KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Domanov I., De Lathauwer L. (2021), **From Computation to Comparison of Tensor Decompositions.** SIAM Journal on Matrix Analysis and Applications, vol. 42(2), 2021, 449-474.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	https://doi.org/10.1137/20M1349370
Journal homepage	https://epubs.siam.org/doi/abs/10.1137/20M1349370
Author contact	ignat.domanov@kuleuven.be, lieven.delathauwer@kuleuven.be
Abstract	Decompositions of higher-order tensors into sums of simple terms are ubiquitous. We show that in order to verify that two tensors are generated by the same (possibly scaled) terms it is not necessary to compute the individual decompositions. In general the explicit computation of such a decomposition may have high complexity and can be ill-conditioned. We now show that under some assumptions the verification can be reduced to a comparison of both the column and row spaces of the corresponding matrix representations of the tensors. We consider rank-1 terms as well as low multilinear rank terms (also known as block terms) and show that the number of the terms and their multilinear rank can be inferred as well. The comparison relies only on numerical linear algebra and can be done in a numerically reliable way. We also illustrate how our results can be applied to solve a multilabel classification problem that appears in the context of blind source separation.
IR	url in Lirias https://lirias.kuleuven.be/3367084

(article begins on next page)

From computation to comparison of tensor decompositions*

Ignat Domanov† and Lieven De Lathauwer†

Abstract. Decompositions of higher-order tensors into sums of simple terms are ubiquitous. We show that in order to verify that two tensors are generated by the same (possibly scaled) terms it is not necessary to compute the individual decompositions. In general the explicit computation of such a decomposition may have high complexity and can be ill-conditioned. We now show that under some assumptions the verification can be reduced to a comparison of both the column and row spaces of the corresponding matrix representations of the tensors. We consider rank-1 terms as well as low multilinear rank terms (also known as block terms) and show that the number of the terms and their multilinear rank can be inferred as well. The comparison relies only on numerical linear algebra and can be done in a numerically reliable way. We also illustrate how our results can be applied to solve a multi-label classification problem that appears in the context of blind source separation.

Key words. multilinear algebra, higher-order tensor, multi-label classification, multilinear rank, canonical polyadic decomposition, PARAFAC, block term decomposition

AMS subject classifications. 15A23, 15A69

1. Introduction.

Decompositions of tensors of order N (i.e., N-way arrays of real or complex numbers) into a sum of simple terms are ubiquitous. The most common simple term is a rank-1 tensor, i.e. a nonzero tensor whose columns (resp. rows, fibers, etc.) are proportional. The corresponding decomposition into a minimal number of terms is known as Canonical Polyadic Decomposition (CPD).

It is well-known that for $N = 2$, that is, in the matrix case, the decomposition in a minimal number of rank-1 terms is not unique unless the matrix itself is rank-1: indeed, any factorization $\mathbf{A} = \mathbf{X}^{(1)}\mathbf{X}^{(2)T}$ with full column rank factors $\mathbf{X}^{(1)} = [\mathbf{x}_1^{(1)} \ \dots \ \mathbf{x}_R^{(1)}]$ and $\mathbf{X}^{(2)} = [\mathbf{x}_1^{(2)} \ \dots \ \mathbf{x}_R^{(2)}]$ generates a valid decomposition $\mathbf{A} = \mathbf{x}_1^{(1)}\mathbf{x}_1^{(2)T} + \cdots + \mathbf{x}_R^{(1)}\mathbf{x}_R^{(2)T}$, where R is the rank of $\mathbf{A}$, and this decomposition is not unique. On the other hand, if $\mathbf{X}^{(1)}$ and/or $\mathbf{X}^{(2)}$ are subject to constraints (e.g., triangularity or orthogonality), then the decomposition can be unique, but from an application point of view the imposed constraints can be unrealistic and the rank-1 terms not interpretable as meaningful "data components". In contrast, for $N \geq 3$, that is, in the higher order tensor case, the unconstrained CPD is easily unique (see, for instance, [8, 9, 21, 22] and the references therein). Its uniqueness properties make the CPD a fundamental tool for unique retrieval of data components, latent variable analysis, independent

*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).

component analysis, etc., with countless applications in chemometrics [6], telecommunication, array processing, machine learning, etc. [10, 11, 30, 32].

The higher order setting actually allows the recovery of terms that are more general than rank-1 terms. A MultiLinear (ML) rank-$(L_1, L_2, \dots)$ term is a tensor whose columns (resp. rows, fibers, etc.) form a matrix of rank L_1 (resp. L_2, L_3, etc.). Like CPD, a decomposition into a sum of ML rank-$(L_1, L_2, \dots)$ terms (also known as block term decomposition) is unique under reasonably mild assumptions (see [13, 23, 24] and the references therein), so that it has found applications in wireless communication [16], blind signal separation [14, 20], etc.

Tensor decompositions can be considered as tools for *data analysis* that allow one to break a single (tensor) data set into small interpretable components. It is known that, in general, the explicit computation of the CPD and the decomposition into a sum of ML rank-$(L_1, L_2, \dots)$ terms may have high complexity and can be ill-conditioned [1, 2, 5]. In other words, the mildness of the uniqueness conditions comes with a numerical and a computational cost.

In this paper we consider tensor decompositions from a different perspective that is closer to *pattern recognition*. Namely, we consider the following "tensor similarity" problem:

- *How to verify that two $I_1 \times \cdots \times I_N$ tensors are generated by the same (possibly scaled) rank-1 terms?*
- *More generally, how to verify that two $I_1 \times \cdots \times I_N$ tensors are generated by the same (possibly scaled) ML rank-$(L_1, L_2, \dots)$ terms?*

For brevity, our presentation will be in terms of the more general variant. The simpler (C)PD variant will follow as a special case (see, for instance, Theorem 2.1).

An obvious approach would be to compute the decompositions of all tensors and then to compare them. This has two drawbacks. First, as mentioned above, the explicit computation of the decompositions may have high complexity and can be ill-conditioned. Second, the approach may fail if the tensors are generated by the same (possibly scaled) terms in cases where the decompositions are not unique.

In this paper we will not compute the tensor decompositions. We will pursue a different approach, starting from the following trivial observation: if

(1.1) a tensor $\mathcal{B}$ is a sum of (possibly scaled) terms from the decomposition of a tensor $\mathcal{A}$,

then

$$\text{col}(\mathbf{B}_{(S^c;S)}) \subseteq \text{col}(\mathbf{A}_{(S^c;S)}) \text{ for all proper subsets } S \text{ of } \{1, \dots, N\}, \tag{1.2}$$

where $col(\cdot)$ denotes the column space of a matrix, S^c denotes the complement of the set S, and $\mathbf{A}_{(S^c;S)}$ denotes the $(\prod_{n\in S^c} I_n) \times (\prod_{n\in S} I_n)$ matrix representation of $\mathcal{A}$ (see subsection 4.2 for a formal definition of $\mathbf{A}_{(S^c;S)}$). Actually we will explain that (1.2) implies (1.1) (in a way that requires some more technical detail). A clear advantage of the approach based on the implication (1.2)⇒(1.1) is that the conditions in (1.2) rely only on numerical linear algebra and can be verified in a numerically reliable way. While the implication (1.1)⇒(1.2) is trivial, the implication (1.2)⇒(1.1) is not.

The main contribution of this paper is to show that, with some technicalities, (1.2) implies (1.1). As a matter of fact, we will need only N conditions in (1.2) for this, namely the

conditions

$$\operatorname{col}(\mathbf{B}_{(n^c;n)}) \subseteq \operatorname{col}(\mathbf{A}_{(n^c;n)}), \quad n \in \{1,\dots,N\}, \tag{1.3}$$

and we will show that the $\frac{I_1\cdots I_N}{I_n} \times I_n$ matrices $\mathbf{A}_{(n^c;n)}$ and $\mathbf{B}_{(n^c;n)}$ in (1.3) can be used to compute the number of terms in the decompositions of $\mathcal{A}$ and $\mathcal{B}$ as well as their multilinear ranks. We also consider a more general case where the inclusions in (1.3) are only known to hold for some n in $\{1,\dots,N\}$.

It is well known that in the case of CPD i) each of the subspaces $\operatorname{col}(\mathbf{A}_{(n^c;n)})$ determines the number of rank-1 terms in the CPD of $\mathcal{A}$ (i.e., the rank of $\mathcal{A}$) and ii) that the inclusion $\operatorname{col}(\mathbf{B}_{(n^c;n)}) \subseteq \operatorname{col}(\mathbf{A}_{(n^c;n)})$ in (1.3) implies that the rank-1 terms in the CPD of $\mathcal{A}$ and $\mathcal{B}$ can be matched so that their fibers are proportional in all modes that are complementary to n [7, Proposition 14.45], [31, Theorem 3.1.1.1], [26, Theorem 2.4]. At first sight it may seem that this implies that if all N inclusions in (1.3) hold, then i) the number of rank-1 terms needed to generate (with tensor-specific scaling coefficients) both $\mathcal{A}$ and $\mathcal{B}$, also just equals the rank of $\mathcal{A}$ so that ii) the fibers of the properly matched rank-1 terms are proportional in all modes. Put simply, it may seem that if all inclusions in (1.3) hold, then the tensor $\mathcal{B}$ consists of the sum of the rank-1 terms in a CPD of $\mathcal{A}$, possibly scaled. However, this is not correct. In Appendix A we give counterexamples for tensors of order three. Thus, (1.3) (or (1.2)) does not necessarily imply (1.1) in the case of CPD. There are two ways to change our view. A first way is to impose extra conditions. A second way is to consider terms that can be more general than just rank-1. In Theorems 2.1, 4.1, and 4.3 below we present such conditions and we replace the rank-1 assumption by a low ML rank assumption. Framed like this, (1.3) (or (1.2)) actually does imply (1.1). Note that the decomposition into a sum of low ML rank terms is a nontrivial extension of the CPD. While in the case of the CPD the rank-1 structure of the terms is assumed beforehand and the number of terms is a characteristic of the tensor (i.e., equals its rank), the ML rank values in the decomposition of a tensor into a sum of ML rank-$(L_1, L_2, \dots)$ terms are not known in advance, and in general, more than one combination of ML rank values and number of terms is possible. The new Theorems 2.1, 4.1, and 4.3 also imply a procedure to compute the number of terms and their ML rank values in the "similarity" setting.

It is also worth noting that the conditions

$$\operatorname{row}(\mathbf{B}_{(n^c;n)}) \subseteq \operatorname{row}(\mathbf{A}_{(n^c;n)}), \quad n \in \{1,\dots,N\}, \tag{1.4}$$

in which $row(\cdot)$ denotes the row space of a matrix, are more relaxed than the conditions in (1.3) (see Statement 1 of Lemma 3.2 below) and in general do not imply (1.1). For instance, if $\frac{I_1\cdots I_N}{I_n} \geq I_n$, then the conditions $\operatorname{row}(\mathbf{B}_{(n^c;n)}) = \operatorname{row}(\mathbf{A}_{(n^c;n)})$ $(= \mathbb{F}^{I_n})$, $n \in \{1,\dots,N\}$ hold for any generic tensors $\mathcal{A}$ and $\mathcal{B}$ (no matter whether they are generated by the same (possibly scaled) terms or not).

We will also explain that the remaining $2^N - 2 - N$ conditions in (1.2) are redundant, i.e., that the N conditions in (1.3) imply all $2^N - 2$ conditions in (1.2). (A fortiori, (1.1) follows from the N conditions in (1.3), as mentioned under the "main contribution" above.)

Prior work on tensor similarity is limited to [36]. Both the present paper and [36] originated from the technical report [15]. The theoretical contributions of [36] related to the implication

(1.3)⇒ (1.1) rely on prior knowledge on the decompositions of $\mathcal{A}$ and $\mathcal{B}$[1] and can be summarized as follows: if $N = 3$ and (1.3) holds with "$\subseteq$" replaced by "$=$", then $\mathcal{A}$ and $\mathcal{B}$ are generated by the same (possibly scaled) terms. The results obtained in the current paper imply that the prior knowledge on the decompositions is not needed. Further, [36] presents applications in the context of emitter movement detection and fluorescence data analysis.

The paper is organized as follows. In subsections 2.1 and 2.2 we introduce tensor related notations and formalize the problem statement, respectively. Section 3 contains preliminary results. In subsection 3.1, for the convenience of the reader, we remind the primary decomposition theorem and the Jordan canonical form. Subsection 3.2 contains an auxiliary result about the simultaneous compression of tensors $\mathcal{A}$ and $\mathcal{B}$ for which the first $\hat{N}$ inclusions in (1.3) hold (Lemma 3.2). The main results are given in section 4. In subsection 4.1 we establish connections between the terms in the decompositions of tensors $\mathcal{A}$ and $\mathcal{B}$ that satisfy the conditions in (1.3) (Theorems 4.1 and 4.3). In subsection 4.2 we show that the N conditions in (1.3) imply the $2^N - 2$ conditions in (1.2) (Corollary 4.5). In section 5 we illustrate how our results can be applied to solve a multi-label classification problem that appears in the context of blind source separation. Appendix A contains some numerical examples that illustrate a particular advantage of using the decomposition into a sum of ML rank-$(L_1, L_2, \cdot)$ terms over the CPD when we deal with the implication (1.3)⇒(1.1).

2. Basic definitions and problem statement.

2.1. Basic definitions.

Matrix representations. Let $1 \leq n \leq N$. *A mode-n matrix representation* of a tensor $\mathcal{A} \in \mathbb{F}^{I_1 \times \cdots \times I_N}$ is a matrix $\mathbf{A}_{(n^c;n)} \in \mathbb{F}^{\frac{I_1 \cdots I_N}{I_n} \times I_n}$ whose columns are the vectorized mode-n slices (see Figure 2.1 (top)) of $\mathcal{A}$. Using Matlab colon notation, the columns of $\mathbf{A}_{(n^c;n)}$ are the vectorized $I_1 \times \cdots \times I_{n-1} \times 1 \times I_{n+1} \times \cdots \times I_N$ tensors $\mathcal{A}(:, \ldots, :, 1, :, \ldots, :), \ldots, \mathcal{A}(:, \ldots, :, I_n, :, \ldots, :)$. Formally,

$$\text{(2.1)} \qquad \text{the } \Big(1 + \sum_{\substack{k=1 \\ k \neq n}}^{N} (i_k - 1) \prod_{\substack{l=1 \\ l \neq n}}^{k-1} I_l, i_n\Big)\text{th entry of } \mathbf{A}_{(n^c;n)} = \text{ the } (i_1, \ldots, i_N)\text{th entry of } \mathcal{A}.$$

For instance, the mode-1 matrix representation $\mathbf{A}_{(2,3;1)}$ of an $I_1 \times I_2 \times I_3$ tensor $\mathcal{A}$ is the $I_2 I_3 \times I_1$ matrix whose columns are the vectorized matrices $\mathbf{A}(1, :, :), \ldots, \mathbf{A}(I_1, :, :)$. It can also be verified that the rows of $\mathbf{A}_{(2,3;1)}$ are the transposed columns of $\mathcal{A}$, i.e., the transposed columns of $\mathbf{A}(:, 1, :), \ldots, \mathbf{A}(:, I_2, :)$ or $\mathbf{A}(:, :, 1), \ldots, \mathbf{A}(:, :, I_3)$ (see Figure 2.1 (top)).

