \mathbf{B})$.

We now consider the case $k'_{\mathbf{B}} > 2$.

(i) Suppose that $\Phi(\mathbf{A}, \mathbf{B})\mathbf{f} = \mathbf{0}$ for some $(\sum\limits_{r_1<r_2} L_{r_1} L_{r_2}) \times 1$ vector $\mathbf{f}$. We represent $\mathbf{f}$ as $\mathbf{f} = [\mathbf{f}_{1,2}^T\ \dots\ \mathbf{f}_{R-1,R}^T]^T$, where $\mathbf{f}_{r_1,r_2} \in \mathbb{F}^{L_{r_1} L_{r_2}}$. Then $\Phi(\mathbf{A}, \mathbf{B})\mathbf{f} = \mathbf{0}$ is equivalent to

$$\sum_{r_1<r_2} (\mathbf{a}_{r_1} \wedge \mathbf{a}_{r_2}) \otimes (\mathbf{B}_{r_1} \wedge \mathbf{B}_{r_2})\mathbf{f}_{r_1,r_2} = \mathbf{0}. \tag{F.3}$$

We can further rewrite (F.3) in matrix form as

$$\begin{aligned} \mathbf{O} &= \sum_{r_1<r_2} (\mathbf{B}_{r_1} \wedge \mathbf{B}_{r_2})\mathbf{f}_{r_1,r_2}(\mathbf{a}_{r_1} \wedge \mathbf{a}_{r_2})^T \\ &= \mathcal{C}'_2(\mathbf{B}) \operatorname{blockdiag}(\mathbf{f}_{1,2}, \dots, \mathbf{f}_{R-1,R})\mathcal{C}_2(\mathbf{A})^T. \end{aligned} \tag{F.4}$$

(ii) Let us for now assume that the last $r_{\mathbf{A}}$ columns of $\mathbf{A}$ are linearly independent. We show that $\mathbf{f}_{k'_{\mathbf{B}}-1,k'_{\mathbf{B}}} = \mathbf{0}$. Let us set

$$s_1 := L_1 + \dots + L_{k'_{\mathbf{B}}-2}, \quad s_2 := L_{k'_{\mathbf{B}}-1} + L_{k'_{\mathbf{B}}}, \quad s_3 := L_{k'_{\mathbf{B}}+1} + \dots + L_R.$$

By definition of $k'_{\mathbf{B}}$, the matrix $\mathbf{X} := \begin{bmatrix}\mathbf{B}_1 & \dots & \mathbf{B}_{k_{\mathbf{B}}}\end{bmatrix}$ has full column rank. Hence, $\mathbf{X}^\dagger\mathbf{X} = \mathbf{I}_{s_1+s_2}$, where $\mathbf{X}^\dagger$ denotes the Moore–Penrose pseudo-inverse of $\mathbf{X}$. Denoting $\mathbf{Y} := [\mathbf{O}_{s_2\times s_1}\ \mathbf{I}_{s_2}]\mathbf{X}^\dagger$, we have

$$\begin{aligned} \mathbf{YB} =& [\mathbf{O}_{s_2\times s_1}\ \mathbf{I}_{s_2}]\mathbf{X}^\dagger[\mathbf{X}\ \mathbf{B}_{k'_{\mathbf{B}}+1}\ \dots\ \mathbf{B}_R] \\ =& [\mathbf{O}_{s_2\times s_1}\ \mathbf{I}_{s_2}][\mathbf{I}_{s_1+s_2}\ \boxplus_{(s_1+s_2)\times s_3}] = [\mathbf{O}_{s_2\times s_1}\ \mathbf{I}_{s_2}\ \boxplus_{s_2\times s_3}] \\ =& \left[\mathbf{O}_{s_2\times L_1}\ \dots\ \mathbf{O}_{s_2\times L_{k'_{\mathbf{B}}-2}}\ \begin{bmatrix}\mathbf{I}_{L_{k'_{\mathbf{B}}-1}} \\ \mathbf{O}_{L_{k'_{\mathbf{B}}}\times L_{k'_{\mathbf{B}}-1}}\end{bmatrix}\ \begin{bmatrix}\mathbf{O}_{L_{k'_{\mathbf{B}}-1}\times L_{k'_{\mathbf{B}}}} \\ \mathbf{I}_{L_{k'_{\mathbf{B}}}}\end{bmatrix}\ \boxplus_{s_2\times s_3}\right], \end{aligned}$$

where $\boxplus_{p\times q}$ denotes a $p\times q$ matrix that is not further specified. From the definition of the matrix $\mathcal{C}'_2(\cdot)$ it follows that $\mathcal{C}'_2(\mathbf{YB})$ consists of $(R-1)+(R-2)+\dots+(R-k'_{\mathbf{B}}+2)$ zero blocks followed by the nonzero block $\mathbf{G} := \begin{bmatrix}\mathbf{I}_{L_{k'_{\mathbf{B}}-1}} \\ \mathbf{O}_{L_{k'_{\mathbf{B}}}\times L_{k'_{\mathbf{B}}-1}}\end{bmatrix} \wedge \begin{bmatrix}\mathbf{O}_{L_{k'_{\mathbf{B}}-1}\times L_{k'_{\mathbf{B}}}} \\ \mathbf{I}_{L_{k'_{\mathbf{B}}}}\end{bmatrix}$ and some other blocks. One can easily verify that $\mathbf{G}$ is formed by distinct columns of the $C_{s_2}^2 \times C_{s_2}^2$ identity matrix, implying that $\mathbf{G}$ has full column rank. Multiplying (F.4) by $\mathcal{C}_2(\mathbf{Y})$, applying statement 4) of Lemma F.1 and taking into account that the first $(R-1)+(R-2)+\dots+(R-k'_{\mathbf{B}}+2)$ blocks of $\mathcal{C}'_2(\mathbf{YB})$ are zero, we obtain

(F.5)

$$\begin{aligned} \mathbf{O} &= \mathcal{C}_2(\mathbf{Y})\mathbf{O} = \mathcal{C}_2(\mathbf{Y})\mathcal{C}'_2(\mathbf{B}) \operatorname{blockdiag}(\mathbf{f}_{1,2}, \dots, \mathbf{f}_{R-1,R})\mathcal{C}_2(\mathbf{A})^T \\ &= \mathcal{C}'_2(\mathbf{YB}) \operatorname{blockdiag}(\mathbf{f}_{1,2}, \dots, \mathbf{f}_{R-1,R})\mathcal{C}_2(\mathbf{A})^T \\ &= [\mathbf{G} \boxplus\ \dots\ \boxplus] \operatorname{blockdiag}(\mathbf{f}_{k'_{\mathbf{B}}-1,k'_{\mathbf{B}}}, \dots, \mathbf{f}_{R-1,R})[\mathbf{a}_{k'_{\mathbf{B}}-1} \wedge \mathbf{a}_{k'_{\mathbf{B}}}\ \dots\ \mathbf{a}_{R-1} \wedge \mathbf{a}_R]^T, \end{aligned}$$

where $\boxplus$ denotes a block of the matrix $\mathcal{C}'_2(\mathbf{YB})$. From the definition of $\mathcal{C}_2(\cdot)$ it follows that $[\mathbf{a}_{k'_{\mathbf{B}}-1} \wedge \mathbf{a}_{k'_{\mathbf{B}}}\ \dots\ \mathbf{a}_{R-1} \wedge \mathbf{a}_R] = \mathcal{C}_2([\mathbf{a}_{k'_{\mathbf{B}}-1}\ \dots\ \mathbf{a}_R])$. Since the last $r_{\mathbf{A}}$ columns of $\mathbf{A}$ are linearly independent and $r_{\mathbf{A}} \geq R - k'_{\mathbf{B}} + 2$ it follows that the vectors $\mathbf{a}_{k'_{\mathbf{B}}-1}, \dots, \mathbf{a}_R$

are also linearly independent. Hence, by Lemma F.1 the matrix $\mathcal{C}_2([\mathbf{a}_{k'_{\mathbf{B}}-1}\ \dots\ \mathbf{a}_R])$ has full column rank. Hence (F.5) is equivalent to

$$\mathbf{O} = [\mathbf{G} \boxplus\ \dots\ \boxplus]\,\text{blockdiag}(\mathbf{f}_{k'_{\mathbf{B}}-1,k'_{\mathbf{B}}},\dots,\mathbf{f}_{R-1,R}),$$

implying that $\mathbf{G}\mathbf{f}_{k'_{\mathbf{B}}-1,k'_{\mathbf{B}}} = \mathbf{0}$. Since $\mathbf{G}$ has full column rank, it follows that $\mathbf{f}_{k'_{\mathbf{B}}-1,k'_{\mathbf{B}}} = \mathbf{0}$.

(iii) We show that $\mathbf{f}_{r_1,r_2} = \mathbf{0}$ for all $1 \le r_1 < r_2 \le R$. Since $k_{\mathbf{A}} \ge 2$, the vectors $\mathbf{a}_{r_1}, \mathbf{a}_{r_2}$ are linearly independent. Let us extend two vectors $\mathbf{a}_{r_1}, \mathbf{a}_{r_2}$ to a basis of range($\mathbf{A}$) by adding $r_{\mathbf{A}}-2$ linearly independent columns of $\mathbf{A}$. It is clear that there exists an $R\times R$ permutation matrix $\mathbf{\Pi}$ such that the last $r_{\mathbf{A}}$ columns of $\mathbf{A\Pi}$ coincide with the chosen basis. Moreover, since $k'_{\mathbf{B}}-1 \ge R-r_{\mathbf{A}}+1$ we can choose $\mathbf{\Pi}$ such that the $(k'_{\mathbf{B}}-1)$th and $k'_{\mathbf{B}}$th columns of $\mathbf{A\Pi}$ are equal to $\mathbf{a}_{r_1}$ and $\mathbf{a}_{r_2}$, respectively. We can now reason as under (ii) for $\mathbf{A\Pi}$ and $\mathbf{B\Pi}$ to obtain that $\mathbf{f}_{r_1,r_2} = \mathbf{0}$.

(iv) From (iii) we immediately obtain that $\mathbf{f} = \mathbf{0}$. Hence, $\Phi(\mathbf{A},\mathbf{B})$ has full column rank.

Now we prove that (2.16) implies (2.17). Substituting $\mathbf{E}_r = \mathbf{B}_r\mathbf{C}_r^T$ in the expressions for $\mathbf{F}$, we obtain that $\mathbf{F} = [\mathbf{B}_{r_1}\ \mathbf{B}_{r_2}\ \dots\ \mathbf{B}_{r_{R-r_{\mathbf{A}}+2}}]\,\text{blockdiag}(\mathbf{C}_{r_1}^T,\mathbf{C}_{r_2}^T,\dots, \mathbf{C}_{r_{R-r_{\mathbf{A}}+2}}^T)$, implying that $r_{[\mathbf{B}_{r_1}\ \mathbf{B}_{r_2}\ \dots\ \mathbf{B}_{r_{R-r_{\mathbf{A}}+2}}]} \ge r_{\mathbf{F}}$. Hence, by (2.16), $k'_{\mathbf{B}} \ge R-r_{\mathbf{A}}+2$. Since $k_{\mathbf{A}}\ge 2$, the result follows from the first part of statement 4). ◻

Proof of statement 5) of Lemma 3.1. Assume that $(\mathbf{a}_1\otimes\mathbf{B}_1)\mathbf{f}_1+\dots+(\mathbf{a}_R\otimes\mathbf{B}_R)\mathbf{f}_R = \mathbf{0}$ for some vectors $\mathbf{f}_r\in\mathbb{F}^{L_r}$. It is sufficient to prove that all vectors $\mathbf{f}_r$ are zero. We rewrite the identity $(\mathbf{a}_1\otimes\mathbf{B}_1)\mathbf{f}_1+\dots+(\mathbf{a}_R\otimes\mathbf{B}_R)\mathbf{f}_R = \mathbf{0}$ in the matrix form $[\mathbf{a}_1\ \dots\ \mathbf{a}_R][\mathbf{B}_1\mathbf{f}_1\ \dots\ \mathbf{B}_R\mathbf{f}_R]^T = \mathbf{O}$. Then from statements 2) and 3) of Lemma F.1 and from the definition of the second compound matrix it follows that

$$\begin{aligned}\mathcal{C}_2(\mathbf{O}) &= \mathcal{C}_2([\mathbf{a}_1\ \dots\ \mathbf{a}_R][\mathbf{B}_1\mathbf{f}_1\ \dots\ \mathbf{B}_R\mathbf{f}_R]^T) = \mathcal{C}_2([\mathbf{a}_1\ \dots\ \mathbf{a}_R])\mathcal{C}_2([\mathbf{B}_1\mathbf{f}_1\ \dots\ \mathbf{B}_R\mathbf{f}_R])^T\\ &= \sum_{1\le r_1<r_2\le R}(\mathbf{a}_{r_1}\wedge\mathbf{a}_{r_2})\left(\mathbf{B}_{r_1}\mathbf{f}_{r_1}\wedge\mathbf{B}_{r_2}\mathbf{f}_{r_2}\right)^T\\ &= \sum_{1\le r_1<r_2\le R}(\mathbf{a}_{r_1}\wedge\mathbf{a}_{r_2})\left((\mathbf{B}_{r_1}\wedge\mathbf{B}_{r_2})(\mathbf{f}_{r_1}\otimes\mathbf{f}_{r_2})\right)^T,\end{aligned}$$

which can be rewritten in vectorized form as $\mathbf{0} = \Phi(\mathbf{A},\mathbf{B})[(\mathbf{f}_1\otimes\mathbf{f}_2)^T\ \dots\ (\mathbf{f}_{R-1}\otimes\mathbf{f}_R)^T]^T$. Since the matrix $\Phi(\mathbf{A},\mathbf{B})$ has full column rank, it follows easily that at least $R-1$ of the vectors $\mathbf{f}_1,\dots,\mathbf{f}_R$ are zero. We assume w.l.o.g. that the last $R-1$ vectors are zero. Then $\mathbf{0} = (\mathbf{a}_1\otimes\mathbf{B}_1)\mathbf{f}_1$, which implies that $\mathbf{f}_1$ is also zero. ◻

Appendix G. Proofs of Lemmas 4.1 and 4.2.

Proof of Lemma 4.1. 1) Assume that $\mathbf{N}\mathbf{f} = \mathbf{0}$, where $\mathbf{f} = [\mathbf{f}_1^T\ \dots\ \mathbf{f}_R^T]^T$ and $\mathbf{f}_r\in\mathbb{F}^{d_r}$. Then, by construction of $\mathbf{N}_r$,

$$\mathbf{0} = \mathbf{C}^T\mathbf{N}\mathbf{f} = \text{blockdiag}(\mathbf{C}_1^T\mathbf{N}_1,\dots,\mathbf{C}_R^T\mathbf{N}_R)\mathbf{f} = [(\mathbf{C}_1^T\mathbf{N}_1\mathbf{f}_1)^T\ \dots\ (\mathbf{C}_R^T\mathbf{N}_R\mathbf{f}_R)^T]^T,$$

implying that $\mathbf{C}_r^T\mathbf{N}_r\mathbf{f}_r = \mathbf{0}$ for $r = 1,\dots,R$. Hence,

$$\text{(G.1)}\qquad \mathbf{C}^T(\mathbf{N}_r\mathbf{f}_r) = (\mathbf{0},\dots,\mathbf{0},\mathbf{C}_r^T\mathbf{N}_r\mathbf{f}_r,\mathbf{0},\dots,\mathbf{0}) = \mathbf{0},\qquad r = 1,\dots,R.$$

By (1.5) and (2.14), $\mathbf{C}^T$ has full column rank. Since $\mathbf{N}_r$ also has full column rank, it follows from (G.1) that $\mathbf{f}_r = \mathbf{0}$ for $r = 1,\dots,R$. Hence we must have $\mathbf{f} = \mathbf{0}$. Thus the matrix $\mathbf{N}$ has full column rank.

2) It follows from statement 1) that $[\mathbf{N}_1 \otimes \mathbf{N}_1 \ \dots \ \mathbf{N}_R \otimes \mathbf{N}_R]$ has full column rank. Obviously, $\operatorname{blockdiag}(\mathbf{M}_1, \dots, \mathbf{M}_R)$ has full column rank. Since $\mathbf{W} = [\mathbf{N}_1 \otimes \mathbf{N}_1 \ \dots \ \mathbf{N}_R \otimes \mathbf{N}_R] \operatorname{blockdiag}(\mathbf{M}_1, \dots, \mathbf{M}_R)$, it also has full column rank.

3) Since, by (2.14), $r_{\mathbf{T}_{(3)}} = K$ and, by (1.5), $\mathbf{T}_{(3)} = [\mathbf{a}_1 \otimes \mathbf{I}_J \ \dots \ \mathbf{a}_R \otimes \mathbf{I}_J][\mathbf{E}_1^T \ \dots \ \mathbf{E}_R^T]^T$, it follows that the $JR \times K$ matrix $[\mathbf{E}_1^T \ \dots \ \mathbf{E}_R^T]^T$ has full column rank. Hence for any r the columns of $[\mathbf{E}_1^T \ \dots \ \mathbf{E}_R^T]^T \mathbf{N}_r = [\mathbf{O} \ \dots \mathbf{O} \ (\mathbf{E}_r \mathbf{N}_r)^T \ \mathbf{O} \ \dots \ \mathbf{O}]^T$ are nonzero. Assume that $\mathbf{O} = \alpha_1 \mathbf{E}_1 + \dots + \alpha_R \mathbf{E}_R$ for some $\alpha_1, \dots, \alpha_R \in \mathbb{F}$. Then for any r, $\mathbf{O} = (\alpha_1 \mathbf{E}_1 + \dots + \alpha_R \mathbf{E}_R)\mathbf{N}_r = \alpha_r \mathbf{E}_r \mathbf{N}_r$. Since $\mathbf{E}_r \mathbf{N}_r$ is not the zero matrix, it follows that $\alpha_r = 0$. Thus, the matrices $\mathbf{E}_1, \dots, \mathbf{E}_R$ are linearly independent. □

Proof of Lemma 4.2. By (1.3),

$$\mathbf{T}_{(1)} = [\operatorname{vec}(\mathbf{E}_1) \ \dots \ \operatorname{vec}(\mathbf{E}_R)]\mathbf{A}^T = [\operatorname{vec}(\tilde{\mathbf{E}}_1) \ \dots \ \operatorname{vec}(\tilde{\mathbf{E}}_{\tilde{R}})]\tilde{\mathbf{A}}^T, \tag{G.2}$$

where $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}_1 \ \dots \ \tilde{\mathbf{a}}_{\tilde{R}}]$.