Mode-n product. If for some tensor $\mathcal{D} \in \mathbb{F}^{I_1 \times \ldots I_{n-1} \times L_n \times I_{n+1} \times I_N}$ and matrix $\mathbf{X}^{(n)} \in \mathbb{F}^{I_n \times L_n}$,

$$\text{(2.2)} \qquad \mathbf{A}_{(n^c;n)} = \mathbf{D}_{(n^c;n)} \mathbf{X}^{(n)T},$$

[1] Namely, the working assumption in [36] is that both tensors $\mathcal{A}$ and $\mathcal{B}$ admit decompositions of the same type (CPD, decomposition in ML rank-$(L, L, 1)$ terms, decomposition in ML rank-$(L, L, \cdot)$ terms), that the decompositions include the same number of terms, and that in the latter two decomposition types the terms of $\mathcal{A}$ and $\mathcal{B}$ can be matched so that their ML ranks are equal.

i.e., if the mode-n fibers of $\mathcal{A}$ are obtained by multiplying the corresponding mode-n fibers of $\mathcal{D}$ by $\mathbf{X}^{(n)}$, then we say that $\mathcal{A}$ is *the mode-n product* of a $\mathcal{D}$ and $\mathbf{X}^{(n)}$ and write $\mathcal{A} = \mathcal{D} \bullet_n \mathbf{X}^{(n)}$. It can be easily verified that the remaining $N-1$ matrix representations of $\mathcal{A}$ can be factorized as

$$\mathbf{A}_{(k^c;k)} = \left(\bigotimes_{l=1,l\neq k}^{n-1} \mathbf{I}_{I_l} \otimes \mathbf{X}^{(n)} \otimes \bigotimes_{l=n+1,l\neq k}^{N} \mathbf{I}_{I_l} \right) \mathbf{D}_{(k^c;k)}, \qquad k \in \{1,\ldots,N\} \setminus \{n\}. \tag{2.3}$$

where $\mathbf{I}_{I_l}$ and "$\otimes$" denote the $I_l \times I_l$ identity matrix and the Kronecker product, respectively. Figure 2.1 (bottom) illustrates the mode-1 product of a third-order tensor and matrix.

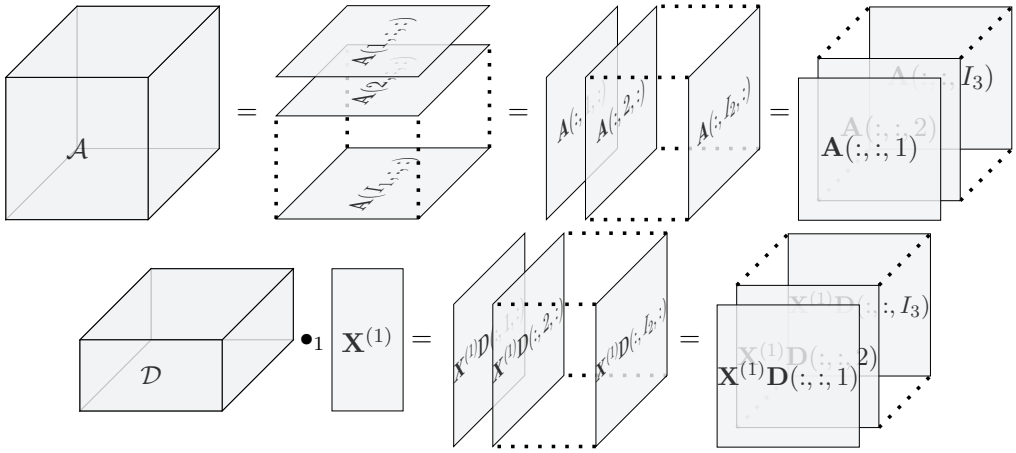


Figure 2.1. *Representations of an $I_1 \times I_2 \times I_3$ tensor $\mathcal{A}$ as a set of mode-n slices, $n = 1,2,3$ (top) and mode-1 product of a tensor $\mathcal{D}$ and matrix $\mathbf{X}^{(1)}$ (bottom). The columns of $\mathcal{D} \bullet_1 \mathbf{X}^{(1)}$ are obtained from the columns of $\mathcal{D}$ by multiplying them with $\mathbf{X}^{(1)}$.*

Several products in the same mode or across modes. It easily follows from (2.2) that for compatible matrix and tensor dimensions,

$$\left(\left(\left(\mathcal{D} \bullet_n \mathbf{X}_1^{(n)}\right) \bullet_n \mathbf{X}_2^{(n)}\right) \cdots \bullet_n \mathbf{X}_k^{(n)}\right) = \mathcal{D} \bullet_n \left(\mathbf{X}_k^{(n)} \cdots \mathbf{X}_1^{(n)}\right).$$

Let $\hat{N} \leq N$ and

$$\mathcal{D} \in \mathbb{F}^{L_1 \times \cdots \times L_{\hat{N}} \times I_{\hat{N}+1} \times \cdots \times I_N}, \qquad \mathbf{X}^{(1)} \in \mathbb{F}^{I_1 \times L_1}, \ldots, \mathbf{X}^{(\hat{N})} \in \mathbb{F}^{I_{\hat{N}} \times L_{\hat{N}}}. \tag{2.4}$$

For products across different modes, we have

$$\begin{aligned} \mathcal{D} \bullet_1 \mathbf{X}^{(1)} \cdots \bullet_N \mathbf{X}^{(N)} := \left(\left(\left(\mathcal{D} \bullet_1 \mathbf{X}^{(1)}\right) \bullet_2 \mathbf{X}^{(2)}\right) \cdots \bullet_N \mathbf{X}^{(N)}\right) = \\ \left(\left(\left(\mathcal{D} \bullet_{i_1} \mathbf{X}^{(i_1)}\right) \bullet_{i_2} \mathbf{X}^{(i_2)}\right) \cdots \bullet_{i_N} \mathbf{X}^{(i_N)}\right) \end{aligned} \tag{2.5}$$

for any permutation $i_1, \ldots, i_N$ of $1, \ldots, N$. It follows from (2.2), (2.3), and (2.5), that the matrix representations of $\mathcal{A} = \mathcal{D} \bullet_1 \mathbf{X}^{(1)} \cdots \bullet_N \mathbf{X}^{(N)}$ are given by

$$\mathbf{A}_{(n^c;n)} = \left(\bigotimes_{k=1,k\neq n}^{N} \mathbf{X}^{(k)} \right) \mathbf{D}_{(n^c;n)} \mathbf{X}^{(n)T}, \qquad n \in \{1, \ldots, N\}. \tag{2.6}$$

If $\mathcal{A} = \mathcal{D} \bullet_1 \mathbf{X}^{(1)} \cdots \bullet_{\hat{N}} \mathbf{X}^{(\hat{N})}$ with $\hat{N} < N$, then the identities in (2.6) hold with $\mathbf{X}^{(\hat{N}+1)} = \mathbf{I}_{I_{\hat{N}+1}}, \ldots, \mathbf{X}^{(N)} = \mathbf{I}_{I_N}$. That is,

$$\mathbf{A}_{(n^c;n)} = \left(\bigotimes_{k=1,k\neq n}^{\hat{N}} \mathbf{X}^{(k)} \otimes \bigotimes_{k=\hat{N}+1}^{N} \mathbf{I}_{I_k} \right) \mathbf{D}_{(n^c;n)} \mathbf{X}^{(n)T}, \qquad n \in \{1, \ldots, \hat{N}\}, \tag{2.7}$$

$$\mathbf{A}_{(n^c;n)} = \left(\bigotimes_{k=1}^{\hat{N}} \mathbf{X}^{(k)} \otimes \bigotimes_{k=\hat{N}+1,k\neq n}^{N} \mathbf{I}_{I_k} \right) \mathbf{D}_{(n^c;n)}, \qquad n \in \{\hat{N}+1, \ldots, N\}. \tag{2.8}$$

ML rank of a tensor. By definition,

$$\mathcal{A} \text{ is ML rank-}(L_1, \ldots, L_{\hat{N}}, \cdot, \ldots, \cdot) \overset{\text{def}}{\Longleftrightarrow} r_{\mathbf{A}_{(n^c;n)}} = L_n, \quad n \in \{1, \ldots, \hat{N}\}, \quad 2 \leq \hat{N} \leq N,$$

that is, L_n is the dimension of the subspace spanned by the mode-n fibers of $\mathcal{A}$. It can be shown that $\mathcal{A}$ is ML rank-$(L_1, \ldots, L_{\hat{N}}, \cdot, \ldots, \cdot)$ if and only if it admits the factorization $\mathcal{A} = \mathcal{D} \bullet_1 \mathbf{X}^{(1)} \cdots \bullet_{\hat{N}} \mathbf{X}^{(\hat{N})}$ such that $\mathcal{D}$, $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(\hat{N})}$ have dimensions as in (2.4) and $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(\hat{N})}, \mathbf{D}_{(1^c;1)}, \ldots, \mathbf{D}_{(\hat{N}^c;\hat{N})}$ have full column rank. In this paper we assume that the tensor dimensions have been permuted so that we can just specify the rank values for the first $\hat{N}$ matrix representations of $\mathcal{A}$. A special case of the factorization $\mathcal{A} = \mathcal{D} \bullet_1 \mathbf{X}^{(1)} \cdots \bullet_{\hat{N}} \mathbf{X}^{(\hat{N})}$, where $\hat{N} = N$, $\mathbf{X}^{(n)}$ equals the "U" factor in the compact Singular Value Decomposition (SVD) of $\mathbf{A}_{(n^c;n)}$, and $\mathcal{D} = \mathcal{A} \bullet_1 \mathbf{X}^{(1)H} \cdots \bullet_N \mathbf{X}^{(N)H}$ is known as the MLSVD of $\mathcal{A}$ and is used for the compression of an $I_1 \times \cdots \times I_N$ tensor to the size $L_1 \times \cdots \times L_N$ [17]. By setting $\mathbf{X}^{(n)}$ equal to the identity matrix for $n = \hat{N}+1, \ldots, N$, we compress only along the first $\hat{N}$ dimensions.

ML rank-$(L_{1r}, \ldots, L_{\hat{N}r}, \cdot, \ldots, \cdot)$ decomposition of a tensor. In this paper we consider the decomposition of $\mathcal{A}$ into a sum of ML rank-$(L_{1r}, \ldots, L_{\hat{N}r}, \cdot, \ldots, \cdot)$ terms:

$$\mathcal{A} = \sum_{r=1}^{R} \mathcal{D}_r \bullet_1 \mathbf{X}_r^{(1)} \cdots \bullet_{\hat{N}} \mathbf{X}_r^{(\hat{N})}, \quad 2 \leq \hat{N} \leq N, \tag{2.9}$$

$$\mathcal{D}_r \in \mathbb{F}^{L_{1r} \times \cdots \times L_{\hat{N}r} \times I_{\hat{N}_r+1} \times \cdots \times I_N}, \quad \mathbf{X}_r^{(n)} \in \mathbb{F}^{I_n \times L_{nr}}, \quad n \in \{1, \ldots, \hat{N}\}, \quad r \in \{1, \ldots, R\}.$$

In our derivation we will also use a matricized version of (2.9). It can be obtained as follows. First, we call

$$\mathbf{X}^{(n)} := [\mathbf{X}_1^{(n)} \ \ldots \ \mathbf{X}_R^{(n)}] \in \mathbb{F}^{I_n \times \sum_{r=1}^{R} L_{nr}}, \qquad n \in \{1, \ldots, \hat{N}\}, \tag{2.10}$$

the *concatenated factor matrices of* $\mathcal{A}$. If further we set

$$\mathbf{X}^{(n)} := [\mathbf{I}_{I_n} \ \dots \ \mathbf{I}_{I_n}] \in \mathbb{F}^{I_n \times RI_n}, \qquad n \in \{\hat{N}+1, \dots, N\}, \tag{2.11}$$

then, by (2.6), we can express (2.9) in a matricized way as

$$\begin{aligned} \mathbf{A}_{(n^c;n)} = \sum_{r=1}^{R} \left(\bigotimes_{l=1, l\neq n}^{N} \mathbf{X}_r^{(l)} \right) \mathbf{D}_{r(n^c;n)} \mathbf{X}_r^{(n)T} = \\ \left(\bigodot_{l=1, l\neq n}^{N} \mathbf{X}^{(l)} \right) \mathrm{Bdiag}(\mathbf{D}_{1(n^c;n)}, \dots, \mathbf{D}_{R(n^c;n)}) \mathbf{X}^{(n)T}, \qquad n \in \{1, \dots, N\}, \end{aligned} \tag{2.12}$$

where

$$\bigodot_{l=1, l\neq n}^{N} \mathbf{X}^{(l)} := \left[\bigotimes_{l=1, l\neq n}^{N} \mathbf{X}_1^{(l)} \ \dots \ \bigotimes_{l=1, l\neq n}^{N} \mathbf{X}_R^{(l)} \right] \tag{2.13}$$

and $\mathrm{Bdiag}(\mathbf{D}_{1(n^c;n)}, \dots, \mathbf{D}_{R(n^c;n)})$ denotes a block-diagonal matrix with the matrices $\mathbf{D}_{1(n^c;n)}$, $\dots$, $\mathbf{D}_{R(n^c;n)}$ on the diagonal.

Note that (2.9) captures several well-studied decompositions as special cases (see also the introduction). If $\hat{N} = N$ and $L_{1r} = \cdots = L_{Nr} = 1$ for all r, then all terms in (2.9) are rank-1 tensors, so (2.9) reduces to a polyadic decomposition of $\mathcal{A}$. It can easily be verified that if $\hat{N} = 2$, $N = 3$, and $L_{1r} = 1$ for all r, then the ML rank-$(1, L_{2r}, \cdot)$ terms in (2.9) are actually ML rank-$(1, L_{2r}, L_{2r})$ terms.[2] Thus, (2.9) reduces to the decomposition into a sum of ML rank-$(1, L_{2r}, L_{2r})$ terms. Finally, if $\hat{N} = 2$ and $N = 3$, then (2.9) is a tensor reformulation of the joint block diagonalization problem. Namely, (2.9) means that the frontal slices of $\mathcal{A}$ can simultaneously be factorized as

$$\mathbf{A}(:,:,i) = \mathbf{X}^{(1)} \, \mathrm{Bdiag}(\mathbf{D}_1(:,:,i), \dots, \mathbf{D}_R(:,:,i)) \mathbf{X}^{(2)T}, \quad i = 1, \dots, I_3,$$

where $\mathbf{D}_r(:,:,i) \in \mathbb{F}^{L_{1r} \times L_{2r}}$.

2.2. Problem statement. Assume that a tensor $\mathcal{B} \in \mathbb{F}^{I_1 \times \cdots \times I_N}$ consists of the same ML rank-$(L_{1r}, \dots, L_{\hat{N}r}, \cdot, \dots, \cdot)$ terms as $\mathcal{A}$, but possibly differently scaled:

$$\mathcal{B} = \sum_{r=1}^{R} \lambda_r \mathcal{D}_r \bullet_1 \mathbf{X}_r^{(1)} \cdots \bullet_{\hat{N}} \mathbf{X}_r^{(\hat{N})}, \qquad \lambda_1 \cdots \lambda_R \neq 0. \tag{2.14}$$

[2]Indeed, since the column rank is equal to 1, a ML rank-$(1, L_{2r}, \cdot)$ $(I_1 \times I_2 \times I_3)$ term consists of scaled versions of the same $(I_2 \times I_3)$ matrix. Since column rank and row rank of the latter matrix coincide, the ML rank-$(1, L_{2r}, \cdot)$ term is necessary ML rank-$(1, L_{2r}, L_{2r})$.

Then by (2.12),

$$\mathbf{B}_{(n^c;n)} = \left(\bigodot_{k=1,k\neq n}^{N} \mathbf{X}^{(k)}\right) \mathrm{Bdiag}(\lambda_1 \mathbf{D}_{1(n^c;n)}, \ldots, \lambda_R \mathbf{D}_{R(n^c;n)}) \mathbf{X}^{(n)T} = \left(\bigodot_{k=1,k\neq n}^{N} \mathbf{X}^{(k)}\right) \mathrm{Bdiag}(\mathbf{D}_{1(n^c;n)}, \ldots, \mathbf{D}_{R(n^c;n)}) \, \mathrm{Bdiag}(\lambda_1 \mathbf{I}_{L_{n1}}, \ldots, \lambda_R \mathbf{I}_{L_{nR}}) \mathbf{X}^{(n)T}. \tag{2.15}$$

Assume that $\hat{N} \geq 2$ and that the matrices

$$\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(\hat{N})} \text{ have full column rank.} \tag{2.16}$$

It can be easily shown[3] that the matrices in (2.13) have full column rank for all n. Hence, by (2.12) and (2.15), the column spaces of the first $\hat{N}$ matrix representations of $\mathcal{A}$ and $\mathcal{B}$ coincide:

$$\mathrm{col}(\mathbf{A}_{(n^c;n)}) = \mathrm{col}(\mathbf{B}_{(n^c;n)}), \qquad n \in \{1, \ldots, \hat{N}\}. \tag{2.17}$$

If we further limit[4] ourselves to the case where the matrices

$$\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(\hat{N})} \text{ are square and nonsingular,} \tag{2.18}$$

then, obviously,

$$\mathbf{B}_{(n^c;n)} = \mathbf{A}_{(n^c;n)} \mathbf{M}_n, \qquad n \in \{1, \ldots, \hat{N}\}, \tag{2.19}$$

where

$$\mathbf{M}_n = \left(\mathbf{X}^{(n)T}\right)^{-1} \mathrm{Bdiag}(\lambda_1 \mathbf{I}_{L_{n1}}, \ldots, \lambda_R \mathbf{I}_{L_{nR}}) \mathbf{X}^{(n)T}, \qquad n \in \{1, \ldots, \hat{N}\}. \tag{2.20}$$

Thus, if (2.9), (2.14), and (2.18) hold, then the column spaces of the first $\hat{N}$ matrix representations of $\mathcal{A}$ and $\mathcal{B}$ coincide, the matrices $\mathbf{M}_n := \mathbf{A}^{\dagger}_{(n^c;n)} \mathbf{B}_{(n^c;n)}$ have the same spectrum $\lambda_1, \ldots, \lambda_R \in \mathbb{F}$ and can be diagonalized, $n = 1, \ldots, \hat{N}$. Moreover, the concatenated factor matrices $\mathbf{X}^{(n)}$ and the sizes of blocks L_{nr} (and hence the overall decompositions of $\mathcal{A}$ and $\mathcal{B}$) can be recovered from the EVDs of $\mathbf{M}_1, \ldots, \mathbf{M}_{\hat{N}}$.

In this paper we consider the inverse problem: we assume that the column spaces of the first $\hat{N}$ matrix representations of $\mathcal{A}$ and $\mathcal{B}$ coincide and we investigate how the ML rank decompositions $\mathcal{A}$ and $\mathcal{B}$ relate to each other. (A version of Theorem 2.1 in which (2.18) and (2.19) hold for $\hat{N}$ values arbitrary chosen from $\{1, \ldots, N\}$ can be obtained by permuting the tensor dimensions.) In particular, we obtain the following result.

[3] Indeed, the result holds since, by assumption (2.16), the first $\hat{N} - 1$ factors $\mathbf{X}^{(l)}$ have full column rank and, by construction, the remaining factors do not have zero columns.

[4] Lemma 3.2 below implies that assumption (2.16) can always be replaced by assumption (2.18). Computationally, this can be done by Multilinear Singular Value Decomposition (MLSVD) [17, 34, 35].

Theorem 2.1. *Let $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ and $2 \leq \hat{N} \leq N$. Assume that $\mathbf{A}_{(n^c;n)}$ and $\mathbf{B}_{(n^c;n)}$ have full column rank for each $n \in \{1 \ldots, \hat{N}\}$, that (2.19) holds and that at least one of the matrices $\mathbf{M}_1, \ldots, \mathbf{M}_{\hat{N}}$ can be diagonalized.*[5] *Then the following statements hold.*

1. *The matrices $\mathbf{M}_1, \ldots, \mathbf{M}_{\hat{N}}$ have the same spectrum.*
2. *All matrices $\mathbf{M}_1, \ldots, \mathbf{M}_{\hat{N}}$ can be diagonalized.*
3. *Let the distinct eigenvalues of $\mathbf{M}_n$ be $\lambda_1, \ldots, \lambda_R$ with respective multiplicities $L_{n1}, \ldots, L_{nR}$ and let $\mathbf{X}_n \in \mathbb{C}^{I_n \times I_n}$ be a nonsingular matrix such that (2.20) holds. Then $\mathcal{A}$ and $\mathcal{B}$ admit the ML rank-$(L_{1r}, \ldots, L_{\hat{N}r}, \cdot, \ldots, \cdot)$ decompositions in (2.9) and (2.14), respectively. In particular, if $L_{nr} = 1$ for all n and r, then $\mathcal{A}$ and $\mathcal{B}$ are generated by the same (possibly scaled) R rank-1 terms.*

Proof. The proof follows from Theorem 4.3 below. ∎

The theorem can be used as follows. First, the matrices $\mathbf{M}_1, \ldots, \mathbf{M}_{\hat{N}}$ are found from the sets of linear equations (2.19). (If any of the sets of linear equations does not have a solution, then $\mathcal{B}$ is not of the form (2.14), i.e., it cannot be generated by terms from the decomposition of $\mathcal{A}$.) The number of terms R is found as the number of distinct eigenvalues of $\mathbf{M}_n$, $1 \leq n \leq \hat{N}$. The distinct eigenvalues themselves correspond to the scaling factors λ_r in (2.14). Both R and the eigenvalues λ_r are necessarily the same for all $\mathbf{M}_n$, but the multiplicities can be different. The multiplicity of λ_r in the EVD of $\mathbf{M}_n$ corresponds to the nth entry L_{nr} in the ML rank of the rth term, so that to apply the theorem we should necessarily have that $L_{n1} + \cdots + L_{nR} = I_n$ for $1 \leq n \leq \hat{N}$. (Recall from Footnote 4 on p. 8 that this means that for the given tensors we should have $L_{n1} + \cdots + L_{nR} \leq I_n$ for $1 \leq n \leq \hat{N}$.) The larger $\hat{N}$, the more the terms are specified. The minimal value for $\hat{N}$ is 2, since a decomposition in ML rank-$(L_{1r}, \cdot, \ldots, \cdot)$ terms is meaningless.

So far, we have explained the use of the theorem for decompositions that are exact. Obviously, the theorem also suggests a procedure for approximate decompositions (of noisy tensors). The equations in (2.19) may be solved in least squares sense. The eigenvalues λ_{nr} of the matrices $\mathbf{M}_1, \ldots, \mathbf{M}_{\hat{N}}$ may be averaged over n to obtain estimates of λ_r. The values L_{nr}, $1 \leq n \leq \hat{N}$, $1 \leq r \leq R$ may be estimated by assessing how close the eigenvalues λ_{rn} are to the averaged values λ_r.

3. Preliminaries.

3.1. Primary decomposition theorem and the Jordan canonical form.

In this subsection we recall known results that will be used in section 4. Recall that the minimal polynomial $q(x)$ of a matrix $\mathbf{M} \in \mathbb{F}^{I \times I}$ is the polynomial of least degree over $\mathbb{F}$ whose leading coefficient is 1 and such that $q(\mathbf{M}) = \mathbf{O}$. It is well known that the minimal polynomial does not depend of $\mathbb{F}$, is unique, and that the set of its zeros coincides with the set of the eigenvalues of the matrix (in the case $\mathbb{F} = \mathbb{R}$ both sets can be empty, namely, when the minimal polynomial does not have real roots). Recall also that a non-constant polynomial is irreducible over $\mathbb{F}$ if its coefficients belong to $\mathbb{F}$ and it cannot be factorized into the product of two non-constant

[5] The assumption on diagonalization will later be relaxed by using the Jordan canonical form in Theorem 4.3.

polynomials with coefficients in $\mathbb{F}$. For instance, the minimal polynomials of the matrices

$$\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \text{ and } \mathbf{I}_I$$

are $x^2 - x$, $x^2 - 1$, $x^2 + 1$, x^2, and $x - 1$, respectively. The matrix $\mathbf{I}_I$ has a single eigenvalue 1 of multiplicity I which corresponds to a single root of $x - 1$ of multiplicity 1. The polynomial $x^2 + 1$ is irreducible over $\mathbb{R}$ and is reducible over $\mathbb{C}$, $x^2 + 1 = (x + i)(x - i)$, which agrees with the fact that the matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ does not have eigenvalues over $\mathbb{R}$ but has two eigenvalues $-i$ and i over $\mathbb{C}$. It is well known that any polynomial with leading coefficient 1 can be factorized as

$$q(x) = p_1(x)^{\mu_1} \cdots p_R(x)^{\mu_R}$$

where p_r are distinct irreducible polynomials and $\mu_r \geq 1$. Since in this paper $\mathbb{F}$ is either $\mathbb{C}$ or $\mathbb{R}$, we have that

$$p_1, \ldots, p_R \in \{x - \lambda : \ \lambda \in \mathbb{C}\}, \text{ if } \mathbb{F} = \mathbb{C},$$
$$p_1, \ldots, p_R \in \{x - \lambda : \ \lambda \in \mathbb{R}\} \cup \{x^2 + 2ax + a^2 + b^2 : \ a, b \in \mathbb{R} \text{ and } b > 0\}, \text{ if } \mathbb{F} = \mathbb{R}.$$

The following theorem implies that the minimal polynomial of a matrix can be used to construct a basis in which that matrix has block-diagonal form.

Theorem 3.1 (Primary decomposition theorem [12, pp.196–197]). *Let $\mathbf{M} \in \mathbb{F}^{I \times I}$ and let*

$$q(x) = p_1(x)^{\mu_1} \cdots p_R(x)^{\mu_R}$$

be the minimal polynomial of $\mathbf{M}$, factorized into powers of distinct polynomials $p_r(x)$ that are irreducible (over $\mathbb{F}$). Then the subspaces

$$E_r := \operatorname{Null}\left(p_r(\mathbf{M})^{\mu_r}\right), \quad 1 \leq r \leq R$$

are invariant for $\mathbf{M}$, i.e., $\mathbf{M}E_r \subseteq E_r$ and we have

$$\mathbb{F}^I = E_1 \oplus \cdots \oplus E_R, \tag{3.1}$$

where "$\oplus$" denotes the direct sum of subspaces.

Decomposition (3.1) in Theorem 3.1 implies that the matrix $\mathbf{M}$ is similar to a block-diagonal matrix. Indeed, let $L_r = \dim E_r$ and let the columns of $\mathbf{S}_r \in \mathbb{F}^{I \times L_r}$ form a basis of E_r, $r = 1, \ldots, R$. Then by (3.1), the columns of $\mathbf{S} := [\mathbf{S}_1 \ \ldots \ \mathbf{S}_R]$ form a basis of the entire space $\mathbb{F}^I$, implying that $\mathbf{S}$ is nonsingular. Since $\mathbf{M}E_r \subseteq E_r$ it follows that there exists a unique matrix $\mathbf{T}_r \in \mathbb{F}^{L_r \times L_r}$ such that $\mathbf{M}\mathbf{S}_r = \mathbf{S}_r\mathbf{T}_r$, $r = 1, \ldots, R$. Hence $\mathbf{M}[\mathbf{S}_1 \ \ldots \ \mathbf{S}_R] = [\mathbf{S}_1\mathbf{T}_1 \ \ldots \ \mathbf{S}_R\mathbf{T}_R]$ or

$$\mathbf{M} = \mathbf{S} \operatorname{Bdiag}(\mathbf{T}_1, \ldots, \mathbf{T}_R)\mathbf{S}^{-1}, \quad \mathbf{S} = [\mathbf{S}_1 \ \ldots \ \mathbf{S}_R], \quad \mathbf{S}_r \in \mathbb{F}^{I \times L_r}.$$

It is well-known that each of the matrices $\mathbf{T}_r$ can further be reduced to Jordan canonical form by a similarity transform. Namely, if $p_r(x)^{\mu_r} = (x - \lambda)^{\mu_r}$ with $\lambda \in \mathbb{F}$, then $\mathbf{T}_r$ is similar to

$J(\lambda, n_{r1}) \oplus \cdots \oplus J(\lambda, n_{rk_r})$, where $J(\lambda, n)$ denotes the $n \times n$ Jordan block with λ on the main diagonal:

$$\begin{bmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda \end{bmatrix}.$$

If $\mathbb{F} = \mathbb{R}$ and $p_r(x)^{\mu_r} = (x^2 + 2ax + a^2 + b^2)^{\mu_r}$ with $a, b \in \mathbb{R}$ and $b > 0$, then $\mathbf{T}_r$ is similar to $C(a, b, n_{r1}) \oplus \cdots \oplus C(a, b, n_{rk_r})$, where $C(a, b, n)$ denotes the $2n \times 2n$ block matrix of the form

$$\begin{bmatrix} C(a,b) & \mathbf{I}_2 & 0 & \dots & 0 \\ 0 & C(a,b) & \mathbf{I}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{I}_2 \\ 0 & 0 & 0 & \dots & C(a,b) \end{bmatrix}, \qquad C(a,b) = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

It is known that the values $n_{r_1}, \dots, n_{rk_r}$ are uniquely determined by $\mathbf{T}_r$ up to permutation, in particular, $\max(n_{r_1}, \dots, n_{rk_r}) = \mu_r$. Thus, the Jordan canonical form is unique up to permutation of its blocks. For more details on the Jordan canonical form we refer to [27, Chapter 3].

3.2. An auxiliary result about simultaneous compression of a pair of tensors. Let $\mathcal{A}, \mathcal{B} \in \mathbb{F}^{I_1 \times \cdots \times I_N}$. It is clear that the conditions

$$\operatorname{col}(\mathbf{B}_{(n^c;n)}) \subseteq \operatorname{col}(\mathbf{A}_{(n^c;n)}), \quad n \in \{1, \dots, \hat{N}\}. \tag{3.2}$$

can be rewritten as

$$\mathbf{B}_{(n^c;n)} = \mathbf{A}_{(n^c;n)} \mathbf{M}_n, \qquad n \in \{1, \dots, \hat{N}\}, \tag{3.3}$$

in which $\mathbf{M}_n \in \mathbb{F}^{I_n \times I_n}$ is not necessarily unique. The goal of the following lemma is to show that (3.3) can further be reduced to the case where the matrices $\mathbf{A}_{(n^c;n)}$ do have full column rank, so $\mathbf{M}_n$ can be uniquely recovered as $\mathbf{M}_n = \mathbf{A}^{\dagger}_{(n^c;n)} \mathbf{B}_{(n^c;n)}$. In subsection 4.1 we will use $\mathbf{M}_1, \dots, \mathbf{M}_{\hat{N}}$ to establish connections between the terms in the decompositions of $\mathcal{A}$ and $\mathcal{B}$.

Lemma 3.2. *Let $\tilde{\mathcal{A}}, \tilde{\mathcal{B}} \in \mathbb{F}^{\tilde{I}_1 \times \cdots \times \tilde{I}_N}$, $N \geq \hat{N} \geq 2$ and let $\tilde{\mathcal{A}}$ be ML rank-$(I_1, \dots, I_{\hat{N}}, \cdot, \dots, \cdot)$. Assume that*

$$\operatorname{col}(\tilde{\mathbf{B}}_{(n^c;n)}) \subseteq \operatorname{col}(\tilde{\mathbf{A}}_{(n^c;n)}), \quad n \in \{1, \dots, \hat{N}\}. \tag{3.4}$$

Let also the rows of $\mathbf{U}_n \in \mathbb{F}^{I_n \times \tilde{I}_n}$ form an orthonormal basis of the row space of $\tilde{\mathbf{A}}_{(n^c;n)}$, $n \in \{1, \dots, \hat{N}\}$[6] and

$$\mathcal{A} := \tilde{\mathcal{A}} \bullet_1 \mathbf{U}_1^* \cdots \bullet_{\hat{N}} \mathbf{U}_{\hat{N}}^*, \qquad \mathcal{B} := \tilde{\mathcal{B}} \bullet_1 \mathbf{U}_1^* \cdots \bullet_{\hat{N}} \mathbf{U}_{\hat{N}}^*. \tag{3.5}$$

Then the following statements hold.