Case 1: condition b) *holds.* Then, $\mathbf{A}$ has full column rank. Hence, by (G.2),

$$[\operatorname{vec}(\mathbf{E}_1) \ \dots \ \operatorname{vec}(\mathbf{E}_R)] = [\operatorname{vec}(\tilde{\mathbf{E}}_1) \ \dots \ \operatorname{vec}(\tilde{\mathbf{E}}_{\tilde{R}})](\mathbf{A}^\dagger \tilde{\mathbf{A}})^T.$$

Since any column of $\tilde{\mathbf{A}}$ is a column of $\mathbf{A}$, each column of $\mathbf{A}^\dagger \tilde{\mathbf{A}}$ contains at most one nonzero entry. Since $\mathbf{E}_1, \dots, \mathbf{E}_R$ are nonzero matrices, it follows that the columns of $(\mathbf{A}^\dagger \tilde{\mathbf{A}})^T \in \mathbb{F}^{\tilde{R} \times R}$ are also nonzero, which is possible only if $\tilde{R} = R$ and $\tilde{\mathbf{A}} = \mathbf{AP}$ for some $R \times R$ permutation matrix $\mathbf{P}$. Hence, by (G.2), $[\operatorname{vec}(\mathbf{E}_1) \ \dots \ \operatorname{vec}(\mathbf{E}_R)] = [\operatorname{vec}(\tilde{\mathbf{E}}_1) \ \dots \ \operatorname{vec}(\tilde{\mathbf{E}}_{\tilde{R}})]\mathbf{P}^T$. Thus, the decompositions coincide up to permutation of summands. It is also clear that the matrices $\mathbf{E}_1, \dots, \mathbf{E}_R$ can be computed by solving the system of linear equations $[\operatorname{vec}(\mathbf{E}_1) \ \dots \ \operatorname{vec}(\mathbf{E}_R)]\mathbf{A}^T = \mathbf{T}_{(1)}$.

Case 2: condition c) *holds.* To prove statement 1) it is sufficient to show that the matrices $\mathbf{E}_1, \dots, \mathbf{E}_R$ can be computed by EVD up to scaling. Indeed, if $\mathbf{E}_r = x_r \hat{\mathbf{E}}_r$ and the matrices $\hat{\mathbf{E}}_r$ are known, then, by (1.3), the scaling factors x_r can be found as from the linear equation $[\mathbf{a}_1 \otimes \operatorname{vec}(\hat{\mathbf{E}}_1) \ \dots \ \mathbf{a}_r \otimes \operatorname{vec}(\hat{\mathbf{E}}_R)][x_1 \ \dots \ x_r]^T = \operatorname{vec}(\mathbf{T}_{(1)})$.

We choose arbitrary integers $r_1, \dots, r_{R - r_{\mathbf{A}} + 2}$ such that $1 \le r_1 < \dots < r_{R - r_{\mathbf{A}} + 2} \le R$ and show that the matrices $\mathbf{E}_{r_1}, \dots, \mathbf{E}_{r_{R - r_{\mathbf{A}} + 2}}$ can be computed by EVD up to scaling. We set

$$\Omega = \{r_1, \dots, r_{R - r_{\mathbf{A}} + 2}\} \quad \text{and} \quad \{p_1, \dots, p_{r_{\mathbf{A}} - 2}\} = \{1, \dots, R\} \setminus \Omega. \tag{G.3}$$

Since $k_{\mathbf{A}} = r_{\mathbf{A}}$, it follows that the intersection of the null space of the $(r_{\mathbf{A}} - 2) \times I$ matrix $[\mathbf{a}_{p_1} \ \dots \ \mathbf{a}_{p_{r_{\mathbf{A}} - 2}}]^T$ and the column space of $\mathbf{A}$ is two-dimensional. Let the intersection be spanned by the vectors $\mathbf{h}_{\Omega,1}, \mathbf{h}_{\Omega,2} \in \mathbb{F}^I$, where here and later in the proof the subindex "Ω" indicates that a quantity depends on $r_1, \dots, r_{R - r_{\mathbf{A}} + 2}$. Then again, since $k_{\mathbf{A}} = r_{\mathbf{A}}$, it follows that

$$\text{any two columns of } \begin{bmatrix} \mathbf{h}_{\Omega,1}^T \mathbf{a}_{r_1} & \dots & \mathbf{h}_{\Omega,1}^T \mathbf{a}_{r_{R - r_{\mathbf{A}} + 2}} \\ \mathbf{h}_{\Omega,2}^T \mathbf{a}_{r_1} & \dots & \mathbf{h}_{\Omega,2}^T \mathbf{a}_{r_{R - r_{\mathbf{A}} + 2}} \end{bmatrix} \text{ are linearly independent.} \tag{G.4}$$

Let $\mathcal{Q}_\Omega$ denote the $2 \times J \times K$ tensor such that $\mathbf{Q}_{\Omega(1)} = \mathbf{T}_{(1)}[\mathbf{h}_{\Omega,1} \ \mathbf{h}_{\Omega,2}]$. Then, by (1.3),

$$\mathcal{Q}_\Omega = \sum_{r=1}^{R} \begin{bmatrix} \mathbf{h}_{\Omega,1}^T \mathbf{a}_r \\ \mathbf{h}_{\Omega,2}^T \mathbf{a}_r \end{bmatrix} \circ \mathbf{E}_r = \sum_{k=1}^{R - r_{\mathbf{A}} + 2} \begin{bmatrix} \mathbf{h}_{\Omega,1}^T \mathbf{a}_{r_k} \\ \mathbf{h}_{\Omega,2}^T \mathbf{a}_{r_k} \end{bmatrix} \circ \mathbf{E}_{r_k} = \sum_{k=1}^{R - r_{\mathbf{A}} + 2} \begin{bmatrix} \mathbf{h}_{\Omega,1}^T \mathbf{a}_{r_k} \\ \mathbf{h}_{\Omega,2}^T \mathbf{a}_{r_k} \end{bmatrix} \circ (\mathbf{B}_{r_k} \mathbf{C}_{r_k}^T), \tag{G.5}$$

where $\mathbf{B}_{r_k} \in \mathbb{F}^{J\times L_{r_k}}$ and $\mathbf{C}_{r_k} \in \mathbb{F}^{K\times L_{r_k}}$ denote full column rank matrices such that $\mathbf{E}_{r_k} = \mathbf{B}_{r_k}\mathbf{C}_{r_k}^T$. Since condition c) in Theorem 2.5 is equivalent to condition c) in Theorem 2.6, it follows that $k'_{\mathbf{B}} \geq R - r_{\mathbf{A}} + 2$ and $k'_{\mathbf{C}} \geq R - r_{\mathbf{A}} + 2$. Hence,

$$\text{(G.6)} \qquad [\mathbf{B}_{r_1} \ \dots \ \mathbf{B}_{r_{R-r_{\mathbf{A}}+2}}] \ \text{and} \ [\mathbf{C}_{r_1} \ \dots \ \mathbf{C}_{r_{R-r_{\mathbf{A}}+2}}] \ \text{have full column rank.}$$

Hence, by Theorem 1.4, the decomposition of $\mathcal{Q}_\Omega$ into a sum of max ML rank-$(1, L_{r_k}, L_{r_k})$ terms is unique and can be computed by EVD. Thus, the matrices $\mathbf{E}_{r_1}, \dots, \mathbf{E}_{r_{R-r_{\mathbf{A}}+2}}$ can be computed by EVD up to scaling. Since the indices $r_1, \dots,$ $r_{R-r_{\mathbf{A}}+2}$ were chosen arbitrary, it follows that all matrices $\mathbf{E}_{r_1}, \dots, \mathbf{E}_{r_{R-r_{\mathbf{A}}+2}}$ can be computed by EVD up to scaling. The overall procedure is summarized in steps $11-18$ of Algorithm 2.1.

Now we prove statement 2). First we show that $\tilde{R} = R$ and that the $\tilde{\mathbf{E}}_1, \dots, \tilde{\mathbf{E}}_R$ involves the same matrices as $\mathbf{E}_1, \dots, \mathbf{E}_R$. Similarly to (G.5) we obtain that

$$\text{(G.7)} \qquad \mathcal{Q}_\Omega = \sum_{r=1}^{\tilde{R}} \begin{bmatrix} \mathbf{h}_{\Omega,1}^T \tilde{\mathbf{a}}_r \\ \mathbf{h}_{\Omega,2}^T \tilde{\mathbf{a}}_r \end{bmatrix} \circ \tilde{\mathbf{E}}_r.$$

It is clear that there exist $C_R^{R-r_{\mathbf{A}}+2}$ sets Ω of the form (G.3). Thus, by (G.5) and (G.7), we obtain a system of $C_R^{R-r_{\mathbf{A}}+2}$ identities:

$$\text{(G.8)} \quad \mathcal{Q}_\Omega = \sum_{k=1}^{R-r_{\mathbf{A}}+2} \begin{bmatrix} \mathbf{h}_{\Omega,1}^T \mathbf{a}_{r_k} \\ \mathbf{h}_{\Omega,2}^T \mathbf{a}_{r_k} \end{bmatrix} \circ \mathbf{E}_{r_k} = \sum_{r=1}^{\tilde{R}} \begin{bmatrix} \mathbf{h}_{\Omega,1}^T \tilde{\mathbf{a}}_r \\ \mathbf{h}_{\Omega,2}^T \tilde{\mathbf{a}}_r \end{bmatrix} \circ \tilde{\mathbf{E}}_r, \ 1 \leq r_1 < \dots < r_{R-r_{\mathbf{A}}+2} \leq R.$$