[6]For instance, one can take $\mathbf{U}_n$ equal to the transpose of the "U" factor in the compact SVD of $\tilde{\mathbf{A}}^T_{(n^c;n)}$. In this case, (3.6) implements a standard compression by multilinear singular value decomposition [17, 34, 35], in which the compression matrices are obtained from $\mathcal{A}$.

1. *For all* $k \in \{1,\dots,N\}$, *the row space of* $\tilde{\mathbf{A}}_{(k^c;k)}$ *contains the row space of* $\tilde{\mathbf{B}}_{(k^c;k)}$.
2. $\tilde{\mathcal{A}}$ *and* $\tilde{\mathcal{B}}$ *can be recovered from* $\mathcal{A}$ *and* $\mathcal{B}$, *respectively, as*

$$\tilde{\mathcal{A}} = \mathcal{A} \bullet_1 \mathbf{U}_1^T \cdots \bullet_{\hat{N}} \mathbf{U}_{\hat{N}}^T, \qquad \tilde{\mathcal{B}} = \mathcal{B} \bullet_1 \mathbf{U}_1^T \cdots \bullet_{\hat{N}} \mathbf{U}_{\hat{N}}^T. \tag{3.6}$$

3. $\mathcal{A}, \mathcal{B} \in \mathbb{F}^{I_1 \times \cdots \times I_{\hat{N}} \times \tilde{I}_{\hat{N}+1} \times \cdots \times \tilde{I}_N}$, $\mathcal{A}$ *is ML rank-*$(I_1,\dots,I_{\hat{N}},\cdot,\dots,\cdot)$, *and the ML rank of* $\mathcal{B}$ *equals the ML rank of* $\tilde{\mathcal{B}}$.

Proof. 1. Recall that (2.2) is equivalent to any identity in (2.3). Hence if (2.2) holds for $n = 1$ and $n = 2$, then, by (2.3), the row space of $\mathbf{D}_{(k^c;k)}$ contains the row space of $\mathbf{A}_{(k^c;k)}$ for $k \in \{2,\dots,N\}$ and for $k \in \{1,3,\dots,N\}$, respectively, i.e., for all k. To complete the proof one should replace $\mathcal{D}$ and $\mathcal{A}$ in (2.2) and (2.3) by $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$, respectively.

2. Since the rows of $\mathbf{U}_n$ form an orthonormal basis of the row space of $\tilde{\mathbf{A}}_{(n^c;n)}$, it follows that $\tilde{\mathbf{A}}_{(n^c;n)}\mathbf{U}_n^H\mathbf{U}_n = \tilde{\mathbf{A}}_{(n^c;n)}$ or $\tilde{\mathcal{A}} \bullet_n (\mathbf{U}_n^T\mathbf{U}_n^*) = \tilde{\mathcal{A}}$, $n \in \{1,\dots,\hat{N}\}$. Hence

$$\mathcal{A} \bullet_1 \mathbf{U}_1^T \cdots \bullet_{\hat{N}} \mathbf{U}_{\hat{N}}^T = (\tilde{\mathcal{A}} \bullet_1 \mathbf{U}_1^* \cdots \bullet_{\hat{N}} \mathbf{U}_{\hat{N}}^*) \bullet_1 \mathbf{U}_1^T \cdots \bullet_{\hat{N}} \mathbf{U}_{\hat{N}}^T = \tilde{\mathcal{A}} \bullet_1 (\mathbf{U}_1^T\mathbf{U}_1^*) \cdots \bullet_{\hat{N}} (\mathbf{U}_{\hat{N}}^T\mathbf{U}_{\hat{N}}^*) = \tilde{\mathcal{A}}.$$

By statement 1, the identity for $\tilde{\mathcal{B}}$ can be proved in a similar way.

3. From (2.2), (3.5), and (3.6) it follows that

$$r_{\mathbf{A}_{(n^c;n)}} \le r_{\tilde{\mathbf{A}}_{(n^c;n)}} \le r_{\mathbf{A}_{(n^c;n)}}, \quad r_{\mathbf{B}_{(n^c;n)}} \le r_{\tilde{\mathbf{B}}_{(n^c;n)}} \le r_{\mathbf{B}_{(n^c;n)}}, \quad n = 1,\dots,\hat{N}$$

implying that $r_{\mathbf{A}_{(n^c;n)}} = r_{\tilde{\mathbf{A}}_{(n^c;n)}} = I_n$ and $r_{\mathbf{B}_{(n^c;n)}} = r_{\tilde{\mathbf{B}}_{(n^c;n)}}$ for $n = 1,\dots,\hat{N}$. ∎

4. Main results.

4.1. Connections between tensors $\mathcal{A}$ and $\mathcal{B}$ that satisfy the first $\hat{N}$ conditions in (1.3).

To simplify the presentation throughout this subsection we assume that the first $\hat{N}$ matrix representations of $\mathcal{A}$ have full column rank. The general case follows from Lemma 3.2 above. Also, to keep the presentation and derivation of results easy to follow, we first consider the particular case where $\mathcal{A}$ and $\mathcal{B}$ are third-order tensors (i.e., $N = 3$) that satisfy only the first two conditions (i.e., $\hat{N} = 2$) in

$$\operatorname{col}(\mathbf{B}_{(2,3;1)}) \subseteq \operatorname{col}(\mathbf{A}_{(2,3;1)}),\ \operatorname{col}(\mathbf{B}_{(1,3;2)}) \subseteq \operatorname{col}(\mathbf{A}_{(1,3;2)}),\ \operatorname{col}(\mathbf{B}_{(1,2;3)}) \subseteq \operatorname{col}(\mathbf{A}_{(1,2;3)}). \tag{4.1}$$

The case where all three conditions in (4.1) hold (i.e., $N = \hat{N} = 3$) and the general case $N \ge 3$, $N \ge \hat{N} \ge 2$ will be covered by Theorem 4.3 below.

It is worth noting that the following theorem not only presents conditions that guarantee that $\mathcal{A}$ and $\mathcal{B}$ are generated by the same (possibly scaled) terms but also implies a procedure to compute the number of terms R and their ML rank values (see similar discussion after Theorem 2.1). To apply the theorem we should necessarily have that $L_{n1} + \cdots + L_{nR} = I_n$ for $n = 1, 2$.

Theorem 4.1. *Let tensors* $\mathcal{A}, \mathcal{B} \in \mathbb{F}^{I_1 \times I_2 \times I_3}$. *Assume that*

$$\mathbf{A}_{(2,3;1)} \text{ and } \mathbf{A}_{(1,3;2)} \text{ have full column rank} \tag{4.2}$$

and that there exist matrices $\mathbf{M}_1 \in \mathbb{F}^{I_1 \times I_1}$ *and* $\mathbf{M}_2 \in \mathbb{F}^{I_2 \times I_2}$ *such that*

$$\mathbf{B}_{(2,3;1)} = \mathbf{A}_{(2,3;1)}\mathbf{M}_1 \quad \text{and} \quad \mathbf{B}_{(1,3;2)} = \mathbf{A}_{(1,3;2)}\mathbf{M}_2. \tag{4.3}$$

Then the following statements hold.

1. *The matrices* $\mathbf{M}_1$ *and* $\mathbf{M}_2$ *have the same minimal polynomial* $q(x)$.
2. *Consider the factorization* $q(x) = p_1(x)^{\mu_1} \cdots p_R(x)^{\mu_R}$ *with distinct polynomials* $p_r(x)$ *that are irreducible (over* $\mathbb{F}$*) and set*

$$L_{1r} := \dim(\mathrm{Null}\,(p_r(\mathbf{M}_1)^{\mu_r})),\ L_{2r} := \dim(\mathrm{Null}\,(p_r(\mathbf{M}_2)^{\mu_r})) \quad 1 \le r \le R.$$

Let also

$$\mathbf{M}_1 = \mathbf{S}_1\,\mathrm{Bdiag}(\mathbf{T}_{11}, \ldots, \mathbf{T}_{1R})\mathbf{S}_1^{-1}, \quad \mathbf{S}_1 = [\mathbf{S}_{11}\ \ldots\ \mathbf{S}_{1R}], \quad \mathbf{S}_{1r} \in \mathbb{F}^{I_1 \times L_{1r}}, \tag{4.4}$$

$$\mathbf{M}_2 = \mathbf{S}_2\,\mathrm{Bdiag}(\mathbf{T}_{21}, \ldots, \mathbf{T}_{2R})\mathbf{S}_2^{-1}, \quad \mathbf{S}_2 = [\mathbf{S}_{21}\ \ldots\ \mathbf{S}_{2R}], \quad \mathbf{S}_{2r} \in \mathbb{F}^{I_2 \times L_{2r}} \tag{4.5}$$

be the primary decompositions of $\mathbf{M}_1$ *and* $\mathbf{M}_2$*, respectively, such that the minimal polynomials of* $\mathbf{T}_{1r}$ *and* $\mathbf{T}_{2r}$ *are equal to* $p_r(x)^{\mu_r}$ *for each* $r = 1, \ldots, R$*. Then the matrices*

$$\mathbf{D}_i := \mathbf{S}_1^T \mathbf{A}_i \mathbf{S}_2, \qquad \mathbf{A}_i := \mathbf{A}(:,:,i), \qquad i = 1, \ldots, I_3 \tag{4.6}$$

are block-diagonal, $\mathbf{D}_i = \mathrm{Bdiag}(\mathbf{D}_{i,11}, \ldots, \mathbf{D}_{i,RR})$*,* $\mathbf{D}_{i,rr} \in \mathbb{F}^{L_{1r} \times L_{2r}}$ *and*

$$\mathbf{T}_{1r}^T \mathbf{D}_{i,rr} = \mathbf{D}_{i,rr}\mathbf{T}_{2r}, \qquad i = 1, \ldots, I_3, \quad r = 1, \ldots, R. \tag{4.7}$$

3. *Let* $\mathcal{D}_r \in \mathbb{F}^{L_{1r} \times L_{2r} \times I_3}$ *denote a tensor with frontal slices* $\mathbf{D}_{1,rr}, \ldots, \mathbf{D}_{I_3,rr} \in \mathbb{F}^{L_{1r} \times L_{2r}}$ *and let*

$$\mathbf{S}_1^{-T} =: \mathbf{X}^{(1)} = [\mathbf{X}_1^{(1)}\ \ldots\ \mathbf{X}_R^{(1)}], \quad \mathbf{X}_r^{(1)} \in \mathbb{F}^{I_1 \times L_{1r}},$$
$$\mathbf{S}_2^{-T} =: \mathbf{X}^{(2)} = [\mathbf{X}_1^{(2)}\ \ldots\ \mathbf{X}_R^{(2)}], \quad \mathbf{X}_r^{(2)} \in \mathbb{F}^{I_2 \times L_{2r}}.$$

Then the tensors $\mathcal{A}$ *and* $\mathcal{B}$ *admit decompositions into ML rank-*$(L_{1r}, L_{2r}, \cdot)$ *terms which are connected as follows:*

$$\mathcal{A} = \sum_{r=1}^{R} \mathcal{D}_r \bullet_1 \mathbf{X}_r^{(1)} \bullet_2 \mathbf{X}_r^{(2)} =: \sum_{r=1}^{R} \mathcal{A}_r, \tag{4.8}$$

$$\mathcal{B} = \sum_{r=1}^{R} (\mathcal{D}_r \bullet_1 \mathbf{T}_{1r}^T) \bullet_1 \mathbf{X}_r^{(1)} \bullet_2 \mathbf{X}_r^{(2)} =: \sum_{r=1}^{R} \mathcal{B}_r, \tag{4.9}$$

and

$$\mathcal{D}_r \bullet_1 \mathbf{T}_{1r}^T = \mathcal{D}_r \bullet_2 \mathbf{T}_{2r}^T \qquad r = 1, \ldots, R. \tag{4.10}$$

4. *If* $I_1 = I_2$ *and if there exists a linear combination of* $\mathbf{A}_1, \ldots, \mathbf{A}_{I_3}$ *that is nonsingular, then* $\mathbf{M}_1$ *is similar to* $\mathbf{M}_2$.
5. *If* $\mathbf{M}_1$ *is similar to* $\mathbf{M}_2$*, then* $L_{1r} = L_{2r}$ *for all* r *and the matrices* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *in* (4.4) *and* (4.5) *can be chosen such that* $\mathbf{T}_{1r} = \mathbf{T}_{2r}$ *for all* r.

6. *If, for some r, the matrix $\mathbf{T}_{1r}$ (or $\mathbf{T}_{2r}$) is a scalar multiple of the identity matrix, i.e., if $\mathbf{T}_{1r} = \lambda_r \mathbf{I}_{L_{1r}}$ (or $\mathbf{T}_{2r} = \lambda_r \mathbf{I}_{L_{2r}}$), then $\mathcal{A}_r = \lambda_r \mathcal{B}_r$.*
7. *If $\mathbf{T}_{1r} = \lambda_r \mathbf{I}_{L_{1r}}$ (or $\mathbf{T}_{2r} = \lambda_r \mathbf{I}_{L_{2r}}$) for all r, then $\mathcal{A}$ and $\mathcal{B}$ consist of the same ML rank-$(L_{1r}, L_{2r}, \cdot)$ terms, possibly differently scaled.*

Proof. 1. To prove that the minimal polynomials of $\mathbf{M}_1$ and $\mathbf{M}_2$ coincide, it is sufficient to show that a polynomial $q(x)$ annihilates $\mathbf{M}_1$ if and only if $q(x)$ annihilates $\mathbf{M}_2$. By (4.3), $\mathcal{B} = \mathcal{A} \bullet_1 \mathbf{M}_1^T = \mathcal{A} \bullet_2 \mathbf{M}_2^T$. Since, by (2.1),

$$\mathbf{A}_{(2,3;1)} = [\mathbf{A}_1 \ \dots \ \mathbf{A}_{I_3}]^T \text{ and } \mathbf{A}_{(1,3;2)} = [\mathbf{A}_1^T \ \dots \ \mathbf{A}_{I_3}^T]^T, \tag{4.11}$$

it follows that

$$(\mathbf{B}_i =)\mathbf{M}_1^T \mathbf{A}_i = \mathbf{A}_i \mathbf{M}_2, \qquad i \in \{1, \dots, I_3\}. \tag{4.12}$$

Hence for any $k \geq 1$,

$$\begin{aligned}(\mathbf{M}_1^T)^k \mathbf{A}_i =& (\mathbf{M}_1^T)^{k-1} \mathbf{M}_1^T \mathbf{A}_i = (\mathbf{M}_1^T)^{k-1} \mathbf{A}_i \mathbf{M}_2 = \\ & (\mathbf{M}_1^T)^{k-2} \mathbf{M}_1^T \mathbf{A}_i \mathbf{M}_2 = (\mathbf{M}_1^T)^{k-2} \mathbf{A}_i \mathbf{M}_2^2 = \dots = \mathbf{A}_i \mathbf{M}_2^k,\end{aligned}$$

implying that for any polynomial q,

$$q(\mathbf{M}_1)^T \mathbf{A}_i = \mathbf{A}_i q(\mathbf{M}_2), \qquad i \in \{1, \dots, I_3\}. \tag{4.13}$$

It follows from (4.11) that (4.13) is equivalent to

$$\mathbf{A}_{(1,3;2)} q(\mathbf{M}_2) = \operatorname{Bdiag}(q(\mathbf{M}_1)^T, \dots, q(\mathbf{M}_1)^T) \mathbf{A}_{(1,3;2)} \tag{4.14}$$

and to

$$\mathbf{A}_{(2,3;1)} q(\mathbf{M}_1) = \operatorname{Bdiag}(q(\mathbf{M}_2)^T, \dots, q(\mathbf{M}_2)^T) \mathbf{A}_{(2,3;1)}. \tag{4.15}$$

Assume that q annihilates $\mathbf{M}_1$. Then, by (4.14), $\mathbf{A}_{(1,3;2)} q(\mathbf{M}_2) = \mathbf{O}$. Since $\mathbf{A}_{(1,3;2)}$ has full column rank, it follows that q annihilates $\mathbf{M}_2$. On the other hand, if q annihilates $\mathbf{M}_2$, then by (4.15), $\mathbf{A}_{(2,3;1)} q(\mathbf{M}_1) = \mathbf{O}$. Since $\mathbf{A}_{(2,3;1)}$ has full column rank, it follows that q annihilates $\mathbf{M}_1$. Thus, the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ have the same minimal polynomial.

2. By (4.4), (4.5), and (4.12),

$$\begin{aligned}(\mathbf{S}_1 \operatorname{Bdiag}(\mathbf{T}_{11}, \dots, \mathbf{T}_{1R}) \mathbf{S}_1^{-1})^T \cdot \mathbf{A}_i = & \\ \mathbf{A}_i \cdot \mathbf{S}_2 \operatorname{Bdiag}(\mathbf{T}_{21}, \dots, \mathbf{T}_{2R}) \mathbf{S}_2^{-1}, & \qquad i \in \{1, \dots, I_3\}.\end{aligned}$$

Hence

$$\begin{aligned}\operatorname{Bdiag}(\mathbf{T}_{11}^T, \dots, \mathbf{T}_{1R}^T) \mathbf{S}_1^T \mathbf{A}_i \mathbf{S}_2 = & \\ \mathbf{S}_1^T \mathbf{A}_i \mathbf{S}_2 \operatorname{Bdiag}(\mathbf{T}_{21}, \dots, \mathbf{T}_{2R}), & \qquad i \in \{1, \dots, I_3\}.\end{aligned} \tag{4.16}$$

Let

$$\mathbf{S}_1^T \mathbf{A}_i \mathbf{S}_2 =: \mathbf{D}_i = (\mathbf{D}_{i,r_1 r_2})_{r_1, r_2 = 1}^R$$

denote a block matrix with $\mathbf{D}_{i,r_1r_2} \in \mathbb{F}^{L_{1r_1} \times L_{2r_2}}$. It is clear that (4.16) can be rewritten as

$$\mathbf{T}_{1r_1}^T \mathbf{D}_{i,r_1r_2} = \mathbf{D}_{i,r_1r_2} \mathbf{T}_{2r_2}, \qquad r_1, r_2 = 1, \dots, R, \qquad i \in \{1, \dots, I_3\}, \tag{4.17}$$

implying that (4.7) holds.

Now we show that $\mathbf{D}_i$ is a block diagonal matrix, i.e., that $\mathbf{D}_{i,r_1r_2} = \mathbf{O}$ for $r_1 \neq r_2$. Let $p_r(x)^{\mu_r}$ denote the minimal polynomial of $\mathbf{T}_{1r}$ (or $\mathbf{T}_{2r}$). Then, by (4.17), $\mathbf{O} = \left(\mathbf{T}_{1r_1}^k\right)^T \mathbf{D}_{i,r_1r_2} = \mathbf{D}_{i,r_1r_2} \mathbf{T}_{2r_2}^k$ for all $k \geq 1$, implying that

$$\mathbf{O} = \left(p_{r_1}(\mathbf{T}_{1r_1})^{\mu_{r_1}}\right)^T \mathbf{D}_{i,r_1r_2} = \mathbf{D}_{i,r_1r_2} p_{r_1}(\mathbf{T}_{2r_2})^{\mu_{r_1}} \tag{4.18}$$

for all $r_1, r_2 = 1, \dots, R$ and $i \in \{1, \dots, I_3\}$. Let $r_1 \neq r_2$. To prove that $\mathbf{D}_{i,r_1r_2} = \mathbf{O}$, it is sufficient to show that the matrix $p_{r_1}(\mathbf{T}_{2r_2})^{\mu_{r_1}}$ is nonsingular. Since the polynomials $p_{r_1}(x)^{\mu_{r_1}}$ and $p_{r_2}(x)^{\mu_{r_2}}$ are relatively prime, it follows from the Euclidean algorithm that there exist polynomials $f(x)$ and $g(x)$ such that $1 = p_{r_1}(x)^{\mu_{r_1}} f(x) + p_{r_2}(x)^{\mu_{r_2}} g(x)$ for all $x \in \mathbb{F}$. Hence

$$\mathbf{I} = p_{r_1}(\mathbf{T}_{2r_2})^{\mu_{r_1}} f(\mathbf{T}_{2r_2}) + p_{r_2}(\mathbf{T}_{2r_2})^{\mu_{r_2}} g(\mathbf{T}_{2r_2}) = p_{r_1}(\mathbf{T}_{2r_2})^{\mu_{r_1}} f(\mathbf{T}_{2r_2}).$$

Thus, $p_{r_1}(\mathbf{T}_{2r_2})^{\mu_{r_1}}$ is nonsingular.

3. By (4.6),

$$\mathbf{A}_i = \mathbf{S}_1^{-T} \mathbf{D}_i \mathbf{S}_2^{-1} = \mathbf{X}^{(1)} \mathbf{D}_i \mathbf{X}^{(2)T}, \qquad i = 1, \dots, I_3 \tag{4.19}$$

which is equivalent to (4.8). Since, by (4.3), $\mathbf{B}_i = \mathbf{M}_1^T \mathbf{A}_i$, it follows from (4.4) and (4.19) that

$$\begin{aligned} \mathbf{B}_i = \mathbf{M}_1^T \mathbf{A}_i = (\mathbf{S}_1 \operatorname{Bdiag}(\mathbf{T}_{11}, \dots, \mathbf{T}_{1R}) \mathbf{S}_1^{-1})^T \mathbf{S}_1^{-T} \mathbf{D}_i \mathbf{S}_2^{-1} = \\ \mathbf{S}_1^{-T} \operatorname{Bdiag}(\mathbf{T}_{11}^T, \dots, \mathbf{T}_{1R}^T) \mathbf{S}_1^T \mathbf{S}_1^{-T} \mathbf{D}_i \mathbf{S}_2^{-1} = \\ \mathbf{X}^{(1)} \operatorname{Bdiag}(\mathbf{T}_{11}^T \mathbf{D}_{i,11}, \dots, \mathbf{T}_{1R}^T \mathbf{D}_{i,RR}) \mathbf{X}^{(2)T}, \qquad i = 1, \dots, I_3, \end{aligned} \tag{4.20}$$

which is equivalent to (4.9). Finally, identity (4.10) is equivalent to (4.7).

4. Let the linear combination $t_1\mathbf{A}_1 + \cdots + t_{I_3}\mathbf{A}_{I_3}$ be nonsingular. Then, by (4.12),

$$\mathbf{M}_2 = (t_1\mathbf{A}_1 + \cdots + t_{I_3}\mathbf{A}_{I_3})^{-1} \mathbf{M}_1^T (t_1\mathbf{A}_1 + \cdots + t_{I_3}\mathbf{A}_{I_3}),$$

i.e., $\mathbf{M}_2$ is similar to $\mathbf{M}_1^T$. Since any matrix is similar to its transpose [27, Section 3.2.3], it follows that $\mathbf{M}_2$ is similar to $\mathbf{M}_1$.

5. We choose $\mathbf{S}_1$ such that the matrices $\mathbf{T}_{11}, \dots, \mathbf{T}_{1R}$ in (4.4) are in the Jordan canonical form. Since similar matrices have the same Jordan canonical form, the matrix $\mathbf{M}_2$ is similar to $\operatorname{Bdiag}(\mathbf{T}_{11}, \dots, \mathbf{T}_{1R})$, i.e., there exists $\mathbf{S}_2$ such that (4.5) holds for $\mathbf{T}_{11} = \mathbf{T}_{21}, \dots, \mathbf{T}_{1R} = \mathbf{T}_{2R}$.

6. and 7. follow from (4.9). ∎

Example 4.2. This example illustrates that although the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ in Theorem 4.1 have the same minimal polynomial they are not necessarily similar. Let the frontal slices of $\mathcal{A} \in \mathbb{C}^{3\times3\times4}$ have the following nonzero pattern:

$$\begin{bmatrix} 0 & * & * \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix}$$

It is clear that any linear combination of the frontal slices of $\mathcal{A}$ is singular so the assumption in statement 4 of Theorem 4.1 does not hold. We choose the values "$*$" (e.g., generic values) such that $\mathbf{A}_{(2,3;1)}$ and $\mathbf{A}_{(1,3;2)}$ have full column rank. It is clear that $\mathcal{A}$ is the sum of a ML rank-$(1,2,\cdot)$ and a ML rank-$(2,1,\cdot)$ term. More precisely, $\mathcal{A}$ is the sum of a ML rank-$(1,2,2)$ and a ML rank-$(2,1,2)$ term. Let $\mathbf{M}_1 := \mathrm{diag}(\lambda_2, \lambda_1, \lambda_1)$ and $\mathcal{B} = \mathcal{A} \bullet_1 \mathbf{M}_1^T$. One can easily verify that $\mathcal{B} = \mathcal{A} \bullet_2 \mathbf{M}_2^T$, where $\mathbf{M}_2 = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_2)$. Thus, if $\lambda_1 \neq \lambda_2$, then $\mathbf{M}_1$ and $\mathbf{M}_2$ have the same minimal polynomial but are not similar.

Now we consider the general case, that is, we assume that $\mathcal{A}$ and $\mathcal{B}$ are tensors of order $N \geq 3$ and satisfy (3.2) for $N \geq \hat{N} \geq 2$. First we extend the notion of block diagonal matrices to tensors. Let the numbers $L_{n1}, \ldots, L_{nR}$ sum up to I_n for each $n = 1, \ldots, \hat{N}$. Consider the partition of $\{1, \ldots, I_n\}$ into consecutive blocks $V_{n1}, \ldots, V_{nR}$ of length $L_{n1}, \ldots, L_{nR}$, respectively, so $V_{n1} = \{1, \ldots, L_{n1}\}, \ldots, V_{nR} = \{I_n - L_{nR} + 1, \ldots, L_{nR}\}$. If the condition

$$(\mathcal{D})_{i_1,\ldots,i_N} = 0 \text{ for } (i_1, \ldots, i_{\hat{N}}) \notin \bigcup_{r=1}^{R} (V_{1r} \times \cdots \times V_{\hat{N}r}) \tag{4.21}$$

holds, then we say that $\mathcal{D}$ is a block diagonal tensor and write $\mathcal{D} = \mathrm{Bdiag}(\mathcal{D}_1, \ldots, \mathcal{D}_R)$, where $\mathcal{D}_r := \mathcal{D}(V_{1r}, \ldots, V_{\hat{N}r}, :, \ldots, :) \in \mathbb{F}^{L_{1r} \times \cdots \times L_{\hat{N}r} \times I_{\hat{N}+1} \times \cdots \times I_N}$ denote the diagonal blocks. For instance, statement 2 of Theorem 4.1 means that if $\mathcal{D}$ is the $I_1 \times I_2 \times I_3$ tensor formed by the $I_1 \times I_2$ matrices $\mathbf{D}_i$ in (4.6), i.e., if $\mathcal{D} := \mathcal{A} \bullet_1 \mathbf{S}_1 \bullet_2 \mathbf{S}_2$, then $\mathcal{D} = \mathrm{Bdiag}(\mathcal{D}_1, \ldots, \mathcal{D}_R)$, where the diagonal blocks $\mathcal{D}_r \in \mathbb{F}^{L_{1r} \times L_{2r} \times I_3}$ are defined in statement 3 of Theorem 4.1.

The following result generalizes Theorem 4.1 for $N \geq 3$ and $N \geq \hat{N} \geq 2$. The proof is obtained by applying Theorem 4.1 to the $I_i \times I_j \times \frac{\prod I_n}{I_i I_j}$ reshapings of $\mathcal{A}$ and $\mathcal{B}$.

Theorem 4.3. *Let tensors $\mathcal{A}, \mathcal{B} \in \mathbb{F}^{I_1 \times \cdots \times I_N}$ and let $N \geq 3$, $N \geq \hat{N} \geq 2$. Assume that for each $n \in \{1 \ldots, \hat{N}\}$,*

$$\mathbf{A}_{(n^c;n)} \text{ has full column rank} \tag{4.22}$$

and that there exists matrix $\mathbf{M}_n \in \mathbb{F}^{I_n \times I_n}$ such that

$$\mathbf{B}_{(n^c;n)} = \mathbf{A}_{(n^c;n)} \mathbf{M}_n. \tag{4.23}$$

Then the following statements hold.

1. *The matrices $\mathbf{M}_1, \ldots, \mathbf{M}_{\hat{N}}$ have the same minimal polynomial $q(x)$.*
2. *Consider the factorization $q(x) = p_1(x)^{\mu_1} \cdots p_R(x)^{\mu_R}$ with distinct polynomials $p_r(x)$ that are irreducible (over $\mathbb{F}$) and set*

$$L_{nr} := \dim(\mathrm{Null}\,(p_r(\mathbf{M}_n)^{\mu_r})), \qquad 1 \leq r \leq R, \quad 1 \leq n \leq \hat{N}.$$

Let also

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{S}_1\, \mathrm{Bdiag}(\mathbf{T}_{11}, \ldots, \mathbf{T}_{1R}) \mathbf{S}_1^{-1}, \quad \mathbf{S}_1 = [\mathbf{S}_{11} \ \ldots \ \mathbf{S}_{1R}], \quad \mathbf{S}_{1r} \in \mathbb{F}^{I_1 \times L_{1r}}, \\ &\vdots \\ \mathbf{M}_{\hat{N}} &= \mathbf{S}_{\hat{N}}\, \mathrm{Bdiag}(\mathbf{T}_{\hat{N}1}, \ldots, \mathbf{T}_{\hat{N}R}) \mathbf{S}_{\hat{N}}^{-1}, \quad \mathbf{S}_{\hat{N}} = [\mathbf{S}_{\hat{N}1} \ \ldots \ \mathbf{S}_{\hat{N}R}], \quad \mathbf{S}_{\hat{N}r} \in \mathbb{F}^{I_{\hat{N}} \times L_{\hat{N}r}} \end{aligned} \tag{4.24}$$

be the primary decompositions of $\mathbf{M}_1, \dots, \mathbf{M}_{\hat{N}}$*, respectively, such that the minimal polynomials of* $\mathbf{T}_{1r}, \dots, \mathbf{T}_{\hat{N}r}$ *are equal to* $p_r(x)^{\mu_r}$ *for each* $r = 1, \dots, R$*. Then the tensor*

$$\mathcal{D} := \mathcal{A} \bullet_1 \mathbf{S}_1 \cdots \bullet_{\hat{N}} \mathbf{S}_{\hat{N}}$$

is block-diagonal (see (4.21)),

$$\mathcal{D} = \operatorname{Bdiag}(\mathcal{D}_1, \dots, \mathcal{D}_R), \qquad \mathcal{D}_r \in \mathbb{F}^{L_{1r} \times \cdots \times L_{\hat{N}r} \times I_{\hat{N}+1} \times \cdots \times I_N}$$

and

$$\mathcal{D}_r \bullet_1 \mathbf{T}_{1r}^T = \cdots = \mathcal{D}_r \bullet_{\hat{N}} \mathbf{T}_{\hat{N}r}^T \qquad r = 1, \dots, R. \tag{4.25}$$

3. *Let*

$$\mathbf{S}_n^{-T} =: \mathbf{X}^{(n)} = [\mathbf{X}_1^{(n)} \ \dots \ \mathbf{X}_R^{(n)}], \quad \mathbf{X}_r^{(n)} \in \mathbb{F}^{I_n \times L_{nr}}.$$