Hence, by (1.5) and (G.5), system (G.8) can be rewritten in matrix form as

$$\text{(G.9)} \quad \mathbf{Q}_{\Omega(3)} = \left[\begin{bmatrix} \mathbf{h}_{\Omega,1}^T \mathbf{a}_{r_1} \\ \mathbf{h}_{\Omega,2}^T \mathbf{a}_{r_1} \end{bmatrix} \otimes \mathbf{B}_{r_1} \ \dots \ \begin{bmatrix} \mathbf{h}_{\Omega,1}^T \mathbf{a}_{r_{R-r_{\mathbf{A}}+2}} \\ \mathbf{h}_{\Omega,2}^T \mathbf{a}_{r_{R-r_{\mathbf{A}}+2}} \end{bmatrix} \otimes \mathbf{B}_{r_{R-r_{\mathbf{A}}+2}}\right] [\mathbf{C}_{r_1} \ \dots \ \mathbf{C}_{r_{R-r_{\mathbf{A}}+2}}]^T =$$
$$\sum_{r=1}^{\tilde{R}} \begin{bmatrix} \mathbf{h}_{\Omega,1}^T \tilde{\mathbf{a}}_r \\ \mathbf{h}_{\Omega,2}^T \tilde{\mathbf{a}}_r \end{bmatrix} \otimes \tilde{\mathbf{E}}_r, \qquad 1 \leq r_1 < \dots < r_{R-r_{\mathbf{A}}+2} \leq R.$$

From (G.4), (G.6) and the first identity in (G.9), it follows that $\mathbf{Q}_{\Omega(3)}$ has rank $L_{r_1} + \dots + L_{r_{R-r_{\mathbf{A}}+2}}$. Since the rank is subadditive, it follows from (G.9), that

$$\text{(G.10)} \quad L_{r_1} + \dots + L_{r_{R-r_{\mathbf{A}}+2}} \leq \sum_{r=1}^{\tilde{R}} r\left(\begin{bmatrix} \mathbf{h}_{\Omega,1}^T \tilde{\mathbf{a}}_r \\ \mathbf{h}_{\Omega,2}^T \tilde{\mathbf{a}}_r \end{bmatrix}\right) r_{\tilde{\mathbf{E}}_r}, \ 1 \leq r_1 < \dots < r_{R-r_{\mathbf{A}}+2} \leq R,$$

where $r(\mathbf{f})$ denotes the rank of a 2×1 matrix $\mathbf{f}$: $r(\mathbf{0}) = 0$ and $r(\mathbf{f}) = 1$, if $\mathbf{f} \neq 0$. It is clear that for each r there exist exactly $C_{R-1}^{R-r_{\mathbf{A}}+1}$ subsets $\{r_1, \dots, r_{R-r_{\mathbf{A}}+2}\} \subset \{1, \dots, R\}$ that contain r. Hence each L_r appears in exactly $C_{R-1}^{R-r_{\mathbf{A}}+1}$ inequalities in (G.10). Since $\tilde{\mathbf{a}}_1 = \mathbf{a}_r$ for some r, it follows that the term $r\left(\begin{bmatrix} \mathbf{h}_{\Omega,1}^T \tilde{\mathbf{a}}_1 \\ \mathbf{h}_{\Omega,2}^T \tilde{\mathbf{a}}_1 \end{bmatrix}\right) r_{\tilde{\mathbf{E}}_1} =$ $r\left(\begin{bmatrix} \mathbf{h}_{\Omega,1}^T \mathbf{a}_r \\ \mathbf{h}_{\Omega,2}^T \mathbf{a}_r \end{bmatrix}\right) r_{\tilde{\mathbf{E}}_1}$ appears in the same $C_{R-1}^{R-r_{\mathbf{A}}+1}$ inequalities as L_r, implying, by the construction of $\mathbf{h}_{\Omega,1}$ and $\mathbf{h}_{\Omega,2}$, that $\begin{bmatrix} \mathbf{h}_{\Omega,1}^T \mathbf{a}_r \\ \mathbf{h}_{\Omega,2}^T \mathbf{a}_r \end{bmatrix} \neq \mathbf{0}$. Thus, $r_{\tilde{\mathbf{E}}_1}$ appears in exactly

$C_{R-1}^{R-r_{\mathbf{A}}+1}$ inequalities in (G.10). In the same fashion one can prove that each of the values $1 \cdot r_{\tilde{\mathbf{E}}_2}, \dots, 1 \cdot r_{\tilde{\mathbf{E}}_{\tilde{R}}}$ appears in (G.10) exactly $C_{R-1}^{R-r_{\mathbf{A}}+1}$ times. Thus, summing all inequalities in (G.10) and taking into account that $\tilde{R} \le R$ and $r_{\tilde{\mathbf{E}}_r} \le L_r$ for all r we obtain

$$\text{(G.11)}\quad (L_1 + \dots + L_R)C_{R-1}^{R-r_{\mathbf{A}}+1} \le (r_{\tilde{\mathbf{E}}_1} + \dots + r_{\tilde{\mathbf{E}}_{\tilde{R}}})C_{R-1}^{R-r_{\mathbf{A}}+1} \le (L_1 + \dots + L_{\tilde{R}})C_{R-1}^{R-r_{\mathbf{A}}+1} \le (L_1 + \dots + L_R)C_{R-1}^{R-r_{\mathbf{A}}+1}.$$

Hence $\tilde{R} = R$ and $r_{\tilde{\mathbf{E}}_r} = L_r$ for all r.

To complete the proof of statement 2) we need to show that the terms $\tilde{\mathbf{a}}_1 \circ \tilde{\mathbf{E}}_1, \dots, \tilde{\mathbf{a}}_R \circ \tilde{\mathbf{E}}_R$ coincide with the terms $\mathbf{a}_1 \circ \mathbf{E}_1, \dots, \mathbf{a}_R \circ \mathbf{E}_R$. If we assume that at least one of the inequalities in (G.10) is strict, then the first inequality in (G.11) should also be strict, which is not possible. Thus, (G.10) holds with "$\le$" replaced by "$=$". Hence, by Theorem 1.4, the two decompositions of $\mathcal{Q}_\Omega$ in (G.8) coincide up to permutation of their terms. This readily implies that the matrices $\tilde{\mathbf{E}}_1, \dots, \tilde{\mathbf{E}}_R$ coincide with $\lambda_1\mathbf{E}_1, \dots, \lambda_R\mathbf{E}_R$ for some $\lambda_1, \dots, \lambda_R \in \mathbb{F}\setminus\{0\}$, i.e., there exists an $R\times R$ permutation matrix $\mathbf{P}$ such that

$$\text{(G.12)}\qquad [\operatorname{vec}(\tilde{\mathbf{E}}_1)\ \dots\ \operatorname{vec}(\tilde{\mathbf{E}}_{\tilde{R}})] = [\operatorname{vec}(\mathbf{E}_1)\ \dots\ \operatorname{vec}(\mathbf{E}_R)]\operatorname{diag}(\lambda_1, \dots, \lambda_R)\mathbf{P}.$$

Substituting (G.12) in (G.2) we obtain that

$$\text{(G.13)}\quad [\operatorname{vec}(\mathbf{E}_1)\ \dots\ \operatorname{vec}(\mathbf{E}_R)]\mathbf{A}^T = [\operatorname{vec}(\mathbf{E}_1)\ \dots\ \operatorname{vec}(\mathbf{E}_R)]\operatorname{diag}(\lambda_1, \dots, \lambda_R)\mathbf{P}\tilde{\mathbf{A}}^T.$$

Since the matrices $\mathbf{E}_1, \dots, \mathbf{E}_R$ are linearly independent, it follows from (G.13) that $\mathbf{A}^T = \operatorname{diag}(\lambda_1, \dots, \lambda_R)\mathbf{P}\tilde{\mathbf{A}}^T$. Hence $\mathbf{A} = \tilde{\mathbf{A}}\mathbf{P}^T\operatorname{diag}(\lambda_1, \dots, \lambda_R)$. Since any column of $\tilde{\mathbf{A}}$ is a column of $\mathbf{A}$ and since $k_{\mathbf{A}} = r_{\mathbf{A}} \ge 2$, it follows that $\lambda_1 = \dots = \lambda_R = 1$. Hence $\tilde{\mathbf{A}} = \mathbf{AP}$ and, by (G.12), $[\operatorname{vec}(\tilde{\mathbf{E}}_1)\ \dots\ \operatorname{vec}(\tilde{\mathbf{E}}_{\tilde{R}})] = [\operatorname{vec}(\mathbf{E}_1)\ \dots\ \operatorname{vec}(\mathbf{E}_R)]\mathbf{P}$, i.e., the terms $\tilde{\mathbf{a}}_1 \circ \tilde{\mathbf{E}}_1, \dots, \tilde{\mathbf{a}}_R \circ \tilde{\mathbf{E}}_R$ coincide with the terms $\mathbf{a}_1 \circ \mathbf{E}_1, \dots, \mathbf{a}_R \circ \mathbf{E}_R$. ∎

Appendix H. Proof of Theorem 2.17. The following theorem complements results on uniqueness[15] presented in subsection 2.5.1 and will be used in the proof of Theorem 2.17. Namely, we will show that Theorem 2.17 is the generic counterpart of Theorem H.1.

THEOREM H.1. *Let $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ admit decomposition (1.2) with $\mathbf{a}_r \ne \mathbf{0}$ and $r_{\mathbf{B}_r} = r_{\mathbf{C}_r} = L_r$ for all r. Assume that the matrix $\mathbf{C}$ has full column rank and that the matrices $\mathbf{A}$ and $\mathbf{B}$ satisfy the following assumption:*

(H.1) *if at least two of the vectors $\mathbf{g}_1 \in \mathbb{C}^{L_1}, \dots, \mathbf{g}_R \in \mathbb{C}^{L_R}$ are nonzero, then the rank of $\mathbf{a}_1(\mathbf{B}_1\mathbf{g}_1)^T + \dots + \mathbf{a}_R(\mathbf{B}_R\mathbf{g}_R)^T$ is at least 2.*

Then the decomposition of $\mathcal{T}$ into a sum of max ML rank-$(1, L_r, L_r)$ terms is unique.