Then the tensors $\mathcal{A}$ *and* $\mathcal{B}$ *admit decompositions into ML rank-*$(L_{1r}, \dots, L_{\hat{N}r}, \cdot, \dots, \cdot)$ *terms which are connected as follows:*

$$\mathcal{A} = \sum_{r=1}^{R} \mathcal{D}_r \bullet_1 \mathbf{X}_r^{(1)} \cdots \bullet_{\hat{N}} \mathbf{X}_r^{(\hat{N})} =: \sum_{r=1}^{R} \mathcal{A}_r, \tag{4.26}$$

$$\mathcal{B} = \sum_{r=1}^{R} (\mathcal{D}_r \bullet_1 \mathbf{T}_{1r}^T) \bullet_1 \mathbf{X}_r^{(1)} \cdots \bullet_{\hat{N}} \mathbf{X}_r^{(\hat{N})} =: \sum_{r=1}^{R} \mathcal{B}_r, \tag{4.27}$$

in which the tensors $\mathcal{D}_r$ *satisfy the identities in (4.25).*

4. *Let* $\mathbf{A}_{ij,k}$*,* $k = 1, \dots, (I_1 \cdots I_N)/(I_i I_j)$ *denote the* $I_i \times I_j$ *slices of* $\mathcal{A}$*, that is,* $\mathbf{A}_{ij,k} \in \mathbb{F}^{I_i \times I_j}$ *is obtained from* $\mathcal{A}$ *by fixing all indices but* i *and* j*. If* $I_i = I_j$ *and if there exists a linear combination of* $\mathbf{A}_{ij,k}$ *that is nonsingular, then* $\mathbf{M}_i$ *is similar to* $\mathbf{M}_j$*.*
5. *If* $\mathbf{M}_i$ *is similar to* $\mathbf{M}_j$*, then* $L_{ir} = L_{jr}$ *for all* r *and the matrices* $\mathbf{S}_i$ *and* $\mathbf{S}_j$ *in (4.24) can be chosen such that* $\mathbf{T}_{ir} = \mathbf{T}_{jr}$ *for all* r*.*
6. *If, for some* r*, there exists* n *such that the matrix* $\mathbf{T}_{nr}$ *is a scalar multiple of the identity matrix, i.e., if* $\mathbf{T}_{nr} = \lambda_r \mathbf{I}_{L_{nr}}$*, then* $\mathcal{A}_r = \lambda_r \mathcal{B}_r$*.*
7. *If for each* r *there exists* n_r *such that* $\mathbf{T}_{n_r r} = \lambda_r \mathbf{I}_{L_{n_r r}}$*, then* $\mathcal{A}$ *and* $\mathcal{B}$ *consist of the same ML rank-*$(L_{1r}, \dots, L_{\hat{N}r}, \cdot, \dots, \cdot)$ *terms, possibly differently scaled.*

Proof. Let $1 \le i < j \le \hat{N}$. We reshape $\mathcal{A}$ and $\mathcal{B}$ into the $I_i \times I_j \times \frac{\prod I_n}{I_i I_j}$ tensors $\mathcal{A}^{ij}$ and $\mathcal{B}^{ij}$ such that

$$\mathbf{A}^{ij}_{(2,3;1)} = \mathbf{A}_{(i^c;i)}, \quad \mathbf{B}^{ij}_{(2,3;1)} = \mathbf{B}_{(i^c;i)}, \quad \mathbf{A}^{ij}_{(1,3;2)} = \mathbf{A}_{(j^c;j)}, \quad \mathbf{B}^{ij}_{(1,3;2)} = \mathbf{B}_{(j^c;j)}. \tag{4.28}$$

Then, by (4.22) and (4.28), the first two matrix representations of $\mathcal{A}^{ij}$ have full column rank and, by (4.23) and (4.28),

$$\mathbf{B}^{ij}_{(2,3;1)} = \mathbf{A}^{ij}_{(2,3;1)} \mathbf{M}_i \text{ and } \mathbf{B}^{ij}_{(1,3;2)} = \mathbf{A}^{ij}_{(1,3;2)} \mathbf{M}_j.$$

Thus $\mathcal{A}^{ij}$ and $\mathcal{B}^{ij}$ satisfy the assumptions in Theorem 4.1. We leave it to the reader to show that the statements in Theorem 4.3 can be obtained from the corresponding statements of Theorem 4.1 by applying it to all pairs $(\mathcal{A}^{ij}, \mathcal{B}^{ij})$, where $1 \le i < j \le \hat{N}$. ∎

4.2. Redundancy of conditions in (1.2). In this subsection we explain that if $\mathrm{col}(\mathbf{B}_{(n^c;n)}) \subseteq \mathrm{col}(\mathbf{A}_{(n^c;n)})$, then for any subset $S \subsetneq \{1,\dots,N\}$ that contains n we also have that $\mathrm{col}(\mathbf{B}_{(S^c;S)}) \subseteq \mathrm{col}(\mathbf{A}_{(S^c;S)})$ (Lemma 4.4). Hence the N conditions in (1.3) imply the 2^N-2 conditions in (1.2) (Corollary 4.5).

Let us first formally define generalized matrix representations. Let $\mathcal{A} \in \mathbb{F}^{I_1\times\cdots\times I_N}$, let S be a proper subset of $\{1,\dots,N\}$ and let S^c denote the complement of S. A *mode-S slice* of $\mathcal{A}$ is a subtensor obtained from $\mathcal{A}$ by fixing the indices in S. It is clear that $\mathcal{A}$ has $\prod_{n\in S} I_n$ mode-S slices. *A mode-S matrix representation* of $\mathcal{A}$ is a matrix $\mathbf{A}_{(S^c;S)} \in \mathbb{F}^{(\prod_{n\notin S} I_n)\times(\prod_{n\in S} I_n)}$ whose columns are the vectorized mode-S slices of $\mathcal{A}$. Formally, if we follow the conventions that

$$S=\{q_1,\dots,q_{N-k}\} \text{ with } q_1<\cdots<q_{N-k} \quad \text{and} \quad S^c=\{p_1,\dots,p_k\} \text{ with } p_1<\cdots<p_k, \tag{4.29}$$

then

$$\text{the } (ind_{i_{p_1},\dots,i_{p_k}}^{I_{p_1}\times\cdots\times I_{p_k}}, ind_{i_{q_1},\dots,i_{q_{N-k}}}^{I_{q_1}\times\cdots\times I_{q_{N-k}}})\text{th entry of the matrix } \mathbf{A}_{(S^c;S)} \text{ is equal to } a_{i_1\dots i_N}, \tag{4.30}$$

where

$$ind_{i_{p_1},\dots,i_{p_k}}^{I_{p_1}\times\cdots\times I_{p_k}} := 1+\sum_{u=1}^{k}(i_{p_u}-1)\prod_{s=1}^{u-1} I_{p_s}$$

denotes the linear index corresponding to the element in the $(i_{p_1},\dots,i_{p_k})$ position of an $I_{p_1}\times\cdots\times I_{p_k}$ tensor. If $S=\{n\}$, then $\mathbf{A}_{(S^c;S)}$ coincides with the mode-n matrix representation $\mathbf{A}_{(n^c;n)}$ introduced earlier in (2.1). It is easy to show that, if for two $I_1\times\cdots\times I_N$ tensors $\mathcal{A}$ and $\mathcal{B}$ the identity $\mathbf{B}_{(n^c;n)}=\mathbf{A}_{(n^c;n)}\mathbf{M}_n$ holds for some n, then for any subset S that contains n there exists a matrix $\mathbf{M}_S$ such that $\mathbf{B}_{(S^c;S)}=\mathbf{A}_{(S^c;S)}\mathbf{M}_S$. Indeed, the matrices $\mathbf{B}_{(n^c;n)}$ and $\mathbf{A}_{(n^c;n)}$ can be simultaneously reshaped into the matrices $\mathbf{B}_{(S^c;S)}$ and $\mathbf{A}_{(S^c;S)}$, respectively, so that the kth column of $\mathbf{B}_{(n^c;n)}$ (resp. $\mathbf{A}_{(n^c;n)}$) is reshaped into a group of $(\prod_{l\in S} I_l)/I_n$ columns of $\mathbf{B}_{(S^c;S)}$ (resp. $\mathbf{A}_{(S^c;S)}$) whose indices are determined by k. Since the kth column of $\mathbf{B}_{(n^c;n)}$ is a linear combination of columns of $\mathbf{A}_{(n^c;n)}$, it follows that each of the $(\prod_{l\in S} I_l)/I_n$ corresponding columns of $\mathbf{B}_{(S^c;S)}$ is a linear combination of the $(\prod_{l\in S} I_l)/I_n$ corresponding columns of $\mathbf{A}_{(S^c;S)}$. Thus, $\mathbf{B}_{(S^c;S)}=\mathbf{A}_{(S^c;S)}\mathbf{M}_S$ holds for some matrix $\mathbf{M}_S$. More in detail, we show in the following lemma that the matrix $\mathbf{M}_S$ coincides up to column and row permutation with the direct sum of $\mathbf{M}_n$ multiple times with itself.

Lemma 4.4. *Let $N\geq 4$, and let $\mathcal{A},\mathcal{B}\in\mathbb{F}^{I_1\times\cdots\times I_N}$ be such that $\mathbf{B}_{(n^c;n)}=\mathbf{A}_{(n^c;n)}\mathbf{M}_n$ for some $\mathbf{M}_n\in\mathbb{F}^{I_n\times I_n}$. Let S and S^c be as in (4.29) and let $n\in S$, that is, $q_l=n$ for some $l\in\{1,\dots,N-k\}$. Then $\mathbf{B}_{(S^c;S)}=\mathbf{A}_{(S^c;S)}\mathbf{M}_S$, where*

$$\mathbf{M}_S=\left(\bigotimes_{v=1}^{l-1}\mathbf{I}_{I_{q_v}}\right)\otimes\mathbf{M}_n\otimes\left(\bigotimes_{v=l+1}^{N-K}\mathbf{I}_{I_{q_v}}\right) \tag{4.31}$$

or $\mathbf{M}_S=\mathbf{I}_K\otimes\mathbf{M}_n\otimes\mathbf{I}_L$, where $K=\prod_{v=1}^{l-1} I_{q_v}$ and $L=\prod_{v=l+1}^{N-k} I_{q_v}$.

Proof. Let $\delta(i,j)$ denote the Kronecker delta symbol, i.e., $\delta(i,j)=1$ for $i=j$ and $\delta(i,j)=0$ for $i\neq j$. One can easily verify that

$$
\begin{aligned}
&(\mathbf{M}_S)_{ind^{I_{q_1}\times\cdots\times I_{q_{N-k}}}_{i_{\tilde{q}_1},\ldots,i_{\tilde{q}_{N-k}}},ind^{I_{q_1}\times\cdots\times I_{q_{N-k}}}_{i_{q_1},\ldots,i_{q_{N-k}}}} = \\
&\delta(i_{\tilde{q}_1},i_{q_1})\cdots\delta(i_{\tilde{q}_{l-1}},i_{q_{l-1}})\cdot(\mathbf{M}_n)_{\tilde{q}_l,n}\cdot\delta(i_{\tilde{q}_{l+1}},i_{q_{l+1}})\cdots\delta(i_{\tilde{q}_{N-k}},i_{q_{N-k}}) = \\
&\begin{cases}(\mathbf{M}_n)_{\tilde{q}_l,n}, & \text{if } i_{\tilde{q}_1}=i_{q_1},\ldots,i_{\tilde{q}_{l-1}}=i_{q_{l-1}},i_{\tilde{q}_{l+1}}=i_{q_{l+1}},\ldots,i_{\tilde{q}_{N-k}}=i_{q_{N-k}}\\ 0, & \text{otherwise.}\end{cases}
\end{aligned}
\tag{4.32}
$$

Hence

$$
\begin{aligned}
&\left(\mathbf{B}_{(S^c;S)}\right)_{ind^{I_{p_1}\times\cdots\times I_{p_k}}_{i_{p_1},\ldots,i_{p_k}},ind^{I_{q_1}\times\cdots\times I_{q_{N-k}}}_{i_{q_1},\ldots,i_{q_{N-k}}}} \overset{(4.30)}{=} (\mathcal{B})_{i_1,\ldots,i_N} \overset{(4.30)}{=} \left(\mathbf{B}_{(n^c;n)}\right)_{ind^{I_1\times\cdots\times I_{n-1}\times I_{n+1}\cdots\times I_N}_{i_1,\ldots,i_{n-1},i_{n+1},\ldots,i_N},i_n} = \\
&\left(\mathbf{A}_{(n^c;n)}\mathbf{M}_n\right)_{ind^{I_1\times\cdots\times I_{n-1}\times I_{n+1}\cdots\times I_N}_{i_1,\ldots,i_{n-1},i_{n+1},\ldots,i_N},i_n} = \sum_{\tilde{q}_l=1}^{n}(\mathcal{A})_{i_1,\ldots,i_{n-1},i_{\tilde{q}_l},i_{n+1},\ldots,i_N}(\mathbf{M}_n)_{\tilde{q}_l,n} \overset{(4.30)}{=} \\
&\sum_{\tilde{q}_l=1}^{n}\left(\mathbf{A}_{(S^c;S)}\right)_{ind^{I_{p_1}\times\cdots\times I_{p_k}}_{i_{p_1},\ldots,i_{p_k}},ind^{I_{q_1}\times\cdots\times I_{q_{N-k}}}_{i_{q_1},\ldots,i_{q_{l-1}},i_{\tilde{q}_l},i_{q_{l+1}}\ldots,i_{q_{N-k}}}}(\mathbf{M}_n)_{\tilde{q}_l,n} \overset{(4.32)}{=} \\
&\sum_{\tilde{q}_1=1}^{q_1}\cdots\sum_{\tilde{q}_{N-k}=1}^{q_{N-k}}\left(\mathbf{A}_{(S^c;S)}\right)_{ind^{I_{p_1}\times\cdots\times I_{p_k}}_{i_{p_1},\ldots,i_{p_k}},ind^{I_{q_1}\times\cdots\times I_{q_{N-k}}}_{i_{\tilde{q}_1},\ldots,i_{\tilde{q}_{N-k}}}}(\mathbf{M}_S)_{ind^{I_{q_1}\times\cdots\times I_{q_{N-k}}}_{i_{\tilde{q}_1},\ldots,i_{\tilde{q}_{N-k}}},ind^{I_{q_1}\times\cdots\times I_{q_{N-k}}}_{i_{q_1},\ldots,i_{q_{N-k}}}} = \\
&\left(\mathbf{A}_{(S^c;S)}\mathbf{M}_S\right)_{ind^{I_{p_1}\times\cdots\times I_{p_k}}_{i_{p_1},\ldots,i_{p_k}},ind^{I_{q_1}\times\cdots\times I_{q_{N-k}}}_{i_{q_1},\ldots,i_{q_{N-k}}}}.
\end{aligned}
$$

■

The following corollary follows from Lemma 4.4 and states that 2^N-2-N conditions in (1.2) are redundant.

Corollary 4.5. *Let $N\geq 4$, and let $\mathcal{A},\mathcal{B}\in\mathbb{F}^{I_1\times\cdots\times I_N}$. If (1.3) holds, then (1.2) also holds.*

5. Illustration: classification of linear mixtures of signals. A basic problem in signal processing is to assess whether two observed signals involve the same underlying signal "components". Typically, the component signals manifest themselves with a different amplitude in the observed signals. If moreover the component signals are by themselves unknown, which is the case in many applications, the problem can be very challenging. As a preview, in Figure 5.3 it may a priori not be obvious to establish which displayed signals are generated by the same components up to scaling.

One of the possible applications is in underdetermined Blind Source Separation (BSS). In BSS, the task is to recover sources from a set of their linear mixtures [11]. Often, sources are sparsely combined in the observed mixed signals [25], i.e., the number of sources is large but each mixture contains a small number of sources. This means that the mixing matrix is sparse and has many more columns than rows. BSS problems that involve a wide mixing matrix are called underdetermined and are generally much harder to solve than overdetermined BSS problems (involving a mixing matrix that is square or tall). As a preprocessing step one can first try to solve the following multi-label classification problem: mixture i belongs to the same class as mixture j if mixture i is generated by (some of) the sources that appear in mixture j.