Proof. Since $\mathbf{C}$ has full column rank we have that $K \ge \sum L_r$. By statement 1) of Theorem 2.4, we can assume that $K = \sum L_r$, i.e., that $\mathbf{C}$ is square and nonsingular.

[15] It can be shown that if $\mathbf{C}$ has full column rank, then Theorem H.1 guarantees uniqueness under more relaxed assumptions than Theorem 2.6. On the other hand, assumption (H.1) in Theorem H.1 is not easy to verify for particular $\mathbf{A}$ and $\mathbf{B}$ and Theorem H.1 does not come with an EVD-based algorithm.

i) First we reformulate assumption (H.1). Such reformulation will immediately imply that

$$k_{\mathbf{A}} \geq 2 \ \text{ and matrix } [\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R] \text{ has full column rank.} \tag{H.2}$$

If the rank of $\mathbf{a}_1(\mathbf{B}_1\mathbf{g}_1)^T + \cdots + \mathbf{a}_R(\mathbf{B}_R\mathbf{g}_R)^T$ is less than 2, then there exist vectors $\mathbf{z} \in \mathbb{F}^I$ and $\mathbf{y} \in \mathbb{F}^J$ such that

$$\mathbf{a}_1(\mathbf{B}_1\mathbf{g}_1)^T + \cdots + \mathbf{a}_R(\mathbf{B}_R\mathbf{g}_R)^T = \mathbf{z}\mathbf{y}^T. \tag{H.3}$$

Transposing and vectorizing both sides of (H.3) we obtain that $(\mathbf{a}_1 \otimes \mathbf{B}_1)\mathbf{g}_1 + \cdots + (\mathbf{a}_R \otimes \mathbf{B}_R)\mathbf{g}_R = \mathbf{z} \otimes \mathbf{y}$. Hence assumption (H.1) can be reformulated as follows:

$$\begin{aligned}&\text{the identity } \ (\mathbf{a}_1 \otimes \mathbf{B}_1)\mathbf{g}_1 + \cdots + (\mathbf{a}_R \otimes \mathbf{B}_R)\mathbf{g}_R = \mathbf{z} \otimes \mathbf{y} \ \text{ holds}\\ &\text{only if at most one of } \ \mathbf{g}_1, \dots, \mathbf{g}_R \text{ is nonzero.}\end{aligned} \tag{H.4}$$

One can now easily derive (H.2) from (H.4).

ii) Now we prove uniqueness. Let $\mathcal{T} = \sum_{r=1}^{\widehat{R}} \widehat{\mathbf{a}}_r \circ (\widehat{\mathbf{B}}_r \widehat{\mathbf{C}}_r^T)$, where $\widehat{R} \leq R$, $\widehat{\mathbf{a}}_r \neq \mathbf{0}$, $\widehat{\mathbf{B}}_r \in \mathbb{F}^{J \times \widehat{L}_r}$ and $\widehat{\mathbf{C}}_r \in \mathbb{F}^{K \times \widehat{L}_r}$ have full column rank, and $\widehat{L}_r \leq L_r$ for $r = 1, \dots, \widehat{R}$. Then, by (1.5),

$$[\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R]\mathbf{C}^T = \mathbf{T}_{(3)} = [\widehat{\mathbf{a}}_1 \otimes \widehat{\mathbf{B}}_1 \ \dots \ \widehat{\mathbf{a}}_{\widehat{R}} \otimes \widehat{\mathbf{B}}_{\widehat{R}}]\widehat{\mathbf{C}}^T. \tag{H.5}$$

Since, by (H.2), $[\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R]$ has full column rank and since $\mathbf{C}$ is a nonsingular matrix, it follows from (H.5) that $r_{\mathbf{T}_{(3)}} = \sum L_r$. Hence the matrices $[\widehat{\mathbf{a}}_1 \otimes \widehat{\mathbf{B}}_1 \ \dots \ \widehat{\mathbf{a}}_{\widehat{R}} \otimes \widehat{\mathbf{B}}_{\widehat{R}}]$ and $\widehat{\mathbf{C}}$ are at least rank-$\sum L_r$, implying that $\sum\limits_{r=1}^{\widehat{R}} \widehat{L}_r \geq \sum\limits_{r=1}^{R} L_r$. On the other hand, since $\widehat{R} \leq R$ and $\widehat{L}_r \leq L_r$ for $r = 1, \dots, \widehat{R}$, we also have that $\sum\limits_{r=1}^{\widehat{R}} \widehat{L}_r \leq \sum\limits_{r=1}^{R} L_r$. Hence $\sum\limits_{r=1}^{\widehat{R}} \widehat{L}_r = \sum\limits_{r=1}^{R} L_r$ which is possible only if $\widehat{R} = R$ and $\widehat{L}_r = L_r$ for all r. Multiplying (H.5) by $\widehat{\mathbf{C}}^{-T}$ we obtain that

$$[\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R]\mathbf{G} = [\widehat{\mathbf{a}}_1 \otimes \widehat{\mathbf{B}}_1 \ \dots \ \widehat{\mathbf{a}}_R \otimes \widehat{\mathbf{B}}_R], \tag{H.6}$$

where $\mathbf{G} = \mathbf{C}^T\widehat{\mathbf{C}}^{-T}$ is a $\sum L_r \times \sum L_r$ nonsingular matrix. Let $\mathbf{g}_1 = [\mathbf{g}_{1,1}^T \ \dots \ \mathbf{g}_{1,R}^T]^T$ and $\mathbf{g}_2 = [\mathbf{g}_{2,1}^T \ \dots \ \mathbf{g}_{2,R}^T]^T$ be columns of $\mathbf{G}$, where $\mathbf{g}_{1,r}, \mathbf{g}_{2,r} \in \mathbb{F}^{L_r}$. Then, by assumption (H.1), at most one of the vectors $\mathbf{g}_{1,1}, \dots, \mathbf{g}_{1,R}$ is nonzero. Since $\mathbf{G}$ is nonsingular we have that exactly one of the vectors $\mathbf{g}_{1,1}, \dots, \mathbf{g}_{1,R}$ is nonzero. Let $\mathbf{g}_{1,i} \neq \mathbf{0}$. Similarly, we also have that exactly one of the vectors $\mathbf{g}_{2,1}, \dots, \mathbf{g}_{2,R}$ is nonzero. Let $\mathbf{g}_{2,j} \neq \mathbf{0}$. We claim that if $\mathbf{g}_1$ and $\mathbf{g}_2$ are columns of the same block $\mathbf{G}_r \in \mathbb{F}^{\sum L_r \times L_r}$ of $\mathbf{G} = [\mathbf{G}_1 \ \dots \ \mathbf{G}_R]$, then $i = j$. Indeed, by (H.5),

$$(\mathbf{a}_i \otimes \mathbf{B}_i)\mathbf{g}_{1,i} = \widehat{\mathbf{a}}_r \otimes \mathbf{y}_1 \ \text{ and } \ (\mathbf{a}_j \otimes \mathbf{B}_j)\mathbf{g}_{2,j} = \widehat{\mathbf{a}}_r \otimes \mathbf{y}_2, \tag{H.7}$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are columns of $\widehat{\mathbf{B}}_r$. It follows from (H.7) that $\mathbf{a}_i$ and $\mathbf{a}_j$ are proportional to $\widehat{\mathbf{a}}_r$. Since, by (H.2), $k_{\mathbf{A}} \geq 2$, it follows that $i = j$. Thus, in the partition $\mathbf{G}_r = [\mathbf{G}_{1r}^T \ \dots \ \mathbf{G}_{Rr}^T]^T$ with $\mathbf{G}_{1r} \in \mathbb{F}^{L_1 \times L_r}, \dots \mathbf{G}_{Rr} \in \mathbb{F}^{L_R \times L_r}$, exactly one block is nonzero. Since $\mathbf{G} = [\mathbf{G}_1 \ \dots \ \mathbf{G}_R]$ is nonsingular, it follows that the nonzero block of $\mathbf{G}_r$ is square, i.e. $L_r \times L_r$, and nonsingular, $r = 1, \dots, R$. Hence $\mathbf{G}$ can be reduced to block diagonal form by permuting its blocks $\mathbf{G}_1, \dots, \mathbf{G}_R$. Let $\mathbf{P}$ denote a permutation matrix such that $\mathbf{GP} = \text{blockdiag}(\tilde{\mathbf{G}}_{11}, \dots, \tilde{\mathbf{G}}_{RR})$ with nonsingular $\tilde{\mathbf{G}}_{rr} \in \mathbb{F}^{L_r \times L_r}$. It is clear that multiplication of the right hand side of (H.6) by $\mathbf{P}$ corresponds to a permutation of the summands in $\mathcal{T} = \sum_{r=1}^{R} \widehat{\mathbf{a}}_r \circ (\widehat{\mathbf{B}}_r \widehat{\mathbf{C}}_r^T)$. Thus, the terms in $\mathcal{T} = \sum_{r=1}^{R} \widehat{\mathbf{a}}_r \circ (\widehat{\mathbf{B}}_r \widehat{\mathbf{C}}_r^T)$ can can be permuted so that (H.6) holds for $\mathbf{G} = \text{blockdiag}(\tilde{\mathbf{G}}_{11}, \dots, \tilde{\mathbf{G}}_{RR})$. Hence (H.6) reduces to the R identities

$$(\mathbf{a}_r \otimes \mathbf{B}_r)\tilde{\mathbf{G}}_{rr} = \widehat{\mathbf{a}}_r \otimes \widehat{\mathbf{B}}_r, \qquad r = 1, \dots, R$$

which imply that $\widehat{\mathbf{a}}_r$ is proportional to $\mathbf{a}_r$ and that the column space of $\widehat{\mathbf{B}}_r$ coincides with the column space of $\mathbf{B}_r$. In other words, we have shown that $\widehat{\mathbf{a}}_r$ and $\widehat{\mathbf{B}}_r$ in $\mathcal{T} = \sum_{r=1}^{R} \widehat{\mathbf{a}}_r \circ (\widehat{\mathbf{B}}_r \widehat{\mathbf{C}}_r^T)$ can be chosen to be equal to $\mathbf{a}_r$ and $\mathbf{B}_r$, respectively. Since the matrix $[\mathbf{a}_1 \otimes \mathbf{B}_1 \ \dots \ \mathbf{a}_R \otimes \mathbf{B}_R]$ has full column rank, we also have from (H.5) that $\widehat{\mathbf{C}} = \mathbf{C}$. ∎

Proof of Theorem 2.17. If $I \geq R$, then the result follows from Theorem 1.9. So, throughout the proof we assume that $I < R$.