In this way the initial underdetermined BSS problem with many sources can be replaced by a set of smaller overdetermined BSS problems.

In this section we explain how Theorem 4.1 can be used to solve the multi-label classification problem. Our derivation is valid under the assumption that the sources can simultaneously be mapped (i.e., "tensorized") into low ML rank tensors and that the mapping, so called *tensorization*, is linear. Such mappings are known [19, 4, 29] for sources that can be modeled as exponential polynomials (*Hankelization*), rational functions (*Löwnerization*), and periodic signals (*Segmentation*), among others. To demonstrate the approach we confine ourselves to exponential polynomials.

To solve the multi-label classification problem, we do not use more prior knowledge about the sources than that they can be (approximately) modeled as exponential polynomials (with a mild bound on the value L_s in (5.2) that will be introduced in the next subsection).

5.1. Exponential polynomials and Hankelization mapping. A univariate exponential polynomial is a function of the form

$$s(t) = \sum_{f=1}^{F} p_f(t) a_f^t, \tag{5.1}$$

where $p_1, \dots, p_F$ are non-zero polynomials in one variable and $a_1, \dots, a_F \in \mathbb{C} \setminus \{0\}$. Let T_s denote the sampling time and let N be the number of sampling points. It can be shown [14, 19] that for any positive integers I_1, I_2, I_3 that sum up to $N+2$ and are greater than or equal to $L_{\mathbf{s}}$, the vector $\mathbf{s} = [s_1 \ \dots \ s_N]^T := [s(0) \ \dots \ s((N-1)T_s)]^T$ can be mapped to an $I_1 \times I_2 \times I_3$ ML rank-$(L_{\mathbf{s}}, L_{\mathbf{s}}, L_{\mathbf{s}})$ tensor $\mathcal{S}$, where the value

$$L_{\mathbf{s}} := F + \sum_{f=1}^{F} \deg p_f \tag{5.2}$$

does not depend on I_1, I_2, I_3. The mapping $H : \mathbf{s} \mapsto \mathcal{S}$, $H : \mathbb{C}^N \to \mathbb{C}^{I_1 \times I_2 \times I_3}$ is given by [14, 19]

$$(\mathcal{S})_{i_1 i_2 i_3} = s_{i_1+i_2+i_3-2} = s((i_1+i_2+i_3-3)T_s),$$

where $1 \le i_1 \le I_1$, $1 \le i_2 \le I_2$, $1 \le i_3 \le I_3$. Since $(\mathcal{S})_{i_1 i_2 i_3}$ depends only on $i_1+i_2+i_3$, the mapping H was called "Hankelization" in [19]. It is worth noting that if $I_1 = I_2 = I_3$, then $\mathcal{S}$ is a fully symmetric tensor, implying that $\mathbf{S}_{(2,3;1)} = \mathbf{S}_{(1,3;2)} = \mathbf{S}_{(1,2;3)}$.

It is clear that H is a linear mapping, so if $\mathbf{y} = [y_1 \ \dots \ y_N]^T := [y(0) \ \dots \ y((N-1)T_s)]^T$ is a linear mixture of sampled sources of the form (5.1)

$$y(t) = \mathbf{g}_1 s_1(t) + \dots \mathbf{g}_R s_r(t), \qquad t = 0, T_s, \dots, (N-1)T_s \tag{5.3}$$

and $\min(I_1, I_2, I_3) \ge \max L_{\mathbf{s}_r}$, then, by (5.2),

$$\mathcal{Y} := H(\mathbf{y}) = \mathbf{g}_1 H(\mathbf{s}_1) + \dots + \mathbf{g}_R H(\mathbf{s}_R) = \mathbf{g}_1 \mathcal{S}_1 + \dots + \mathbf{g}_R \mathcal{S}_R \tag{5.4}$$

is a decomposition of $\mathcal{Y}$ into a sum of ML rank-$(L_{\mathbf{s}_r}, L_{\mathbf{s}_r}, L_{\mathbf{s}_r})$ terms.

5.2. Example. We generate 25 mixtures

$$y_j(t) = \mathbf{g}_{1j}s_1(t) + \dots \mathbf{g}_{8j}s_8(t), \qquad j = 1, \dots, 25 \tag{5.5}$$

of 8 exponential polynomials

$$\begin{aligned}
&s_1(t) = 3 \cdot 2^{-\frac{t}{5}}, && s_2(t) = 3\cos(\pi t + \frac{1}{2}), && s_3(t) = 3\cos(2\pi t + 2),\\
&s_4(t) = 3\cos(3\pi t - 2), && s_5(t) = (5-t)\cos(10\pi t + \frac{1}{2}), && s_6(t) = (5-t)\cos(12\pi t - \frac{3}{2}),\\
&s_7(t) = t\cos(8\pi t + 1), && s_8(t) = t\cos(14\pi t - \frac{1}{2}).
\end{aligned}$$

The coefficients $\mathbf{g}_{ij}$ are generated randomly[7] so that for each $j = 1, \dots, 25$ at least three and at most six of $\mathbf{g}_{1j}, \dots, \mathbf{g}_{8j}$ are zero. The nonzero coefficients $\mathbf{g}_{ij}$ are randomly chosen from $[-2.5, -0.5] \cup [0.5, 2.5]$. We thus obtain that

$$[y_1(t) \ \dots y_{25}(t)] = [s_1(t) \ \dots s_8(t)]\mathbf{G}, \tag{5.6}$$

where $\mathbf{G} = (\mathbf{g}_{ij})$ is an 8×25 sparse matrix. The nonzero pattern of $\mathbf{G}$ is shown in Figure 5.1. By way of example, the mixtures $y_1(t)$, $y_4(t)$, $y_8(t)$, and $y_{19}(t)$ were generated as

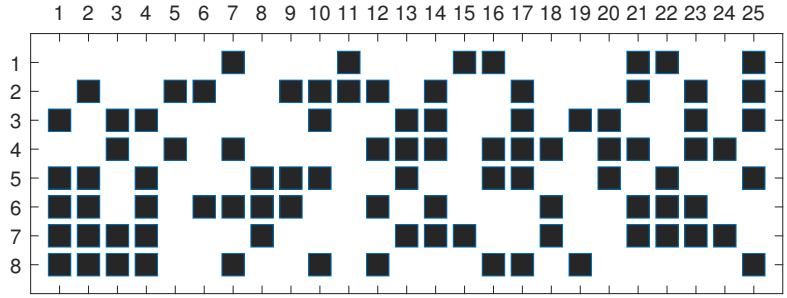


Figure 5.1. *The nonzero pattern of the matrix* $\mathbf{G}$.

$$y_1(t) = 2.22s_3(t) - 1.95s_5(t) - 2.38s_6(t) - 2.39s_7(t) + 1.77s_8(t), \tag{5.7}$$

$$y_4(t) = -0.55s_3(t) - 2.07s_5(t) + 0.50s_6(t) + 2.41s_7(t) - 1.90s_8(t), \tag{5.8}$$

$$y_8(t) = + 1.16s_5(t) + 0.94s_6(t) + 1.35s_7(t), \tag{5.9}$$

$$y_{19}(t) = 0.69s_3(t) - 0.68s_8(t). \tag{5.10}$$

We consider a noisy sampled (with $T_s = 0.05$ and $N = 100$) version of (5.6):

$$[\mathbf{y}_1^n \ \dots \ \mathbf{y}_{25}^n] := [\mathbf{y}_1 \ \dots \ \mathbf{y}_{25}] + \sigma\mathbf{N} = [\mathbf{s}_1 \ \dots \mathbf{s}_8]\mathbf{G} + \sigma\mathbf{N}, \tag{5.11}$$

[7] The numerical experiments in the example 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).

in which the entries of the 100×25 matrix $\mathbf{N}$ are independently drawn from the standard normal distribution $N(0,1)$ and $\sigma = 0.1\frac{\|[\mathbf{y}_1 \ \cdots \ \mathbf{y}_{25}]\|_F}{\|\mathbf{N}\|_F}$, where $\|\cdot\|_F$ denotes the Frobenius norm.[8] The sampled sources $\mathbf{s}_1, \ldots, \mathbf{s}_8$ and noisy sampled mixtures $\mathbf{y}_1^n$, $\mathbf{y}_4^n$, $\mathbf{y}_8^n$, $\mathbf{y}_{19}^n$ are shown in Figures 5.2 and 5.3, respectively. We now use Theorem 4.1 to verify whether the pair of

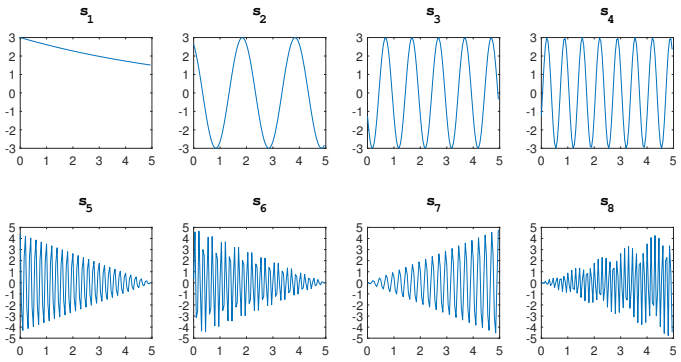


Figure 5.2. *Sampled source signals* $\mathbf{s}_1, \ldots, \mathbf{s}_8$*; the sampling interval is 0.05.*

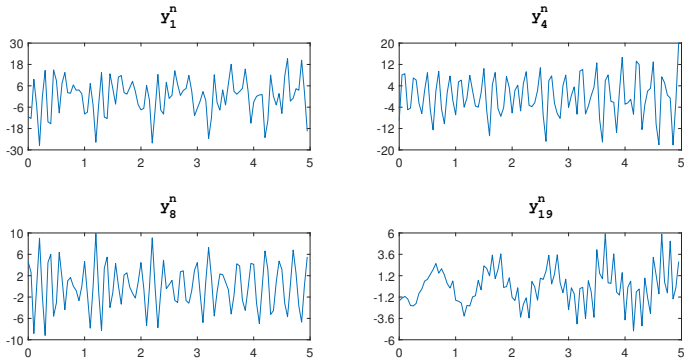


Figure 5.3. *Noisy sampled mixtures* $\mathbf{y}_1^n$, $\mathbf{y}_4^n$, $\mathbf{y}_8^n$, $\mathbf{y}_{19}^n$ *(see (5.7)–(5.10))*

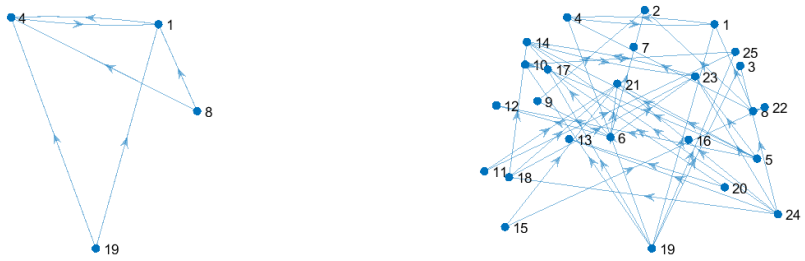
.

mixtures $(y_i(t), y_j(t))$ is generated by the same subset of sources, $1 \leq i, j \leq 25$[9], $i \neq j$. For visualization purposes it is convenient to associate the mixtures $y_1(t), \ldots, y_{25}(t)$ with vertices of a directed graph: a directed edge from vertex i to vertex j indicates that $y_i(t)$ is generated

[8]Note that the matrix pencil based algorithm in [14] can be used to estimate the matrix $\mathbf{G}$ and the sources $\mathbf{s}_1, \ldots, \mathbf{s}_8$ only for much smaller values of σ.

[9]Note that in contrast to BSS, we do not work with the full matrix $\mathbf{G}$ but only with pairs of its columns. The number of mixtures 25 is just chosen to illustrate the approach for a large number of pairs (namely, $25^2 - 25 = 600$), making the results very convincing.

by sources that also appear in $y_j(t)$. For instance, the subgraph corresponding to the mixtures $y_1(t)$, $y_4(t)$, $y_8(t)$, and $y_{19}(t)$ is shown in Figure 5.4a. In this example we show how to recover the overall graph in Figure 5.4b based only on the (observed) vectors $\mathbf{y}_1^n, \dots, \mathbf{y}_{25}^n$ and without estimating $\mathbf{s}_1, \dots, \mathbf{s}_8$. Let $H : \mathbb{C}^{100} \to \mathbb{C}^{34\times 34\times 34}$ denote the Hankelization mapping. By (5.4)



(a) (b)

Figure 5.4. *The subgraph corresponding to the mixtures $y_1(t)$, $y_4(t)$, $y_8(t)$, $y_{19}(t)$ (left) and the graph corresponding to all mixtures $y_1(t), \dots, y_{25}(t)$ (right); a directed edge from vertex i to vertex j indicates that mixture $y_i(t)$ is generated by sources that also appear in mixture $y_j(t)$.*

and (5.11), we have that

$$\mathcal{Y}_i^n := H(\mathbf{y}_i^n) = \mathbf{g}_{i1}H(\mathbf{s}_1) + \cdots + \mathbf{g}_{i8}H(\mathbf{s}_8) + \sigma H(\mathbf{n}_i) = \mathbf{g}_{i1}\mathcal{S}_1 + \cdots + \mathbf{g}_{18}\mathcal{S}_R + \sigma H(\mathbf{n}_i)$$

is an approximate decomposition of $\mathcal{Y}_i^n$ into a sum of ML rank-$(L_{\mathbf{s}_r}, L_{\mathbf{s}_r}, L_{\mathbf{s}_r})$ terms. One can easily verify that the exact values of $L_{\mathbf{s}_1}, \dots, L_{\mathbf{s}_8}$ are $1, 2, 2, 2, 4, 4, 4, 4$, respectively. For instance, since

$$s_8(t) = t\frac{1}{2}(e^{(14\pi t - 0.5)i} + e^{-(14\pi t - 0.5)i}) = (\frac{1}{2}e^{-0.5it^1})e^{14\pi t} + (\frac{1}{2}e^{0.5it^1})e^{-14\pi t},$$

we get, by (5.2), that $L_{\mathbf{s}_8} = 2+(1+1) = 4$. On the other hand, it can be verified that although the tensors $\mathcal{S}_5, \dots, \mathcal{S}_8$ are ML rank-$(4, 4, 4)$, they can be approximated by ML rank-$(2, 2, 2)$ tensors with a relative error less than 0.061, which is below the noise level.

To verify whether $y_i(t)$ is generated by sources that appear in $y_j(t)$, it is sufficient to show that $\mathcal{Y}_i^n$ is generated by ML rank terms that appear in $\mathcal{Y}_j^n$, which, by Theorem 4.1, is reduced to verifying that the column space of $\mathbf{Y}^n_{i\,(2,3;1)}$ is contained in the column space of $\mathbf{Y}^n_{j\,(2,3;1)}$.[10] To compare the column spaces we proceeded as follows. For each $i = 1, \dots, 25$ we computed the first r_i singular vectors $\mathbf{u}_{i1}, \dots, \mathbf{u}_{ir_i}$ of $\mathbf{Y}^n_{i\,(2,3;1)}$, where the rank r_i of $\mathbf{Y}^n_{i\,(2,3;1)}$ was estimated as the largest index k such that the ratio of the kth and the $(k+1)$st singular value of $\mathbf{Y}^n_{i\,(2,3;1)}$ is greater than a certain threshold τ. (We chose $\tau = 2.3$.) Then for all

[10]Note that we use only the part of Theorem 4.1 that guarantees that (1.2) implies (1.1). This part relies only on the comparison of column spaces; the computation of the matrices $\mathbf{M}_i$ is not necessary. One could actually go further, i.e., compute the matrices $\mathbf{M}_i$, their spectra, the number of low ML rank terms and their ML rank values. This is possible because even for two mixtures that contain all 8 sources (which is never the case in the example) we have that $L_1 + \cdots + L_8 = 1+2+2+2+4+4+4+4 = 23 \leq 34 = I_n$ for $n = 1, 2, 3$.