By definition set

$$\text{(H.8)} \qquad W_{\mathbf{A},\mathbf{B},\mathbf{C}} := \{(\mathbf{A}, \mathbf{B}, \mathbf{C}) : \ \text{the assumptions in Theorem H.1 do not hold}\}.$$

We show that $\mu\{W_{\mathbf{A},\mathbf{B},\mathbf{C}}\} = 0$, where μ denotes a measure on $\mathbb{F}^{I \times R} \times \mathbb{F}^{J \times \sum L_r} \times \mathbb{F}^{K \times \sum L_r}$ that is absolutely continuous with respect to the Lebesgue measure. Obviously, $W_{\mathbf{A},\mathbf{B},\mathbf{C}} = W_{\mathbf{C}} \cup W_{\mathbf{A},\mathbf{B}}$, where

$$W_{\mathbf{C}} := \{(\mathbf{A}, \mathbf{B}, \mathbf{C}) : \ \mathbf{C} \text{ does not have full column rank}\} \ \text{ and}$$
$$W_{\mathbf{A},\mathbf{B}} := \{(\mathbf{A}, \mathbf{B}, \mathbf{C}) : \ \text{assumption (H.1) does not hold}\}.$$

It is clear that, by the assumption $\sum L_r \leq K$ in (2.49), $\mu\{W_{\mathbf{C}}\} = 0$, so we need to show that $\mu\{W_{\mathbf{A},\mathbf{B}}\} = 0$. Since (H.1) does not depend on $\mathbf{C}$, we have $W_{\mathbf{A},\mathbf{B}} = W \times \mathbb{F}^{J \times \sum L_r}$, where

$$W := \{(\mathbf{A}, \mathbf{B}) : \ \text{assumption (H.1) does not hold}\}$$

is a subset of $\mathbb{F}^{I \times R} \times \mathbb{F}^{J \times \sum L_r}$. From Fubini's theorem [23, Theorem C, p.148] it follows that $\mu\{W_{\mathbf{A},\mathbf{B}}\} = 0$ if and only if $\mu_1\{W\} = 0$, where μ_1 is a measure on $\mathbb{F}^{I \times R} \times \mathbb{F}^{J \times \sum L_r}$ that is absolutely continuous with respect to the Lebesgue measure. Since $R > I$ and $J \geq L_{R-1} + L_R$ $(= \max_{1 \leq i < j \leq R}(L_i + L_j))$, it follows that

$$\mu_1\{(\mathbf{A}, \mathbf{B}) : \ k_{\mathbf{A}} < I \ \text{ or } \ k'_{\mathbf{B}} < 2\} = 0.$$

Hence we can assume w.l.o.g. that

$$\text{(H.9)} \qquad W = \{(\mathbf{A}, \mathbf{B}) : \ \text{assumption (H.1) does not hold}, \ k_{\mathbf{A}} = I, \ \text{and} \ k'_{\mathbf{B}} \geq 2\}.$$

The remaining part of the proof is based on a well-known algebraic geometry based method. In [19] we have explained the method and used it to study generic uniqueness of CPD and INDSCAL. We have explained in [19] that to prove that $\mu_1\{W\} = 0$, it is sufficient to show that for $\mathbb{F} = \mathbb{C}$ the Zariski closure $\overline{W}$ of W is not the entire space $\mathbb{C}^{I \times R} \times \mathbb{C}^{J \times \sum L_r}$, which is equivalent to $\dim \overline{W} \leq IR + J \sum L_r - 1$. To estimate the dimension of $\overline{W}$ we will take the following four steps (for a detailed explanation of the steps and examples see [19]; also, for $L_1 = \cdots = L_r = 1$, the overall derivation is similar to the proof of Lemma 2.5 in [33]). To simplify the presentation of the steps, we omit mentioning the isomorphism between $\mathbb{C}^{k\times l}\times\mathbb{C}^{m\times n}$ and $\mathbb{C}^{kl+mn}$; for instance, we consider W as a subset of $\mathbb{C}^{d_1}$, where $d_1 = IR + J\sum L_r$. In the first step we parameterize W. Namely, we construct a subset $\widehat{Z} \subseteq \mathbb{C}^{d_1+I+J+\sum L_r}$ and a projection $\pi : \mathbb{C}^{d_1+I+J+\sum L_r} \to \mathbb{C}^{d_1}$ such that $W = \pi(\widehat{Z})$. In step 2 we represent $\widehat{Z}$ as a finite union of subsets $Z^{l_1,\dots,l_I}_{r_1,\dots,r_I}$ such that each $Z^{l_1,\dots,l_I}_{r_1,\dots,r_I}$ is the image of a Zariski open subset of $\mathbb{C}^{d_1-d_2+1}$ under a rational mapping, where $d_2 := (I-1)(J-1)-\sum L_r$ is nonnegative by (2.49). In step 3 we show that $\dim(Z^{l_1,\dots,l_I}_{r_1,\dots,r_I}) = d_1 - d_2 + 1$ and that $\dim(\pi(Z^{l_1,\dots,l_I}_{r_1,\dots,r_I})) \le d_1 - d_2 - 1$. Finally, in step 4 we conclude that $\dim \overline{W} = \dim(\pi(\widehat{Z})) \le \max(\dim(\pi(Z^{l_1,\dots,l_I}_{r_1,\dots,r_I}))) = d_1 - d_2 - 1 \le d_1 - 1$.

Step 1. Let $\omega(\mathbf{g}_1,\dots,\mathbf{g}_R)$ denote the number of nonzero vectors in the set $\{\mathbf{g}_1,\dots,\mathbf{g}_R\}$. We claim that if assumption (H.1) does not hold, $k_{\mathbf{A}} = I$, and $k'_{\mathbf{B}} \ge 2$, then $\omega(\mathbf{g}_1,\dots,\mathbf{g}_R) \ge I$. Indeed, if $I > \omega(\mathbf{g}_1,\dots,\mathbf{g}_R) \ge 2$, then by the Frobenius inequality,

$$1 \ge r_{\mathbf{a}_1(\mathbf{B}_1\mathbf{g}_1)^T+\cdots+\mathbf{a}_R(\mathbf{B}_R\mathbf{g}_R)^T} = r_{\mathbf{A}\,\mathrm{blockdiag}(\mathbf{g}_1^T,\dots,\mathbf{g}_R^T)\mathbf{B}^T} \ge$$
$$r_{\mathbf{A}\,\mathrm{blockdiag}(\mathbf{g}_1^T,\dots,\mathbf{g}_R^T)} + r_{\mathrm{blockdiag}(\mathbf{g}_1^T,\dots,\mathbf{g}_R^T)\mathbf{B}^T} - r_{\mathrm{blockdiag}(\mathbf{g}_1^T,\dots,\mathbf{g}_R^T)} =$$
$$\omega(\mathbf{g}_1,\dots,\mathbf{g}_R) + r_{[\mathbf{B}_1\mathbf{g}_1\ \dots\ \mathbf{B}_R\mathbf{g}_r]} - \omega(\mathbf{g}_1,\dots,\mathbf{g}_R) \ge 2,$$

which is a contradiction. Hence, W in (H.9) can be expressed as

$$W = \Big\{(\mathbf{A},\mathbf{B}) :\ \text{there exist } \mathbf{g}_1 \in \mathbb{C}^{L_1},\dots,\mathbf{g}_R \in \mathbb{C}^{L_R},\ \mathbf{z} \in \mathbb{C}^I,\ \text{and } \mathbf{y} \in \mathbb{C}^J$$
$$\text{such that } \mathbf{a}_1(\mathbf{B}_1\mathbf{g}_1)^T + \cdots + \mathbf{a}_R(\mathbf{B}_R\mathbf{g}_R)^T = \mathbf{z}\mathbf{y}^T, \tag{H.10}$$
$$k_{\mathbf{A}} = I,\ k'_{\mathbf{B}} \ge 2,\ \text{and} \tag{H.11}$$
$$\omega(\mathbf{g}_1,\dots,\mathbf{g}_R) \ge I\Big\}. \tag{H.12}$$