$i, j = 1, \ldots, 25$ and $i \neq j$ we concluded that the column space of $\mathbf{Y}^n_{i\,(2,3;1)}$ is contained in the column space of $\mathbf{Y}^n_{j\,(2,3;1)}$ if the (r_j+1)st singular value of the matrix $[\mathbf{u}_{i1} \ \ldots \ \mathbf{u}_{ir_i} \ \mathbf{u}_{j1} \ \ldots \ \mathbf{u}_{jr_j}]$ was less than 0.1 (and in this case we plotted the directed edge from vertex i to vertex j). The resulting directed graph is shown in Figure 5.4b. The same graph can be obtained directly from the nonzero pattern of the matrix $\mathbf{G}$ which means that all 43 (out of the possible 600) edges of the graph were detected correctly and no superfluous edges were added.

6. Conclusion. An obvious requirement for a tensor $\mathcal{B}$ to be the sum of (possibly scaled) terms from the decomposition of a tensor $\mathcal{A}$, is that its column (row, fiber, ...) space is a subspace of the corresponding space of tensor $\mathcal{A}$. Formally, this means that $\mathrm{row}(\mathbf{B}_{(n^c;n)}) \subseteq \mathrm{row}(\mathbf{A}_{(n^c;n)})$ should hold for all $n \in \{1, \ldots, N\}$. However, this is only a *necessary* condition. Switching to the column spaces, we have shown in this paper that

$$\mathrm{col}(\mathbf{B}_{(n^c;n)}) \subseteq \mathrm{col}(\mathbf{A}_{(n^c;n)}), \qquad n \in \{1, \ldots, N\} \tag{6.1}$$

is a *sufficient* condition for $\mathcal{B}$ to be generated by (possibly scaled) terms from the decomposition of $\mathcal{A}$. The number or terms and their "type" (namely, their ML rank) follow from the analysis as well. As the derivation relies only on linear algebra, it bypasses the typical difficulties in the computation of CPD and BTD, such as NP-hardness and possible ill-conditioning. We believe that this paper introduces a new tool that will prove important for tensor-based pattern recognition and machine learning, in a similar way as (explicit) tensor decompositions have proven to be fundamental tools for data analysis. We have illustrated the practical use of the new tool in a new clustering-based scheme for sparse underdetermined BSS.

An interesting topic of further study would be to investigate partially shared structure of $\mathcal{A}$ and $\mathcal{B}$, in the sense that $\mathcal{A}$ and $\mathcal{B}$ share some but not all terms. We will also derive more detailed information from the actual principal angles and associated directions between the subspaces obtained from $\mathcal{A}$ and $\mathcal{B}$. Another topic of further study is the generalization to "flower", "butterfly" and related decompositions [3, 4, 33].

Appendix A. The goal of this appendix is to illustrate a particular advantage of using the decomposition into a sum of ML rank-$(L_1, L_2, \cdot)$ terms over the CPD when we deal with the implication (1.3)⇒(1.1). Namely, we show that in the case where the terms in (1.1) are rank-1, (1.3) does not necessarily imply (1.1) but if the rank-1 assumption on the terms is replaced by the low ML rank assumption, then (1.3) does imply (1.1). We contrast our approach with the kind of results that can be derived from [7, Proposition 14.45], [31, Theorem 3.1.1.1], [26, Theorem 2.4].

Example A.1. We consider the $2 \times 2 \times 2$ tensors $\mathcal{A}, \mathcal{B}$ with frontal slices

$$\mathcal{A}(:,:,1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \mathcal{A}(:,:,2) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \ \mathcal{B}(:,:,1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \ \mathcal{B}(:,:,2) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{A.1}$$

Since

$$\mathbf{B}_{(2,3;1)} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}^T \cdot \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \mathbf{A}_{(2,3;1)}\mathbf{M}_1,$$

$$\mathbf{B}_{(1,3;2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}^T \cdot \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \mathbf{A}_{(1,3;2)}\mathbf{M}_2,$$

$$\mathbf{B}_{(1,2;3)} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}^T \cdot \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \mathbf{A}_{(1,2;3)}\mathbf{M}_3,$$

the column spaces of the corresponding matrix representations of $\mathcal{A}$ and $\mathcal{B}$ coincide. The result in [7, Proposition 14.45], [31, Theorem 3.1.1.1], [26, Theorem 2.4] only implies that $\mathcal{A}$ and $\mathcal{B}$ share the factor matrices in the first two modes, last two modes, and in the first and third mode, so, that $\mathcal{A}$ and $\mathcal{B}$ can be decomposed as

$$\mathcal{A} = \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes ? + \mathbf{a}_2 \otimes \mathbf{b}_2 \otimes ? + \cdots + \mathbf{a}_R \otimes \mathbf{b}_R \otimes ?, \tag{A.2}$$

$$\mathcal{B} = \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes ? + \mathbf{a}_2 \otimes \mathbf{b}_2 \otimes ? + \cdots + \mathbf{a}_R \otimes \mathbf{b}_R \otimes ?, \tag{A.3}$$

or as

$$\mathcal{A} = ? \otimes \mathbf{c}_1 \otimes \mathbf{d}_1 + ? \otimes \mathbf{c}_2 \otimes \mathbf{d}_2 + \cdots + ? \otimes \mathbf{c}_R \otimes \mathbf{d}_R, \tag{A.4}$$

$$\mathcal{B} = ? \otimes \mathbf{c}_1 \otimes \mathbf{d}_1 + ? \otimes \mathbf{c}_2 \otimes \mathbf{d}_2 + \cdots + ? \otimes \mathbf{c}_R \otimes \mathbf{d}_R, \tag{A.5}$$

or as

$$\mathcal{A} = \mathbf{e}_1 \otimes ? \otimes \mathbf{f}_1 + \mathbf{e}_2 \otimes ? \otimes \mathbf{f}_2 + \cdots + \mathbf{e}_R \otimes ? \otimes \mathbf{f}_R, \tag{A.6}$$

$$\mathcal{B} = \mathbf{e}_1 \otimes ? \otimes \mathbf{f}_1 + \mathbf{e}_2 \otimes ? \otimes \mathbf{f}_2 + \cdots + \mathbf{e}_R \otimes ? \otimes \mathbf{f}_R. \tag{A.7}$$

Note that, in general, the result in [7, Proposition 14.45], [31, Theorem 3.1.1.1], [26, Theorem 2.4] does not imply a procedure to compute R nor the vectors $\mathbf{a}_1, \ldots, \mathbf{f}_R$. Since $2 \times 2 \times 2$ tensors are well studied, it can be inferred from known results (see, for instance,[18]) that in this example $R = 3$. On the other hand, even in this small example decompositions (A.2)–(A.7) are not unique and the border rank of $\mathcal{A}$ and $\mathcal{B}$ is 2 [18], which means that the rank-3 tensors $\mathcal{A}$ and $\mathcal{B}$ can be arbitrary well approximated by tensors of rank 2. Further, again since we are dealing with $2 \times 2 \times 2$ tensors, it can be shown that $\mathcal{A}$ and $\mathcal{B}$ cannot be generated by the same three (possibly scaled) rank-1 tensors, that is,

$$\mathcal{A} = \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 + \mathbf{a}_2 \otimes \mathbf{b}_2 \otimes \mathbf{c}_2 + \mathbf{a}_3 \otimes \mathbf{b}_3 \otimes \mathbf{c}_3, \tag{A.8}$$

$$\mathcal{B} = \lambda_1\mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 + \lambda_2\mathbf{a}_2 \otimes \mathbf{b}_2 \otimes \mathbf{c}_2 + \lambda_3\mathbf{a}_3 \otimes \mathbf{b}_3 \otimes \mathbf{c}_3 \tag{A.9}$$

is not possible.[11] Hence, the number of rank-1 terms that (possibly scaled) generate both $\mathcal{A}$ and $\mathcal{B}$, is at least 4. (Obviously, 4 is also sufficient.) Thus, the minimal number of rank-1 tensors that generate both $\mathcal{A}$ and $\mathcal{B}$ is greater than the individual rank of $\mathcal{A}$.

Example A.2. Let $\mathcal{D}_1 = \mathcal{D}_2 = \mathcal{D}_3 \in \mathbb{F}^{2\times 2\times 2}$ be the same as $\mathcal{A}$ in (A.1), that is,

$$\mathcal{D}_r(:,:,1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \mathcal{D}_r(:,:,2) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \ r = 1, 2, 3.$$

[11] Indeed, if we assume that (A.8) and (A.9) hold, then for $(\alpha, \beta) \in \{(\lambda_1, 1), (\lambda_2, 1), (\lambda_3, 1)\}$ the tensor $\alpha\mathcal{A} - \beta\mathcal{B}$ is at most rank-2. Since $\alpha\mathcal{A} - \beta\mathcal{B} = \mathcal{B} \bullet_1 \begin{bmatrix} \alpha - \beta & \alpha \\ 0 & \alpha - \beta \end{bmatrix}$ and $\mathcal{B}$ is rank-3 tensor, it follows that $\alpha\mathcal{A} - \beta\mathcal{B}$ is at most rank-2 only if $\alpha = \beta$. Hence, $\lambda_1 = 1$, $\lambda_2 = 1$, and $\lambda_3 = 1$, implying that $\mathcal{A} = \mathcal{B}$, which is a contradiction.

and let

$$\mathcal{A} = \sum_{r=1}^{R} \mathcal{D}_r \bullet_1 \mathbf{X}_r^{(1)} \bullet_2 \mathbf{X}_r^{(2)} \bullet_3 \mathbf{X}_r^{(3)},$$

where $\mathbf{X}_r^{(1)}, \mathbf{X}_r^{(2)}, \mathbf{X}_r^{(3)} \in \mathbb{F}^{6\times 2}$ are such that the matrices $[\mathbf{X}_1^{(1)}\ \mathbf{X}_2^{(1)}\ \mathbf{X}_3^{(1)}]$, $[\mathbf{X}_1^{(2)}\ \mathbf{X}_2^{(2)}\ \mathbf{X}_3^{(2)}]$, and $[\mathbf{X}_1^{(3)}\ \mathbf{X}_2^{(3)}\ \mathbf{X}_3^{(3)}] \in \mathbb{F}^{6\times 6}$ are nonsingular, so $\mathcal{A} \in \mathbb{F}^{6\times 6\times 6}$ is a sum of three ML rank-$(2,2,2)$ terms. Let also $\mathcal{B} \in \mathbb{F}^{6\times 6\times 6}$ be a tensor such that assumption (4.3) in Theorem 4.1 holds. The result in [7, Proposition 14.45], [31, Theorem 3.1.1.1], [26, Theorem 2.4] only implies that (A.4)–(A.7) hold for some R and for some vectors $\mathbf{c}_1, \ldots, \mathbf{f}_R \in \mathbb{F}^6$. Since $\mathcal{A}$ is the sum of three ML rank-$(2,2,2)$ terms and each of the terms is a rank-3 tensor, it follows from [28] (Strassen's direct sum conjecture) that $\mathcal{A}$ is rank-9 tensor, that is $R = 9$. Note that even in this small scale example it is very hard to estimate R from the column spaces of the matrix representations of $\mathcal{A}$ or $\mathcal{B}$. The reason is that both $\mathcal{A}$ and $\mathcal{B}$ can be arbitrary well approximated by tensors of rank-6. (Indeed each ML rank-$(2,2,2)$ term can be arbitrary well approximated by tensors of rank-2.)

On the other hand, Theorem 4.1 says under what assumptions on $\mathbf{M}_1$ and $\mathbf{M}_2$ both tensors $\mathcal{A}$ and B can be decomposed into a sum of three ML rank-$(2,2,\cdot)$ terms and how the ML rank-$(2,2,\cdot)$ terms of $\mathcal{A}$ and $\mathcal{B}$ relate to each other.

REFERENCES

[1] C. Beltrán, P. Breiding, and N. Vannieuwenhoven, *The average condition number of most tensor rank decomposition problems is infinite*, arxiv:1903.05527, (2019).

[2] C. Beltrán, P. Breiding, and N. Vannieuwenhoven, *Pencil-based algorithms for tensor rank decomposition are not stable.*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 739–773.

[3] M. Boussé, O. Debals, and L. De Lathauwer, *Tensor-based large-scale blind system identification using segmentation*, IEEE Transactions on Signal Processing, 65 (2017), pp. 5770–5784.

[4] M. Boussé, O. Debals, and L. De Lathauwer, *A tensor-based method for large-scale blind source separation using segmentation*, IEEE Transactions on Signal Processing, 65 (2017), pp. 346–358.

[5] P. Breiding and N. Vannieuwenhoven, *On the average condition number of tensor rank decompositions.*, IMA Journal of Numerical Analysis, (2019).

[6] 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.

[7] P. Buergisser, M. Clausen, and M. A. Shokrollahi, *Algebraic Complexity Theory*, Springer Publishing Company, Incorporated, 1st ed., 2010.

[8] 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.

[9] 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.

[10] 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.

[11] P. Comon and C. Jutten, eds., *Handbook of Blind Source Separation, Independent Component Analysis and Applications*, Academic Press, Oxford, UK, 2010.

[12] C. W. Curtis, *Linear Algebra: An Introductory Approach*, Springer, New York, NY, 1984.

[13] 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.

[14] 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.

[15] L. De Lathauwer, *Characterizing higher-order tensors by means of subspaces*, STADIUS Res. Division, Dept. Elect. Eng., Katholieke Univ. Leuven, Leuven, Belgium, Tech. Rep. 11-32, (2011).

[16] 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.

[17] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.

[18] V. de Silva and L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127.

[19] O. Debals, *Tensorization and applications in blind source separation*, PhD thesis, Faculty of Engineering, KU Leuven (Leuven, Belgium), August 2017.

[20] 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.

[21] 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.

[22] 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.

[23] I. Domanov and L. De Lathauwer, *On uniqueness and computation of the decomposition of a tensor into multilinear rank-$(1, L_r, L_r)$ terms*, SIAM J. Matrix Anal. Appl., 41 (2018), pp. 747–803.

[24] 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).

[25] D. Donoho, *Sparse components of images and optimal atomic decompositions*, Constructive Approximation, 17 (2001), pp. 353–382.

[26] S. Friedland, *On the generic and typical ranks of 3-tensors*, Linear Algebra and its Applications, 436 (2012), pp. 478–497.

[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 2012.

[28] J. Ja' Ja' and J. Takche, *On the validity of the direct sum conjecture*, SIAM Journal on Computing, 15 (1986), pp. 1004–1020.

[29] B. N. Khoromskij, *Tensor Numerical Methods in Scientific Computing*, De Gruyter, 2018.

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

[31] J. M. Landsberg, *Tensors: geometry and applications*, volume 128 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2012.

[32] 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.

[33] L. Sorber, M. Van Barel, and L. De Lathauwer, *Optimization-based algorithms for tensor decompositions: canonical polyadic decomposition, decomposition in rank-$(L_r, L_r, 1)$ terms and a new generalization*, SIAM J. Optim., 23 (2013), pp. 695–720.

[34] L. R. Tucker, *The extension of factor analysis to three-dimensional matrices, in Contributions to Mathematical Psychology, H. Gulliksen and N. Frederiksen, eds.*, Holt, Rinehart & Winston, New York, 1966.

[35] L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311.

[36] F. Van Eeghem, O. Debals, and L. De Lathauwer, *Tensor similarity in two modes*, IEEE Transactions on Signal Processing, 66 (2018), pp. 1273–1285.