It is clear that $W = \pi(\widehat{Z})$, where

$$\widehat{Z} = \Big\{(\mathbf{A},\mathbf{B},\mathbf{g}_1,\dots,\mathbf{g}_R,\mathbf{z},\mathbf{y}) : \text{(H.10)–(H.12) hold}\Big\}$$

is a subset of $\mathbb{C}^{I\times R}\times\mathbb{C}^{J\times\sum L_r}\times\mathbb{C}^{L_1}\times\cdots\times\mathbb{C}^{L_R}\times\mathbb{C}^I\times\mathbb{C}^J$ and π is the projection onto the first two factors

$$\pi : \mathbb{C}^{I\times R}\times\mathbb{C}^{J\times\sum L_r}\times\mathbb{C}^{L_1}\times\cdots\times\mathbb{C}^{L_R}\times\mathbb{C}^I\times\mathbb{C}^J \to \mathbb{C}^{I\times R}\times\mathbb{C}^{J\times\sum L_r}.$$

Step 2. Let $g_{l,r}$ denote the lth entry of $\mathbf{g}_r$. Since

$$\omega(\mathbf{g}_1,\dots,\mathbf{g}_R) \ge I \Leftrightarrow \mathbf{g}_{r_1} \ne \mathbf{0},\dots,\mathbf{g}_{r_I} \ne \mathbf{0} \text{ for some } 1 \le r_1 < \cdots < r_I \le R$$

and since

$$\mathbf{g}_{r_1} \ne \mathbf{0},\dots,\mathbf{g}_{r_I} \ne \mathbf{0} \Leftrightarrow g_{l_1,r_1}\cdots g_{l_I,r_I} \ne 0 \text{ for some } 1 \le l_1 \le L_{r_1},\dots,1 \le l_I \le L_{r_I},$$

we obtain that

$$\widehat{Z} = \bigcup_{1\le r_1<\cdots<r_I\le R}\ \bigcup_{1\le l_1\le L_{r_1},\dots,1\le l_I\le L_{r_I}}$$
$$\Big\{(\mathbf{A},\mathbf{B},\mathbf{g}_1,\dots,\mathbf{g}_R,\mathbf{z},\mathbf{y}) : \text{(H.10)–(H.11) hold and } g_{l_1,r_1}\cdots g_{l_I,r_I} \ne 0\Big\}.$$

Let $\mathbf{A}_{r_1,\dots,r_I}$ denote the submatrix of $\mathbf{A}$ formed by columns $r_1, \dots, r_I$. Since (H.11) is more restrictive than the condition $\det(\mathbf{A}_{r_1,\dots,r_I}) \neq 0$, it follows that

$$\widehat{Z} \subseteq \bigcup_{1\le r_1<\dots<r_I\le R} \; \bigcup_{1\le l_1\le L_{r_1},\dots,1\le l_I\le L_{r_I}} Z^{l_1,\dots,l_I}_{r_1,\dots,r_I},$$

where

$$Z^{l_1,\dots,l_I}_{r_1,\dots,r_I} = \Big\{(\mathbf{A}, \mathbf{B}, \mathbf{g}_1, \dots, \mathbf{g}_R, \mathbf{z}, \mathbf{y}) : \text{(H.10) holds,} \quad \det(\mathbf{A}_{r_1,\dots,r_I}) \neq 0, \; g_{l_1,r_1} \cdots g_{l_I,r_I} \neq 0\Big\}.$$

We show that each subset $Z^{l_1,\dots,l_I}_{r_1,\dots,r_I}$ can be represented as the image of a Zariski open subset $Y^{l_1,\dots,l_I}_{r_1,\dots,r_I}$ of $\mathbb{C}^{IR+J\sum L_r+\sum L_r-IJ+I+J}$ under a rational map $\phi^{l_1,\dots,l_I}_{r_1,\dots,r_I}$, $Z^{l_1,\dots,l_I}_{r_1,\dots,r_I} = \phi^{l_1,\dots,l_I}_{r_1,\dots,r_I}(Y^{l_1,\dots,l_I}_{r_1,\dots,r_I})$. To simplify the presentation we restrict ourselves to the case $r_1 = 1, \dots, r_I = I$ and $l_1 = \dots = l_I = 1$. The general case can be proved in the same way. Let $\mathbf{A} = [\mathbf{A}_1 \; \mathbf{A}_2]$ with $\mathbf{A}_1 \in \mathbb{F}^{I\times I}$ and $\mathbf{A}_2 \in \mathbb{F}^{I\times(R-I)}$, so that $\mathbf{A}_1 = \mathbf{A}_{1\dots1}$. By (H.10),

$$\text{(H.13)} \qquad [\mathbf{B}_1\mathbf{g}_1 \; \dots \; \mathbf{B}_I\mathbf{g}_I] = [\mathbf{y}\mathbf{z}^T - [\mathbf{B}_{I+1}\mathbf{g}_{I+1} \; \dots \; \mathbf{B}_R\mathbf{g}_R]\mathbf{A}_2^T]\mathbf{A}_1^{-T}.$$

Let $\mathbf{B}_r = [\mathbf{b}_{1,r} \; \mathbf{B}_{2,r}]$ and $\mathbf{g}_r = [g_{1,r} \; \mathbf{g}_{2,r}^T]^T$, so

$$\text{(H.14)} \quad [\mathbf{B}_1\mathbf{g}_1 \; \dots \; \mathbf{B}_I\mathbf{g}_I] = [\mathbf{b}_{1,1} \dots \mathbf{b}_{1,I}] \operatorname{diag}(g_{1,1}, \dots, g_{1,I}) + [\mathbf{B}_{2,1}\mathbf{g}_{2,1} \; \dots \; \mathbf{B}_{2,I}\mathbf{g}_{2,I}].$$

Then, by (H.13) and (H.14),

$$\text{(H.15)} \qquad \begin{aligned}[\mathbf{b}_{1,1} \dots \mathbf{b}_{1,I}] = \big([\mathbf{y}\mathbf{z}^T - [\mathbf{B}_{I+1}\mathbf{g}_{I+1} \; \dots \; \mathbf{B}_R\mathbf{g}_R]\mathbf{A}_2^T]\mathbf{A}_1^{-T} - \\ [\mathbf{B}_{2,1}\mathbf{g}_{2,1} \; \dots \; \mathbf{B}_{2,I}\mathbf{g}_{2,I}]\big) \operatorname{diag}(g_{1,1}^{-1}, \dots, g_{1,I}^{-1}),\end{aligned}$$

so the entries of $\mathbf{b}_{1,1} \dots \mathbf{b}_{1,I}$ are rational functions of the entries of $\mathbf{A}$, $\mathbf{B}_{2,1}, \dots, \mathbf{B}_{2,I}$, $\mathbf{B}_{I+1}, \dots, \mathbf{B}_R$, $\mathbf{g}_1, \dots, \mathbf{g}_R$, $\mathbf{z}$, and $\mathbf{y}$. It is clear that

$$\begin{aligned} Y^{1,\dots,1}_{1,\dots,I} := \Big\{([\mathbf{A}_1 \; \mathbf{A}_2], [\mathbf{B}_{2,1} \; \dots \; \mathbf{B}_{2,I} \; \mathbf{B}_{I+1} \; \dots \; \mathbf{B}_R], \mathbf{g}_1, \dots, \mathbf{g}_R, \mathbf{z}, \mathbf{y}) : \\ \det(\mathbf{A}_1) \neq 0, \; g_{1,1} \cdots g_{1,I} \neq 0\Big\}\end{aligned}$$

is a Zariski open subset of $\mathbb{C}^{I\times R} \times \mathbb{C}^{J\times\left(\sum\limits_{r=1}^{I}(L_r-1)+\sum\limits_{r=I+1}^{R} L_r\right)} \times \mathbb{C}^{L_1} \times \dots \times \mathbb{C}^{L_R} \times \mathbb{C}^I \times \mathbb{C}^J$ and that $Z^{1,\dots,1}_{1,\dots,I} = \phi^{1,\dots,1}_{1,\dots,I}(Y^{1,\dots,1}_{1,\dots,I})$, where the rational mapping

$$\begin{aligned}\phi^{1,\dots,1}_{1,\dots,I} : ([\mathbf{A}_1 \; \mathbf{A}_2], [\mathbf{B}_{2,1} \; \dots \; \mathbf{B}_{2,I} \; \mathbf{B}_{I+1} \; \dots \; \mathbf{B}_R], \mathbf{g}_1, \dots, \mathbf{g}_R, \mathbf{z}, \mathbf{y}) \to \\ ([\mathbf{A}_1 \; \mathbf{A}_2], [[\mathbf{b}_{1,1} \; \mathbf{B}_{2,1}] \; \dots \; [\mathbf{b}_{1,I} \; \mathbf{B}_{2,I}] \; \mathbf{B}_{I+1} \; \dots \; \mathbf{B}_R], \mathbf{g}_1, \dots, \mathbf{g}_R, \mathbf{z}, \mathbf{y}) = \\ (\mathbf{A}, \mathbf{B}, \mathbf{g}_1, \dots, \mathbf{g}_R, \mathbf{z}, \mathbf{y})\end{aligned}$$

is defined by (H.15).

Step 3. In this step we prove that $\dim(\pi(Z^{l_1,\dots,l_I}_{r_1,\dots,r_I})) \le IR + J\sum L_r - 1$. W.l.o.g. we restrict ourselves again to the case $r_1 = 1, \dots, r_I = I$ and $l_1 = \dots = l_I = 1$. Since the dimension of the image $\phi^{1,\dots,1}_{1,\dots,I}(Y^{1,\dots,1}_{1,\dots,I})$ cannot exceed the dimension of $Y^{1,\dots,1}_{1,\dots,I}$ and since $Y^{1,\dots,1}_{1,\dots,I}$ is a Zariski open subset we have

$$\dim(Z^{1,\dots,1}_{1,\dots,I}) \leq^{16} \dim(Y^{1,\dots,1}_{1,\dots,I}) = IR + J(-I + \sum_{r=1}^{R} L_r) + L_1 + \dots + L_r + I + J. \tag{H.16}$$

Let $f:\ Z^{1,\dots,1}_{1,\dots,I} \to \mathbb{C}^{I\times R} \times \mathbb{C}^{J\times \sum L_r}$ denote the restriction of π to $Z^{1,\dots,1}_{1,\dots,I}$:

$$f:\ (\mathbf{A}, \mathbf{B}, \mathbf{g}_1, \dots, \mathbf{g}_R, \mathbf{z}, \mathbf{y}) \to (\mathbf{A}, \mathbf{B}), \qquad (\mathbf{A}, \mathbf{B}, \mathbf{g}_1, \dots, \mathbf{g}_R, \mathbf{z}, \mathbf{y}) \in Z^{1,\dots,1}_{1,\dots,I}.$$

From the definition of $Z^{1,\dots,1}_{1,\dots,I}$ it follows that if $(\mathbf{A}, \mathbf{B}, \mathbf{g}_1, \dots, \mathbf{g}_R, \mathbf{z}, \mathbf{y}) \in Z^{1,\dots,1}_{1,\dots,I}$, then $(\mathbf{A}, \mathbf{B}, \alpha\beta\mathbf{g}_1, \dots, \alpha\beta\mathbf{g}_R, \alpha\mathbf{z}, \beta\mathbf{y}) \in Z^{1,\dots,1}_{1,\dots,I}$ for any nonzero $\alpha, \beta \in \mathbb{C}$. Hence for any $(\mathbf{A}, \mathbf{B}) \in f(Z^{1,\dots,1}_{1,\dots,I})$ we have that

$$f^{-1}((\mathbf{A}, \mathbf{B})) \supseteq \{(\mathbf{A}, \mathbf{B}, \alpha\beta\mathbf{g}_1, \dots, \alpha\beta\mathbf{g}_R, \alpha\mathbf{z}, \beta\mathbf{y}):\ \alpha \neq 0,\ \beta \neq 0\},$$

implying that

$$\dim(f^{-1}(\mathbf{A}, \mathbf{B})) \geq \dim\{(\alpha\mathbf{z}, \beta\mathbf{y}):\ \alpha \neq 0,\ \beta \neq 0\} = 2, \tag{H.17}$$

where $f^{-1}(\cdot)$ denotes the preimage. From the fiber dimension theorem [30, Theorem 3.7, p. 78], (H.16), (H.17), and the assumption $\sum L_r \leq (I-1)(J-1)$ in (2.49) it follows that

$$\begin{aligned}\dim(f(Z^{1,\dots,1}_{1,\dots,I})) &\leq \dim(Z^{1,\dots,1}_{1,\dots,I}) - \dim(f^{-1}(\mathbf{A}, \mathbf{B})) = \\ &IR + J\sum_{r=1}^{R} L_r - 1 + \sum_{r=1}^{R} L_r - (I-1)(J-1) \leq IR + J\sum_{r=1}^{R} L_r - 1.\end{aligned}$$

Since $\pi(Z^{1,\dots,1}_{1,\dots,I}) = f(Z^{1,\dots,1}_{1,\dots,I})$, we have that $\dim(\pi(Z^{1,\dots,1}_{1,\dots,I})) \leq IR + J\sum\limits_{r=1}^{R} L_r - 1$.

Step 4. Finally, we have that $\dim \overline{W} = \dim(\pi(\widehat{Z})) \leq \max(\dim(\pi(Z^{l_1,\dots,l_I}_{r_1,\dots,r_I}))) \leq IR + J\sum L_r - 1$. ∎

REFERENCES

[1] E. Angelini, C. Bocci, and L. Chiantini, *Real identifiability vs. complex identifiability*, Linear and Multilinear Algebra, 66 (2018), pp. 1257–1267.

[2] C. Beltrán, P. Breiding, and N. Vannieuwenhoven, *Pencil-based algorithms for tensor rank decomposition are not stable*, arXiv:1807.04159, (2018).

[3] C. Bocci, L. Chiantini, and G. Ottaviani, *Refined methods for the identifiability of tensors*, Ann. Mat. Pura. Appl., 193 (2014), pp. 1691–1702.

[4] R. Bro, R. A. Harshman, N. D. Sidiropoulos, and M. E. Lundy, *Modeling multi-way data with linearly dependent loadings*, Journal of Chemometrics, 23 (2009), pp. 324–340.

[5] Y. Cai and C. Liu, *An algebraic approach to nonorthogonal general joint block diagonalization*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 50–71.

[6] O. Cherrak, H. Ghennioui, N. Thirion-Moreau, and E. H. Abarkan, *Preconditioned optimization algorithms solving the problem of the non unitary joint block diagonalization: application to blind separation of convolutive mixtures*, Multidim. Syst. Sign. Process., 29 (2018), pp. 1373–1396.

[16] It can be proved that actually "=" holds but in the sequel we will only need "≤".

[7] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, *An algorithm for generic and low-rank specific identifiability of complex tensors*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1265–1287.

[8] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, *Effective criteria for specific identifiability of tensors and forms*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 656–681.

[9] A. Cichocki, D. Mandic, C. Caiafa, A.-H. Phan, G. Zhou, Q. Zhao, and L. De Lathauwer, *Tensor decompositions for signal processing applications. From two-way to multiway component analysis*, IEEE Signal Process. Mag., 32 (2015), pp. 145–163.

[10] *Handbook of Blind Source Separation, Independent Component Analysis and Applications*, Academic Press, Oxford, UK, 2010.

[11] L. De Lathauwer, *A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 642–666.

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

[13] L. De Lathauwer, *Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1451–1474.

[14] L. De Lathauwer and A. de Baynast, *Blind deconvolution of DS-CDMA signals by means of decomposition in rank-$(1, L, L)$ terms*, IEEE Trans. Signal Process., 56 (2008), pp. 1562–1571.

[15] O. Debals, M. Van Barel, and L. De Lathauwer, *Löwner-based blind signal separation of rational functions with applications*, IEEE Trans. Signal Process., 64 (2016), pp. 1909–1918.

[16] I. Domanov and L. De Lathauwer, *On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part I: Basic results and uniqueness of one factor matrix*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 855–875.

[17] I. Domanov and 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 (2013), pp. 876–903.

[18] I. Domanov and L. De Lathauwer, *Canonical polyadic decomposition of third-order tensors: reduction to generalized eigenvalue decomposition*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 636–660.

[19] I. Domanov and L. De Lathauwer, *Generic uniqueness conditions for the canonical polyadic decomposition and INDSCAL*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1567–1589.

[20] I. Domanov and L. De Lathauwer, *Generic uniqueness of a structured matrix factorization and applications in blind source separation*, IEEE J. Sel. Topics Signal Process., 10 (2016), pp. 701–711.

[21] I. Domanov, N. Vervliet, and L. De Lathauwer, *Decomposition of a tensor into multilinear rank-$(M_r, N_r, \cdot)$ terms*, Internal Report 18-51, ESAT-STADIUS, KU Leuven (Leuven, Belgium), (2018).

[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 4th ed., 2013.

[23] P. R. Halmos, *Measure theory*, Springer-Verlag, New-York, 1974.

[24] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1990.

[25] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Review, 51 (2009), pp. 455–500.

[26] X. Liu, T. Jiang, L. Yang, and H. Zhu, *Paralind-based identifiability results for parameter estimation via uniform linear array*, EURASIP Journal on Advances in Signal Processing, 2012 (2012), p. 154.

[27] C. Mueller-Smith and P. Spasojević, *Column-wise symmetric block partitioned tensor decomposition*, in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 2956–2960.

[28] Y. Nakatsukasa, , T. Soma, and A. Uschmajew, *Finding a low-rank basis in a matrix subspace*, Mathematical Programming, 162 (2017), pp. 325–361.

[29] D. Nion and L. De Lathauwer, *A link between the decomposition of a third-order tensor in rank-$(L, L, 1)$ terms and joint block diagonalization*, in 2009 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2009, pp. 89–92.

[30] D. Perrin, *Algebraic Geometry. An Introduction*, Springer-Verlag London, 2008.

[31] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, *Tensor decomposition for signal processing and machine learning*, IEEE Trans. Signal Process., 65 (2017), pp. 3551–3582.

[32] M. Sørensen, I. Domanov, and L. De Lathauwer, *Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-*$(L_{r_n}, L_{r_n}, 1)$ *terms—Part II: Algorithms*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1015–1045.
[33] V. Strassen, *Rank and optimal computation of generic tensors*, Linear Algebra Appl., 52–53 (1983), pp. 645–685.
[34] J. M. F. ten Berge, *Partial uniqueness in CANDECOMP/PARAFAC*, Journal of Chemometrics, 18 (2004), pp. 12–16.
[35] A.-J. Van Der Veen and A. Paulraj, *An analytical constant modulus algorithm*, IEEE Trans. Signal Process., 44 (1996), pp. 1136–1155.
[36] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, *Tensorlab 3.0*, Mar. 2016, https://www.tensorlab.net. Available online.
[37] M. Yang, *On partial and generic uniqueness of block term tensor decompositions*, Annali Dell'Universita'Di Ferrara, 60 (2014), pp. 465–493.
[38] M. Yang, D. Che, W. Liu, Z. Kang, C. Peng, M. Xiao, and Q. Cheng, *On identifiability of 3-tensors of multilinear rank* $(1, L_r, L_r)$, Big Data & Information Analytics, 1 (2016), pp. 391–401.